Protein and RNA Analysis of Extracellular Vesicle Subtypes

Wittaya Suwakulsiri

Bachelor of Science (Biology)

Master of Science (Chemical Biology)

A thesis submitted in total fulfilment of the requirements for the degree of

Doctor of Philosophy

Department of Biochemistry and Genetics

La Trobe Institute for Molecular Sciences

College of Science Health & Engineering

La Trobe University

Victoria, Australia

February 2021

Table of contents

Abstract

Extracellular vesicles (EVs) contain DNA, RNA, proteins and lipids that are crucial in intercellular communication. Cells release at least three major EV classes – namely, exosomes (Exos), shed microvesicles (sMVs) and shed midbody remnants (sMB-Rs). While purification, biochemical/ biophysical characterisation procedures, and biological activities of exosomes are advanced, our knowledge of sMVs and sMB-Rs is lacking. The overarching aim of my thesis was to undertake a comparative proteomic and transcriptomic analysis of Exos, sMVs and sMB-Rs secreted from human colorectal cancer (CRC) cells using mass spectrometry and next-generation RNA sequencing. To this end, I developed a purification strategy involving a combination of differential ultracentrifugation and buoyant density gradient centrifugation (iodixanol, OptiPrep™) centrifugation to isolate and purify Exos (50-200 nm diameter, buoyant density 1.10 g/mL), sMVs (100-500 nm diameter, buoyant density 1.10 g/mL) and sMB-Rs (100-500 nm diameter, buoyant density 1.15 g/mL) from the culture medium of the isogenic human CRC cancer cell lines SW480 (primary tumor) and SW620 (lymph node-metastatic tumor). To obtain large (mg) quantities of EVs for comprehensive proteomic/ transcriptomic profiling of the three EV classes, I used continuous culture in a CELLine™ Biorector device. Comparative protein/ transcript analyses whowed that exosomal proteins such as CD63, CD81, ALIX, TSG101 and protein-coding transcripts (mRNA) involved in exosome biogenesis/ release (ARRDC3-201, BICD2-201, CAV2-201, for example) are highly-enriched (log2 fold change >1

or <-1, pvalue <0.05) in Exos, and metabolic enzymes (e.g., DTYMK, IMPA1 and MRI) and membrane-associated proteins (SLC29A2) are selectively-enriched in sMVs. A salient finding was that principal component analysis and hierarchical clustering analysis showed that the proteome and and transcriptome of sMB-Rs are highly dissimilar to Exos and sMVs (which are similar to each other). sMB-Rs are highly enriched with midbody components, mitochondrial proteins, histone subunits and RNA-binding proteins (RBPs) such as ribonucleoproteins, RNA stress/granule proteins, splicing factors, and translation initiation factors at the protein level. The transcriptome of sMB-Rs reveals highly-enriched mitochondrial transcripts, lncRNA/pseudogene transcripts that are predicted to bind to 27 RBPs (from ENCORI, RBP-target database). Interestingly, 19/27 RBPs were identified in the sMB-R proteome. Importantly, RNA-Seq analysis revealed 770 fusion genes – predominantly, in sMB-Rs - including fusion genes from the tumour suppressor gene (MSH2) such as PLAGL1-MSH2, METRNL-MSH2, and oncogenes (CDK, RAS) such as RRAS-BMP8B and CDK-STX4. Interestingly, the fusion genes NSD1-ZNF346 and CPS1-ATXN10 are only detected in SW620-EVs, with selective enrichment in sMB-Rs. These findings suggest that sMB-Rs-derived fusion genes might act as liquid biopsy targets for metastatic cancer detection. Expression of cancer progression-related transcripts in Exos, sMVs and sMB-Rs positively correlate with their proteomic profiles. Moreover, SW620-EV classes contain enriched non-coding transcripts that are known to promote cancer progression. Findings from this thesis provide a better understanding

of cargo molecules selectively enriched in the three EV classes and lay the groundwork for a better understanding of the role of EVs in intercellular communication and CRC progression. Many of the findings have potential for targeted EV-based liquid biopsies and clinical application.

Statement of authorship

The work described in the manuscripts included this thesis was performed entirely by myself except for the following collaborative work described: a published review article in *Nature Review Clinical Oncology* in **Chapter 1**, my contribution is in the scoping and preparation of Tables 1, 2 and 3 along with Dr. Rong Xu and Dr. Alin Rai. **Chapter 2**, Dr. David Greening assisted with sample preparation for mass spectrometry and protein identification, Dr. Alin Rai assisted with the migration/invasion experiment. **Chapter 3** (prepared manuscript for publication), Dr. David Greening assisted with sample preparation for mass spectrometry and protein identification, Dr. Alin Rai assisted with isolation and purification of shed midbody remnants. **Chapter 4** (prepared manuscript for publication), Dr. Rong Xu assisted with RNA isolation from EV samples and qRT-PCR technique for my published article in *Methods in Molecular Biology* (inserted as a part of **Chapter 4**), Dr. Maoshan Chen assisted contact with Beijing Genomic Institute, China, for RNA sequencing of EV samples. Professor Dr. Richard Simpson assisted with experimental plan and every manuscript preparation. I estimate that my contribution to **Chapter 2** to be greater than 85%, **Chapter 3** to be greater than 90% and my contribution to **Chapter 4** to be greater than 95%, therefore, achieving an overall contribution of my work in the thesis to be greater than 90%.

This thesis includes work by the author that has been published as described in the text. Except where reference is made in the text of the thesis,

this thesis contains no other material published elsewhere or extracted in the whole or in part from a thesis accepted for the award of any other degree or diploma. No other person's work has been used without due acknowledgment in the main text of the thesis. This thesis has not been submitted for the award of any degree or diploma in any other tertiary institution.

This work was supported by a La Trobe University Postgraduate Research Scholarship and a La Trobe University Full Fee Research Scholarship.

Wittaya Suwakulsiri

February 2021

Preface

Published articles and submitted manuscripts during PhD candidature

1. **Suwakulsiri, W**., Rai, A., Xu, R., Chen, M., Greening, D. and Simpson, R. Proteomic profiling reveals key cancer progression modulators in shed microvesicles released from isogenic human primary and metastatic colorectal cancer cell lines. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*. **12**, 140171 (2019). PMID: 30502510.

2. Chen, M., Xu, R., Rai, A., **Suwakulsiri, W**., Izumikawa, K., Ishikawa, H., Greening, D., Takahashi, N. and Simpson, R. Distinct shed microvesicle and exosome microRNA signatures reveal diagnostic markers for colorectal cancer. *PLoS One*. (2019). PMID: 30608951 (Published online: https://doi.org/10.1371/journal.pone.0210003).

3. Xu, R., Rai, A., Chen, M., **Suwakulsiri, W**., Greening, D. and Simpson, R. Extracellular vesicles in cancer—implications for future improvements in cancer care. *Nature Reviews Clinical Oncology*. **15**, 617-638 (2018). PMID: 29795272.

4. **Suwakulsiri, W.**, Chen, M., Greening, D., Xu, R. and Simpson, R. Analysis of annotated and unannotated long non-coding RNAs from exosome subtypes using next-generation RNA sequencing. *Methods in Molecular Biology*. **2254**, 195-218 (2021). PMID: 33326077.

5. Rai, A., Greening, D., Xu, R., **Suwakulsiri, W**. and Simpson, R. Exosomes Derived from the Human Primary Colorectal Cancer Cell Line SW480 Orchestrate Fibroblast-Led Cancer Invasion. *Proteomics*. 2020. PMID: 32438511 (Published online: doi: 10.1002/pmic.202000016).

6. Rai, A., Greening, D., Xu, R., Chen, M., **Suwakulsiri, W**. and Simpson, R. Secreted midbody remnants are a class of extracellular vesicles molecularly distinct from exosomes and microparticles, *Nature Communications Biology* (2021).

Manuscripts in preparation for publication

1. **Suwakulsiri, W**., Rai, A., Greening, D., Xu, R. and Simpson, R. Comparative proteomic analysis of Exos, sMVs and sMB-Rs released from the isogenic colorectal cell lines SW480 and SW620. (Encompassing Chapter 3 of this thesis. Manuscript in preparation for submission to *Proteomics*).

2. **Suwakulsiri, W**., Rai, A., Greening, D., Chen, M., Xu, R. and Simpson, R. Comparative transcriptomic and fusion gene analysis of Exos, sMVs and sMB-Rs released from the isogenic colorectal cell lines SW480 and SW620. (Encompassing Chapter 4 of this thesis. Manuscript in preparation for submission to *Scientific Reports*).

Acknowledgement

There are many people who have given me support during my PhD study and I would very much like to express my sincere gratitude to all of them.

The first person is my supervisor, Professor Dr. Richard Simpson. I first met him two years in a scientific conference before I have started my PhD. I was impressed by his presentation and accomplishments in Science. During my PhD study, he has given me endless support and guidance. His ability to supervise and inspire me has been excellent. He is considerate and compassionate, especially in a very hard time, COVID19 pandemic. He is open minded to my ideas and allow me to push my research boundaries and learn new things. He makes me confident that I am stepping on the right steppingstone. Apart from research, we share similar political beliefs and values as well as interesting TV show/documentaries, cooking, baking and more. I truly have enjoyed these years and friendship with him. I am sure the friendship will be for evermore.

The second person is my co-supervisor, Dr. David Greening who has supported me even before I came to Australia. He always makes sure that myself and everybody in the laboratory are going well both on- and off-campus. His assistance with mass spectrometry is significant for my PhD study. I have also learned to distinguish between ch- and sh- sounds from him.

The third person is Dr. Rong Xu. He is one of the colleges and a very good friend of mine. I remember I had my first dinner in Australia with him and

his wife, Jennifer. They both are very supportive and understanding. Dr. Rong Xu has helped me settled down in Australia as well as with experiments. He always offered his support for issues that I have had during my PhD study. His generosity towards me is truly remarkable.

The fourth person is Dr. Alin Rai who has taught me almost every experiment, step by step. He always shares knowledge, jokes, laughs and companionship which are very important when experiments went wrong.

I would like to thank Adnan Shafiq for friendship during my PhD study and Dr. Maoshan Chen who assisted with Beijing Genomic Institute (BGI) regarding RNA sequencing of my samples. I appreciate assistance from friends and researchers in LIMS, Christina from Chen laboratory for flow cytometry training and antibodies for my side project, Dr. Peter from microscopy platform, La Trobe university for fluorescence microscopy training, Dr. Georgia from Hulett laboratory for sharing THP-1 cell line, Pamali from Mathivanan laboratory for sharing supplies in cell culture and EV isolation, Tao and Shane from Perugini laboratory and Dr. Andrew from genomic platform, La Trobe university for troubleshooting in LIMS-HPC.

I thank La Trobe university for granting scholarships for international students, LIMS for traveling grants to present my research in Adelaide, Japan and New Zealand and Graduate Research School (GRS) for their assistance during my PhD study.

My life outside university has been wonderful because of friendship from Anthony, David, Jeffrey, Kittridge, Matt, Sean, and Yvonne. Thanks for putting events on my vacant weekend and long holidays. I specially thank Ian, Susan, Melissa, and James for friendship and endless support since I came to Australia, I eat more greens because of you guys.

To my family in Thailand, I thank them for allowing me to do whatever I want to achieve in life. I specially give a big hug to my mother and brother. I could write another 1,000 words to explain how much they support me but keep it short, I know they are the only two people who are sadder if I am sad and happier if I am happy.

Finally, I would not be writing ‘Acknowledgement’ without support from Michael who has put up with my music whether with the highest or lowest note, for three summers now.

List of figures

Chapter 1

Chapter 3

Chapter 4

Chapter 5

List of tables

Chapter 1

Chapter 3

Chapter 4

Abbreviations

5T4	5T4 oncotrophoblast glycoprotein
A33	A33 antigen
A4F	Asymmetrical-flow field-flow fractionation
ABCB	ATP-binding cassette sub-family B
ADAM	Disintegrin and metalloproteinase domain-containing protein
AGO2	Argonaute 2
ALG-2	Alpha-1,3/1,6-mannosyltransferase
ALIX	ALG-2 interacting protein X
ALYREF	Aly/REF export factor
ANAX	Annexin
AP2A1	AP-2 complex subunit alpha-1
APC	Adenomatous polyposis coli
ARF	ADP-ribosylation factor
ARHGAP	Rho GTPase-activating protein
ARRDC1	Arrestin domain-containing protein 1
ATPs	ATP synthase subunits
AURB	Aurora kinase B
AXL	AXL oncogene
BAX	Bcl-2-like protein 4
BICD2	Bicaudal D homolog 2
BMAL1	Brain and muscle ARNT-like 1

BMMSC	Bone marrow mesenchymal stem cell
BRAF	v-Raf murine sarcoma viral oncogene homolog B
BSG	Basigin
BTG2	BTG family member 2
CACNA1G	Calcium channel, voltage-dependent, T type, alpha 1G subunit
CacyBP	Calcyclin-binding protein
CAF	Cancer associated fibroblast
CANX	Calnexin
CAV2	Caveolin-2
CCA	Cholangiocarcinoma
CD	Antigen/tetraspanin CD
CDC42	CDC42 small effector protein 1
CDK	Cyclin-dependent kinase
cDNA	Complementary deoxyribonucleic acid
CEACAM1	Carcinoembryonic antigen-related cell adhesion molecule 1
CEP55	Centrosomal protein of 55 kDa
CHMP	Charged multivesicular body protein
CIMP	CpG island methylator phenotype
CIN	Chromosomal instability
circRNA	circular RNA
CLDN	Claudin

CM	Culture medium
COPB	Coatomer subunit beta
CpG	Cytosine followed by a guanine nucleotide
CRC	Colorectal cancer
CRY1	Cryptochrome-1, AtCry
CTN	Catenin
CXCR4	C-X-C chemokine receptor type 4
dATP	Deoxyadenosine triphosphate
DAVID	Database for annotation, visualization and integrated discovery
DBP	DEAD box protein
dCTP	Deoxycytosine triphosphate
DEL-1	E2F-like protein
dGTP	Deoxyguanine triphosphate
DLGAP1	Disks large-associated protein 1
DNA	Deoxyribonucleic acid
DNAJC27	DnaJ homolog subfamily C member 27
DOCK1	Dedicator of cytokinesis protein 1
DTYMK	Thymidylate kinase
dUTP	Deoxyuridine triphosphate
EAP	Endoplasmic reticulum-arrested pen
ECM	Extracellular matrix
EDIL3	EGF-like repeat and discoidin I-like

	domain-containing protein 3
EGF/EGFR	Epidermal growth factor/
	Epidermal growth factor receptor
EIF	Eukaryotic translation initiation factor
ELAVL1	Elav-like generic protein
EMMPRIN	Extracellular matrix metalloproteinase inducer
ENCORI	Encyclopedia of RNA interactomes
EPCAM	Epithelial cell adhesion molecule
ERBB	Proto-oncogene c-ErbB
ERK	Extracellular signal-regulated kinase
eRNA	Endogenous RNA
ERP44	Endoplasmic reticulum resident protein 44
ESCRT	Endosomal sorting complex required for transport
EV	Extracellular vesicle
Exo	Exosome
FAK	Focal adhesion kinase
FAP	Familial adenomatous polyposis
FAP	Familial adenomatous polyposis coli
FAS	Apoptosis-mediating surface antigen FAS
FGFR	Fibroblast growth factor receptor
FHL1	Four and a half LIM domains protein 1
FPKM	Fragments per kilobase of
	transcript per million mapped reads

FUS	Fusion 1 protein
G6PD	Glucose-6-phosphate 1-dehydrogenase
GAG	Envelope glycoprotein
GBC	Gallbladder carcinoma
GDF15	Growth/differentiation factor-15
gDNA	Genomic deoxy ribonucleic acid
GeL-LC	Gel-based separation liquid chromatography
GLC	Glucan endo-1,3-beta-glucosidase
GO	Gene ontology
GOLT1A	Golgi transport 1 homolog A
GRCh38	Genome reference consortium human build 38
GSE1	Genetic suppressor element 1
GSEA	Gene set enrichment analysis
HDAC1	Histone deacetylase 1
HER2	HMG2-induced ER-remodeling protein 2
HGS	Hepatocyte growth factor-regulated tyrosine
HIST	Histone subunit
HNPCC	Hereditary Nonpolyposis Colorectal Cancer
HNPCC	Hereditary non-polyposis colon cancer
HNRNP	Heterogeneous nuclear ribonucleoprotein
HRS	Transcription factor HRS1
HSPA5	Heat shock protein 70 family protein 5
ICAM1	Intercellular adhesion molecule 1

IDH1	Isocitrate dehydrogenase 1
IGF2	Insulin-like growth factor II
IGF2BP	Insulin-like growth factor 2 mRNA-binding protein
ILV	Intraluminal vesicles
IMPA1	Inositol monophosphatase 1
ITG	Integrin
ITS	Insulin transferrin selenium
JAG1	Protein jagged-1
JNK	Stress-activated protein kinase jnk
KEGG	Kyoto encyclopedia of genes and genomes
KIF	Kinesin-like protein
KISS	Kisspeptins receptor
KRAS	GTPase KRas
L1CAM	Neural cell adhesion molecule L1
LAMP1	Lysosome-associated membrane glycoprotein 1
LAMTOR	Late endosomal/lysosomal adaptor and MAPK and MTOR activator
LARP1	La ribonucleoprotein domain family member 1
LCK	Leukocyte C-terminal Src kinase
LFQ	Label-free quantification
lncRNA	Long non-coding RNA
LOH	Loss of heterozygosity
LRBA	Lipopolysaccharide-responsive

	and beige-like anchor protein
MACC1	Metastasis-associated in colon cancer protein 1
MAGOH	Protein mago nashi homolog
MAPK	Mitogen-activated protein kinase
MAPK	Mitogen-activated protein kinase
MARCKS	Macrophage myristoylated
	alanine-rich C kinase substrate
MB	Midbody
MDK	Midkine
MDSC	Myeloid-derived suppressor cell
MET	Hepatocyte growth factor receptor
METTL3	Methyltransferase-like protein 3
MGST1	Microsomal glutathione S-transferase 1
MHC	HLA class II histocompatibility antigen
MIF	Macrophage migration inhibitory factor
miRNA	MicroRNA
MKLP1	Mitotic kinesin-like protein 1
MLH1	mutL homolog 1
MMP	Matrix metalloproteinase
MMR	DNA mismatch repair system
MRI1	Methylthioribose-1-phosphate isomerase
mRNA	Messenger RNA
MS/MS	Tandem mass spectrometry

MSC	Mesenchymal stem cell
MSH	mutS homolog
MSI	Microsatellite instability
MSS	Microsatellite stable
MT	Mitochondrial
mTOR	Mammalian target of rapamycin
MUC1	Mucin-1
mutL	DNA mismatch repair protein mutL
mutS	DNA mismatch repair protein mutS
MVB	Multivesicular body/ Multivesicular body sorting factor
N4BP2L2	NEDD4-binding protein 2-like 2
ncRNA	Non protein-coding RNA
NEUROG1	Neurogenin 1
NLR	Nucleotide-binding domain, leucine rich repeat containing receptor
NOTCH	Neurogenic locus notch homolog protein
NTA	Nanoparticle tracking analysis
OCLN	Occludin
OSGIN2	Oxidative stress-induced growth inhibitor 2
OST	Oligosaccharyl transferase
P/S	Penicillin/Streptomycin
PAK1	p21-activated kinase 1

PARP1	Poly ADP-ribose polymerase 1
PCA	Principle component analysis
PCR	Polymerase Chain Reaction
PDIA	Protein disulfide-isomerase A
pDNA	Plasmid deoxyribonucleic acid
PER3	Period circadian protein homolog 3
P-gp	P glycoprotein
PININ	Nuclear protein SDK3
piRNA	Piwi-interacting RNA
PKP	Plastidial pyruvate kinase
PLD	Phospholipase D
PLK1	Serine/threonine-protein kinase/ Polo-like kinase
PMS	Mismatch repair system component
POSTN	Periostin
PRC1	Protein regulator of cytokinesis 1
PRCC	Proline-rich protein
PRKCA	Protein kinase C alpha type
PRP43	Pre-mRNA-splicing factor ATP-dependent RNA helicase PRP43
PRP8BP	Prp8-binding protein
PTB	Polypyrimidine tract-binding protein
PTEN	Phosphatidylinositol 3,4,5-trisphosphate

	3-phosphatase and dual-specificity protein phosphatase
PTPRZ1	Receptor-type tyrosine-protein phosphatase zeta
PVRL	Poliovirus receptor-related protein
RAB	Ras-related protein Rab
RACGAP1	Rac GTPase-activating protein 1
RAE1	mRNA export factor (Rae1 protein homolog)
RANGAP	Ran GTPase-activating protein 1
RAP2A	Ras-related protein Rap-2a
RAS	Proto-oncogene protein p21
RASA	Ras GTPase-activating protein
RBM	RNA-binding motif protein
RBM44	RNA binding motif protein 44
RBP	RNA binding protein
RHOA	Ras-like GTP-binding protein rhoA
RICTOR	Rapamycin-insensitive companion of mTOR
RIPK2	Receptor-interacting
	serine/threonine-protein kinase 2
RNA	Ribonucleic acid
RNPS1	RNA-binding protein with serine-rich domain 1
ROS	Reactive oxygen species
RPMI-1640	Roswell park memorial institute culture medium 1640
RRAS	Ras-related protein R-Ras2
rRNA	Ribosomal RNA

RUNX3	Runt-related transcription factor 3
RUSC1	RUN and SH3 domain-containing protein 1
S100A	S100 calcium-binding protein
SCAF1	SR-related C-terminal domain-associated factor 1
SCAP	Apical papilla
SDCBP	Syndecan-binding protein
SDS-PAGE	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
SEC13	GATOR complex protein SEC13
SETD5	Histone-lysine N-methyltransferase SETD5
SF1	Splicing factor
SFPQ	Splicing factor, proline- and glutamine-rich
SH3D19	SH3 domain-containing protein 19
SLC	Solute carrier family
SMAD	Mothers against decapentaplegic homolog
SMARCA5	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5
sMB-Rs	Shed midbody remnant
sMVs	Shed microvesicle
SNAP	Synaptosomal-associated protein
snoRNA	Small nucleolar RNA
snRNA	Small nuclear RNA
SNRNP	U1 small nuclear ribonucleoprotein

SOCS1	Suppressor of cytokine signaling 1
SRC	Proto-oncogene c-Src
STAM	Signal transducing adapter molecule
STAT	Signal transducer and activator of transcription
STRING	Search tool for the retrieval of interacting genes/proteins
STX7	Syntaxin-7
TARDBP	TAR DNA-binding protein
TCGA	The Cancer Genome Atlas
TEM	Transmission electron microscopy
TEX14	Testis-expressed protein 14
TGFBR2	TGF-beta receptor type-2
TGF-β1	Transforming growth factor beta1
TGM2	Tissue transglutaminase 2
TIMP	Tissue inhibitor of metalloproteinases
TKT	Transketolase
TM4SF1	Transmembrane 4 L6 family member 1
TNC	Tenascin
TNIK	TRAF2 and NCK-interacting protein kinase
TOMM	Translocase of outer mitochondrial membrane
TP53	Tumour suppressor 53
TP53RK	TP53-regulating kinase
tRNA	Transfer ribonucleic acid RNA

TSG101	Tumor susceptibility gene 101 protein
TSPAN	Tetraspanin
TTC28	Tetratricopeptide repeat protein 28, TPR repeat protein 28
U2AF	U2 auxiliary factor
UBC9	Ubiquitin carrier protein 9
UGGT	UDP-glucose:glycoprotein glucosyltransferase
UPLC	Ultra performance liquid chromatography
VAMP	Vesicle-associated membrane protein
VDAC	Voltage-dependent anion-selective channel protein
VIM	Vimentin
VPS	Vacuolar protein sorting/targeting protein
WNT	Wingless and Int-1
YBX1	Y-box protein-1
YRNA	Ro-associated Y RNA
YWHAZ	14-3-3 protein zeta/delta
ZNF	Zinc finger protein

Chapter 1

Introduction

1.1 Colorectal cancer and genetic instability

Colorectal cancer (CRC) is the third highest cause of death and one of the prevalent cancers globally[1-3] due, primarily, to the difficulties of early diagnosis, treatment, as well as CRC metastasis. It has been estimated that CRC stage I patients have greater than 90% 5-year survival rate, whereas the 5-year survival rate is less than 10% for metastatic cancer patients[2]. CRC begins with the development of aberrant crypt foci, which progress to advanced adenoma and then to carcinoma or invasive cancer that further spreads to regional lymph nodes and distant sites (**Fig.1.1**)[1, 4, 5]. Genetic instability events are involved in the transformation of normal colonic epithelial cells to carcinomas[6]

Genetic instability such as chromosomal instability (CIN) and microsatellite instability (MSI) can drive CRC development by specific genetic alterations[7]. CIN is the most commonly occurring genetic instability in CRC[7, 8]. CIN results from defects in chromosome segregation and cell cycle control that causes aneuploidy with abnormal chromosome numbers distributed to daughter cells[9]. Abnormal centrosome numbers can also result in multiple spindle fibers during mitosis, causing unequal distribution of chromosomes[10]. Loss of heterozygosity (LOH) is a hallmark of CIN-positive tumours whereby 25%-30% of alleles are deleted [11]. The consequences of LOH are chromosome deletion and chromosome recombination[9]. CIN can alter a wild-type copy of tumour suppressor genes such as TP53, BAX[12], and also the

APC (adenomatous polyposis coli) gene that is a regulator of the Wnt signaling pathway; mutations of these genes cause rapid growth of normal colon epithelium to carcinoma[12, 13] (**Fig. 1.2**).

Hereditary nonpolyposis colorectal cancer (HNPCC) – also known as Lynch Syndrome – and Familial Adenomatous Polyposis Coli (FAP) are two hereditary disorders known to predispose the affected individual to the development of CRC[14]. Both HNPCC and FAP arise from germline mutations – HNPCC from mutations in the DNA mismatch repair (MMR) genes MSH2, MSH6, MLH1 and PMS2[14-18], and mutation of the APC gene in the case of FAP[19, 20]. In the case of HNPCC -, lack of DNA repair mechanisms mediates epigenetic changes such as aberrant DNA hypo- and hyper-methylation[12, 21] that lead to microsatellite instability (MSI). The identification of specific patterns of chromosome instability (loss and gain) during colon adenoma to carcinoma process and the demonstration that CIN is an early event in tumor formation that increases during tumor progression indicate that CIN is pathogenetic in CRC[22]. In the case of FAP, mutation of the APC gene leads to accumulation of nuclear β-catenin (a part of Wnt/β-catenin pathway) that activates genes related to cell survival and proliferation such as c-MYC, c-JUN, CCDN1, EGFR, CD44 and CD133 via transcription factors (T-cell factor (TCF) and lymphoid enhancer factor (LEF))[23, 24].

MSI is detected in about 15% of CRCs as a result of aberrant DNA mismatch repair system (MMR system) and DNA hypermethylation[25]. MSI is identified by frameshift mutations of microsatellite repeats in the genome[26].

The cause of MSI in CRC is due, predominantly, by the mutation in at least six MMR genes - PMS1, PMS2, MLH1, MSH2, MSH3 and MSH6[27]. One abnormal mechanism of MMR system is epigenetic silencing of MLH1 that leads to aberrant methylation[28]. This methylation typically involves BRAF mutation (V600E)[29]. The activation of oncogenes such as RAS, BRAF drives CRC progression via MAPK signaling pathway[30] as well as the phosphatidylinositol-3-kinase gene through PI3K signaling pathway[31] (**Fig. 1.2**). MSI tumours positively correlate with high levels of CpG island methylation (CpG island methylator phenotype (CIMP))[29].

CpG islands are DNA regions that contain a high frequency of cytosine nucleotides followed by a guanine nucleotide[32]. CpG islands appears to overlap with the promoter region of 60–70% of genes and tend to be protected from methylation[33]. In normal physiological conditions, cytosine is usually methylated outside exons. However, in CRC the genome, there are many aberrantly methylated cytosines within promoter-associated CpG islands, so called CpG island methylator phenotype (CIMP), which regulate epigenetic silencing of gene expression[33, 34] (**Fig. 1.3**). The mechanism of CIMP is not well understood but it is observed that 20% of CRCs are positive to CIMP[35] and 15% of CIMP appear to have abnormal methylation on MLH1 gene[34]. BRAF, NEUROG1, SOCS1, RUNX3, IGF2 and CACNA1G genes are reported to be candidates for determining CIMP status[36].

There is a chicken-and-egg question concerning the timeline of genome instability in CRC. Three parallel pathways for CIMP have been proposed by

Cheng *et al.,*[37]. – i) CIMP-positive CRC cases (10-20%) with BRAF and KRAS mutations and lower APC and p53 mutations. An aberrant MLH1 mutation and MSI appear to be detected in this group, - ii) CIMP-positive CRC cases with KRAS and APC mutations that occur in 10-30% of CRC cases, and - iii) CRC cases with the APC mutation and are less associated with CIMP (50-70% of CRC cases) and tend to further have CIN and P53 mutation[38]. Overall, there are overlapping causes of CIN, MSI and CIMP in CRC including defects in the MMR system, inherited germ-line defects, genetic alterations, and aberrant methylation (**Fig. 1.4**).

While genome instability and genetic alterations initiate CRC development, more recently the microenvironment of cancer has been considered as another crucial element for cancer growth and progression[39].

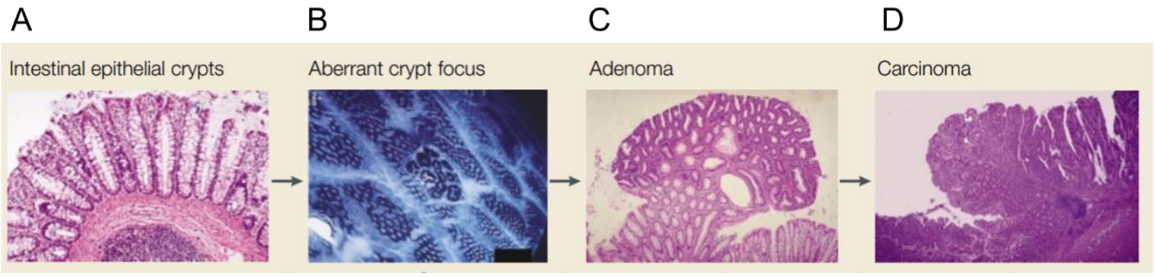


Figure 1.1 Histopathology of colorectal cancer. (A) Throughout the large intestine, a monolayer of epithelial cells lines tubular glands or crypts, **(B)** The earliest manifestations of colorectal neoplasia are the aberrant crypt foci (ACF), **(C)** A benign tumour mass that protrudes into the lumen from the intestinal epithelium, **(D)** The epithelium is organized in multiple layers, nuclei are enlarged, and their alignment at the basal membrane is lost. Figure reproduced with permission from Nature Reviews Cancer[40].

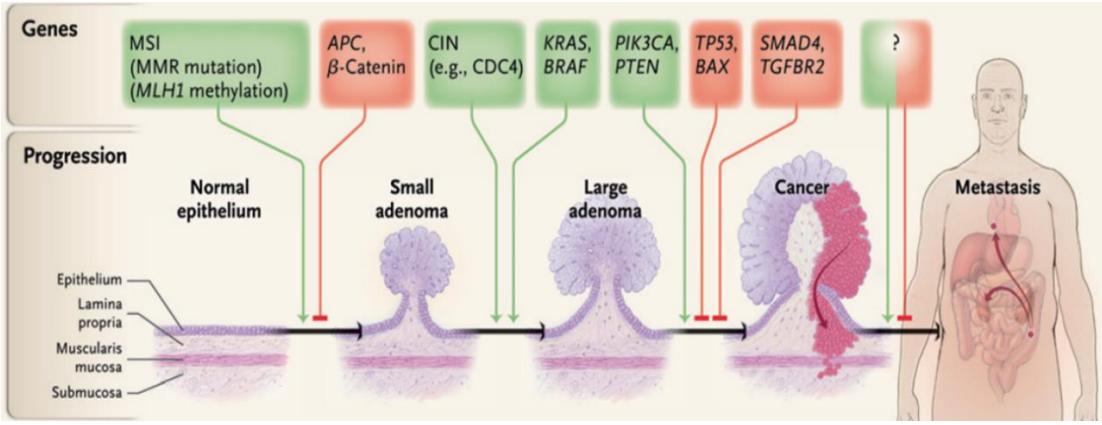


Figure 1.2 Colorectal cancer progression is driven by genetic instability. The microsatellite instability (MSI) pathway is initiated by mismatch-repair (MMR) gene mutation or by aberrant MLH1 methylation and is further associated with downstream mutations in TGFBR2 and BAX. Aberrant MLH1 methylation and BRAF mutation are each associated with the serrated adenoma pathway (CIN is chromosomal instability, EGFR is epidermal growth factor receptor, 15-PGDH is 15-prostaglandin dehydrogenase, and TGF-β is transforming growth factor β.). Figure reproduced with permission from The new England journal of medicine[41], Copyright Massachusetts Medical Society.

Figure 1.3 CpG island DNA hypermethylation and global DNA hypomethylation in colorectal cancer as compared with normal colonic epithelium. Unmethylated CpG islands within the promoter region of genes are correlated with an open chromatin structure (euchromatin), whereas methylated CpG islands are correlated with a condensed, closed chromatin structure (heterochromatin) and transcriptional silencing. Normal colonic epithelium generally has unmethylated CpG islands in the promoter regions of genes, whereas aberrant hypermethylation of promoter associated CpG islands is a hallmark of neoplasms. In addition to the aberrant local hypermethylation seen in colorectal cancers, global hypomethylation at LINE-1 sequences is also observed, which has been shown to be associated with genomic instability. Interestingly, an inverse association exists between local CpG island hypermethylation and global LINE-1 hypomethylation as colonic neoplasms progress. Figure reproduced with permission from Nature Reviews Gastroenterology & Hepathology[33].

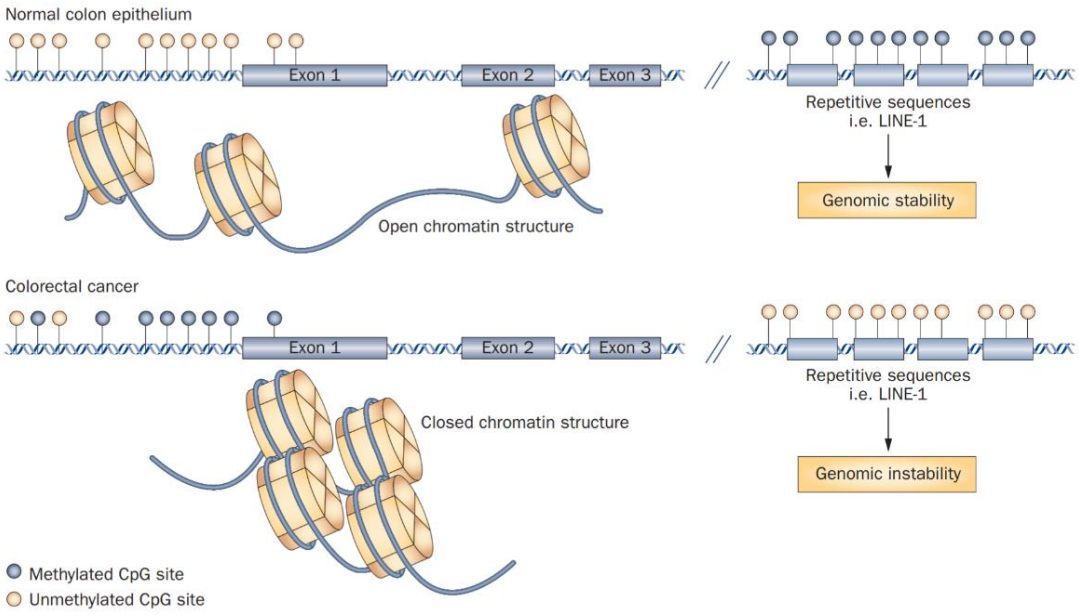
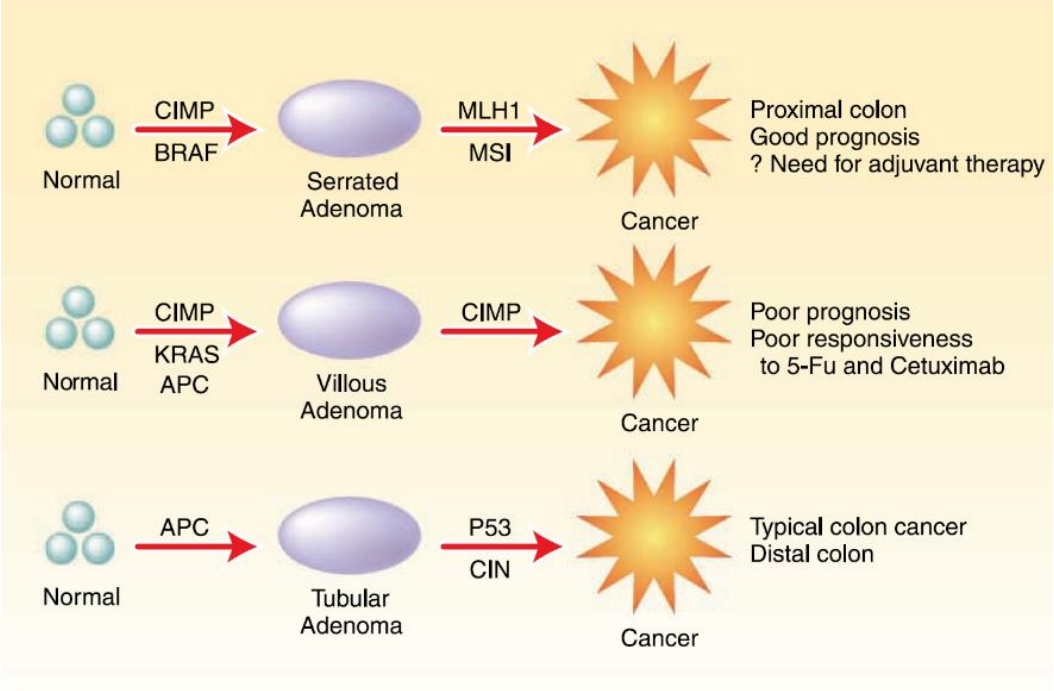

Normal colon epithelium
Exon 1
Exon 2
Exon 3
Open chromatin structure
Repetitive sequences
i.e. LINE-1
Genomic stability
Colorectal cancer
Exon 1
Exon 2
Exon 3
Closed chromatin structure
Repetitive sequences
i.e. LINE-1
Genomic instability
Methylated CpG site
Unmethylated CpG site

Figure 1.4 Multiple parallel pathways to colorectal cancer, with clinical implications for therapy. Colorectal cancer seems to arise from (at least) three distinct parallel modes. The top and bottom pathways are the most homogeneous, with clear distinctions in precursor lesions (serrated vs. tubular adenomas), genetics (BRAF vs. APC and p53 mutations, MSI vs. CIN), epigenetics (CIMP positive vs. negative) and outcome (good vs. average). The middle pathway is more heterogeneous than depicted (or perhaps incompletely understood). It may arise mostly from villous adenomas, but perhaps also from serrated adenomas. It has a different form of CIMP, predominant KRAS but occasional BRAF mutations, usually lacks CIN, and has the worse prognosis, with apparently lower responsiveness to chemotherapy. The prevalence of the three pathways is estimated at 10% to 20% (top pathway), 10% to 30% (middle pathway), and 50% to 70% (bottom pathway). Figure reproduced with permission from Clinical Cancer Research[38]


Normal
CIMP
BRAF
Serrated Adenoma
MLH1
MSI
Cancer
Proximal colon
Good prognosis
? Need for adjuvant therapy
Normal
CIMP
KRAS
APC
Villous Adenoma
CIMP
Cancer
Poor prognosis
Poor responsiveness
to 5-Fu and Cetuximab
Normal
APC
Tubular Adenoma
P53
CIN
Cancer
Typical colon cancer
Distal colon

.

1.2 Tumour microenvironment

Cancer development consists of multi-step processes beginning with genetic changes that involve cellular metabolic alterations, cancerous transformation and metastasis. It has been reported that in the tumour site, cancer cells interact with surroundings (stroma) to support its growth - the so-called tumour microenvironment (**Fig. 1.5**)[42, 43].

Tumours have been described as 'wounds that do not heal'[44] because there is a similarity between tumour progression and chronic inflammation. The purpose of inflammation is to eliminate an aberrant particle or invader. There is recruitment of inflammatory cells such as macrophages, mast cells, and neutrophils that can either support tumour progression or suppress tumour growth (**Fig. 1.5**). As immune evasion is one of the cancer hallmarks[45], tumours can also use infiltrating cells to create a favorable microenvironment that leads to tumour progression[42].

Tumours recruit stromal cells to their environment during inflammation[46, 47]. Stroma cells involved in this process are endothelial cells that comprise circulatory system and further angiogenesis, fibroblasts that provide a link between cancer, inflammation and extracellular matrix remodeling by releasing cytokines and proteases (**Fig. 1.6**) [48] as well as bone marrow-derived cells (BMDCs, macrophage, mast cell, and neutrophils) that respond to inflammation and might not be associated with direct elimination of cancer cells. This is because most tumour-associated antigens are detected

as self-antigens, whereas viral and bacterial antigens are considered as foreign antigens[49]. Moreover, macrophages can also promote tumour cell invasion, migration, and intravasation, by releasing inhibitory cytokines such as reactive oxygen species (ROS), prostaglandins, and interleukin-10 to inhibit lymphocyte function (**Fig. 1.6**)[50]. This creates hypoxic and acidic conditions which in turn regulate lactate metabolism (low oxygen energy production) in cancer. This metabolic shift supports growth of cancer in harsh conditions and also provides building blocks for cell division[46, 51].

In the tumour microenvironment, cancer and stroma cells exchange their genetic information and biomolecules by several ways including:

- i) release of soluble molecules such as circulating RNA in a non-vesicular form. For example, the release of miRNAs that bind to Argonaute2 (Ago2), RNA binding proteins (RBPs) as a ribonucleoprotein complex to prevent RNA degradation[52] as well as immune cells release soluble proteins such as cytokines and antibodies to trigger immune response[53].

- ii) connective structures (gap junction, intercellular bridges). Gap junction supports transfer of small molecules (single-stranded RNA with RBPs, microRNA, pre-miRNA, and proteins) via connexin 43 (integral proteins) between adjacent cells[54]. Intercellular bridges allows transfer of RNAs and proteins[39]. The intercellular bridge is formed by incomplete cytokinesis mediated by testis-expressed protein 14 (TEX14), RNA binding motif protein 44 (RBM44), endosomal sorting complex required for transport (ESCRT) and microtubule-organizing center (MTOC)[54, 55]

- iii) more recently, extracellular vesicles (EVs) have been implicated in exchange (transfer) of genetic information[56]. EVs are evolutionary conserved membranous vesicles secreted by diverse cell types. They play an important role in intercellular communication by transferring genetic materials[57, 58], protein compositions[38, 59, 60], and lipids[60] from their cell of origin to recipient cells[60-63] resulting in cellular alteration[56, 63]. EVs are also found in various body fluids and represent potential liquid biopsy-based targets for disease detection[64-68]. EVs are heterogenous representing different particle size ranges and originate from different biogenesis mechanisms.

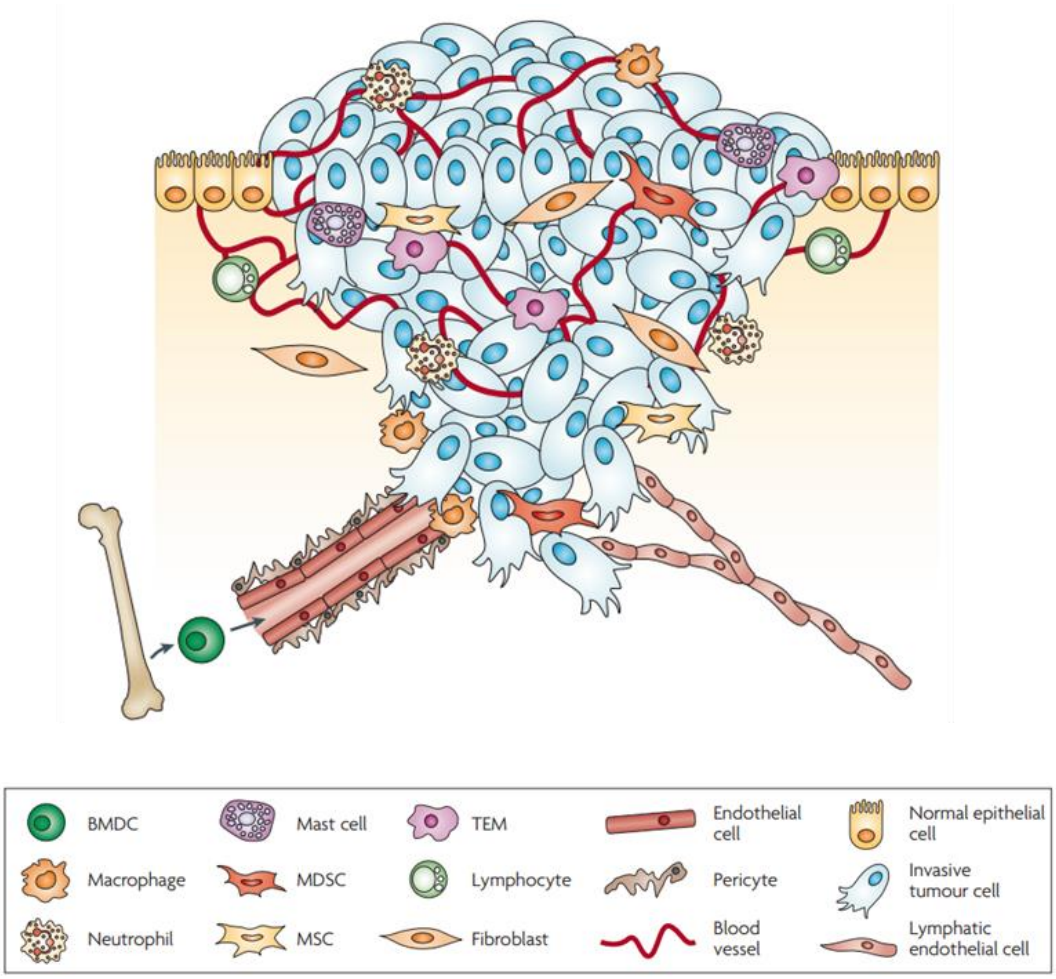


Figure 1.5 Tumour microenvironment. Cancer cells in primary tumours are surrounded by a complex microenvironment comprising numerous cells including endothelial cells of the blood and lymphatic circulation, stromal fibroblasts and a variety of bone marrow-derived cells (BMDCs) including macrophages, myeloid-derived suppressor cells (MDSCs), TIE2-expressing monocytes (TEMs) and mesenchymal stem cells (MSCs). Figure reproduced with permission from Nature Review Cancer[69].

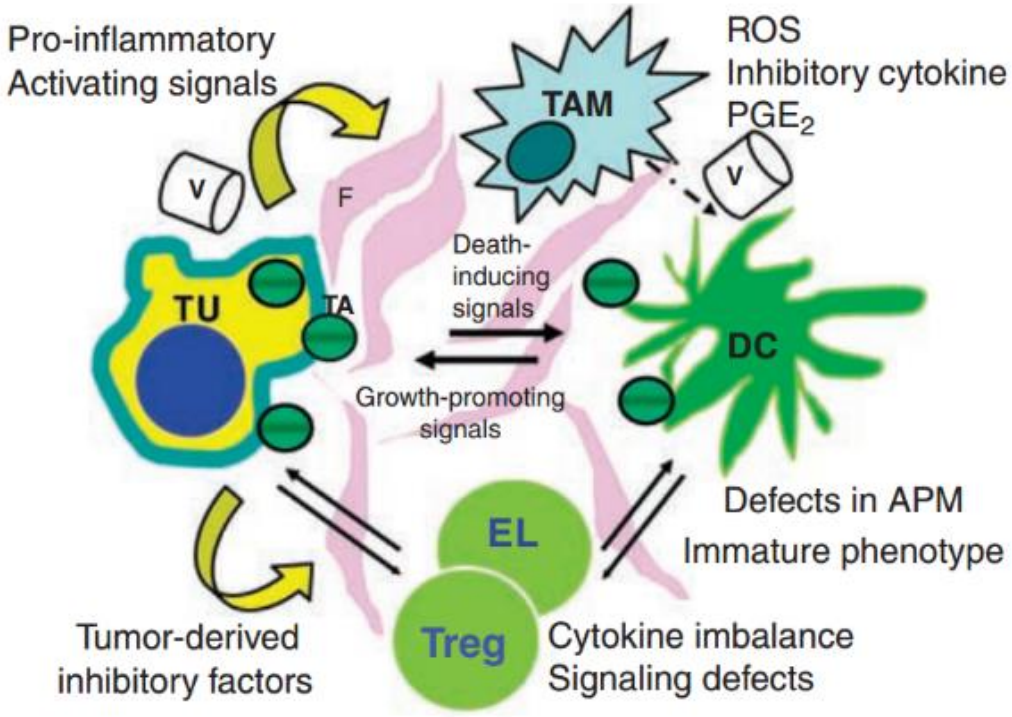


Figure 1.6 Interactions of tumour cell and stromal cells. Interactions of various cells with each other, fibroblasts (F) in the tumour stroma and blood vessels (V) are indicated by arrows. Although the tumour (TU) generates signals inducing dysfunction and death of immune cells, the latter are a source of signals promoting tumour growth. TAM, tumour-associated macrophages; DC, dendritic cells; EL, effector lymphocytes; Treg, regulatory T cells; TA, tumour-derived antigens; ROS, reactive oxygen species; PGE2, prostaglandin E. Figure reproduced with permission from Oncogene[39].

1.3 EV classes

Based on size and mechanism of biogenesis, EVs can be classified into at least three sub classes: exosomes (Exos), shed microvesicles (sMVs), and shed midbody remnants (sMB-Rs)[70] (**Fig. 1.7**).

Exosomes (Exos) are small EVs with the size range of 30-150 nm[57, 68]. The biogenesis of Exos starts with endocytosis when cells internalize material in the extracellular fluid by invagination and inward pinching of the plasma membrane to form internal vesicles - early endosomes. Early endosomes develop into late endosomes (referred to as multivesicular bodies, MVBs) and there is inward budding of MVBs to form intraluminal vesicles (ILVs) mediated by ESCRT-I components (TSG101, VPS28, VPS37A-D, MVB12A, MVB12B, UBAP1), –II components (EAP20, EAP30, EAP45) and -III components (CHMP1A, CHMP1B, CHMP2A, CHMP2B, CHMP3, CHMP4A-C, CHMP5-7, IST1) and ESCRT-associated protein (ALIX)[71]. The multivesicular bodies (MVBs) have two possible fates - degradation by fusing with lysosome, or trafficking to and fusing with the plasma membrane leading to subsequent release of ILVs into the extracellular space as Exos (**Fig. 1.7**)[64]. Silencing of ESCRT-0 (STAM1 and HRS) and ESCRT-I (TSG101) decrease the release of exosomal markers (MHC II and CD63)[72]. Interestingly, CD9, CD63, and CD82 appear to be key mediators for ILV formation as well as neutral sphingomyelinases for exosome release in an ESCRT-independent mechanism[73-75]. There are several techniques for isolating Exos such as

differential ultracentrifugation/ density gradient-based separation[76, 77], flow cytometry, ultrafiltration[77], and asymmetrical-flow field-flow fractionation (A4F), for example [78] – see **Box 1** in Xu et al, **Suwakulsiri, W.**, and Simpson, R, **Chapter 1**, Section 1.4[79]. Tumour-derived Exos play an important role in cancer progression by transferring oncogenic substances such as proteins, mRNA and miRNA to alter the phenotype of recipient cells[72, 80] – see details in **Table 1** in Xu et al, **Suwakulsiri, W.**, and Simpson, R, **Chapter 1**, Section 1.4[79].

Shed microvesicles (sMVs) (also referred to as microparticle or ectosomes) are larger in diameter than Exos with size range of 50-1,500 nm[57, 68]. sMVs are formed by direct budding (blebbing) from the plasma membrane[57] (**Fig. 1.7**). While the mechanism of sMV biogenesis is largely unknown it is thought that translocation of phosphatidylserine to the outer membrane leaflet initiates sMV formation. ADP-ribosylation factor 6 (ARF6) then promotes the activation of phospholipase D (PLD) associated with extracellular signal-regulated kinase (ERK). Then myosin light-chain kinase (MLCK) phosphorylated by ERK promotes release of sMVs (**Fig. 1.8**)[81]. Another proposed biogenesis mechanism for sMVs involves a membrane repair mechanism initiated by damages such as mechanical stress, toxins and ultraviolet radiation[82, 83]. Such damage induce calcium influx that recruits annexin A7 and ESCRT components. Then the damaged membrane portion is shed from the plasma membrane by blebbing[84]. Like Exos, sMVs contain genetic material (miR-1227[85]) and proteins (WNT1/ WNT3A[86], WNT5A[87],

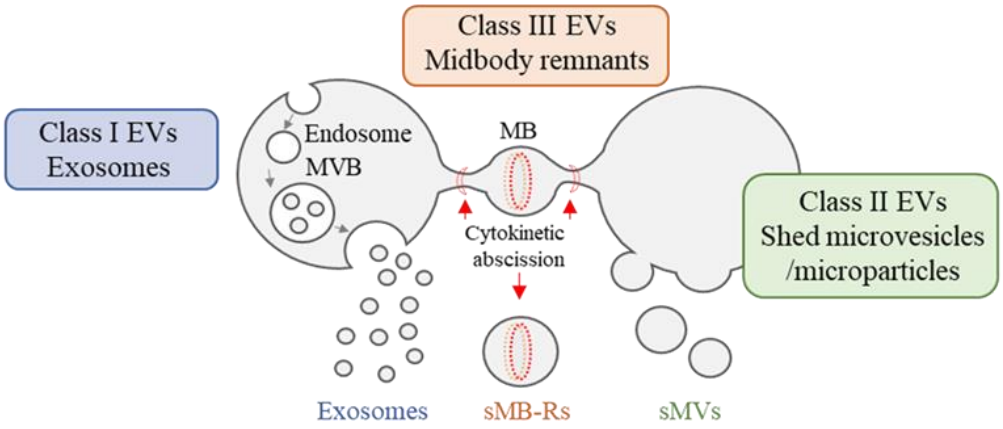
RAB22A[88], ABCB1[89]) which cause phenotypic changes in recipient cells upon uptake[81, 90] (**Table 1.1**). Tumour-derived sMVs have an effect on several stages of tumour progression such as immune evasion, extracellular matrix degradation, angiogenesis, and metastasis[90] by delivering the soluble protein, chemokine/EGF receptors, functional transmembrane proteins[91-94]. Moreover, lipid sphingomyelin from sMVs has been shown to promote angiogenesis by inducing endothelial-cell migration[95]. However, profiling of molecular compositions such as DNAs, RNAs, proteins, and lipids within sMVs is poorly understood.

Midbody remnants (sMB-Rs). During my PhD studies my laboratory published a third class of EV[70] referred to as shed midbody remnants (sMB-Rs) – this manuscript is attached in **Appendix Article 1**; our discovery of sMB-Rs was corroborated by[96] sMB-Rs are large vesicles with a range in diameter size (~1 μm)[97] that form during the final stage of cytokinesis. During cleavage furrow, two daughter cells are connected by an intercellular bridge (**Fig. 1.7**) containing a midbody (MB) and that the intercellular bridge can be cleaved, either on both sides (symmetric abscission) or one side (asymmetric abscission) of the MB[98]. The core structure of MBs consist of centraspindlin complex (KIF4A, KIF23/ MKLP1, RACGAP1) associated with cross-linker protein (PRC1) and kinases (AURB, PLK1) that stabilize the MB structure[99]. Hu *et al*., have characterized KIF4, KIF23, PRC1 as required structural proteins in MBs, anillin and RHOA for MB maturation and PLK1 for cytokinesis and MB assembly[100]. It has been proposed that CEP55 localizes to MBs and

then recruits TSG101, ALIX and other ESCRT machinery subunits (VPS4, CHMP4C)[71] for abscission and release of the MB as a midbody remnant (shed midbody-remnant (sMB-R), also referred to as extracellular post-abscission midbody) into the extracellular space[101, 102]. Proteomic analysis on MBs from Addi et al., Skop et al. revealed consistent profiles that centrasplindlin components, mitochondrial proteins, RBPs, translation initiation factors, traffic transport proteins ESCRT components are enriched in MBs[103, 104]. In sMB-Rs, Petermann et al and Rai et al reported isolation and purification strategies involving a combination of differential centrifugation and density-based separation (sucrose or iodixanol). Petermann et al. report that sMB-Rs can be up taken by cells during interphase and further promote cell proliferation, anchorage-independent growth and cell survival via integrin and EGFR-dependent pathways[96]. Rai et al. demonstrated that sMB-Rs induced fibroblast invasion[70]. However, our understanding of the role of sMB-Rs in the extracellular space is limited.

Figure 1.7 Schematic illustration of three distinct classes of extracellular vesicles – exosomes, shed microvesicles (sMVs) and shed midbody remnants (sMB-Rs). Exosomes (class I EVs) are of endosomal origin (formed by invagination of multivesicular bodies (MVB)), shed microvesicles (sMVs) (class II extracellular vesicles) are formed via direct outward blebbing of plasma membrane and midbody remnants (MB-Rs) (class III extracellular vesicles) are generated by symmetrical cytokinetic abscission of the interconnecting bridge between dividing daughter cells (post completion of cytokinesis). Current understanding of biophysical properties and stereotypic marker proteins for exosomes, sMVs (microparticles) and sMB-Rs are listed in the table. Centraspindlin complex proteins MKLP1 and RACGAP1 enable distinction of sMB-Rs from exosomes and sMVs. (reproduced with permission from Rai, A., et al. **Suwakulsiri, W** and Simpson, R. Nature Communications Biology 2021[70]).

.



	Exosomes	sMB-Rs	sMVs
Biogenesis mechanism	Endosomal pathway	Cytokinesis (cytokinetic abscission)	Plasma membrane blebbing
Stereotypic protein marker	CD63, CD81, ALIX, TSG101	MKLP1 RACGAP1	???
Particle diameter range	30-200 nm	200-600 nm	100-500 nm
Buoyant density	1.08-1.14 g/ml	1.22-1.30 g/ml	1.08-1.14 g/ml

Figure 1.8. Biogenesis of shed microvesicles. Microvesicle arises through outward budding and fission of plasma membrane and is the result of dynamic interplay between phospholipid redistribution and cytoskeletal protein contraction. Membrane budding/vesicle formation is induced by translocation of phosphatidylserine to the outer-membrane leaflet through the activity of amino phospholipid translocases. To enable microvesicle budding, ADP-ribosylation factor 6 (ARF6) initiates a signaling cascade that starts with the activation of phospholipase D (PLD), which recruits the extracellular signal-regulated kinase (ERK) to the plasma membrane. ERK phosphorylates and activates myosin light-chain kinase (MLCK). Phosphorylation and activation of the myosin light chain by MLCK triggers the release of microvesicles. Figure reproduced with permission from Journal of Neuro-Oncology[81] (Open source).

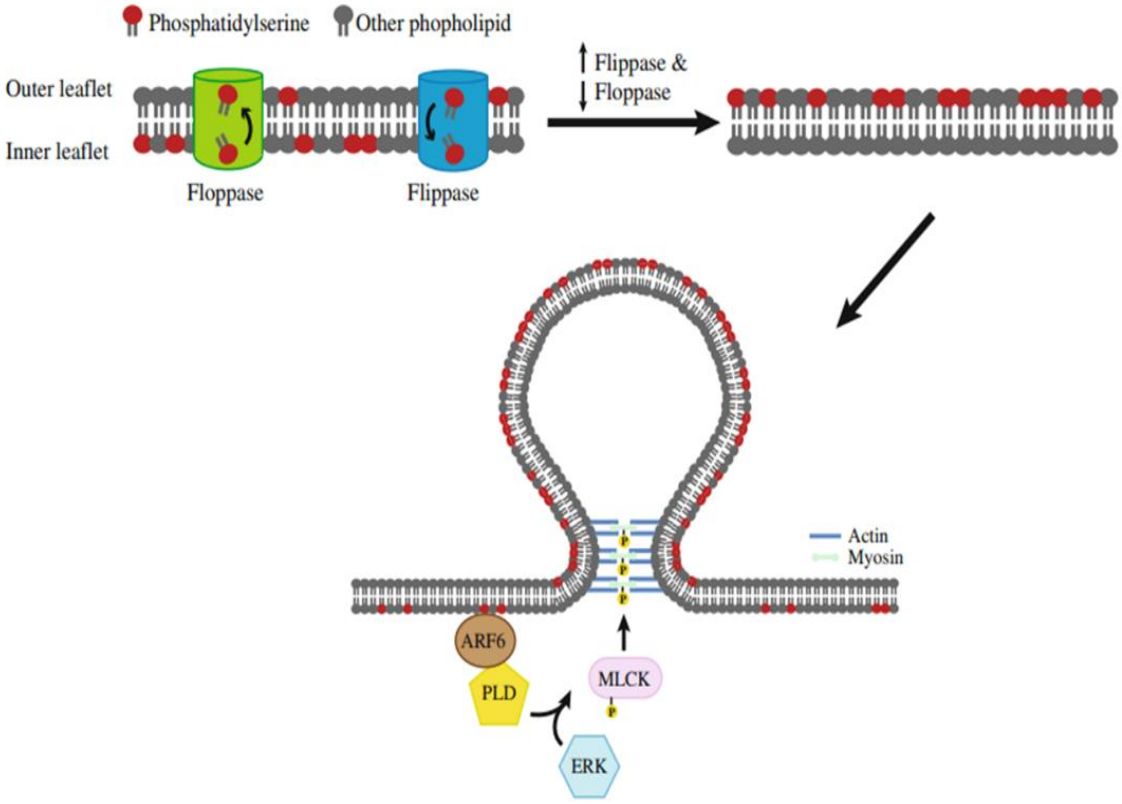

Phosphatidylserine
Other phopholipid
Outer leaflet
Inner leaflet
Floppase
Flippase
Flippase &
Floppase
Actin
Myosin
ARF6
PLD
MLCK
ERK

Table 1.1 Isolation methods and functional characteristics of sMVs and sMB-Rs

EV classes	Isolation techniques	Details of isolation technique	Donor cells/ biofluids	Recipient cells	Cargo	Key findings*
sMVs	Centrifugation (10,000-24,000 x *g*)	10,000 x *g*, 30 min	Human melanoma cell (LOX)	-	MT1-MMP	Protein cargo appears to be selectively sorted into sMVs, and adherence to the extracellular matrix (ECM) is facilitated by microvesicle-associated integrin receptors[99].
		10,000 x *g*, 30 min and filtration	RWPE-2 prostate	CAFs	miR-1227	sMVs derived from RWPE-2 cells enhance migration of cancer-associated fibroblasts (CAFs), an effect That is increased by miR-1227, a miRNA abundant in sMVs produced by RWPE - 2 cells[81].
		10,000 x *g*, 30 min and sucrose gradient separation (0.8 mol/L to 0.25 mol/L)	Blood samples	-	Integrin chains beta 1 and beta3, GPI alpha/beta, and P-selectin	sMVs were shown to bind to factor X and prothrombin [105]
		10,000 x *g*, 30 min	SKBR3B cells	-	-	sMVs showed enrichment of PARP and CYC compared to exosomes[106]
		14,000 x *g*, 35 min	Breast cancer line (SK-BR-3)	Macrophage	WNT5A	sMVs induce production of Wnt 5a in macrophage (MΦ). Active MΦ releases Wnt5 bearing sMVs to promote recipient cell invasion[83]
		10,000 x *g*, 70 min	MDA-MB-231-Rab22A knockout and scramble cell	MCF7 and MDA-MB-231	RAB22A	Incubation of breast cancer cells with sMVs shed by hypoxic breast cancer cells promotes focal adhesion formation, invasion, and metastasis[84].

		10,000 x *g*, 30 min	Human melanoma cell (LOX)		MT1-MMP	VAMP3-shRNA depletes sMVs of MT1-MMP and decreases cell invasiveness when embedded in cross-linked collagen matrices[100].
		14,000 x *g*, 35 min	Breast cancer line (SK-BR-3)	MCF-7	EMMPRIN	EMMPRIN-containing sMVs mediate invasion by activation of the p38/MAPK signaling pathway[101]
		20,000 x *g*, 60 min	PC-3 and LNCaP	PC-3 and LNCaP		sMVs mediated drug delivery enhances the cytotoxicity of Paclitaxel in autologous prostate cancer cells[102]
		20,000 x *g*, 30 min	HEK293FT	HEK293FT and murine 4TI breast cancer	Plasmid DNA (LUC)	De novo reporter protein expression in recipient cells presented only from plasmid DNA (pDNA) after delivery via sMVs[103].
		24,000 x *g*, 60 min	MDR subline of human acute lymphoblastic leukemia (CCRF-CEM VLB_{100})	Drug-sensitive parental cells (CCRF-CEM)	ABCB1 (P-gp)	sMVs derived from drug-resistant cancer cells incorporate cell surface P-gp from their donor cells, effectively bind to drug sensitive recipient cells and transfer functional P-gp[85]
		18,000 x *g*, 30 min	Acute lymphoblastic leukemia cell line CCRF-CEM derivative and MCF-7	Acute lymphoblastic leukemia cell line CCRF-CEM and derivatives	ABCB1 (P-gp) and MRP1	Reprogramming of MRP1 profile in recipient cells via transfer of P-gp from sMVs[104]
	Filtration (>0.22 μm)	sMVs retained on the filter (0.22 μm)	MEFs expressing onco-Dbl MDAMB231	Fibroblasts (NIH-3T3)	Onco-Dbl	Oncogenic transformation gives rise to sMVs, which uniquely contain a signaling protein kinase that confers the transformed phenotype[105].

		sMVs retained on the filter (0.22 µm)	Equine mammary stem/progenitor cell (eMaSC)	Canine mammary stem/progenitor cell (cMaSC)	WNT3A and WNT1	eMaSC-sMVs mediate activation of the Wnt/β-catenin signaling pathway in cMaSCs and significantly improves the ability of cMaSCs to grow as mammospheres[82].
		sMVs retained on the filter (0.22 µm)	Embryonic stem cell	Trophoblast	Laminin and fibronectin	Laminin and fibronectin within sMVs interact with integrins on trophoblast surface, resulting in activation of JNK and FAK (trophoblast migration)[106]
		sMVs retained on the filter (0.22 µm)	MDA-MB-231 carcinoma cells	MCF10a cell		Tumour-derived sMVs prime the early tumour microenvironment by reorganizing ECM, leading to increase cellular traction force, actin-myosin contractility, and FAK activity[107]
		sMVs retained on the filter (0.22 µm)	LIM1863	NIH3T3 fibroblast		SCUF-Exos and SCUF-sMVs isolated from LIM1863 colon cancer cells induce invasion of recipient NIH3T3 cells. Interestingly, the SCUF-sMVs promote invasion to a significantly greater extent (3-fold) than SCUF-Exos[108].
sMB-Rs	Centrifugation	10,000 x *g*	HeLa cells	HeLa cells	MKLP (KIF23)	Post-abscission MBs (sMB-Rs) induced cell proliferation[96]
		10,000 x *g*	SW620 cells	Fibroblasts	MKLP (KIF23), RACGAP1	sMB-Rs promoted fibroblast invasion[70]

*=references

1.4 Role of EVs in cancer

This section is addressed in Xu, R., Rai, A., Chen, M., **Suwakulsiri, W.**, Greening, DW. and Simpson, RJ. '*Extracellular vesicles in cancer—implications for future improvements in cancer care'*. Nature Reviews Clinical Oncology. 15, 617-638 (2018). PMID: 29795272 and reproduced here with permission from the senior author. The article is in **Appendix Article 2.** My contribution to the NRCO review involved scoping and performing the literature review and discussion of Tables 1, 2 and 3).

1.5 EV composition

EVs harbor diverse molecules such as proteins, RNA species (mRNA, miRNA, ncRNA) lipids, metabolites, DNA fragments[68, 79, 107-109]. In the case of protein cargo, EVs typically contain >2000 proteins, and this number will most likely increase as LC-MS/MS technology continues to improve[110]. Of these, an evolutionary-conserved set of proteins have been found to be common to each EV class reflecting their involvement in defining EV shape, rigidity and membrane curvature. In addition to these common proteins, EVs also harbor a small but specific subset of proteins ('protein signatures') reflecting their originating cell type and conditions[59, 111]. In addition to 'protein signatures' EVs contain distinct miRNA signatures[112, 113]. For example, my laboratory reported that LIM1863 colorectal cancer (CRC) cells secrete three distinct vesicle subtypes- two subpopulations of exosomes (apical EpCAM-Exos and basolateral A33-Exos) and shed microvesicles (sMVs) – with distinct protein, transcriptome, and long noncoding signatures[57].

Recent studies highlight the role of EVs in intercellular communication in normal and pathophysiology[68, 79] and their cargo molecules as potential markers of disease [114, 115]. These findings have generated immense interest in the biomedicine field [115, 116] resulting in an exponential increase in molecular data pertaining to EVs from a multitude of cell lines and body fluids.

In 2009 my laboratory published the manually-curated ExoCarta database – a publicly-accessible compendium of exosomal proteins and

RNA[117, 118]. In 2012 the scope of ExoCarta was extended to include exosomal lipids in addition to proteins and RNA species (ExoCarta 2012[118]). Because of the exponential growth in EV compositional data the VesiclePedia database was developed – while this database of EV molecular data is still manually annotated it allows continuous community annotation[119] – a feature that encourages active involvement of EV researchers and to safeguard against future database inactivity. These databases that store EV molecular data allow large-scale bioinformatic analyses feasible and provide a platform for EV-based biomarker studies. More recently, EVpedia – a community web portal for EV research - was constructed[120]. Unlike ExoCarta and Vesiclepedia, EVpedia contains both prokaryote and eukaryote EV data (http://evpedia.info) and contain embedded bioinformatic tools that permit Gene Ontology enrichment analysis, network analysis of vesicular proteins and mRNAs, and set analysis of vesicular datasets by ortholog identification. Recently, Liu and colleagues constructed EVmiRNA – a web-interface database of miRNA profiling in Exos and sMVs[121]. EVmiRNA (http://bioinfo.life.hust.edu.Cn/ EV-miRNA) lists miRNA profiles of EVs from diverse biological sources, specifically expressed miRNAs in different EVs that might be helpful for biomarker identification, and the miRNA annotations include miRNA expression in EVs and TCGA cancer types and functional information (e.g., signaling pathway regulations etc).

One unanswered question in EV biology relates to the spatial distribution of vesicular molecules – i.e., one's ability to distinguish between

luminal and membrane-associated proteins/RNA species (versus integral and non-covalently bound on outer surface of vesicle). (A corollary to is the vexed question of how the many thousands of proteins and RNA species can be accommodated/ packaged within an exosome that is 50-150 nm in diameter upon known protein/ RNA hydrodynamic volume parameters. As a first step towards addressing this question researchers are actively exploring the surface topology of EVs. A knowledge of the surface molecules of EVs (i.e., the surfaceome) is important due to their functional significance (e.g., involvement in targeting/ recognition of recipient cells[122] and cellular uptake[79], immune recognition of EVs by the innate and adaptive immune systems[123]. Important insights into the EV surfaceome have been made recently using proteomic techniques such as mild proteolytic shaving of the vesicle surface[124], combining proteinase treatment and biotin tagging[125], proximity barcoding assay using antibody-DNA conjugates and next-generation sequencing[126]. Interestingly, the surfaceome studies of Xu and colleagues [124] identified several RNA binding proteins/ heterogenous ribonucleoproteins (RBPs/ hnRNPs) and their bound cognate RNA species exposed on the outer surface of colon cancer cell-derived EVs; further studies are required to fully explain the biological function of RNA-loaded RBPs/hnRNPs in recipient cells.

Because of their ability to transfer bioactive protein and RNA cargos and overcome biological barriers (e.g., blood-brain barrier, immune surveillance) EVs are increasingly being investigated as potential therapeutic agents and liquid biopsy-based cancer detection [127]. For reviews of recent

advances in therapeutic applications of EVs, and the role of EVs in facilitating personalized medicine and pharmacogenomics in oncology[115, 116].

1.6 Isogenic human CRC model used in this thesis

Over the past 20 years my laboratory has been addressing the question of EV heterogeneity and functionality. We have reported the isolation, purification, characterization and proteomic profiling (and more recently, miRNA and transcriptomic profiling) of exosomes secreted from several CRC cell lines[57-59, 77, 111, 112]. To extend these studies to include a comparative proteomic and transcriptomic profiling of three EV classes (exosomes, shed microvesicles and shed midbody remnants), EVs from SW480 and SW620 cell lines were used in this thesis. This model is suitable for study of CRC progression because SW480 and SW620 cells were derived from the same CRC patient (50-year-old male) presenting at different CRC stages[128]. SW480 cells were derived from a Dukes' stage B colon carcinoma (primary tumour) six months before SW620 were collected from a Dukes' stage C colon carcinoma (lymph-node CRC metastasis)[129]. SW480 cells display an epithelial-like cell shape while SW620 cells appear to look rounded or fibroblast-like. SW620 cells exhibited higher proliferation index and invasiveness than SW480 cells[128]. Studies on tumour in mice, show that SW620 cells are tumorigenic metastasizing to the liver[128]. At the molecular genetic level, SW480 and SW620 cells are characterized as being microsatellite stable (MSS) containing mutations at KRAS (G12V) and TP53 (R273H and P309S) - however, only SW620 cells are CIMP-positive[130].

Comparative proteomic analysis on SW480 and SW620 cells revealed that during CRC progression, proteins related to loss of focal adhesion (CacyBP) and actin depolymerization (MARCKS, Villin and ATP5J) were up regulated in SW620 compared to SW480 cells[131]. Furthermore, comparative transcriptomic analysis of SW480 and SW620 cells showed a dysregulated profile of circadian genes (BMAL1, CRY1, PER3, DBP) with specific spliceosome components/ splicing factors (PRCC, RBM8A, TARDBP, SF1) in SW620 cells compared to SW480 cells that could affect alternative splicing alterations during CRC development[132].

Exos derived from SW480 and SW620 cells have been shown to contain cancer progression-protein signatures[59, 133, 134]. For example, in SW620-Exos, MET, S100A8, S100A9, TNC, VLC, MARCKS, LCK, VIM are enriched compared to SW480-Exos, while EGFR, CD44, ICAM1, integrins are enriched in SW480-Exos compared to SW620-Exos[59, 134]. Furthermore, the surface proteins on Exos derived from SW480 and SW620 cells can be changeable under hypoxic condition[135]. Nakurte *et al*., identified six proteins (LAMP1, ITGA6, COPB1, TGIFI, SLC3A2, ATP1A1) downregulated in Exos derived from SW480 cells and SW620 cells in hypoxia compared to non-hypoxic conditions[135].

1.7 Thesis aims

Colorectal cancer is a leading cause of cancer death in the Western world[2, 136]. By the age of 70, at least 50% of people in Western populations develop a colorectal tumour, and for about 10% of these individuals, progression to malignancy ensues[137]. Like many cancers, CRC can be viewed as the endpoint of a series of genetic and epigenetic alterations in a cell or group of cells[138-141]. CRC comprises a number of different diseases which often result from mutations to genes such as P53[142], RAS[143] or APC[144], but which can also contain a variety of other mutations to genes such as MLH1[145, 146] or B-RAF[147] that are capable of modulating the phenotype and outcome of the tumour[148]. Several forms of CRC are inherited[149, 150] but the majority arises spontaneously due to somatic mutations[151]. Interestingly, an inherited mutation which predisposes to one type of CRC - namely, familial adenomatous polyposis coli (FAP) - is also associated with more than 80% of spontaneous colorectal cancers[152]. CRC presents in several histological forms and the genetic basis of both the initiation and progression of the different forms is gradually being uncovered[148]. For example, it is clear that hereditary non-polyposis colon cancer (HNPCC) is driven by mutations to a set of mismatch-repair genes[153, 154] and that similar mutations occur in spontaneous colorectal cancers, which lead to genetic instabilities (e.g., micro-satellite instabilities)[155]. In order to improve outcomes for CRC patients, there is a pressing need to identify biomarkers for the early detection (diagnostic markers), prognosis (prognostic indicators), tumour responses (predictive

markers) and disease recurrence (monitoring markers). It has been estimated that 90% of all cancer could be cured if were detected early enough – before invasion or metastasis[156, 157]. Early detection is the single most important factor influencing outcome of CRC patients. If CRC can be identified while still localized, it is more amenable to treatment, thereby preventing not only mortality, but also morbidity and overall associated health costs[158, 159]

Besides cellular genetic instability, it is evident that cancers develop within a complex tissue environment (so called tumor microenvironment) between cancer cells and stromal cells via intercellular communication that supports cancer growth, dissemination and metastasis[39]. Intercellular communication includes direct cell interaction, release of soluble proteins (e.g., growth factors, chemokines) and extracellular vesicles (EVs)[160]. EVs are membrane encapsulated structures released by most cell types that contain bioactive cargo such as DNAs, protein-coding RNAs, non-protein-coding RNAs, proteins and lipids that have been shown to be potential cancer diagnostic candidates in EV-based liquid biopsies[79]. EVs comprise three major classes: Exos, sMVs (also referred as microparticles, ectosomes), and sMB-Rs[70] that differ in size, biochemical properties, and mechanism of biogenesis[161].

My laboratory has been focusing on isolation, purification and characterization of EV subtypes derived from colorectal cancer cell lines such as LIM1215[162], LIM1863[111, 163] and SW480/SW620[59, 164] and their role in colorectal cancer progression[165, 166]. We have established methods used to

isolate and purify Exos, sMVs, and sMB-Rs using a combination of differential ultracentrifugation and isopycnic iodixanol density centrifugation[77, 16370]. EV protein profiles have been identified for many CRC cell line derived Exos, however, there is very little information about protein and RNA profiles for human CRC cell line derived sMVs and sMB-Rs.

The overarching aim of this thesis is to systematically analyze the protein and RNA cargo in EVs (Exos, sMVs, and sMB-Rs) released from the isogenic human CRC cell lines SW480 (derived from primary colorectal cancer tumor) and SW620 (derived from lymph node-metastatic colorectal cancer tumor from the same patient) to achieve understanding of the potential roles of different EVs during cancer progression. This information will allow insights into the biochemistry and potential functions of Exos, sMVs and sMB-Rs in colorectal cancer progression and also provide potential diagnostic markers for clinical purposes.

There are three specific aims in my thesis:

Aim 1: To gain biophysical, biochemical and functional insights of sMVs released from the colorectal cell lines SW480 and SW620.

Specifically, I aim to -

i. *Isolate sMVs from SW480 and SW620 cells using differential centrifugation and density-gradient based separation,*

ii. *Characterize purified SW480 and SW620-drived sMVs using transmission electron microscopy, nanoparticle tracking analysis and Western blot analysis,*

iii. *Identify the proteome of SW480 and SW620 sMVs using GeL-LC MS/MS using and perform bioinformatics analysis, and*

iv. *Assess the invasion capability of NIH3T3 fibroblasts treated SW480 and SW620 sMVs using the Transwell Matrigel™ migration-invasion assay.*

This aim is addressed in **Chapter 2** and incorporates –

Suwakulsiri, W., Rai, A., Xu, R., Chen, M., Greening, D. and Simpson, R. Proteomic profiling reveals key cancer progression modulators in shed microvesicles released from isogenic human primary and metastatic colorectal cancer cell lines. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*. **12**, 140171. (2019). PMID: 30502510.

Aim 2: To analyze and compare the protein components of Exos, sMVs (from Aim 1) and sMB-Rs released from the isogenic colorectal cell lines SW480 and SW620

Specifically, I aim to -

i. *Isolate Exos and sMB-Rs from SW480 and SW620 cells using differential centrifugation and density-gradient based separation,*

ii. *Characterize Exos and sMB-Rs using transmission electron microscopy, nanoparticle tracking analysis and Western blot analysis,*

iii. *Identify the proteome of SW480 and SW620-derived Exos and sMB-Rs using GeL-LC MS/MS and*

iv. *Perform a detailed comparative proteomic analysis of purified Exo, sMB-Rs and sMVs (prepared in Aim 1) secreted from SW480 and SW620.*

Aim 3: To analyze and compare the RNA components in Exos, sMVs and sMB-Rs released from the isogenic colorectal cell lines SW480 and SW620

Specifically, I aim to -

i. *Extract total RNA from SW480- and SW620-derived cells, Exos, sMVs and sMB-Rs purified using the protocols described in Aims 1 and 2,*

ii. *Perform next-generation RNA sequencing analysis of Exos, sMVs and sMB-Rs using the Illumina HiSeq 4000 platform, and*

iii. *Undertake a detailed bioinformatic analysis of transcriptome of SW480- and SW620-derived Exos, sMVs and sMB-Rs.*

Chapter 2

Biophysical, biochemical and functional insights of sMVs released from the isogenic colorectal cell lines SW480 and SW620

2.1 Overview

Extracellular vesicles (EVs) mediate intercellular communication by transferring bioactive cargo such as protein, RNA species, lipids, metabolites, DNA to recipient cells[79] (**Chapter 1**, Section 1.4). EVs are heterogenous and can be categorized into different classes such as exosomes (Exos) and shed microvesicles (sMVs, also referred to as microparticles and microvesicles) based on size and biogenesis[68]. While Exos are relatively homogenous with a size distribution over the range 30 to 150 nm diameter, sMVs are more heterogeneous in size (50 to ~1500 nm diameter)[79]. Exos are formed by inward invagination of late endosomes/multivesicular bodies (~150-200 nm in diameter) mediated by ESCRT machinery (Endosomal Sorting Complexes Required for Transport) such as ALIX and TSG101. Both ALIX and TSG101 along with CD9, CD81, and CD63 are widely used for identification and characterization of Exos[79, 107]. sMVs, on the other hand, are formed by outward blebbing of the cell membrane[79]. To date, there are no internationally accepted protein markers for sMVs.

My laboratory has been focusing on the isolation, purification and characterization of EV classes/ subtypes derived from colorectal cancer (CRC) cell lines such as LIM1215[162], LIM1863[111, 163] and SW480/SW620[59]. Previously, we reported a detailed proteomic analysis of Exos derived from the isogenic human CRC SW480 and SW620 cells[59]. This analysis showed SW480- and SW620-Exos contain distinct cancer progression-related proteins

such as EGFR, CD44, Integrins (ITGA2, ITGA5. ITGA7, ITGB4) in SW480-Exos, and MET, S100A8, S100A9, JAG1, SRC, TNIK in SW620-Exos[59]. These studies led to additional studies investigating the role of SW480-/ SW620-Exos in reprogramming of the fibroblast proteome[165], as well as their role in fibroblast invasion[166]. However, the biochemical characterization and functional role of SW480-/ SW620-sMVs is poorly understood. Hence, in this chapter, I aim to understand protein compositions and possible functional roles of sMVs derived from SW480 and SW620 cells using a combination of differential ultracentrifugation and iodixanol density-based separation technique, mass spectrometry, bioinformatics and biological assays.

For large-scale production of SW480- /SW620-sMVs, SW480 and SW620 cells were continuously cultured in CELLine AD-1000 Bioreactor classic flasks with RPMI-1640 medium supplemented with 0.6% Insulin-Transferrin-Selenium (ITS) and 1% P/S at 37°C and 5% CO_2 atmosphere. sMVs were purified from 180 mL culture media (CM) acquired over 18 days (30 mL CM harvested each day, 6 days for each biological replicate). sMVs were isolated from culture media by a combination of differential ultracentrifugation (500 x *g* for 5 min, 2,000 x *g* for 10 min and 10,000 x *g* for 30 min, respectively). The 10K pellet was resuspended in 500 µL PBS and subjected to isopycnic iodixanol density (100,000 x *g* for 18 h) to purify sMVs. Purified sMVs (at buoyant density 1.10 g/mL) were characterized by Western blot analysis, transmission electron microscopy (TEM) and nanoparticle tracking analysis (NTA). Proteins from sMVs were separated by short-range

SDS-PAGE, the gel bands excised into equal fractions, and subjected to *in gel* proteolytic digestion using trypsin[167]. Generated tryptic peptides were separated using a nanoflow UPLC instrument coupled to an MS-MS instrument. The mass spectrometer was operated in data-dependent mode where the top 10 most abundant precursor ions were selected for MS/MS fragmentation[167]. Ion target values were set for MS/MS scans. Peptides were identified using MaxQuant (v1.6.0.1)[168] with Andromeda (v1.5.6) against UniProt human sequence database (May-2017). Peptides were searched with minimum peptide length 6, and a false discovery rate 1% for protein identification. Label-free quantification (LFQ) intensity values were normalized by protein length. Fold change and p-value were calculated using edgeR software package[169]. Differentially expressed proteins were identified using the criteria of LFQ fold change ratio <-2 (enriched in SW480), >2 (enriched in SW620) with p-value < 0.05. Gene Ontology (GO) enrichment analysis and protein-protein interaction was acquired by DAVID Bioinformatics Resources (v6.8)[170] and STRING (v10.5)[171], respectively.

sMVs exhibited distinct biochemical and biophysical properties when compared to Exos. Unlike Exos, Western blot analysis of sMVs (secreted from both SW480 and SW620 cells) showed that they are ALIX$^-$, TSG101$^-$, CD9$^-$, and CD63$^-$. TEM revealed sMVs display round-like membranous vesicle structures 100-500 nm in size, and NTA showed heterogenous particle mean diameters of 350±28.4 nm for SW480-sMVs and 337.1±9.3 nm for SW620-sMVs. Label-free quantitative mass spectrometry identified 1295 and 1300

proteins in SW480- and SW620-sMVs, respectively. Inspection of these data revealed that 834 proteins (≥-2 and ≤2 LFQ ratio) are common to both SW480- and SW620-sMVs. GO analysis revealed that these 834 common proteins are related to the following processes – 'cell adhesion/adhesion molecules' (PKP2/3, CLDND1, OCLN, CTN families), 'signal transduction' (KRAS, NRAS, RHOA, MAPK1/2K1), 'RNA binding proteins' (e.g., ribonuclear proteins HNRNPK/Q/E2, YBX-1), 'translation-associated proteins' (ribosomal proteins, translation initiation factors, aminoacyl tRNA ligases), and 'transport vesicle' (VPS29/35, SNAP23). Label-free quantitative mass spectrometry also identified 307 significantly enriched proteins in SW480-sMVs (<-2 LFQ ratio, $p < 0.05$), and 291 significantly enriched proteins in SW620-sMVs. Prominent amongst the 307 proteins enriched in SW480-sMVs are regulatory components associated with signal transduction/basement membrane (e.g., integrins ANXA1, CLDN7)[172-177], cell proliferation (EGFR[178]), pro-tumorigenic signaling (CD44[179]), angiogenesis, vascular remodeling (e.g., NOTCH1/2[180]) as well as EGFR/NOTCH signaling networks from STRING analysis[171]. In the case of the 291 proteins significantly enriched in SW620-sMVs most prominent were mediators of cell invasion (e.g., FGFR4[181, 182]), metastasis (e.g., MACC1[183]), cytoskeleton organisation/polarity (e.g., mTORC2, PRKCA[184, 185]), signal transduction (e.g., PAK1, CLDN1, RASA1, ARHGAP1[186-188]) as well as PRKCA, MACC1, FGFR4 and MTOR/MARCKS signaling networks by STRING analysis[171]. Western blot analysis confirmed the high enrichment of CLDN7, EGFR, CD44, ANAX1 in SW480-sMVs and CLDN1, PAK1 in SW620-

sMVs. A major finding of this chapter was that both SW480-/ SW620-sMVs contain pre-metastatic niche factors (ADAM10 and MIF[189]), cancer biomarkers (CEACAM1[190, 191] and ERZ[192, 193]. Like SW480-/ SW620-Exos, I found that sMV-treated fibroblasts also display cell invasive capability in the Transwell Matrigel™ invasion assay.

In summary, sMVs released by SW480 and SW620 cells are biophysically and molecularly distinct from exosomes. Common proteins identified in both SW480- and SW620-sMVs include proteins involved in cell adhesion, signal transduction, protein translation and ribonucleoproteins. A striking finding is that SW480- and SW620-sMVs contain distinct cancer progression proteins (EGFR, CD44, CLDN7 in SW480-sMVs and CLDN1, MACC1, MET in SW620-sMVs). Interestingly, sMVs derived from SW480 and SW620 cells induced fibroblast invasion, indicating a functional role of sMVs in intercellular communication. Collectively, this study provides, for the first time, molecular insights into sMVs and their possible role in CRC biology.

2.2 Future perspectives

A remaining question concerning EV subtypes (classes) relates to the quest of protein markers that would enable distinction of one EV subtype from another. Previously, ARF1/6 have been shown to be associated with microvesicle biogenesis/shedding[194, 195] and in this chapter I detected ARF1/6 in both SW480-/SW620-sMVs using MS/MS analysis. Unfortunately, time and

lack of a suitable reagent did not permit orthogonal confirmation of ARF1/6 as a stereotypic protein marker for sMVs using Western blot analysis. To investigate whether SW480- and SW620-sMVs exhibit different tumorigenic potential, I would like to perform a tumour xenograft experiment by subcutaneously injecting SW480- and SW620-sMVs into nude mice and monitor tumour growth[196]. In addition, loss function assay by small interfering RNA (siRNA)[197] on cancer-progression protein candidates (EGFR, FGFR4, MACC1) could be performed to understand biological roles of sMVs derived from cancer cell lines (time constraints and lack of resources did not permit me to conduct such studies during my PhD).

This chapter has been published in Biochimica et Biophysica Acta - Proteins and Proteomics, Volume 1867, Issue 12, 2019, 140171. *Proteomic profiling reveals key cancer progression modulators in shed microvesicles released from isogenic human primary and metastatic colorectal cancer cell lines* by **Wittaya Suwakulsiri**, Alin Rai, Rong Xu, Maoshan Chen, David W. Greening and Richard J. Simpson. PMID: 30502510. https://doi.org/10.1016/j.bbapap.2018.11.008. The article is inserted in this thesis.

BBA - Proteins and Proteomics 1867 (2019) 140171

Contents lists available at ScienceDirect

BBA - Proteins and Proteomics

journal homepage: www.elsevier.com/locate/bbapap

Proteomic profiling reveals key cancer progression modulators in shed microvesicles released from isogenic human primary and metastatic colorectal cancer cell lines

Wittaya Suwakulsiri, Alin Rai, Rong Xu, Maoshan Chen, David W. Greening, Richard J. Simpson*

Department of Biochemistry and Genetics, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, Victoria 3086, Australia

ARTICLE INFO

Keywords:
Shed microvesicles
Extracellular vesicles
Metastasis
Colon cancer
Proteomics
Exosomes

ABSTRACT

Extracellular vesicles comprise two main classes - exosomes and shed microvesicles (sMVs). Whilst much is known about exosome cargo content and functionality, sMVs are poorly understood. Here, we describe the large-scale purification of sMVs released from primary (SW480) and metastatic (SW620) human isogenic colorectal cancer (CRC) cell lines using a combination of differential ultracentrifugation and isopycnic iodixanol density centrifugation. The yield of SW480-sMVs and SW620-sMVs was 0.75 mg and 0.80 mg, respectively. Both SW480-/SW620-sMVs are heterogeneous in size (100–600 nm diameter) and exhibit identical buoyant densities (1.10 g/mL). In contrast to exosomes, sMVs are ALIX$^-$, TSG101$^-$, CD63$^-$ and CD9$^-$. Quantitative mass spectrometry identified 1295 and 1300 proteins in SW480-sMVs and SW620-sMVs, respectively. Gene Ontology enrichment analysis identified 'cell adhesion' (CDH1, OCLN, CTN families), 'signalling pathway' (KRAS, NRAS, MAPK1, MAP2K1), and 'translation/RNA related' processes (EIF, RPL, HNRNP families) in both sMV types. Strikingly, SW480- and SW620-sMVs exhibit distinct protein signatures - SW480-sMVs being enriched in ITGA/B, ANXA1, CLDN7, CD44 and EGFR/NOTCH signalling networks, while SW620-sMVs are enriched in PRKCA, MACC1, FGFR4 and MTOR/MARCKS signalling networks. Both SW480- and SW620-sMVs are taken up by NIH3T3 fibroblasts resulting in similar cell invasion capability. This study provides, for the first time, molecular insights into sMVs and CRC biology.

1. Introduction

Extracellular vesicles (EVs) are heterogenous populations of lipid bilayer-membrane vesicles derived from various cell types including cancer cells [1]. EVs have been identified as crucial mediators for intercellular communication by transferring their bioactive cargo such as DNA, RNA species, oncoproteins, and lipids to a multitude of recipient cells [2–4]. EVs can be detected in bodily fluids such as blood, bile, malignant effusions and urine. Accumulating evidence shows that cells release at least two main EV classes exosomes (Exos) and shed microvesicles (sMVs, also referred to as microparticles and microvesicles), and that each EV class contains subtypes [1]. Exosomes and sMVs differ in their size range and mechanism of biogenesis. Exosomes are relatively homogenous with respect size (30 to 150 nm diameter) and are formed by inward invagination of late endosomes/multivesicular bodies. On the other hand, sMVs are more heterogeneous in size (50 to ~2000 nm diameter) and originate from outward budding of the plasma membrane [1]. While the molecular cargo and functionality of exosomes has been studied extensively, our knowledge of sMVs is still in its infancy.

Previously, we reported a comparative proteome analysis of exosomes derived from the isogenic human colorectal cancer cell lines SW480 (from a primary colorectal cancer tumour) and SW620 (lymph node-metastatic colorectal cancer tumour) [5]. Here, we describe the large-scale purification of sMVs from SW480 and SW620 cell culture medium using a combination of differential ultracentrifugation and isopycnic iodixanol density centrifugation. Label-free quantitative mass spectrometry [6] was used to compare the protein profiles of SW480- and SW620-derived sMVs. Our findings reveal that SW480-sMVs are enriched in ITGA/B, ANXA1, CLDN7, CD44 as well as the NOTCH and EGFR signalling networks, whereas SW620-sMVs are enriched in PRKCA, MACC1, and FGFR4 as well as the MTOR and MARCKS signalling networks. Because SW480-sMVs and SW620-sMVs display distinct protein profiles and signalling networks, and differ from their

* Corresponding author at: La Trobe Institute for Molecular Science (LIMS), Room 113, Physical Sciences Building 4, La Trobe University, Bundoora, Victoria 3086, Australia.
E-mail address: Richard.Simpson@latrobe.edu.au (R.J. Simpson).

https://doi.org/10.1016/j.bbapap.2018.11.008
Received 3 August 2018; Received in revised form 5 November 2018; Accepted 23 November 2018
Available online 29 November 2018

exosomal counterparts, our study suggests an important role of sMVs in cancer progression.

2. Materials and methods

2.1. Materials

SW480 cells were from Ludwig Institute for Cancer Research Ltd. (Parkville Branch, Melbourne) and SW620 cells were from Dr. E. Vincan (Peter MacCallum Cancer Centre, Australia). All media and supplements were from Life Technologies (NY, USA). OptiPrep™ was from Axis-Shield PoC (Norway). CELLine AD-1000 Bioreactor classic flasks were from Integra Biosciences. Mouse anti-Alix, anti-CD44, rabbit anti-MET, anti-GAPDH were from Cell Signalling, Sigma-Aldrich (MA, USA), Mouse, anti-EGFR, anti-CD9, anti-CD63, rabbit anti-MET, anti-PAK1, anti-CLDN1, anti-ANXA1 and goat-anti-CLDN7 were from Santa Cruz Biotechnology (CA, USA) and Mouse anti-TSG101 was from BD Transduction Laboratories (NJ, USA).

2.2. Cell culture

Initially, SW480 and SW620 cells were cultured in 75 cm^3 flasks with RPMI-1640 medium supplemented with 10% Fetal Bovine Serum (FBS) and 1% penicillin/streptomycin (P/S) at 37 °C and 5% CO_2 atmosphere. For cell lysate isolation, SW480 and SW620 cells (3×10^5 cells) were cultured in a 15-cm cultured plate with RPMI-1640 medium supplemented with 10% FBS and 1% P/S at 37 °C and 5% CO_2 atmosphere for 48 h.

2.3. Cell proliferation assay

SW480 and SW620 cells (5×10^3 cells) were seeded in 96-well plate in 100 µL RPMI-1640 supplemented with 10% FBS and 1% P/S and then incubated for 24, 72, and 120 h, respectively. Then, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide reagent was added to a final concentration of 50 µg/mL and incubated at 37 °C for 4 h. The CM was removed and further replaced with 200 µL of acidified isopropanol. The 96-well plate was shaken for 30 min and cell proliferation was determined by measuring absorbance at 560/690 nm using SpectraMax M5^e (Molecular Devices, CA, USA).

2.4. Invasion assay

Growth factor reduced Matrigel™ matrix (Corning, NY, USA) (50 µL, 10 mg/mL) was mixed with 50 µL cell suspension which contains 20% FBS, 2% P/S, and 300 SW480 or SW620 cells. The mixture was added into 96-well plate and allowed to solidify at 37 °C and 5% CO_2 atmosphere for 2 h. The wells were then supplemented with 100 µL 10% FBS-containing medium was added into a 96-well plate and further incubated for 120 h. Cells were observed and imaged using Zeiss AxioObserver Z1 microscope.

2.5. Large-scale purification of SW480-sMVs and SW620-sMVs

SW480 and SW620 cells (3×10^7 cell) were transferred into the cultivation chamber of continuous-culture CELLine AD-1000 Bioreactor classic flasks (Integra Biosciences) and cultured with RPMI-1640 medium supplemented with 0.6% Insulin-Transferrin-Selenium (ITS) and 1% P/S at 37 °C and 5% CO_2 atmosphere as described [5]. Growth medium in the upper chamber of CELLine AD-1000 Bioreactor classic flasks was replaced every 5 days. Cells in the Cultivation chamber were allowed to attach for 48 h. Then cell suspension was harvested every 24 h and replaced with fresh RPMI-1640 with 0.6% ITS and 1% P/S. The culture medium (CM) of SW480 and SW620 cells was centrifuged at 500 x g for 5 min (4 °C) and 2000 x *g* for 10 min (4 °C) and stored at −20 °C. SW480- and SW620-derived CM (three independent collections or experimental replicates) were centrifuged at 10,000 x *g* for 30 min to pellet crude sMVs. Both sMV pellets were resuspended with 600 µL of filtered PBS (0.2 µm). The supernatant was further centrifuged at 100,000 x *g* for 1 h to pellet crude exosomes and resuspended with 600 µL of PBS. Approximately 550 µL of crude sMVs and exosome pellet was subjected on top of prepared OptiPrep™ density gradient, and separation performed as previously described [5,7]. Briefly, an OptiPrep™ (iodixanol solution) was prepared by adding 3 mL of 40, 20, 10, and 5% of iodixanol solution to 14 × 89 mm polyallomer tubes (Beckman Coulter). Dilutions were made in 0.25 M sucrose/1 M Tris (pH 7.5) solution. These tubes were centrifuged at 100,000 x *g* for 18 h at 4 °C. Twelve fractions (1 mL for each fraction) were removed from the top. Twelve fractions of sMVs and Exos were washed with PBS by centrifugation at 10,000 x *g* for 30 min (4 °C) for sMVs and 100,000 x *g* for 1 h (4 °C) for Exos. PBS was removed and pellets were resuspended with 150 µL PBS.

2.6. Protein quantification and Western blotting

The protein content of sMV, and Exo preparations was estimated by 1D-SDS-PAGE / SYPRO® Ruby protein staining densitometry as described [8,9]. Briefly, 5 µL sample aliquots were solubilised in SDS sample buffer (2% (*w/v*)) sodium dodecyl sulphate, 125 mM Tris-HCl, pH 6.8, 12.5% (*v*/v) glycerol, 0.02% (w/v) bromophenol blue) with 100 mM Dithiothreitol (DTT) and loaded into 1 mm, 10-well NuPAGE™ 4–12% (w/v) Bis-Tris Precast gels (Life Technologies). Electrophoresis was performed at 150 V for 45 min in NuPAGE™ 1× MES running buffer (Life Technologies) using an XCell Surelock™ gel tank (Life Technologies). After electrophoresis, gels were removed from the tank and fixed in 50 mL fixing solution (40% (*v*/v) methanol, 10% (*v*/v) acetic acid in water) for 30 min on an orbital shaker and stained with SYPRO® Ruby (Life Technologies, NY, USA) for 30 min, followed by destaining twice in 50 mL of 10% (v/v) methanol with 6% (v/v) acetic acid in water for 1 h. Gels were imaged on a Typhoon 9410 variable mode imager (Molecular Dynamics, Sunnyvale, USA), using a green (532 nm) excitation laser and a 610BP30 emission filter at 100 µm resolution. Densitometry quantitation was performed using ImageQuant software (Molecular Dynamics) to determine protein concentration relative to a BenchMark™ Protein Ladder standard of known protein concentration (1.7 µg/µL) (Life Technologies).

For western blotting, after electrophoresis proteins were electrotransferred onto nitrocellulose membranes using the iBlot™ 2.0 Dry Blotting System (Life Technologies). The membranes were probed with primary antibodies according to manufacturer's instructions. The secondary antibodies (IRDye 800 goat anti-mouse IgG or IRDye 700 goat anti-rabbit IgG) were diluted (1:15,000) and the fluorescent signals were detected using the Odyssey Infrared Imaging System, v3.0 (Li-COR Biosciences, Nebraska USA).

2.7. Transmission electron microscopy (TEM)

sMVs and Exos samples (1 µg in 10 µL PBS) were applied to 400 mesh carbon-coated copper grids for 2 min. Excess material was removed by blotting and samples were negatively stained with 10 µL of a 2% uranyl acetate solution for 10 min (ProSciTech, Queensland, Australia). The grids were air dried and viewed using a JEOL JEM-2010 transmission electron microscope operated at 80 kV [5].

2.8. Nanoparticle tracking analysis (NTA)

EV diameter (size) and concentration was determined by NanoSight NS300 system (NanoSight technology, Malvern, UK) equipped with a blue laser (488 nm). Briefly, sMVs and Exos were diluted in PBS (~8×10^8 particles/ml) and loaded into a flow-cell top plate using a syringe pump. Three separate technical replicates (60 s/video) were recorded for each sample and analyzed by NTA software (Build 3.1.45)

[10].

2.9. GeLC-MS/MS

sMV and cell lysate samples (15 μg) were lysed in SDS sample buffer, and proteins separated by short-range SDS-PAGE (10 × 6 mm), and visualized by Imperial Protein Stain (Invitrogen) as described [11]. The samples were excised into equal fractions ($n = 2$), reduced with 2 mM tri(2-carboxyethyl)phosphine hydrochloride (Sigma-Aldrich, C4706) at 22 °C for 4 h on gentle rotation, alkylated by treatment with 25 mM iodoacetamide (Sigma-Aldrich) for 30 min, and digested with 1 μg bovine sequencing grade trypsin (Promega, V5111) at 37 °C for 18 h. Subsequently, peptides were purified and extracted using reverse-phase C18 StageTips (Sep-Park cartridges, Waters, MA) in 85% (*v*/v) acetonitrile (ACN) in 0.5% (v/v) formic acid (FA). Peptides were lyophilised and acidified with buffer containing 0.1% FA, 2% ACN.

Proteomic experiments were performed with MIAPE-compliance [12,13]. A nanoflow UPLC instrument (Ultimate 3000 RSLCnano, Thermo Fisher Scientific) was coupled on-line to an Q-Exactive HF Orbitrap mass spectrometer (Thermo Fisher Scientific) with a nanoelectrospray ion source (Thermo Fisher Scientific). Peptides were loaded (Acclaim PepMap100, 5 mm × 300 μm i.d., μ-Precolumn packed with 5 μm C18 beads, (Thermo Fisher Scientific) and separated (BioSphere C18 1.9 μm 120 Å, 360/75 μm × 400 mm, NanoSeparations) with a 120-min gradient from 2 to 100% (*v*/v) phase B (0.1% (v/v) FA in 80% (v/v) ACN) (2–100% 0.1% FA in acetonitrile (2–40% from 0 to 100 mins, 40–80% from 100 to 110 mins at a flow rate of 250 nL/min operated at 55 °C.

The mass spectrometer was operated in data-dependent mode where the top 10 most abundant precursor ions in the survey scan (350–1500 Th) were selected for MS/MS fragmentation. Survey scans were acquired at a resolution of 60,000 with MS/MS resolution of 15,000. Unassigned precursor ion charge states and singly charged species were rejected, and peptide match disabled. The isolation window was set to 1.4 Th and selected precursors fragmented by high-energy collision dissociation (HCD) with normalised collision energies of 25 with a maximum ion injection time of 110 msec. Ion target values were set to 3e6 and 1e5 for survey and MS/MS scans, respectively. Dynamic exclusion was activated for 30 s. Data was acquired using Xcalibur software v4.0 (Thermo Fisher Scientific).

2.10. Label-free mass spectrometry protein identification

Raw data were pre-processed as described [14] and processed using MaxQuant [15] (v1.6.0.1) with Andromeda (v1.5.6), using a Human-only (UniProt #133,798 entries) sequence database (May-2017). Data were searched as described [13] with a parent tolerance of 10 ppm, fragment tolerance of 0.5 Da and minimum peptide length 6, with false discovery rate 1% at the peptide and protein levels, tryptic digestion with up to two missed cleavages, cysteine carbamidomethylation as fixed modification, and methionine oxidation and protein N-terminal acetylation as variable modifications, and data analyzed with label-free quantitation (LFQ) [16].

2.11. Data analysis

LFQ intensity values were normalised by protein length and fold change ratios calculated. Contaminants, and reverse identification were excluded from further data analysis. Resulting *p*-values were adjusted by the Benjamini-Hochberg multi-test adjustment method for a high number of comparisons [17] and statistics performed as previously described [18]. Differentially expressed proteins were identified using the criteria: ratio < 2.0 (SW480), > 2.0 (SW620) with $p < 0.05$, with identifications in at least 2 biological sample replicates.

2.12. Data visualisation

Visualisation and Integrated Discovery Bioinformatics Resources 6.7 (DAVID) resources (https://david.ncifcrf.gov/) were utilised using recommended analytical parameters [19]. DAVID and UniProt (www.uniprot.org) database resources (biological process) were used for gene ontology enrichment and network analysis, respectively. Protein-protein interactions were analysed using the STRING database (Search Tool for the Retrieval of Interacting Genes/Proteins) version 10.5 (https://string-db.org/) [20]. Venn diagrams were created from a facility from Bioinformatics & Evolutionary genomics (http://bioinformatics.psb.ugent.be/webtools/Venn/). Volcano plots were generated using the "mavolcanoplot" function and run in MATLAB programming language (MATLAB R2018a, MathWorks, NSW, Australia, https://au.mathworks.com).

2.13. Labelling of exosomes and uptake assay

Purified sMVs (~200 μg) in 200 μL PBS were labelled with the fluorescent dye, DiO (Invitrogen) at 1 μM concentration for 15 min at room temperature, as previously described [21]. Fluorescently labelled sMVs were collected at 10000 x *g* (30 min) and subjected to isopycnic (iodixanol-density) ultracentrifugation. NIH3T3 fibroblasts were grown on glass cover-slips (70% confluency) in a 24-well plate. Cells were incubated with DiO-labelled sMVs (5 μg) at 37 °C for 2 h and then washed with PBS and subjected to imaging using fluorescent microscope Zeiss AxioObserver Z1 (Zeiss). Nuclei were stained with Hoechst stain (10 μg/mL, 30 min) prior to imaging.

2.14. Transwell-Matrigel™ invasion assay

Transwell-Matrigel™ invasion assays were performed using growth factor-reduced Matrigel™ matrix (Corning), as previously described [22]. The Transwell inserts with 8 μm pore size (Corning) were coated with Matrigel™ (100 μL of 1 mg/mL) and allowed to polymerize for 4 h (37 °C). NIH3T3 fibroblasts (5 x 10^4 cells) in DMEM (1% P/S) were stimulated with SW480- or SW620-sMVs (30 μg/mL) or PBS alone for 2 h (37 °C). and subsequently overlaid onto Matrigel™-coated inserts. The inserts were nested onto a 24-well plate companion plate (Corning) that contained DMEM (5% FBS, 1% P/S) further supplemented with either sMVs (30 μg/mL) or PBS alone. After incubation for 16 h at 37 °C, inserts were washed, cells were fixed (4% (*v*/v) formaldehyde, 5 min), and the nuclei were stained with Hoechst stain (10 μg/mL). Non-invading cells were removed using cotton swab. Fibroblasts that invaded were imaged using Zeiss AxioObserver Z1 microscope.

3. Results and discussion

3.1. SW480 and SW620 cell line characterisation

To confirm the growth and oncogenic characteristics of primary adenocarcinoma-derived SW480 cell line and metastatic tumour-derived SW620 cell line, functional assays were performed to assess cell proliferation, and invasive capabilities. In accordance with previous reports [5,23] our data showed that SW620 cells, when compared with SW480 cells, displayed fibroblast-like morphology and higher proliferative and invasive capacities (Supplemental Fig. S1A-C). In agreement with the phenotypic changes, SW620 cells were shown to have diminished expression of EGFR and CD44, and elevated expression of oncogenic MET (Supplementary Fig. S1D).

3.2. Isolation and characterisation of SW480- and SW620-derived sMVs

Large quantities sMVs from SW480 and SW620 cells were obtained using a continuous cell culture approach CELLine AD-1000 Bioreactor classic flasks to generate ~180 mL SW480 and SW620 CM, as described

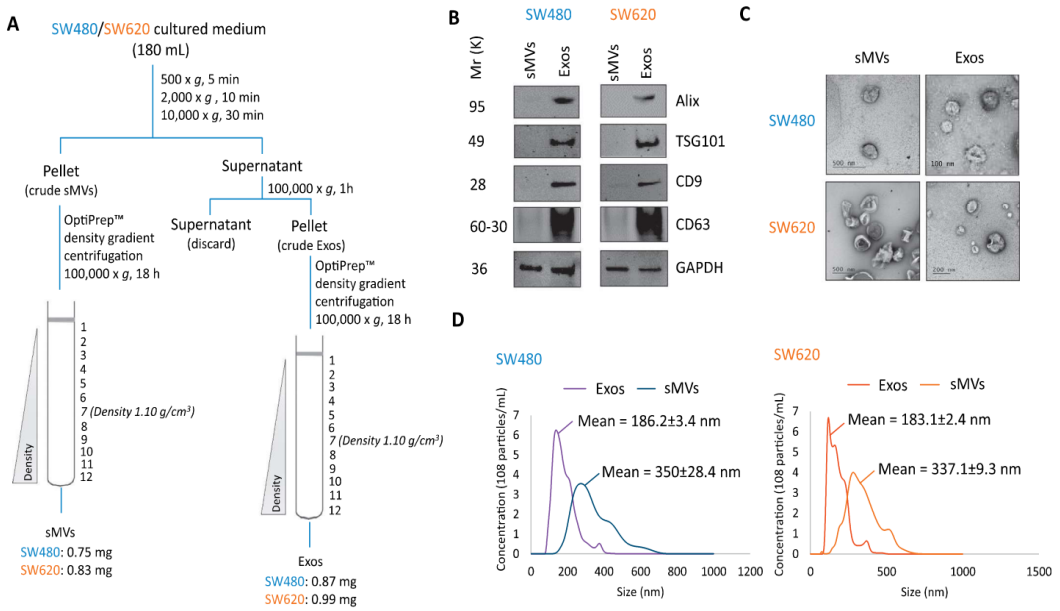


Fig. 1. Isolation, purification, and characterisation of EV subpopulations. (A) The experimental workflow for isolation and purification of shed microvesicle (sMV) and exosome (Exo) derived from SW480 and SW620 cell lines. SW480 and SW620 cells were grown in CELLine AD-1000 Bioreactor classic flask. sMVs were first isolated from the CM by differential centrifugation at 10,000 x g, 30 min. The supernatant was further centrifuged at 100,000 x g for 1 h. Following ultracentrifugation, the SW480 and SW620 pellets were fractionated using a 5–40% OptiPrep™ density gradient (100,000 x g, 18 h). sMVs were washed in 1 mL PBS, 10,000 x g, 30 min to obtain purified sMVs. Exos were washed in 1 mL PBS, 100,000 x g, 1 h to obtain purified Exos. Following ultracentrifugation, densities of 12 fractions from the OptiPrep™ density gradient were determined by absorbance at 244 nm using a molar extinction coefficient of 320 L g^{-1} cm^{-1} (n = 3). (B) Western blot analysis of classical exosomal markers (Alix, TSG101, CD9, and CD63) (Protein 20 μg per lane, n = 3) (C) Transmission electron microscope images of sMV and Exo derived from SW480 and SW620 cell lines. sMV: Scale bar = 500 nm. Exo: Scale bar = 100 (SW480) and 200 (SW620) nm (n = 1). (D) Nanoparticle tracking analysis (NTA) of sMV and Exo derived from SW480 and SW620 cell lines (mean ± SEM, n = 3).

[5]. The workflow used to isolate sMVs from SW480 and SW620 CM using differential centrifugation and isopycnic density (iodixanol) fractionation is outlined in Fig. 1A. Briefly, CM was harvested, and differential centrifugation performed to discard floating cells (500 x g) and cellular debris (2000 x g). For sMVs the resultant supernatant was then subjected to 10,000 x g centrifugation. The pellet was resuspended and subsequently fractionated based upon their buoyant density into 12 fractions using iodixanol density gradient centrifugation (100,000 x g) [9], as outlined in Fig. 1A. SYPRO©-ruby based protein densitometry/quantitation analysis of these fractions for sMVs (as described [5]) revealed that fraction 7 (buoyant density 1.10 g/mL) contained the highest protein yield (Fig. 1A, Supplementary Fig. S2). For the purpose of comparative analysis, we also purified Exos from 10,000 x g supernatant using a combination of ultracentrifugation and iodixanol density gradient centrifugation (100,000 x g) (Fig. 1A). The yield of purified SW480-sMVs and SW620-sMVs from 180 mL of CM was 0.75 mg and 0.83 mg, respectively. Western blot analysis showed that sMVs were $Alix^{-}$, $TSG101^{-}$, $CD9^{-}$, and $CD63^{-}$ [5,7] (Fig. 1B). We investigated the morphology and size distribution of purified sMVs using transmission electron microscopy (TEM) and nanoparticle tracking analysis (NTA) (Fig. 1C-D). TEM revealed sMVs displayed round-like membranous vesicle structures 100–400 nm in size, and NTA showed particle diameters of 170–620 nm (mean 334–350 nm) which is in accordance with the typical size reported for sMVs [1,7]. In comparison, exosomes displayed smaller size distribution range (mean 183–186 nm). These data show that sMVs released by SW480 and SW620 cells are biophysically distinct from exosomes.

3.3. Proteome analysis of SW480- and SW620-derived shed microvesicles

Label-free quantitative mass spectrometry identified [24] 1295 and 1300 proteins in SW480-sMVs (Supplementary Table S1) and SW620-sMVs (Supplementary Table S2), respectively. Inspection of these data revealed 1123 proteins are common to both datasets. Of these, 834 exhibited similar abundances (≥ − 2 and ≤ 2 LFQ ratio) in both SW480- and SW620-sMVs (Fig. 2A) (Supplementary Table S3). Common proteins include key proteins involved in sMVs formation (ARF1/6, BSG, ARRDC1 [25,26], HSP90AA1/B1 [27], RAB22A [28], RHOA [29], and CSE1L [30]) and lipid-raft generation (ATP11C [31], TMEM30A [32], and PLSCR3 [33,34]) (Supplementary Table S4). GO analysis revealed that the 834 common proteins are related to the following processes -"cell adhesion/adhesion molecules" (PKP2/3, CLDND1, OCLN, CTN families), "signal transduction" (KRAS, NRAS, RHOA, MAPK1/2 K1), RNA binding proteins (e.g., ribonuclear proteins HNRNPK/Q/E2, YBX-1), "translation-associated proteins" (ribosomal proteins, translation initiation factors, aminoacyl tRNA ligases), and "transport vesicle" (VPS29/35, SNAP23) (Fig. 2B,C) (Supplementary Table S4,S5). In addition, selective enrichment of proteins associated with the cytoplasm, membrane cytoskeleton, and cell surface (e.g., ARF1/4/6, ACTN1/4, and HLA-A) was common to both SW480- and SW620-sMVs (Fig. 2C) (Supplemental Table S6). Interestingly, 359 of the shared 834 proteins have not been identified in exosomes derived from SW480 and SW620 cells that we previously reported, indicating that sMVs and exosomes have distinct protein profiles [5] (Supplementary Table S3).

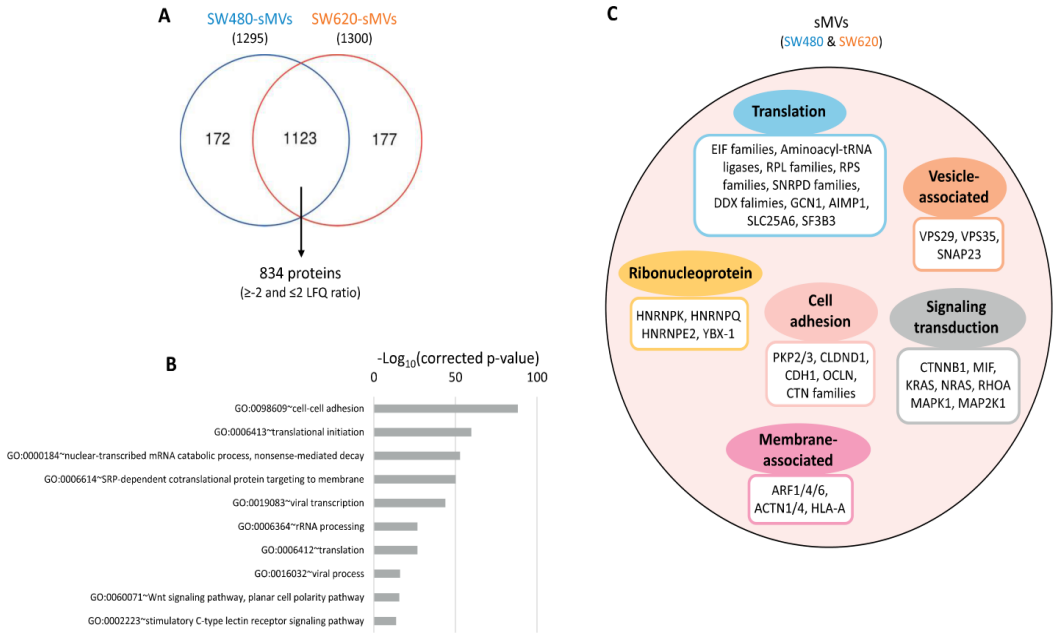


Fig. 2. Proteome analysis of SW480- and SW620-derived purified shed microvesicles.
(A) A Venn diagram of proteins identified in SW480- and SW620-sMVs, 1123 proteins were co-identified in SW480- and SW620-sMVs. Of these, 834 proteins were commonly identified using ≥ -2 and ≤ 2 LFQ ratio cut off. (B) Gene Ontology of biological process of 834 commonly identified proteins in sMVs, data obtained from DAVID (ten lowest corrected-*p*-value GO terms) [19]. (C) sMVs-enriched proteins in different functional categories (normalised LFQ > 50,000).

3.4. Proteins selectively enriched in SW480-sMVs and SW620-sMVs

Inspection of the Venn diagram in Fig. 2A reveals 307 proteins in SW480-sMVs are significantly enriched (≤ -2 and ≥ 2 LFQ ratio, $p < 0.05$) and 172/307 are unique to these vesicles (i.e., not seen in SW620-sMVs). In the case of SW620-MVs, 291 proteins are significantly enriched and 177 of these are unique (not seen in SW480-sMVs). Prominent amongst the 307 proteins enriched in SW480-sMVs are regulatory components associated with signal transduction/basement membrane (e.g., integrins ANXA1, CLDN7) [35–40], cell proliferation (EGFR [41]), pro-tumorigenic signalling (CD44 [42]), angiogenesis and vascular remodelling (e.g., NOTCH1/2 [43]) (Fig. 3A, Table 1). In the case of the 291 proteins enriched in SW620-sMVs most prominent were mediators of cell invasion (e.g., FGFR4 [44,45]), metastasis (e.g., MACC1 [46]), cytoskeleton organisation/polarity (e.g., mTORC2, PRKCA [47,48]), and signal transduction (e.g., PAK1, CLDN1, RASA1, ARHGAP1 [49–51]). Western blot analysis confirmed the differential enrichment in SW480-/SW620-sMVs of CLDN7, CLDN1, EGFR, CD44, ANAX1, and PAK1 (Fig. 3B).

Signalling pathways play a critical role in cancer such as cell proliferation, invasion/migration, angiogenesis, and metastasis [52]. Intriguingly, we found differing signalling protein networks enriched in SW480- and SW620-sMVs. For example, SW620-sMVs, but not SW480-sMVs, contain MTOR signalling network proteins (Fig. 3C). MTOR is a master signalling complex that promotes cancer proliferation, differentiation, metabolism, invasion, and migration [47,53]). mTORC2 (MTOR and RICTOR) (Supplementary Table S7) and proteins in MTOR-down-stream cascade such as PRKCA, kinases, and GTPases that promote actin organisation involved in cell invasion and migration [47,54] were all enriched in SW620-sMVs. Importantly, we found a protein-protein network of MARCKS and PKACA in SW620-sMVs (Fig. 3D). MARCKS is involved in actin cytoskeletal remodelling, cell proliferation and motility [55,56]. Phosphorylation of MARCKS by protein kinase C releases MARCKS from plasma membrane and promotes PI3K/AKT signalling pathways [57]. In colorectal cancer, MARCKS positively regulates metastatic phenotype in vitro and in vivo [58]. The presence of these signalling networks in SW620-sMVs suggests that these vesicles may play a prominent role in signal transduction and cancer invasion/migration.

We also identified protein signalling networks in SW480-sMVs, but not SW620-sMVs. For example, SW480-sMVs contain Notch signalling pathway proteins such as NOTCH1/2 receptors which regulate endothelial cell proliferation and migration, resulting in angiogenesis [43,59]. ADAM17 (positive angiogenic modulator that inhibits expression of anti-angiogenic factor TSP1 [60,61]) and PSEN1 proteases that activate Notch signalling by releasing extracellular and intracellular domains of Notch receptor were also prominent in SW480-sMVs (Fig. 3E) [43,59]. Interestingly, activation of Notch directly promotes artery formation without Dll4-ligand activation, indicating effectiveness of the activated Notch receptor [62]. Importantly, Notch receptor and ITCH (E3 ligase) have been shown to be delivered to recipient cells via microvesicles, resulting in increasing NOTCH-specific gene expression (HES1/5 genes) [63] (Fig. 3E). SW480-sMVs also contain EGFR (an important signalling molecule in cell proliferation, angiogenesis [64–66]) that has been shown to associate with EPCAM [67], ANXA1 [68], ADAM17 [69] (Fig. 3F). Inhibition of EGFR signalling causes decrease of EPCAM intracellular domain shedding, leading to attenuation of colorectal cancer progression [70]. Interestingly, EPCAM subunits such as CLDN7, CD44 were also significantly enriched in SW480-sMVs. β-catenin that binds to the intracellular domain of EPCAM to modulate Wnt signalling [71,72] was also present in SW480-sMVs (Supplementary Table S8). Collectively, these data suggest a possible role of SW480-sMVs in cell proliferation and angiogenesis.

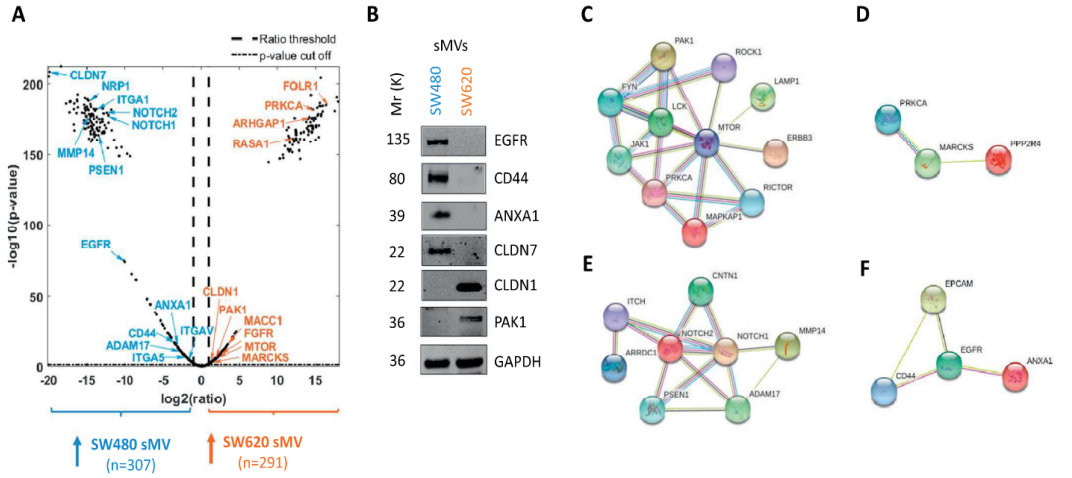


Fig. 3. Enriched signalling protein in sMVs and protein networks identified in sMVs derived from primary and metastatic cell lines (STRING analysis). (A) A volcano plot of proteins in sMVs during cancer progression (291 proteins significantly upregulated in SW620-sMVs (LFQ ratio < -2, $p < 0.05$), 307 proteins significantly downregulated in SW620-sMVs (LFQ ratio < -2, $p < 0.05$)). (B) Western blot analysis of enriched proteins in SW480-sMVs (EGFR, CD44, ANXA1, and CLDN7) and SW620-sMVs (CLDN1 and PAK1) (20 μg per lane, $n = 3$). (C) MTOR protein interaction network in SW620-sMVs. (D) MARCKS protein interaction network in SW620-sMVs. (E) NOTCH protein interaction network in SW480-sMVs. (F) EGFR protein interaction network in SW480-sMVs, (data obtained by STRING).

Table 1
Cancer progression-related enriched proteins in sMVs derived from human colorectal cancer cells.

Protein access	Gene name	Protein description	Ratio (SW620-sMV/SW480-sMV)[a]	p-value
P17252	PRKCA	Protein kinase C alpha type (PKC-A)	25,484.1	3.00E-182
Q07960	ARHGAP1	Rho GTPase-activating protein 1	25,042.5	2.00E-176
E9PGC0	RASA1	Ras GTPase-activating protein 1	4164.1	8.00E-162
Q6ZN28	MACC1	Metastasis-associated in colon cancer protein 1	8.5	2.00E-13
P22455	FGFR4	Fibroblast growth factor receptor 4	4.3	3.00E-07
P17612	PRKACA	cAMP-dependent protein kinase catalytic subunit alpha	3.7	4.00E-06
P11279	LAMP1	Lysosome-associated membrane glycoprotein 1	3.7	4.00E-06
Q13153	PAK1	Serine/threonine-protein kinase PAK 1	3.1	7.00E-05
P42345	MTOR	Mammalian target of rapamycin	2.7	3.00E-04
P23458	JAK1	Tyrosine-protein kinase JAK1	2.7	4.00E-04
Q14644	RASA3	Ras GTPase-activating protein 3	2.7	5.00E-04
Q13464	ROCK1	Rho-associated protein kinase 1	2.5	4.00E-04
Q68EM7	ARHGAP17	Rho GTPase-activating protein 17	2.4	2.00E-03
Q9NRY4	ARHGAP35	Rho GTPase-activating protein 35	2.3	2.00E-03
K7EP40	CLDN7	Claudin (Fragment)	−967,891.5	5.00E-209
E9PEP6	NRP1	Neuropilin1	−28,647.9	1.00E-188
P50281	MMP14	Matrix metalloproteinase-14	−21,885.1	4.00E-179
P56199	ITGA1	Integrin alpha-1	−13,793.4	1.00E-182
P49768	PSEN1	Presenilin-1	−11,146.5	2.00E-167
Q08345	DDR1	Epithelial discoidin domain-containing receptor 1	−10,457.1	1.00E-175
Q15582	TGFBI	Transforming growth factor-beta-induced protein ig-h3	−6738.6	4.00E-166
Q04721	NOTCH2	Neurogenic locus notch homolog protein 2	−5346.1	7.00E-180
P46531	NOTCH1	Neurogenic locus notch homolog protein 1	−4584.9	3.00E-178
A0A024R412	NRP2	Neuropilin2	−2170.5	6.00E-155
P00533	EGFR	Epidermal growth factor receptor	−1044.6	1.00E-74
P26006	ITGA3	Integrin alpha-3	−136.7	2.00E-48
P43121	MCAM	Cell surface glycoprotein MUC18	−15	3.00E-21
P16070	CD44	CD44 antigen	−11.3	4.00E-18
P78536	ADAM17	Disintegrin and metalloproteinase domain-containing protein 17	−5.3	5.00E-10
B5MCA4	EPCAM	Epithelial cell adhesion molecule	−4.4	1.10E-08
P04083	ANXA1	Annexin A1	−4.6	1.00E-08
P18084	ITGB5	Integrin beta-5	−3.2	5.00E-06
P06756	ITGAV	Integrin alpha-V	−2.5	2.00E-04
P08648	ITGA5	Integrin alpha-5	−2.2	1.00E-03

[a] Protein abundance ratio (LFQ ratio) reveals differential protein abundance between SW480 and SW620 sMVs. Positive LFQ values reflect increased protein abundance in SW620 sMVs relative to SW480 sMVs.

Table 2
sMVs contain proteins known to modulate pre-metastatic niche in tumour microenvironment.

Protein Acc.	Gene name	Protein description	SW480-sMV LFQ[a]	SW620-sMV LFQ[b]	Ratio (SW620-sMV/SW480-sMV)[c,*]	Target cell	Function [Reference]
P29966	MARCKS	Myristoylated alanine-rich C-kinase substrate	262,426	602,851	2.3	Immune cells	Induce inflammation in breast cancer tissue [87]
P26447	S100A4	Protein S100-A4	2,820,396	4,292,937	1.5	T-Cell macrophage	Recruitment of T-cells, macrophage motility, promotes metastasis [88]
P14174	MIF	Macrophage migration inhibitory factor	2,329,768	3,091,043	1.3	Macrophage	Macrophage motility [76]
O14672	ADAM10	Disintegrin and metalloproteinase domain-containing protein 10	114,652	92,112	−1.2	Breast cancer cell	Increase breast cancer cell motility [75]
P04083	ANXA1	Annexin A1	151,668	33,306	−4.6	Leukocytes	Potently and specifically inhibits the trans-endothelial migration of leukocytes [89]
P16070	CD44	CD44 antigen	409,137	36,076	−11.3	Pancreatic cancer cell	Metastatic niche formation [90]
P20333	TNFRSF1B	Tumour necrosis factor receptor superfamily member 1B	3182	1	−3181.6	T cells	Activates CD8 T cells [91]
P56199	ITGA1	Integrin alpha-1	13,793	1	−13,793.0	–	Represent cell-ECM interaction, leading to metastasis [92]
P50281	MMP14	Matrix metalloproteinase-14	21,885	1	−21,885.1	Osteoclasts	MMP-14, indirectly modulate TGF-β bioactivity by cleaving the ECM component [93]

[a] LFQ precursor ion intensity (normalised) (LFQ) for SW480 sMVs (n = 3, averaged) (Supplemental Table S1).
[b] LFQ precursor ion intensity (normalised) (LFQ) for SW620 sMVs (n = 3, averaged) (Supplemental Table S1).
[c] Protein abundance ratio (LFQ ratio) reveals differential protein abundance between SW480 and SW620 sMVs. Positive LFQ values reflect increased protein abundance in SW620 sMVs relative to SW480 sMVs.
[*] Differential expression with p-values < .05 as reported in Supplemental Table S1.

3.5. Identification of pre-metastatic niche factors in SW480-sMVs and SW620-sMVs

Establishment of a pre-metastatic niche is required for a survival and outgrowth of disseminated cancer cells at distant sites [73]. Several secreted factors released from stromal and tumour cells such as soluble proteins and cargo within extracellular vesicles enable pre-metastatic niche formation [74]. Interestingly, interrogation of SW480-/SW620-sMVs protein profiles identified several proteins associated with pre-metastatic niche formation – these include MARCKS, S100A4, MIF from metastatic cell-derived sMVs and ADAM10, ANXA1, CD44, TNFRSF1B, ITGA1, and MMP14 from primary cell-derived sMVs (Table 2). Of these, ADAM10 and MIF have been reported elsewhere to be delivered by exosomes [75,76].

Accumulating evidence show that exosomes from highly metastatic cells exhibit greater pre-metastatic niche formation capacity and cancer progression than exosomes secreted from cells with lowly metastatic capability [77-79]. Additionally, exosomes derived from primary tumour cells have been reported to be implicated in cancer metastasis formation [80]. However, the role of sMVs in pre-metastatic niche formation is poorly understood. The findings from our proteome studies reported here reveal that sMVs secreted from both primary and metastatic CRC cells contain proteins previously shown to modulate pre-metastatic niche formation. We hypothesise that sMVs might play a role in pre-metastatic niche formation.

3.6. Colon cancer biomarker identified in SW480- / SW620-sMVs

EVs are receiving much attention as possible targets for cancer diagnostics. Accumulating evidence shows that EVs and their bioactive cargo molecules, found in several bodily fluids such as blood, urine, and saliva, correlate with disease progression [81]. Although several studies indicated that exosomal DNA, RNA, and protein cargo can be serve as possible cancer biomarkers, very little is known about sMV cargo as potential cancer diagnostics. In our study, we identified several known CRC marker proteins found in sMVs. Interestingly, we identified CNN3, a marker for lymph node-metastatic colon cancer which is highly enriched in lymph node-metastatic colon cancer cell compared with primary colon cancer cell [82]. We also identified the well-known CRC markers such as CEACAM1 [83,84] and MUC13 [85] in both SW480-/SW620-sMVs (see Table 3).

3.7. SW480-and SW620-sMVs induce invasion of fibroblasts in vitro

Because sMVs contain proteins implicated in signal transduction [63,86], we next questioned whether sMVs can be taken up by recipient cells and elicit a functional response (Fig. 4). To test this, we labelled SW480- and SW620-sMVs with lipophilic dye DiO and subjected the labelled sMVs (DiO) to OptiPrep™-density gradient centrifugation to remove unbound dye. We incubated NIH3T3 fibroblasts with sMVs (DiO) for 2 h. The cells were then washed and stained with DiI for their plasma membrane and Hoechst stain for their nucleus. Cells were then imaged cells using live fluorescent microscopy (Fig. 4A). Compared to control NIH3T3 fibroblasts treated with PBS vehicle alone, fibroblasts incubated with SW480- or SW620-sMVs (DiO) displayed extensive uptake of sMVs (DiO, green) as distinct puncta (Fig. 4A, insets). This shows that both SW480- and SW620-sMVs were actively taken up by fibroblasts.

Next, we investigated whether sMVs can elicit a functional response in fibroblasts using Transwell Matrigel™ invasion assay (Fig. 4B). NIH3T3 fibroblast were stimulated with SW480- or SW620-sMVs (30 μg/mL) or PBS vehicle alone for 2 h and overlaid onto Transwell-insert (pore size 0.8 μm) coated with Matrigel™ matrix. The cells were incubated for 18 h to allow invasion to occur. Cells in the upper chamber were removed and the cells that had invaded to the lower side of the inset membrane were fixed and stained with Hoechst stain for

Table 3
Known colorectal cancer markers identified in sMVs.

Protein Acc.	Gene name	Protein description [Reference]	SW480-sMV LFQ[a]	SW620-sMV LFQ[b]	Ratio (SW620-sMV/SW480-sMV)[c]
Q9H3R2	MUC13	Mucin-13 [85]	1	48,404	48,404.3*
Q15417	CNN3	Calponin-3 [82]	33,893	391,418	11.6*
Q6ZN28	MACC1	Metastasis-associated in colon cancer protein 1 [94]	4772	40,552	8.5
P29323	EPHB2	Ephrin type-B receptor 2 [95]	33,490	78,319	2.3
P01116	KRAS	GTPase Kras [96]	3,199,224	3,595,732	1.1
P13688	CEACAM1	Carcinoembryonic antigen-related cell adhesion molecule 1 [97]	119,862	118,303	−1.0
P15311	EZR	Ezrin [98]	3,551,877	3,267,406	−1.1
P01111	NRAS	GTPase Nras [99]	320,806	269,658	−1.2
P16070	CD44	CD44 antigen [100]	409,137	36,076	−11.3

[a] LFQ precursor ion intensity (normalised) (LFQ) for SW480 sMVs (n = 3, averaged) (Supplemental Table S1).
[b] LFQ precursor ion intensity (normalised) (LFQ) for SW620 sMVs (n = 3, averaged) (Supplemental Table S1).
[c] Protein abundance ratio (LFQ ratio) reveals differential protein abundance between SW480 and SW620 sMVs. Positive LFQ values reflect increased protein abundance in SW620 sMVs relative to SW480 sMVs.
* Differential expression with *p*-values < .05 as reported in Supplemental Table S1.

their nucleus. Fluorescent microscopy revealed that compared to PBS vehicle treatment, stimulation of fibroblasts with SW480- or SW620-sMVs resulted in significantly greater levels of fibroblasts invasion across Matrigel™ matrix. Although SW620-sMVs promoted greater level of fibroblasts invasion across the matrix compared to SW480-sMVs, the difference did not reach statistical significance. These data show that sMVs secreted from CRC cells can be taken up by a recipient fibroblast cells and elicit a functional response.

4. Concluding remarks

In this study, we developed a large-scale purification method for obtaining mg quantities of highly-purified sMVs secreted from the human isogenic colorectal cancer cell lines SW480 and SW620 to enable biochemical characterisation and functional studies. Label-free quantitative mass spectrometry was used to obtain protein profiles for SW480- and SW620-sMVs. A striking finding was that SW480- and SW620-sMVs have distinct protein signatures that distinguish one sMV from another as well as exosomes derived from the same parental cells. Gene ontology analysis revealed that SW480-/SW620-sMV proteins categorised in biological processes such 'cell adhesion', 'RNA-related', and 'signalling proteins' correlated with known cancer progression biology. Key cancer progression proteins such as MTOR, PRKCA, MACC1, MARCKS and FGFR4 were enriched in SW620-sMV (metastatic cancer cell origin). In contrast, primary cancer cell-derived sMVs (SW480-sMVs) showed a selective enrichment of integrins, ANXA1, CLDN7, CD44, and NOTCH1/2. Proteins known to modulate pre-metastatic niche formation (e.g., ADAM10 and MIF) and colon cancer markers (CEACAM1, MUC13, CNN3) were also identified in SW480-/SW620-sMVs. Collectively, our study provides, for the first time, molecular insights into sMVs and their possible role in CRC biology.

Conflicts of interest

The authors declare no conflict of interest.

Author contribution statement

All authors were involved in Conceptualization and Writing - review & editing. WS, AR, DWG involved in Data curation; Formal analysis.

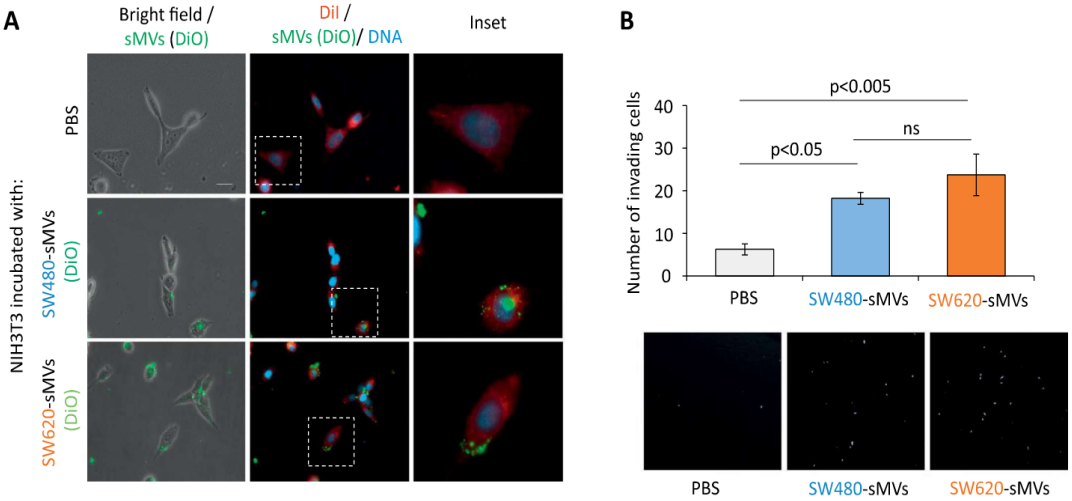


Fig. 4. Primary and metastatic cancer cell-derived shed microvesicles promote fibroblast invasion. (A) Live fluorescence microscopic analysis of NIH3T3 fibroblasts incubated with SW480- or SW620-sMVs (DiO, green). NIH3T3 were stained with DiI (red) and their nuclei with Hoechst (blue). Inset is the higher magnification of the image. Scale bar, 10 μm (B) Transwell Matrigel™ invasion assay of NIH3T3 fibroblasts treated with PBS vehicle alone or SW480- or SW620-sMVs (30 μ/mL). The nuclei of fibroblasts were stained with Hoechst stain (white) and images using fluorescent microscopy and quantified. Representative microscopic images are present in the lower panel ($n = 3$).

Acknowledgements

W.S., A.R., R.X., M.C., D.W.G., and R.J.S. acknowledge funding support from La Trobe University, Melbourne, Australia. W.S is supported by a La Trobe University Postgraduate Scholarship. We acknowledge the La Trobe University-Comprehensive Proteomics Platform for providing infrastructure.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbapap.2018.11.008.

References

[1] R. Xu, A. Rai, M. Chen, W. Suwakulsiri, D.W. Greening, R.J. Simpson, Extracellular vesicles in cancer - implications for future improvements in cancer care, Nat. Rev. Clin. Oncol. 15 (2018) 617–638.
[2] M. Tkach, C. Thery, Communication by Extracellular Vesicles: where we are and where we need to go, Cell 164 (2016) 1226–1232.
[3] W. Guo, Y. Gao, N. Li, F. Shao, C. Wang, P. Wang, Z. Yang, R. Li, J. He, Exosomes: New players in cancer, (Review), Oncol Rep 38 (2017) 665–675.
[4] A. Takahashi, R. Okada, K. Nagao, Y. Kawamata, A. Hanyu, S. Yoshimoto, M. Takasugi, S. Watanabe, M.T. Kanemaki, C. Obuse, E. Hara, Exosomes maintain cellular homeostasis by excreting harmful DNA from cells, Nat. Commun. 8 (2017) 15287.
[5] H. Ji, D.W. Greening, T.W. Barnes, J.W. Lim, B.J. Tauro, A. Rai, R. Xu, C. Adda, S. Mathivanan, W. Zhao, Y. Xue, T. Xu, H.J. Zhu, R.J. Simpson, Proteome profiling of exosomes derived from human primary and metastatic colorectal cancer cells reveal differential expression of key metastatic factors and signal transduction components, Proteomics 13 (2013) 1672–1686.
[6] W. Zhu, J.W. Smith, C.M. Huang, Mass spectrometry-based label-free quantitative proteomics, J Biomed Biotechnol (2010) (2010) 840518.
[7] R. Xu, D.W. Greening, A. Rai, H. Ji, R.J. Simpson, Highly-purified exosomes and shed microvesicles isolated from the human colon cancer cell line LIM1863 by sequential centrifugal ultrafiltration are biochemically and functionally distinct, Methods 87 (2015) 11–25.
[8] D.W. Greening, H. Ji, M. Chen, B.W. Robinson, I.M. Dick, J. Creaney, R.J. Simpson, Secreted primary human malignant mesothelioma exosome signature reflects oncogenic cargo, Sci. Rep. 6 (2016) 32643.
[9] B.J. Tauro, D.W. Greening, R.A. Mathias, H. Ji, S. Mathivanan, A.M. Scott, R.J. Simpson, Comparison of ultracentrifugation, density gradient separation, and immunoaffinity capture methods for isolating human colon cancer cell line LIM1863-derived exosomes, Methods 56 (2012) 293–304.
[10] M. Chen, R. Xu, H. Ji, D.W. Greening, A. Rai, K. Izumikawa, H. Ishikawa, N. Takahashi, R.J. Simpson, Transcriptome and long noncoding RNA sequencing of three extracellular vesicle subtypes released from the human colon cancer LIM1863 cell line, Sci. Rep. 6 (2016) 38397.
[11] M. Alhomrani, J. Correia, M. Zavou, B. Leaw, N. Kuk, R. Xu, M.I. Saad, A. Hodge, D.W. Greening, R. Lim, W. Sievert, The human amnion epithelial cell secretome decreases hepatic fibrosis in mice with chronic liver fibrosis, Front. Pharmacol. 8 (2017) 748.
[12] D.W. Greening, H.P. Nguyen, K. Elgass, R.J. Simpson, L.A. Salamonsen, Human endometrial exosomes contain hormone-specific Cargo modulating trophoblast adhesive capacity: insights into endometrial-embryo interactions, Biol. Reprod. 94 (2016) 38.
[13] S.K. Gopal, D.W. Greening, R.A. Mathias, H. Ji, A. Rai, M. Chen, H.J. Zhu, R.J. Simpson, YBX1/YB-1 induces partial EMT and tumourigenicity through secretion of angiogenic factors into the extracellular microenvironment, Oncotarget 6 (2015) 13718–13730.
[14] V. Gorshkov, T. Verano-Braga, F. Kjeldsen, SuperQuant: a Data Processing Approach to increase Quantitative Proteome Coverage, Anal. Chem. 87 (2015) 6319–6327.
[15] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, Nat. Biotechnol. 26 (2008) 1367–1372.
[16] C.A. Luber, J. Cox, H. Lauterbach, B. Fancke, M. Selbach, J. Tschopp, S. Akira, M. Wiegand, H. Hochrein, M. O'Keeffe, M. Mann, Quantitative proteomics reveals subset-specific viral recognition in dendritic cells, Immunity 32 (2010) 279–289.
[17] S. Kasen, R. Ouellette, P. Cohen, Mainstreaming and postsecondary educational and employment status of a rubella cohort, Am. Ann. Deaf 135 (1990) 22–26.
[18] B.J. Tauro, R.A. Mathias, D.W. Greening, S.K. Gopal, H. Ji, E.A. Kapp, B.M. Coleman, A.F. Hill, U. Kusebauch, J.L. Hallows, D. Shteynberg, R.L. Moritz, H.J. Zhu, R.J. Simpson, Oncogenic H-ras reprograms Madin-Darby canine kidney (MDCK) cell-derived exosomal proteins following epithelial-mesenchymal transition, Mol. Cell. Proteomics 12 (2013) 2148–2159.
[19] W. Huang Da, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, Nat. Protoc. 4 (2009) 44–57.
[20] C. von Mering, M. Huynen, D. Jaeggi, S. Schmidt, P. Bork, B. Snel, STRING: a database of predicted functional associations between proteins, Nucleic Acids Res. 31 (2003) 258–261.
[21] J.L. Hood, H. Pan, G.M. Lanza, S.A. Wickline, Paracrine Induction of Endothelium by Tumor Exosomes, Laboratory Investigation; a Journal of Technical Methods and Pathology, Vol. 89 (2009), pp. 1317–1328.
[22] J.N. Higginbotham, M. Demory Beckler, J.D. Gephart, J.L. Franklin, G. Bogatcheva, G.J. Kremers, D.W. Piston, G.D. Ayers, R.E. McConnell, M.J. Tyska, R.J. Coffey, Amphiregulin exosomes increase cancer cell invasion, Curr. Biol. 21 (2011) 779–786.
[23] R.E. Hewitt, A. McMarlin, D. Kleiner, R. Wersto, P. Martin, M. Tsokos, G.W. Stamp, W.G. Stetler-Stevenson, Validation of a model of colon cancer progression, J. Pathol. 192 (2000) 446–454.
[24] R.J. Simpson, L.M. Connolly, J.S. Eddes, J.J. Pereira, R.L. Moritz, G.E. Reid, Proteomic analysis of the human colon carcinoma cell line (LIM 1215): development of a membrane protein database, Electrophoresis 21 (2000) 1707–1732.
[25] L. Kuo, E.O. Freed, ARRDC1 as a mediator of microvesicle budding, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 4025–4026.
[26] J.F. Nabhan, R. Hu, R.S. Oh, S.N. Cohen, Q. Lu, Formation and release of arrestin domain-containing protein 1-mediated microvesicles (ARMMs) at plasma membrane by recruitment of TSG101 protein, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 4146–4151.
[27] Q. Feng, C. Zhang, D. Lum, J.E. Druso, B. Blank, K.F. Wilson, A. Welm, M.A. Antonyak, R.A. Cerione, A class of extracellular vesicles from breast cancer cells activates VEGF receptors and tumour angiogenesis, Nat. Commun. 8 (2017) 14450.
[28] T. Wang, D.M. Gilkes, N. Takano, L. Xiang, W. Luo, C.J. Bishop, P. Chaturvedi, J.J. Green, G.L. Semenza, Hypoxia-inducible factors and RAB22A mediate formation of microvesicles that stimulate breast cancer invasion and metastasis, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) E3234–E3242.
[29] A.E. Sedgwick, J.W. Clancy, M. Olivia Balmert, C. D'Souza-Schorey, Extracellular microvesicles and invadopodia mediate non-overlapping modes of tumor cell invasion, Sci. Rep. 5 (2015) 14748.
[30] C.F. Liao, S.H. Lin, H.C. Chen, C.J. Tai, C.C. Chang, L.T. Li, C.M. Yeh, K.T. Yeh, Y.C. Chen, T.H. Hsu, S.C. Shen, W.R. Lee, J.F. Chiou, S.F. Luo, M.C. Jiang, CSE1L, a novel microvesicle membrane protein, mediates Ras-triggered microvesicle generation and metastasis of tumor cells, Mol. Med. 18 (2012) 1269–1280.
[31] K. Segawa, S. Kurata, Y. Yanagihashi, T.R. Brummelkamp, F. Matsuda, S. Nagata, Caspase-mediated cleavage of phospholipid flippase for apoptotic phosphatidylserine exposure, Science 344 (2014) 1164–1168.
[32] R. Chen, E. Brady, T.M. McIntyre, Human TMEM30a promotes uptake of antitumor and bioactive choline phospholipids into mammalian cells, J. Immunol. 186 (2011) 3215–3225.
[33] J. Sun, M. Nanjundan, L.J. Pike, T. Wiedmer, P.J. Sims, Plasma membrane phospholipid scramblase 1 is enriched in lipid rafts and interacts with the epidermal growth factor receptor, Biochemistry 41 (2002) 6338–6345.
[34] S. Ory, M. Ceridono, F. Momboisse, S. Houy, S. Chasserot-Golaz, D. Heintz, V. Calco, A.M. Haeberle, F.A. Espinoza, P.J. Sims, Y. Bailly, M.F. Bader, S. Gasman, Phospholipid scramblase-1-induced lipid reorganization regulates compensatory endocytosis in neuroendocrine cells, J. Neurosci. 33 (2013) 3545–3556.
[35] L.E. O'Brien, T.S. Jou, A.L. Pollack, Q. Zhang, S.H. Hansen, P. Yurchenco, K.E. Mostov, Rac1 orientates epithelial apical polarity through effects on basolateral laminin assembly, Nat. Cell Biol. 3 (2001) 831–838.
[36] D.S. Harburger, D.A. Calderwood, Integrin signalling at a glance, J. Cell Sci. 122 (2009) 159–163.
[37] G. van der Horst, L. Bos, M. van der Mark, H. Cheung, B. Heckmann, P. Clement-Lacroix, G. Lorenzon, R.C. Pelger, R.F. Bevers, G. van der Pluijm, Targeting of alpha-v integrins reduces malignancy of bladder carcinoma, PLoS One 9 (2014) e108464.
[38] Y. Li, L. Cai, H. Wang, P. Wu, W. Gu, Y. Chen, H. Hao, K. Tang, P. Yi, M. Liu, S. Miao, D. Ye, Pleiotropic regulation of macrophage polarization and tumorigenesis by formyl peptide receptor-2, Oncogene 30 (2011) 3887–3899.
[39] T. Khau, S.Y. Langenbach, M. Schuliga, T. Harris, C.N. Johnstone, R.L. Anderson, A.G. Stewart, Annexin-1 signals mitogen-stimulated breast tumor cell proliferation by activation of the formyl peptide receptors (FPRs) 1 and 2, FASEB J. 25 (2011) 483–496.
[40] L. Gonzalez-Mariscal, R. Tapia, D. Chamorro, Crosstalk of tight junction components with signaling pathways, Biochim. Biophys. Acta 1778 (2008) 729–756.
[41] S. Sigismund, D. Avanzato, L. Lanzetti, Emerging functions of the EGFR in cancer, Mol. Oncol. 12 (2018) 3–20.
[42] J.M. Louderbough, J.A. Schroeder, Understanding the dual nature of CD44 in breast cancer progression, Mol. Cancer Res. 9 (2011) 1573–1586.
[43] J. Rodriguez-Vita, F. Tetzlaff, A. Fischer, Notch controls endothelial cells, Oncoscience 4 (2017) 45–46.
[44] J. Huang, M. Qiu, L. Wan, G. Wang, T. Huang, Z. Chen, S. Jiang, X. Li, L. Xie, L. Cai, TGF-beta1 Promotes Hepatocellular Carcinoma Invasion and Metastasis via ERK Pathway-Mediated FGFR4 Expression, Cell. Physiol. Biochem. 45 (2018) 1690–1699.
[45] Y. Sun, X. Fan, Q. Zhang, X. Shi, G. Xu, C. Zou, Cancer-associated fibroblasts secrete FGF-1 to promote ovarian proliferation, migration, and invasion through the activation of FGF-1/FGFR4 signaling, Tumour Biol. 39 (2017).
[46] U. Stein, W. Walther, F. Arlt, H. Schwabe, J. Smith, I. Fichtner, W. Birchmeier, P.M. Schlag, MACC1, a newly identified key regulator of HGF-MET signaling, predicts colon cancer metastasis, Nat. Med. 15 (2009) 59–67.
[47] E. Jacinto, R. Loewith, A. Schmidt, S. Lin, M.A. Ruegg, A. Hall, M.N. Hall, Mammalian TOR complex 2 controls the actin cytoskeleton and is rapamycin insensitive, Nat. Cell Biol. 6 (2004) 1122–1128.
[48] N. Angliker, M.A. Ruegg, In vivo evidence for mTORC2-mediated actin cytoskeleton rearrangement in neurons, BioArchitecture 3 (2013) 113–118.
[49] S.T. Sit, E. Manser, Rho GTPases and their role in organizing the actin cytoskeleton, J. Cell Sci. 124 (2011) 679–683.
[50] B. Gong, W.W. Liu, W.J. Nie, D.F. Li, Z.J. Xie, C. Liu, Y.H. Liu, P. Mei, Z.J. Li, MiR-21/RASA1 axis affects malignancy of colon cancer cells via RAS pathways, World J. Gastroenterol. 21 (2015) 1488–1497.
[51] D. Sun, C. Wang, S. Long, Y. Ma, Y. Guo, Z. Huang, X. Chen, C. Zhang, J. Chen, J. Zhang, C/EBP-beta-activated microRNA-223 promotes tumour growth through targeting RASA1 in human colorectal cancer, Br. J. Cancer 112 (2015) 1491–1500.

[52] R. Sever, J.S. Brugge, Signal transduction in cancer, Cold Spring Harb Perspect Med (2015) 5.
[53] H. Populo, J.M. Lopes, P. Soares, The mTOR signalling pathway in human cancer, Int. J. Mol. Sci. 13 (2012) 1886–1918.
[54] D.D. Sarbassov, S.M. Ali, D.H. Kim, D.A. Guertin, R.R. Latek, H. Erdjument-Bromage, P. Tempst, D.M. Sabatini, Rictor, a novel binding partner of mTOR, defines a rapamycin-insensitive and raptor-independent pathway that regulates the cytoskeleton, Curr. Biol. 14 (2004) 1296–1302.
[55] L.W.R. Fong, D.C. Yang, C.H. Chen, Myristoylated alanine-rich C kinase substrate (MARCKS): a multirole signaling protein in cancers, Cancer Metastasis Rev. 36 (2017) 737–747.
[56] R.E. Eckert, L.E. Neuder, J. Park, K.B. Adler, S.L. Jones, Myristoylated alanine-rich C-kinase substrate (MARCKS) protein regulation of human neutrophil migration, Am. J. Respir. Cell Mol. Biol. 42 (2010) 586–594.
[57] C.H. Chen, S. Statt, C.L. Chiu, P. Thai, M. Arif, K.B. Adler, R. Wu, Targeting myristoylated alanine-rich C kinase substrate phosphorylation site domain in lung cancer. Mechanisms and therapeutic implications, Am. J. Respir. Crit. Care Med. 190 (2014) 1127–1138.
[58] K. Rombouts, V. Carloni, T. Mello, S. Omenetti, S. Galastri, S. Madiai, A. Galli, M. Pinzani, Myristoylated Alanine-Rich protein Kinase C Substrate (MARCKS) expression modulates the metastatic phenotype in human and murine colon carcinoma in vitro and in vivo, Cancer Lett. 333 (2013) 244–252.
[59] A. Briot, A. Bouloumie, M.L. Iruela-Arispe, Notch, lipids, and endothelial cells, Curr. Opin. Lipidol. 27 (2016) 513–520.
[60] V. Caolo, G. Swennen, A. Chalaris, A. Wagenaar, S. Verbruggen, S. Rose-John, D.G. Molin, M. Vooijs, M.J. Post, ADAM10 and ADAM17 have opposite roles during sprouting angiogenesis, Angiogenesis 18 (2015) 13–22.
[61] A.J. Groot, M.A. Vooijs, The role of Adams in Notch signaling, Adv. Exp. Med. Biol. 727 (2012) 15–36.
[62] M.E. Pitulescu, I. Schmidt, B.D. Giaimo, T. Antoine, F. Berkenfeld, F. Ferrante, H. Park, M. Ehling, D. Biljes, S.F. Rocha, U.H. Langen, M. Stehling, T. Nagasawa, N. Ferrara, T. Borggrefe, R.H. Adams, Dll4 and Notch signalling couples sprouting angiogenesis and artery formation, Nat. Cell Biol. 19 (2017) 915–927.
[63] Q. Wang, Q. Lu, Plasma membrane-derived extracellular microvesicles mediate non-canonical intercellular NOTCH signaling, Nat. Commun. 8 (2017) 709.
[64] P. Seshacharyulu, M.P. Ponnusamy, D. Haridas, M. Jain, A.K. Ganti, S.K. Batra, Targeting the EGFR signaling pathway in cancer therapy, Expert Opin. Ther. Targets 16 (2012) 15–31.
[65] P. Wee, Z. Wang, Epidermal Growth factor Receptor Cell Proliferation Signaling Pathways, Cancers (Basel) 9 (2017).
[66] H. van Cruijsen, G. Giaccone, K. Hoekman, Epidermal growth factor receptor and angiogenesis: Opportunities for combined anticancer strategies, Int. J. Cancer 117 (2005) 883–888.
[67] Y.T. Hsu, P. Osmulski, Y. Wang, Y.W. Huang, L. Liu, J. Ruan, V.X. Jin, N.B. Kirma, M.E. Gaczynska, T.H. Huang, EpCAM-Regulated Transcription Exerts Influences on Nanomechanical Properties of Endometrial Cancer Cells that Promote Epithelial-to-Mesenchymal transition, Cancer Res. 76 (2016) 6171–6182.
[68] M. Wos, J. Bandorowicz-Pikula, Participation of annexins in endocytosis and EGFR-mediated signal transduction, Postepy Biochem. 60 (2014) 55–61.
[69] R. Mustafi, U. Dougherty, D. Mustafi, F. Ayaloglu-Butun, M. Fletcher, S. Adhikari, F. Sadiq, K. Meckel, H.I. Haider, A. Khalil, J. Pekow, V. Konda, L. Joseph, J. Hart, A. Fichera, Y.C. Li, M. Bissonnette, ADAM17 is a Tumor Promoter and Therapeutic Target in Western Diet-associated Colon Cancer, Clin. Cancer Res. 23 (2017) 549–561.
[70] K.H. Liang, H.C. Tso, S.H. Hung, I.I. Kuan, J.K. Lai, F.Y. Ke, Y.T. Chuang, I.J. Liu, Y.P. Wang, R.H. Chen, H.C. Wu, Extracellular domain of EpCAM enhances tumor progression through EGFR signaling in colon cancer cells, Cancer Lett. 433 (2018) 165–175.
[71] T. Yamashita, A. Budhu, M. Forgues, X.W. Wang, Activation of hepatic stem cell marker EpCAM by Wnt-beta-catenin signaling in hepatocellular carcinoma, Cancer Res. 67 (2007) 10831–10839.
[72] F.Q. Zhou, Y.M. Qi, H. Xu, Q.Y. Wang, X.S. Gao, H.G. Guo, Expression of EpCAM and Wnt/ beta-catenin in human colon cancer, Genet. Mol. Res. 14 (2015) 4485–4494.
[73] R.N. Kaplan, S. Rafii, D. Lyden, Preparing the "soil": the premetastatic niche, Cancer Res. 66 (2006) 11089–11093.
[74] H. Peinado, H. Zhang, I.R. Matei, B. Costa-Silva, A. Hoshino, G. Rodrigues, B. Psaila, R.N. Kaplan, J.F. Bromberg, Y. Kang, M.J. Bissell, T.R. Cox, A.J. Giaccia, J.T. Erler, S. Hiratsuka, C.M. Ghajar, D. Lyden, Pre-metastatic niches: organ-specific homes for metastases, Nat. Rev. Cancer 17 (2017) 302–317.
[75] M. Shimoda, S. Principe, H.W. Jackson, V. Luga, H. Fang, S.D. Molyneux, Y.W. Shao, A. Aiken, P.D. Waterhouse, C. Karamboulas, F.M. Hess, T. Ohtsuka, Y. Okada, L. Ailles, A. Ludwig, J.L. Wrana, T. Kislinger, R. Khokha, Loss of the Timp gene family is sufficient for the acquisition of the CAF-like cell state, Nat. Cell Biol. 16 (2014) 889–901.
[76] B. Costa-Silva, N.M. Aiello, A.J. Ocean, S. Singh, H. Zhang, B.K. Thakur, A. Becker, A. Hoshino, M.T. Mark, H. Molina, J. Xiang, T. Zhang, T.M. Theilen, G. Garcia-Santos, C. Williams, Y. Ararso, Y. Huang, G. Rodrigues, T.L. Shen, K.J. Labori, I.M. Lothe, E.H. Kure, J. Hernandez, A. Doussot, S.H. Ebbesen, P.M. Grandgenett, M.A. Hollingsworth, M. Jain, K. Mallya, S.K. Batra, W.R. Jarnagin, R.E. Schwartz, I. Matei, H. Peinado, B.Z. Stanger, J. Bromberg, D. Lyden, Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver, Nat. Cell Biol. 17 (2015) 816–826.
[77] H. Peinado, M. Aleckovic, S. Lavotshkin, I. Matei, B. Costa-Silva, G. Moreno-Bueno, M. Hergueta-Redondo, C. Williams, G. Garcia-Santos, C. Ghajar, A. Nitadori-Hoshino, C. Hoffman, K. Badal, B.A. Garcia, M.K. Callahan, J. Yuan, V.R. Martins, J. Skog, R.N. Kaplan, M.S. Brady, J.D. Wolchok, P.B. Chapman, Y. Kang, J. Bromberg, D. Lyden, Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET, Nat. Med. 18 (2012) 883–891.
[78] T. Fang, H. Lv, G. Lv, T. Li, C. Wang, Q. Han, L. Yu, B. Su, L. Guo, S. Huang, D. Cao, L. Tang, S. Tang, M. Wu, W. Yang, H. Wang, Tumor-derived exosomal miR-1247-3p induces cancer-associated fibroblast activation to foster lung metastasis of liver cancer, Nat Commun 9 (2018) 191.
[79] M.P. Plebanek, N.L. Angeloni, E. Vinokour, J. Li, A. Henkin, D. Martinez-Marin, S. Filleur, R. Bhowmick, J. Henkin, S.D. Miller, I. Ifergan, Y. Lee, I. Osman, C.S. Thaxton, O.V. Volpert, Pre-metastatic cancer exosomes induce immune surveillance by patrolling monocytes at the metastatic niche, Nat Commun 8 (2017) 1319.
[80] Q. Fu, Q. Zhang, Y. Lou, J. Yang, G. Nie, Q. Chen, Y. Chen, J. Zhang, J. Wang, T. Wei, H. Qin, X. Dang, X. Bai, T. Liang, Primary tumor-derived exosomes facilitate metastasis by regulating adhesion of circulating tumor cells via SMAD3 in liver cancer, Oncogene 37 (2018) 6105–6118.
[81] J. De Toro, L. Herschlik, C. Waldner, C. Mongini, Emerging roles of exosomes in normal and pathological conditions: new insights for diagnosis and therapeutic applications, Front. Immunol. 6 (2015) 203.
[82] C. Nakarai, K. Osawa, M. Akiyama, N. Matsubara, H. Ikeuchi, T. Yamano, S. Hirota, N. Tomita, M. Usami, Y. Kido, Expression of AKR1C3 and CNN3 as markers for detection of lymph node metastases in colorectal cancer, Clin. Exp. Med. 15 (2015) 333–341.
[83] N.P. Crawford, D.W. Colliver, S. Galandiuk, Tumor markers and colorectal cancer: utility in management, J. Surg. Oncol. 84 (2003) 239–248.
[84] Y. Zhang, P. Cai, L. Li, L. Shi, P. Chang, T. Liang, Q. Yang, Y. Liu, L. Wang, L. Hu, Co-expression of TIM-3 and CEACAM1 promotes T cell exhaustion in colorectal cancer patients, Int. Immunopharmacol. 43 (2017) 210–218.
[85] B.K. Gupta, D.M. Maher, M.C. Ebeling, V. Sundram, M.D. Koch, D.W. Lynch, T. Bohlmeyer, A. Watanabe, H. Aburatani, S.E. Puumala, M. Jaggi, S.C. Chauhan, Increased expression and aberrant localization of mucin 13 in metastatic colon cancer, J. Histochem. Cytochem. 60 (2012) 822–831.
[86] V. Muralidharan-Chari, J.W. Clancy, A. Sedgwick, C. D'Souza-Schorey, Microvesicles: mediators of extracellular communication during cancer progression, J. Cell Sci. 123 (2010) 1603–1611.
[87] M. Manai, J. Thomassin-Piana, A. Gamoudi, P. Finetti, M. Lopez, R. Eghozzi, S. Ayadi, O.B. Lamine, M. Manai, K. Rahal, E. Charafe-Jauffret, J. Jacquemier, P. Viens, D. Birnbaum, H. Boussen, M. Chaffanet, F. Bertucci, MARCKS protein overexpression in inflammatory breast cancer, Oncotarget 8 (2017) 6246–6257.
[88] Z.H. Li, N.G. Dulyaninova, R.P. House, S.C. Almo, A.R. Bresnick, S100A4 regulates macrophage chemotaxis, Mol. Biol. Cell 21 (2010) 2598–2610.
[89] A. Walther, K. Riehemann, V. Gerke, A novel ligand of the formyl peptide receptor: annexin I regulates neutrophil extravasation by interacting with the FPR, Mol Cell 5 (2000) 831–840.
[90] T. Jung, D. Castellana, P. Klingbeil, I. Cuesta Hernandez, M. Vitacolonna, D.J. Orlicky, S.R. Roffler, P. Brodt, M. Zoller, CD44v6 dependence of premetastatic niche preparation by exosomes, Neoplasia 11 (2009) 1093–1105.
[91] E.Y. Kim, S.J. Teh, J. Yang, M.T. Chow, H.S. Teh, TNFR2-deficient memory CD8 T cells provide superior protection against tumor cell growth, J. Immunol. 183 (2009) 6051–6057.
[92] C. Yang, M. Zeisberg, J.C. Lively, P. Nyberg, N. Afdhal, R. Kalluri, Integrin alpha1beta1 and alpha2beta1 are the key regulators of hepatocarcinoma cell invasion across the fibrotic matrix microenvironment, Cancer Res. 63 (2003) 8312–8317.
[93] N.A. Dallas, M.J. Gray, L. Xia, F. Fan, G. van Buren 2nd, P. Gaur, S. Samuel, S.J. Lim, T. Arumugam, V. Ramachandran, H. Wang, L.M. Ellis, Neuropilin-2-mediated tumor growth and angiogenesis in pancreatic adenocarcinoma, Clin. Cancer Res. 14 (2008) 8052–8060.
[94] S.H. Ge, X.J. Wu, X.H. Wang, X.F. Xing, L.H. Zhang, Y.B. Zhu, H. Du, B. Dong, Y. Hu, J.F. Ji, Over-expression of Metastasis-associated in Colon Cancer-1 (MACC1) Associates with Better Prognosis of Gastric Cancer patients, Chin. J. Cancer Res. 23 (2011) 153–159.
[95] A.M. Jubb, F. Zhong, S. Bheddah, H.I. Grabsch, G.D. Frantz, W. Mueller, V. Kavi, P. Quirke, P. Polakis, H. Koeppen, EphB2 is a prognostic factor in colorectal cancer, Clin. Cancer Res. 11 (2005) 5181–5187.
[96] D. Dinu, M. Dobre, E. Panaitescu, R. Birla, C. Iosif, P. Hoara, A. Caragui, M. Boeriu, S. Constantinoiu, C. Ardeleanu, Prognostic significance of KRAS gene mutations in colorectal cancer–preliminary study, J Med Life 7 (2014) 581–587.
[97] L. Ohlsson, M.L. Hammarstrom, A. Israelsson, L. Naslund, A. Oberg, G. Lindmark, S. Hammarstrom, Biomarker selection for detection of occult tumour cells in lymph nodes of colorectal cancer patients using real-time quantitative RT-PCR, Br. J. Cancer 95 (2006) 218–225.
[98] F. Liang, Y. Wang, L. Shi, J. Zhang, Association of Ezrin expression with the progression and prognosis of gastrointestinal cancer: a meta-analysis, Oncotarget 8 (2017) 93186–93195.
[99] A. Cercek, M.I. Braghiroli, J.F. Chou, J.F. Hechtman, N. Kemeny, L. Saltz, M. Capanu, R. Yaeger, Clinical Features and Outcomes of patients with Colorectal Cancers Harboring NRAS Mutations, Clin. Cancer Res. 23 (2017) 4753–4760.
[100] F. Jing, H.J. Kim, C.H. Kim, Y.J. Kim, J.H. Lee, H.R. Kim, Colon cancer stem cell markers CD44 and CD133 in patients with colorectal cancer and synchronous hepatic metastases, Int. J. Oncol. 46 (2015) 1582–1588.

Chapter 3

Comparative proteomic analysis of Exos, sMVs and sMB-Rs released from the isogenic colorectal cell lines SW480 and SW620

3.1 Introduction

EVs are evolutionary-conserved, membrane-encapsulated structures containing bioactive cargos such as oncogenic proteins, lipids, metabolites, RNA species (e.g., mRNA, miRNA, lncRNA), and DNA fragments shown to be potential cancer diagnostic candidates in EV-based liquid biopsies [79]. There are three major EV classes based on particle size, biochemical and biophysical properties, and mechanism of biogenesis - exosomes (Exos), shed microvesicles (sMVs), and shed midbody remnants (sMB-Rs). Accumulating evidence indicates the presence of subtypes within the Exos and sMV classes of EVs [79, 161].

My laboratory has been focusing on the isolation, purification and characterization of EV classes/ subtypes derived from colorectal cancer (CRC) cell lines such as LIM1215[162], LIM1863[111, 163] and SW480/SW620[59, 164] and their functional role in colorectal cancer progression[165, 166]. We have established methods for the large-scale purification of Exos, sMVs, and sMB-Rs from cell culture media using a combination of differential ultracentrifugation and isopycnic iodixanol density centrifugation[70, 77, 163]. While EV protein profiles have been determined for many CRC cell line-derived Exos[59, 133, 134, 162] ,sMVs[164] see **Chapter 2**), and sMB-Rs[70, 96, 198], a comparative proteome profiling study of Exos, sMVs and sMB-Rs secreted from the same cell line has not been undertaken. In this chapter, I compare the protein profiles of Exos, sMVs and sMB-Rs released from the isogenic

human CRC cells SW480 and SW620 and perform a detailed bioinformatic analysis in order to gain functional insights. In my next chapter (**Chapter 4**), I extend these studies to compare the RNA profiles from the same EV preparations.

3.2 Methods

3.2.1 Cell culture and large-scale purification of Exos, sMVs and sMB-Rs

SW480 and SW620 cells were cultured in a CELLine™ AD-1000 bioreactor device as described in **Chapter 2**, Sections 2.2 and 2.5. SW480 and SW620 culture media were sequentially centrifuged at 500 x *g* for 5 min (to remove floating cells), 2,000 x *g* for 10 min (to remove apoptotic body debris) and 10,000 x *g* for 30 min at 4 °C. The 10K pellet was subjected to buoyant density (isopycnic iodixanol (OptiPrep™) gradient centrifugation to separate sMVs (low buoyant density, fraction #7, 1.10 g/mL) from sMB-Rs (high buoyant density, fraction #9, 1.15 g/mL) by centrifugation at 100,000 x *g* for 18 h at 4°C. sMVs and sMB-Rs were recovered from fractions #7 and #9, respectively, by centrifugation at 100,000 x *g* for 18 h at 4°C, and the pellets resuspended in PBS (500 µL) and then re-centrifuged at 10,000 x *g* (30 min, 4°C). The pellets were resuspended in PBS (150 µL) for proteome analysis. The 10K supernatant was centrifuged at 100,000 x *g* (1 h, 4°C) to harvest crude Exos. The crude Exos pellet was re-suspended in 500 µL PBS and subjected to OptiPrep™ buoyant density gradient centrifugation as described above and purified Exos harvested (fraction #7 at a buoyant density of 1.10 g/mL) for proteome analysis.

3.2.2 Protein quantification and Western blotting

Protein quantification of EV samples, and Western blot analyses were determined as described in **Chapter 2** (Section 2.6). For Western blot analysis membranes were probed with primary antibodies (anti-mouse ALIX, 1:1,000, Cell Signaling), (anti-mouse TSG101, 1:1,000, BD Bioscience), (anti-rabbit GAPDH, 1:3,000, Cell Signaling), (anti-mouse KIF23, 1:1,000, Santa Cruz Biotechnology), (anti-mouse RACGAP1, 1:1,000, Santa Cruz Biotechnology) according to manufacturer's instructions. The secondary antibodies (IRDye800 goat anti-mouse IgG or IRDye700 goat anti-rabbit IgG) were diluted (1:15,000) and the fluorescent signals were detected using the Odyssey Infrared Imaging System, v3.0 (Li-COR Biosciences, Nebraska USA).

3.2.3 Transmission electron microscopy (TEM)

TEM images of EVs (1 µg in 10 µL PBS) were obtained as described in **Chapter 2** (Section 2.7).

3.2.4 Nanoparticle tracking analysis (NTA)

EV particle diameters were obtained by nanoparticle analysis (NTA) as described in **Chapter 2** (Section 2.8).

3.2.5 Label-free mass spectrometry and protein identification

GeLC-S/MS analysis was performed on EV samples (15 µg) as described in **Chapter 2** (Sections 2.9 and 2.10).

3.2.6 Differential protein enrichment analysis

LFQ intensities of peptide ions identified in Exos, sMV and sMB-R preparations were statistically analysed using the edgeR software package[169]. Briefly, LFQ intensities of each protein were normalized based on 'effective library size'[169] for each sample using a trimmed mean of M-values (TMM)[199]. Pvalues were calculated using Benjamini-Hochberg method[200] and normalized LFQ intensities were presented as log2.

For EV class comparisons 3 specific comparisons were undertaken: -i) SW480-/SW620-Exos vs SW480-/ SW620-sMVs, -ii) SW480-/ SW620-Exos vs SW480-/ SW620-sMB-Rs, and -iii) SW480-/ SW620-sMVs vs SW480-/SW620-sMB-Rs.

For CRC cancer progression-related protein identifications, SW480-combined EVs were compared with SW620-combined EVs.

Highly-enriched proteins in each comparison were identified using the criteria - log2fold change < -1 or > 1 with pvalue < 0.05.

3.2.7 Gene ontology and KEGG pathway analyses

Gene ontology (cellular compartment, specific level 10) and KEGG pathways (organism: hsa, pvalue cutoff: 0.05) were analyzed based on highly-enriched proteins from the statistical analysis comparison (see Section 3.2.6) using clusterProfiler[201] (v.3.11, https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html) in R

3.2.8 Data visualization

Principle component analysis, dot, box, ridge and volcano plots were visualized using ggplot (v.3.3.2, https://ggplot2.tidyverse.org/) in R. Heatmaps were visualized using pheatmap (v.1.0.12, https://www.rdocumentation.org/packages/pheatmap/versions/1.0.12) in R. Venn diagram was generated using a web-based tool (http://www.interactivenn.net/)[202]. KEGG pathway analysis was visualized using pathview (v.3.1.2, http://bioconductor.org/packages/release/bioc/html/pathview.html)[203] in R

3.3 Results

3.3.1 Purification and characterization of Exos, sMVs and sMB-Rs from SW480 and SW620 cell culture media

Large quantities of Exos, sMVs and sMB-Rs were isolated from SW480 and SW620 cells grown in continuous culture using the CELLine™ AD-1000 bioreactor device. The three EV types were purified from 180 mL CM acquired over 18 days (30 mL harvested each day, 6 days for each biological replicate) using a combination of differential centrifugation and isopycnic density (iodixanol) fractionation - see **Fig. 3.1A** for purification pipeline. Using this approach crude sMVs and sMB-Rs were isolated from the 10K pellet and crude Exos from the 100K-10K pellet. sMVs were separated from sMB-Rs by buoyant density (isopycnic iodixanol) gradient centrifugation (**Fig. 3.1B, C**). SDS-PAGE analysis of buoyant density fractions (SYPRO™ Ruby protein gel stain) showed sMVs distribute in fractions #6 and #7 (buoyant density ~1.10 g/mL) and sMB-Rs in fraction #9 (buoyant density ~1.15 g/mL) as evidenced by Western blot analysis using stereospecific marker antibodies for the centraspindlin complex of KIF23/ MKLP1 and RACGAP1, a key component of midbodies[99].

SW480- and SW620-Exos were purified from the 100K-10K pellet by buoyant density (isopycnic iodixanol) gradient centrifugation (fractions #6 and #7, buoyant density 1.08-1.10 g/mL) (**Fig. 3.1D, E**).

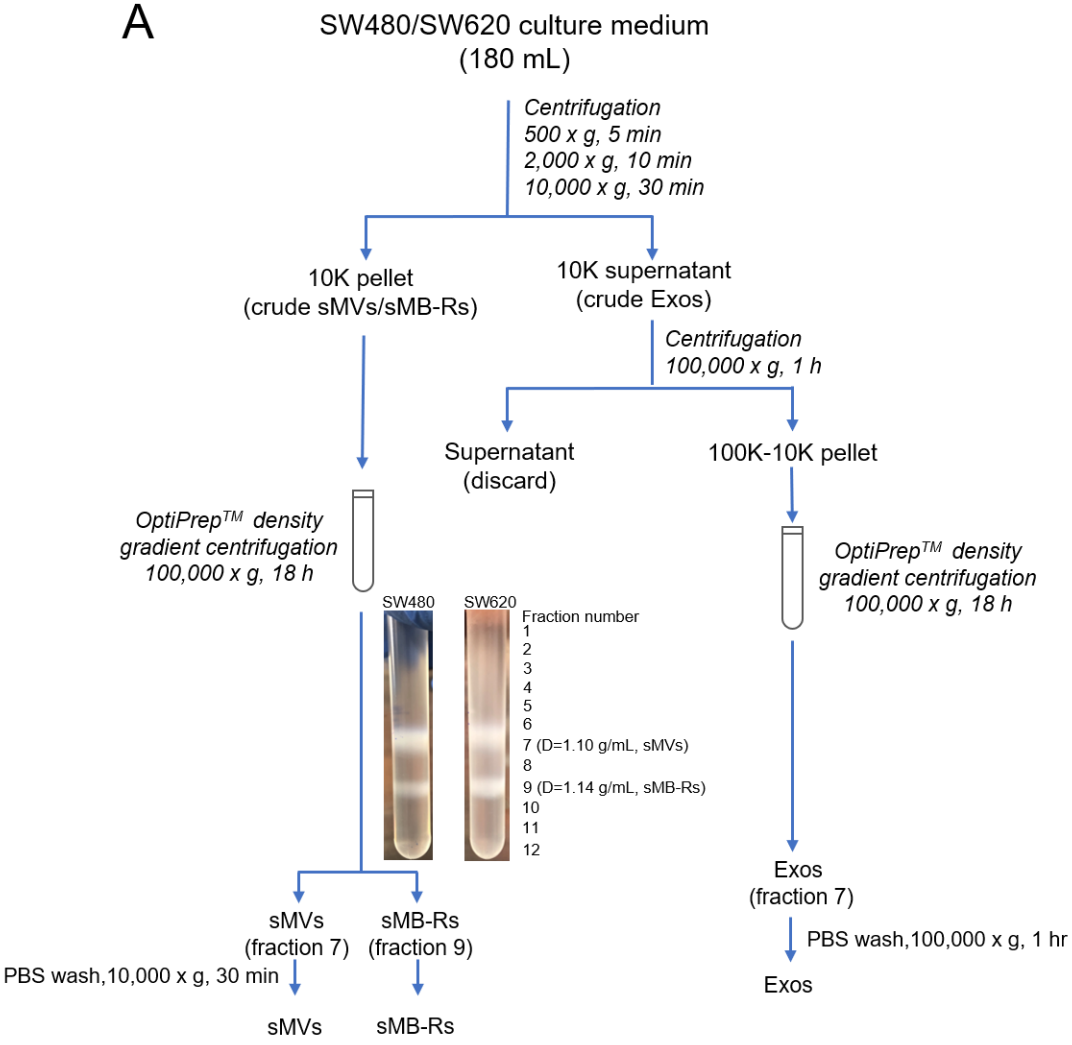
The yields of SW480-/ SW620-derived Exos, sMVs, and sMB-Rs (based on protein quantification, n=3 biological replicates) from 180 mL CM were 868-987 µg, 749-827 µg, and 463-660 µg, respectively.

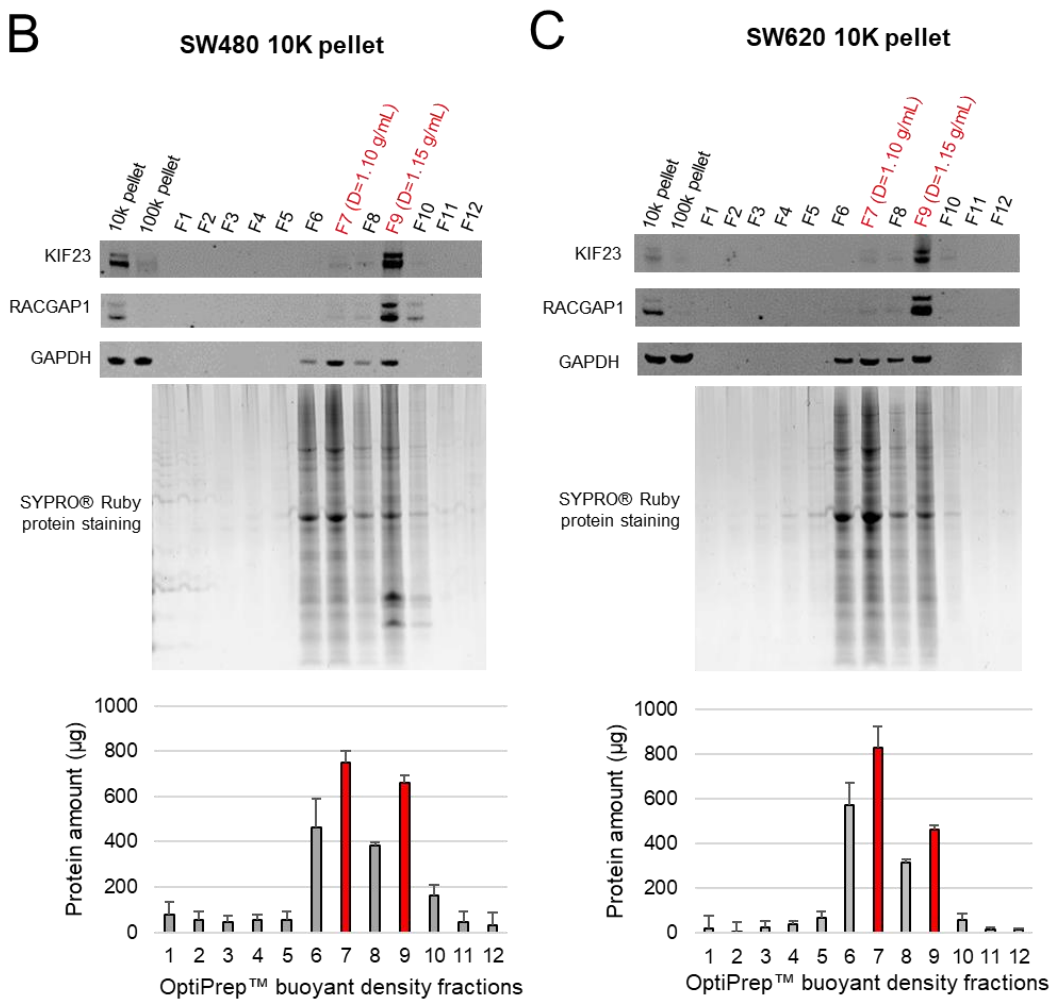
Transmission electron microscopic analysis was used to investigate the morphology and size range of Exos, sMVs, and sMB-Rs (**Fig. 3.1F**. TEM revealed sMVs and sMB-Rs are more ellipsoid in shape and heterogenous in size (100-500 nm diameter) compared to Exos, which displayed round-like structures and a smaller size distribution range (50-200 nm diameter). Nanoparticle tracking analysis (**Fig. 3.1G**) showed mean particle diameters of 186.2±3.4 nm /183.1±2.4 nm, 350±28.4 nm /337.1±9.3 nm and 404.9±13.6 nm /401.2±1.7 nm for SW480/SW620-Exos, sMVs, and sMB-Rs, respectively (biological replicates, n=3).

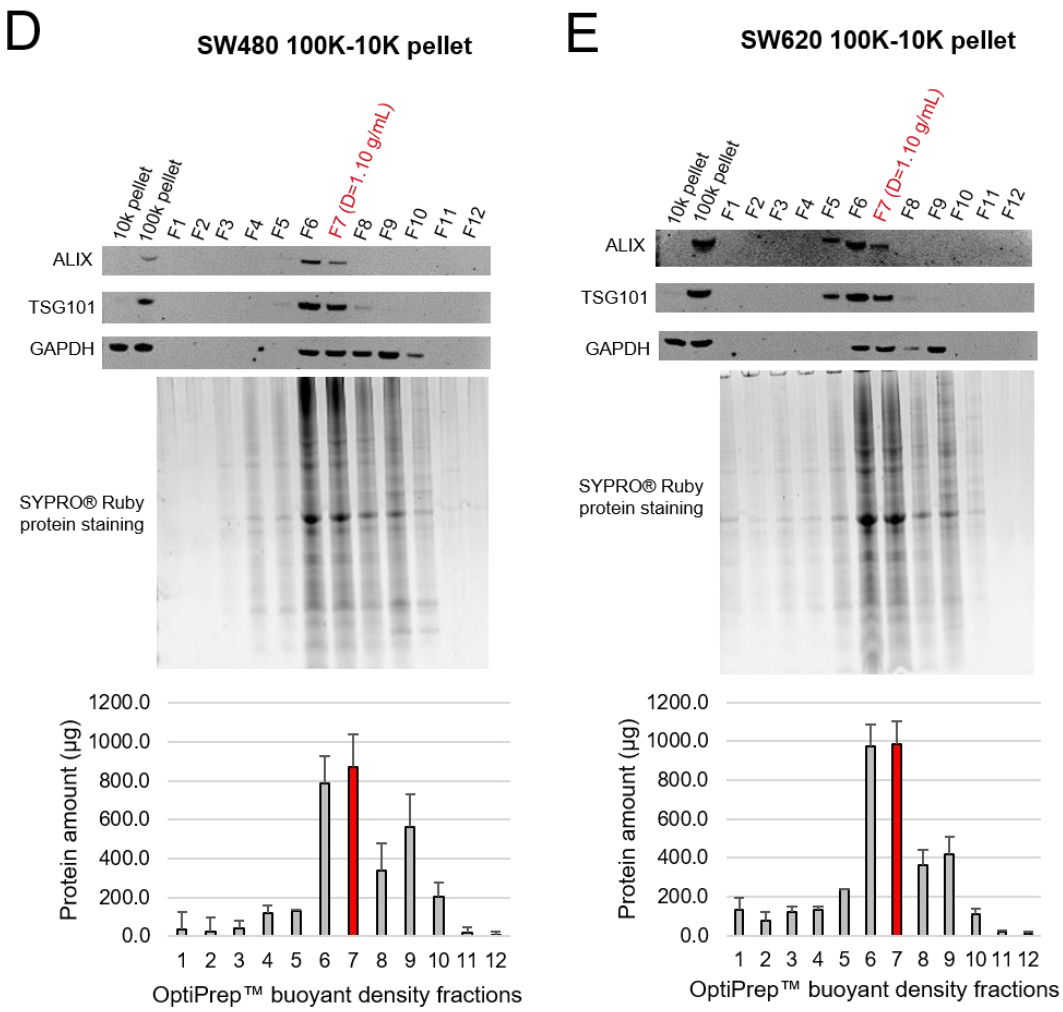
Western blot analysis revealed Exos are $ALIX^+$/ $TSG101^+$/ $CD63^+$/ $CD9^+$/ $CD81^+$, sMVs are $CD9^-$ / $CD81^-$, and sMB-Rs are $KIF23^+$/ $RACGAP1^+$ (**Fig. 3.1 H**).

Figure 3.1 Isolation and characterization of SW480 and SW620 cell-derived exosomes (Exos), shed microvesicles (sMVs), and shed midbody remnants (sMB-Rs). (A) Experimental workflow used for isolation of Exos, sMVs and sMB-Rs from culture media of SW480 and SW620 cells using differential centrifugation in combination with OptiPrep™ density gradient centrifugation. **(B, C)** Western blot analysis of purified SW480-/ SW620-sMVs (fraction #7), -sMB-Rs (fraction #9), crude 10K pellet, crude Exos (100K pellet) using anti-KIF23, anti-RACGAP1 and anti-GAPDH antibodies, protein quantification of OptiPrep™ fractions was performed using SDS-PAGE and SYPRO™ quantitative protein staining, n=3. (**D, E**) Western blot analysis of purified SW480-/ SW620-Exos (fraction #7) using anti-ALIX, anti-TSG101 and anti-GAPDH antibodies, protein quantification of OptiPrep™ fractions was performed using SDS-PAGE and SYPRO™ quantitative protein staining, n=3. (**F**) Transmission electron microscopic analysis of SW480-/ SW620-Exos, -sMVs and -sMB-Rs. (**G**) Nanoparticle tracking analysis (NTA) of SW480-/ SW620-Exos, -sMVs, and -sMB-Rs (mean ± SEM, n=3). (**H**) Western blot analysis of purified SW480-/ SW620-Exos, -sMVs, and sMB-Rs using anti-ALIX, anti-TSG101, anti-CD9, anti-CD63, anti-CD81, anti-KIF23, anti-RACGAP1 and anti-GAPDH antibodies (Protein load: 20 µg protein per lane, n=3).

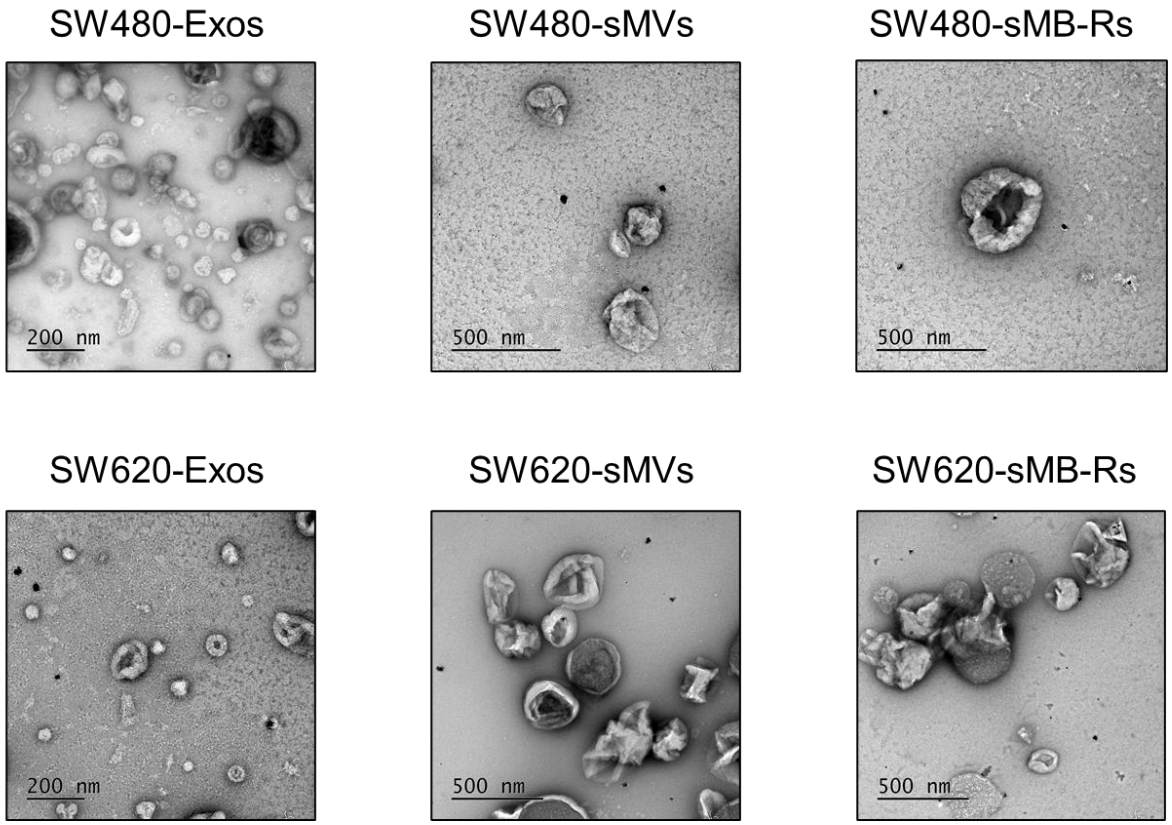
A


SW480/SW620 culture medium
(180 mL)
Centrifugation
500 x g, 5 min
2,000 x g, 10 min
10,000 x g, 30 min
10K pellet
(crude sMVs/sMB-Rs)
10K supernatant
(crude Exos)
Centrifugation
100,000 x g, 1 h
Supernatant
(discard)
100K-10K pellet
OptiPrep™ density
gradient centrifugation
100,000 x g, 18 h
OptiPrep™ density
gradient centrifugation
100,000 x g, 18 h
SW480
SW620
Fraction number
1
2
3
4
5
6
7 (D=1.10 g/mL, sMVs)
8
9 (D=1.14 g/mL, sMB-Rs)
10
11
12
sMVs
(fraction 7)
sMB-Rs
(fraction 9)
PBS wash,10,000 x g, 30 min
sMVs
sMB-Rs
Exos
(fraction 7)
PBS wash,100,000 x g, 1 hr
Exos

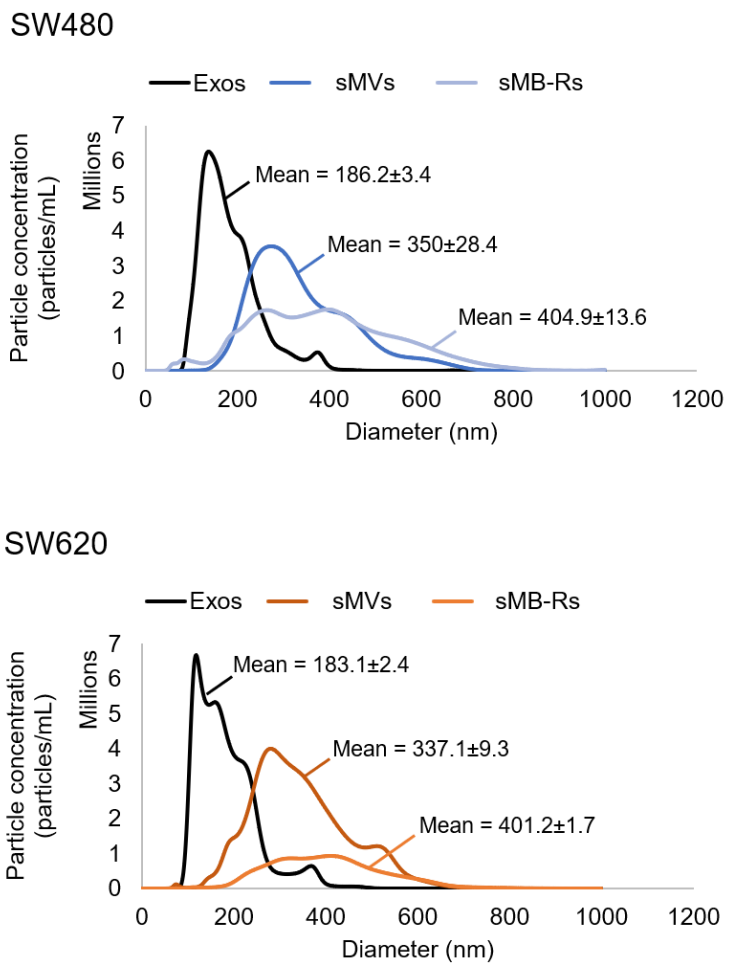

B
SW480 10K pellet
C
SW620 10K pellet
10k pellet
100k pellet
F1
F2
F3
F4
F5
F6
F7 (D=1.10 g/mL)
F8
F9 (D=1.15 g/mL)
F10
F11
F12
KIF23
RACGAP1
GAPDH
SYPRO® Ruby protein staining
Protein amount (µg)
1000
800
600
400
200
0
1 2 3 4 5 6 7 8 9 10 11 12
OptiPrep™ buoyant density fractions


D
SW480 100K-10K pellet
10k pellet
100k pellet
F1
F2
F3
F4
F5
F6
F7 (D=1.10 g/mL)
F8
F9
F10
F11
F12
ALIX
TSG101
GAPDH
SYPRO® Ruby protein staining
Protein amount (µg)
1200.0
1000.0
800.0
600.0
400.0
200.0
0.0
1 2 3 4 5 6 7 8 9 10 11 12
OptiPrep™ buoyant density fractions
E
SW620 100K-10K pellet
10k pellet
100k pellet
F1
F2
F3
F4
F5
F6
F7 (D=1.10 g/mL)
F8
F9
F10
F11
F12
ALIX
TSG101
GAPDH
SYPRO® Ruby protein staining
Protein amount (µg)
1200.0
1000.0
800.0
600.0
400.0
200.0
0.0
1 2 3 4 5 6 7 8 9 10 11 12
OptiPrep™ buoyant density fractions

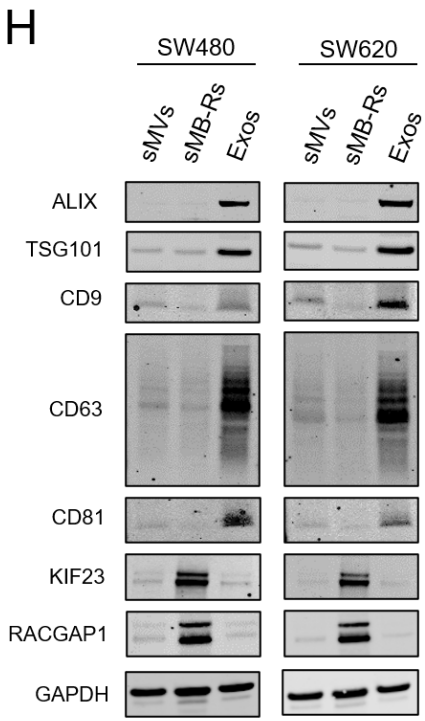
F


SW480-Exos
SW480-sMVs
SW480-sMB-Rs
200 nm
500 nm
500 nm
SW620-Exos
SW620-sMVs
SW620-sMB-Rs
200 nm
500 nm
500 nm

G


SW480
Exos
sMVs
sMB-Rs
Particle concentration (particles/mL)
Millions
Mean = 186.2±3.4
Mean = 350±28.4
Mean = 404.9±13.6
Diameter (nm)
SW620
Exos
sMVs
sMB-Rs
Particle concentration (particles/mL)
Millions
Mean = 183.1±2.4
Mean = 337.1±9.3
Mean = 401.2±1.7
Diameter (nm)

H


SW480
SW620
sMVs
sMB-Rs
Exos
ALIX
TSG101
CD9
CD63
CD81
KIF23
RACGAP1
GAPDH

3.3.2 Exos, sMVs and sMB-Rs secreted by SW480/ SW620 cells are molecularly distinct

Using the label-free MS approach described in **Chapter 2** (Sections 2.9 and 2.10), I compared the protein profiles of Exos, sMVs and sMB-Rs. Overall, 1544, 1527, 1544, 1510, 1562, and 1760 proteins were identified from SW480-Exos, SW620-Exos, SW480-sMVs, SW620-sMVs, SW480-sMB-Rs and SW620-sMB-Rs, respectively (see Venn diagrams, **Fig. 3.2A and B**). A list of top 50 protein identifications is given in **Appendix Table 3.1**. In the case of the three vesicle classes derived from SW480 cells, 101, 76, and 269 proteins were found to be uniquely present (based on presence/ absence of peptide ion intensity) in Exos, sMVs and sMB-Rs, respectively, and 1162 proteins were found to be common to all three EV classes (**Fig. 3A**). For the three classes of EVs secreted from SW620 cells, 186, 55, and 340 proteins are unique to Exos, sMVs, and sMB-Rs, respectively, with 1224 proteins common to the three EV classes (**Fig. 3B**). Principal component analysis (PCA) of these datasets shows that -i) the proteomes of sMB-Rs, Exos and sMVs are dissimilar, regardless of parental cell type, - ii) the proteomes of SW480-/ SW620-sMB-Rs are very dissimilar, -iii) the proteomes of SW480-Exos/ SW480-sMVs are similar, and -iv) the proteomes of SW620-Exos and SW620-sMVs are similar (**Fig. 3.2C**). These differences in proteomes are further highlighted in the proteome clustering plot in **Fig. 3.3D** which shows that the cluster patterns for SW480-/SW620-sMB-Rs proteomes are similar, but

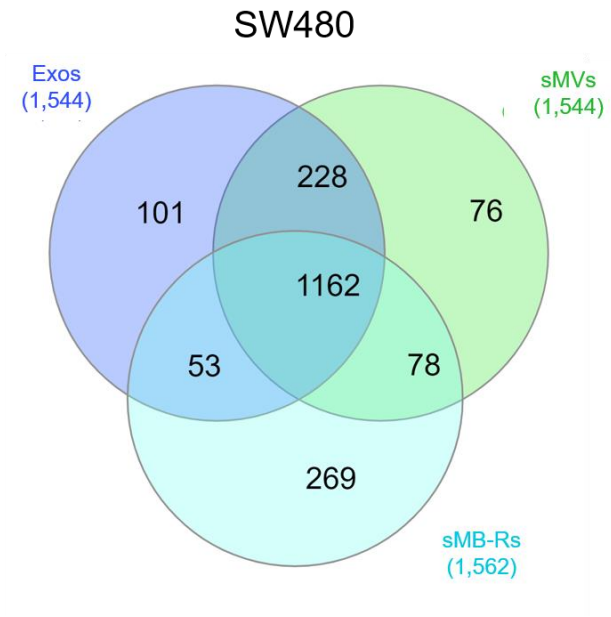
different to SW480-/SW620-Exos and -sMVs cluster patterns that are more alike (see **Appendix Fig. 3.1**, heatmap of total proteome of Exos, sMVs and sMB-Rs).

A list of uniquely-identified proteins found in Exos, sMVs, and sMB-Rs secreted from SW480 and SW620 cells is given in **Appendix Tables 3.2, 3.3 and 3.4, respectively**.

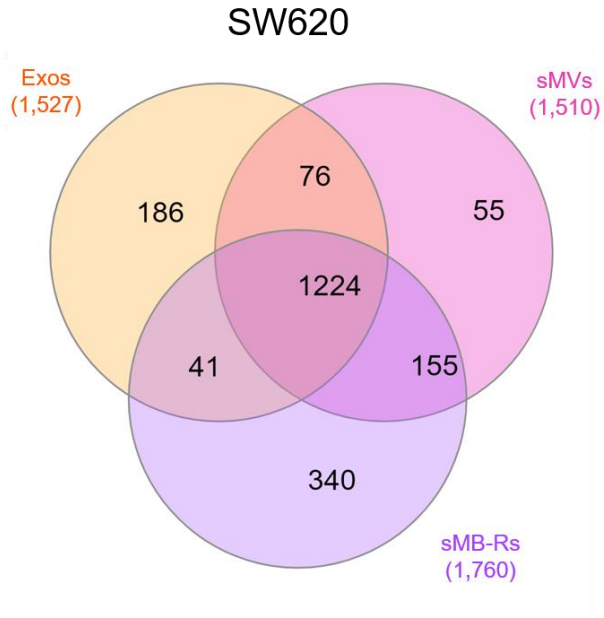
Collectively, these data indicate selective trafficking of many SW480- and SW620-cellular proteins to their respective Exos, sMVs and sMB-Rs.

Figure 3.2 Proteomic profiling of purified Exos, sMVs and sMB-Rs derived from SW480 and SW620 cells. **(A)** A three-way Venn diagram of proteins identified in SW480-Exos, -sMVs and -sMB-Rs reveals 1162 proteins were commonly identified, while 101, 76, and 269 proteins were uniquely identified in SW480-Exos, sMVs, and sMB-Rs, respectively. **(B)** A three-way Venn diagram of proteins identified in SW620-Exos, -sMVs reveals 186, 55, and 340 proteins were uniquely-identified in SW620-Exos, sMVs and sMB-Rs, respectively. **(C)** Principal component analysis (PCA) and **(D)** clustering analysis of total proteomes of Exos, sMVs and sMB-Rs derived from SW480 and SW620 cells.

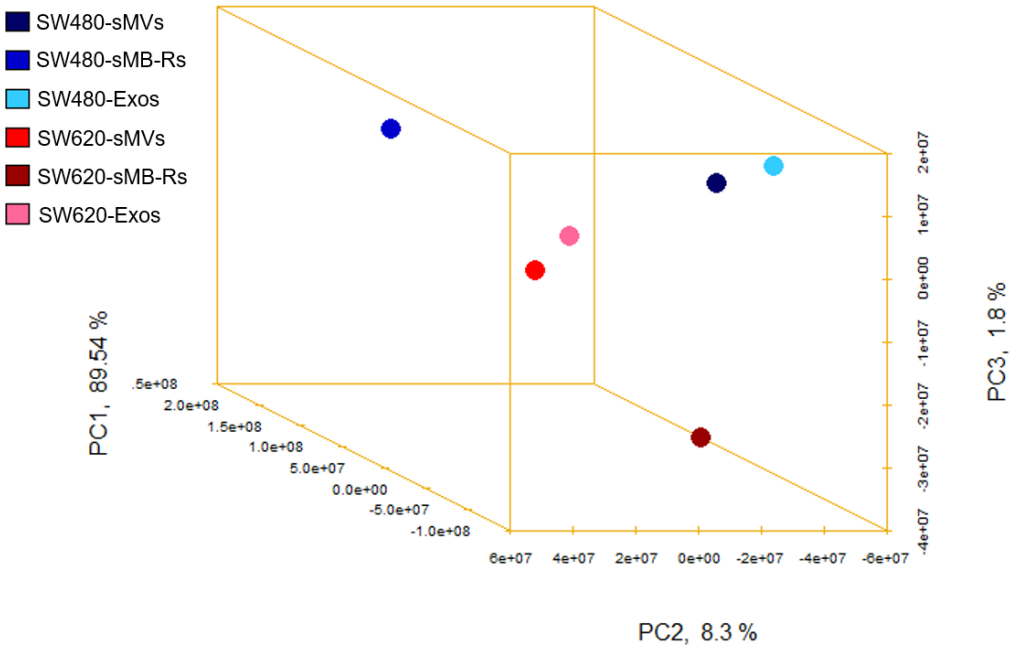
A


SW480
Exos
(1,544)
sMVs
(1,544)
sMB-Rs
(1,562)
101
228
76
1162
53
78
269

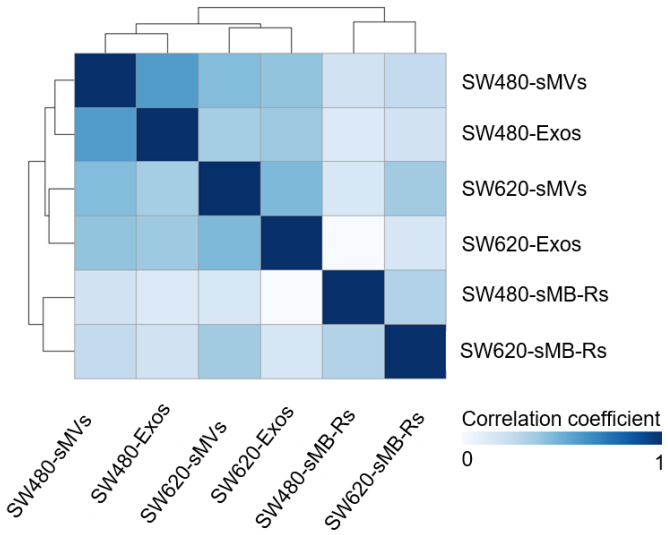
B


SW620
Exos
(1,527)
sMVs
(1,510)
sMB-Rs
(1,760)
186
76
55
1224
41
155
340

C


SW480-sMVs
SW480-sMB-Rs
SW480-Exos
SW620-sMVs
SW620-sMB-Rs
SW620-Exos
PC1, 89.54 %
PC2, 8.3 %
PC3, 1.8 %

D


SW480-sMVs
SW480-Exos
SW620-sMVs
SW620-Exos
SW480-sMB-Rs
SW620-sMB-Rs
Correlation coefficient
0
1

3.3.3 Interrogation of highly-enriched proteins in SW480-/SW620-derived Exos, sMVs and sMB-Rs

To gain insights into similarities and differences between EV classes, I performed a deep interrogation of the protein profiles for SW480-/ SW620-derived Exos, sMVs, and sMB-Rs. For this comparative analysis I performed a differential protein enrichment analysis of the following datasets: - i) combined SW480-/ SW620-Exos versus combined SW480-/ SW620-sMVs, - ii) combined SW480-/ SW620-Exos versus combined SW480-/ SW620-sMB-Rs, and - iii) combined SW480-/ SW620-sMVs versus combined SW480-/ SW620-sMB-Rs. Highly-enriched proteins in each comparison were identified using the criteria: log2fold change < -1.0 or > 1.0 with a pvalue < 0.05 (see Section 3.2.6). A list of highly-enriched protein identifications in Exos, sMVs and sMB-Rs is provided in **Appendix Tables 3.5, 3.6** and **3.7,** respectively.

First, I determined the number of protein identifications highly-enriched in SW480-/SW620-Exos (data combined) when compared to corresponding SW480-/SW620-derived sMVs and the number of proteins enriched in SW480-/SW620-derived Exos compared to sMB-Rs. The criteria I applied for proteins that preferentially traffic into one or another EV class was based on those proteins not being evident, or of very low abundance, in *one* EV class, but not the other *two* EV classes. In **Fig. 3.3A,** it can be seen that there are 95 proteins highly-enriched in Exos when compared to sMVs, and 409 proteins highly-enriched in Exos when Exos are compared to sMB-Rs. When these two datasets are compared (see Venn diagram) it can be seen that 80 proteins

selectively traffic to Exos. A list of the 80 selectively enriched Exos protein identifications is given in **Appendix Table 3.5**. Next, I determined the number of proteins highly-enriched in sMVs compared to Exos (108 proteins) and sMB-Rs (319 proteins) and then compared the two data sets (Venn diagram, **Fig. 3.3B**). This analysis showed that 14 proteins selectively traffic to sMVs. A list of the 14 proteins selectively enriched in sMVs compared to Exos and sMB-Rs is given in **Appendix Table 3.6**. In a third comparative analysis, 604 proteins were found to be highly-enriched in sMB-Rs when sMB-Rs were compared to Exos, and 533 when sMB-R proteins were compared to sMVs (**Fig. 3.3C**); of these, 492 proteins were found to selectively traffic to sMB-Rs (a list of selectively-enriched proteins in sMB-Rs is given in **Appendix Table 3.7**).

Selectively-enriched proteins found in Exos (80), sMVs (14) and sMB-Rs (492) were subjected to GO and KEGG pathway analysis to gain insights into their possible biological roles.

3.3.3.1 Proteins that selectively traffic to Exos

GO terms such as vacuolar membrane and nuclear envelope were identified in all EV subtypes (**Fig. 3.3D**, red stars). GO terms related to endoplasmic reticulum membrane, endosome membrane, coated vesicle membrane, nuclear membrane and secretory granules were co-identified in Exos and sMB-Rs (**Fig. 3.3D**, yellow stars). Exos showed enriched GO terms such as ‘endocytic membrane’, ‘transport vesicle membrane’ and ‘synaptic

vesicles' (**Fig. 3.3D**) and proteins related to exosome biogenesis and trafficking – for example, CD63, CD81, CD82, CHMP4B, VAMP3, VPS25, CD9, SDCBP, ARRDC1, TSG101, CHMP1A, VPS28, VPS37B, VAMP5, TSPAN1, TSPAN6, TSPAN14 (**Fig. 3.3E**, for a list of proteins selectively enriched in Exos, sMVs, and sMB-Rs, see **Table 3.1**).

KEGG pathway analysis of proteins highly-enriched in Exos compared to sMVs (95 proteins) and Exos compared to sMB-Rs (409 proteins) (**Fig. 3.3A**) showed mutual KEGG pathways such as *category-i)* 'Endocytosis' (hsa04144), *category-ii)* 'SNARE interactions in vesicular transport' (hsa04130) and, KEGG *category-iii*) 'Cell adhesion molecules' (hsa04514) enriched in Exos compared to both sMVs and sMB-Rs (**Appendix Fig. 3.2A-F** and **Appendix Table A3.8**).

KEGG *category-i)* 'Endocytosis' (hsa04144) showed enriched ESCRT proteins in Exos (compared to sMVs and sMB-Rs) (VPS37, CHMP1, CHMP2, CHMP3, CHMP4, CHMP5, and Clathrin-dependent/-independent endocytosis (PLD, MHCI, E3 ligase, and PLD, MHCI, E3 ligase, SRC) **(Appendix Fig. 3.2A**, **B)**.

KEGG *category-ii)* 'SNARE interactions in vesicular transport' (hsa04130) displayed enriched VAMP3, VAMP5, VAMP8 and STX7 proteins (**Appendix Fig. 3.2C, D**).

KEGG *category-iii)* 'Cell adhesion molecules' (hsa04514) showed highly-enriched proteins such as MHCI, PVRL1, PVRL2, CD99 in Exos

compared to sMVs and MHCI, PVR, PVRL1, PVRL2, CD99, OCLN, L1CAM in Exos compared to sMB-Rs (**Appendix Fig. 3.2E, F**).

3.3.3.2 Proteins that selectively traffic to sMVs

GO terms such as vacuolar membrane and nuclear envelope were identified sMVs (**Fig. 3.3D**, red stars). Several proteins such as DOCK1, DTYMK, FGFR4, FHL1 (**Fig. 3.3E**), SLC29A2, IMPA1, MRI1 (**Table 3.1**) were enriched in sMVs. However, KEGG pathway analysis on 108 and 319 proteins enriched in sMVs compared to Exos and sMB-Rs, respectively did not show mutual KEGG pathways.

3.3.3.3 Proteins that selectively traffic to sMB-Rs

Highly-enriched proteins identified in SW480-/SW620- sMB-Rs exhibited enriched GO terms such as 'Golgi-associated vesicle membrane', 'Intrinsic component of endoplasmic reticulum membrane' (**Fig. 3.3D**). These identified proteins include AP2A1, CANX, MGST1, PDIA6 (**Fig. 3.3E**), HSPA5 and ERP44 (**Table 3.1**) mitochondrial membrane, inner mitochondrial membrane protein complex such as SLC25A11, SLC25A13 (**Fig. 3.3E**), TOMM22, VDAC1, VDAC2 (**Table 3.1**) and spliceosomal complex, ribonucleoproteins such as SNRNP40, U2AF2, HNRNPAB, HNRNPH3, HNRNPL (**Fig. 3.3E**), SF3B1, SF3B2, SF3B3, U2AF2, U2AF1L5, HNRNPA1, HNRNPK, HNRNPD and HNRNPU (**Table 3.1**). Interestingly, midbody related proteins (AURKB, CEP55, KIF23, PLK1, RACGAP1), histones such as HIST1H4A), HIST1H1C, HIST2H3PS2, HIST2H3A, HIST1H2AC (**Fig. 3.3E,**

Table 3.1) are preeminent. RNA granule proteins such as IGF2BP1, IGF2BP2, FUS and TARDBP (**Table 3.1**) and several translation initiation factors (EIF families) (**Appendix Fig. 3.3**) were also highly-enriched in sMB-Rs.

KEGG pathway analysis on 604 and 533 highly-enriched proteins in sMB-Rs (**Fig. 3.3C**) compared to Exos and sMVs, respectively, showed mutual KEGG pathways such as *category-iv)* 'RNA transport' (hsa03013), *category-v)* 'Spliceosome' (hsa03040), *category-vi)* 'Protein processing in endoplasmic reticulum' (hsa04141), and *KEGG category-vii)* 'Citrate cycle' (TCA cycle) (hsa00020) compared with both Exos and sMVs (**Appendix Fig. 3.2** and **Appendix Table A3.8**).

KEGG *category-iv)* 'RNA transport' (hsa03013) revealed highly-enriched proteins in sMB-Rs such as RAE1, SEC13, UBC9, RANGAP (nuclear pore complex), EIF2, EIF2B, EIF3, EIF5 (translation initiation factor), Y14, MAGOH, EIF4A3, PININ, RNPS1 (exon-junction complex) and FUS, TDP43 (Pre-mRNA processing complex) enriched in sMB-Rs (**Appendix Fig 3.2G**, **H**).

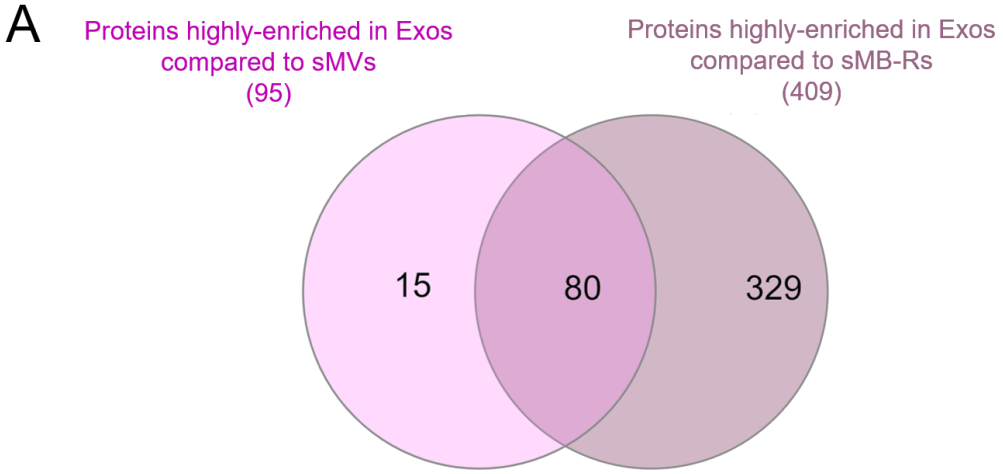
KEGG *category-v)* 'Spliceosome' (hsa03040) showed enriched spliceosome components in sMB-Rs such as SM, U1-70K, U1A, FUS (U1 complex), U2A, SF3B, U2AF, PRP43 (U2 complex), SMU13, SAD1 (U4/U6 complex), SNULL4, PRP8BP (U5 complex) and HNRNP families, SR (common components) (**Appendix Fig. 3.2I, J**).

KEGG *category-vi)* 'Protein processing in endoplasmic reticulum' (hsa04141) demonstrated enriched proteins in sMB-Rs such as OST, CLIMP63, GLCI, GLCII, UGGT, HSP40, PERK, WFS1, SAR1, SEC13/31, NEF, TRAP and BAP31 (**Appendix Fig. 3.2K, L**).

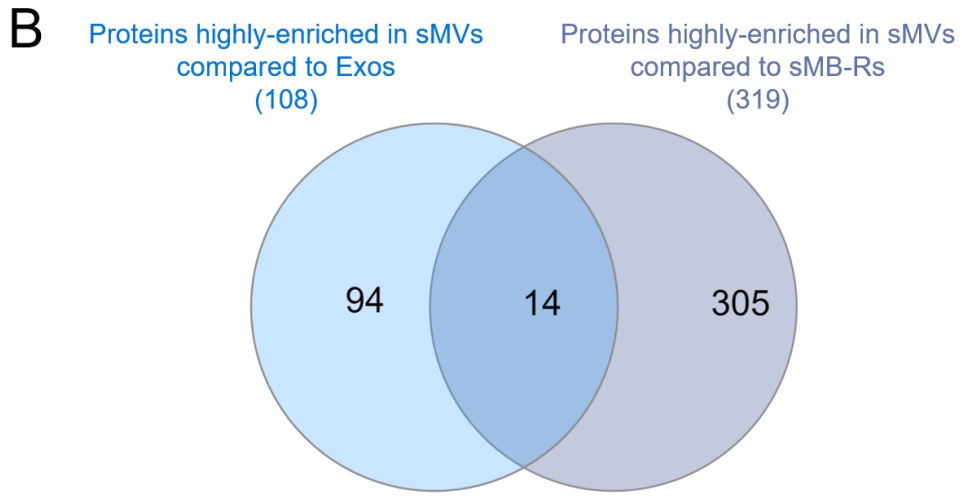
KEGG *category-vii)* 'TCA cycle' (hsa00020) revealed highly-enriched proteins and mitochondrial enzymes such as PCK1 (4.1.1.32), PDC1 (1.2.4.1), LPD1 (1.8.1.4), IDH1 (1.1.1.42), FUM1 (4.2.1.2), DLST (2.3.1.61), and OGDH (1.2.4.2) (**Appendix Fig. 3.2M, N**).

Figure 3.3. Comparative proteomic analysis of Exos, sMVs and sMB-Rs derived from SW480 and SW620 cells. For these analyses EV datasets from SW480/SW620 were combined. **(A)** A two-way Venn diagram of selectively-enriched proteins (log2 fold change >1 or <-1 and pvalue < 0.05) in Exos compared to sMVs and sMB-Rs reveals 80 proteins selectively enriched in Exos. **(B)** Two-way Venn diagram of selectively-enriched proteins in sMVs compared to Exos and sMB-Rs reveals 14 proteins selectively-enriched in sMVs. **(C)** Two-way Venn diagram of selectively-enriched proteins in sMB-Rs compared to Exos and sMVs reveals 492 proteins selectively-enriched in sMB-Rs. **(D)** Identification of enriched gene ontology (GO) term (ranked by protein counts) in Exos, sMVs and sMB-Rs based on selectively-enriched proteins in Exos (80 proteins), sMVs (14 proteins) and sMB-Rs (492 proteins). Red stars indicate commonly-identified GO terms in Exos/sMVs/ and sMB-Rs. Yellow stars indicate commonly-identified GO terms in 2 EV classes. **(E)** Heat map illustration of selectively-enriched proteins in SW480-/SW620-derived Exos, sMVs and sMB-Rs (scale shown is average normalized LFQ subtracted by mean and divided by standard deviation).

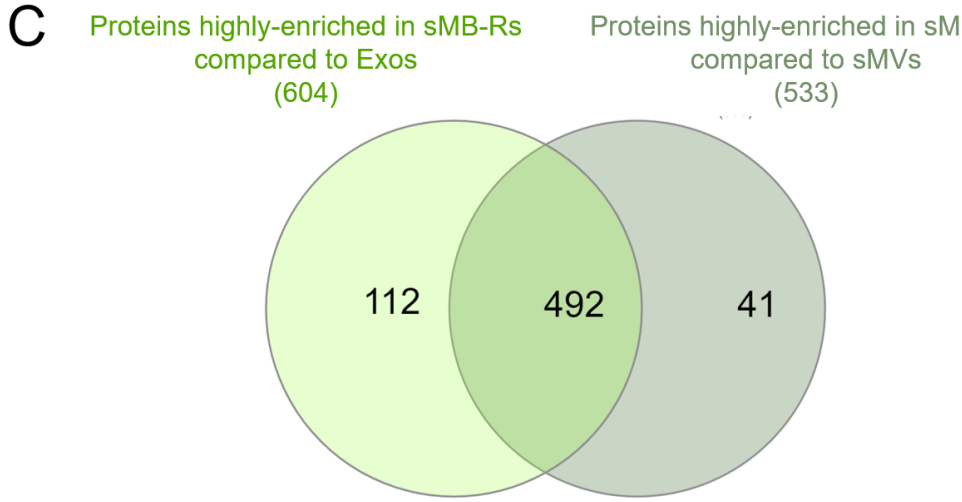
A

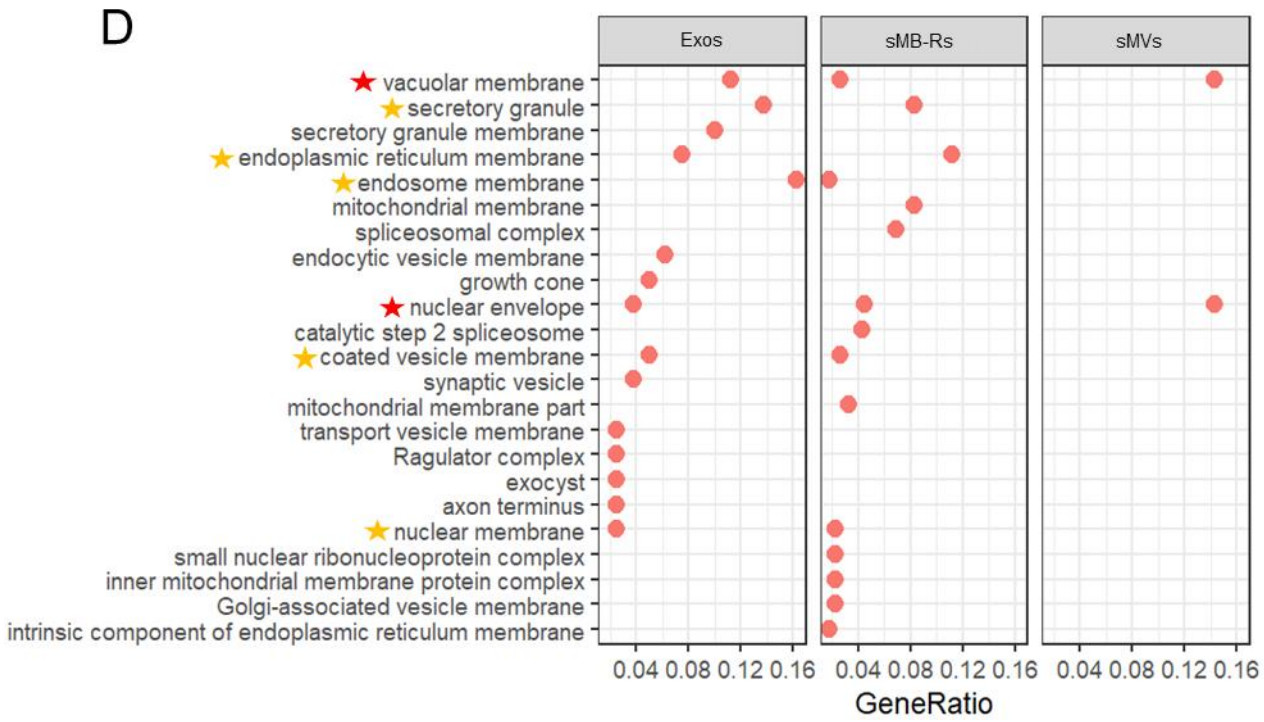

Proteins highly-enriched in Exos compared to sMVs (95)
Proteins highly-enriched in Exos compared to sMB-Rs (409)
15
80
329

B


Proteins highly-enriched in sMVs compared to Exos (108)
Proteins highly-enriched in sMVs compared to sMB-Rs (319)
94
14
305

C


Proteins highly-enriched in sMB-Rs compared to Exos (604)
Proteins highly-enriched in sMB-Rs compared to sMVs (533)
112
492
41


D
Exos
sMB-Rs
sMVs
vacuolar membrane
secretory granule
secretory granule membrane
endoplasmic reticulum membrane
endosome membrane
mitochondrial membrane
spliceosomal complex
endocytic vesicle membrane
growth cone
nuclear envelope
catalytic step 2 spliceosome
coated vesicle membrane
synaptic vesicle
mitochondrial membrane part
transport vesicle membrane
Ragulator complex
exocyst
axon terminus
nuclear membrane
small nuclear ribonucleoprotein complex
inner mitochondrial membrane protein complex
Golgi-associated vesicle membrane
intrinsic component of endoplasmic reticulum membrane
0.04 0.08 0.12 0.16
0.04 0.08 0.12 0.16
0.04 0.08 0.12 0.16
GeneRatio

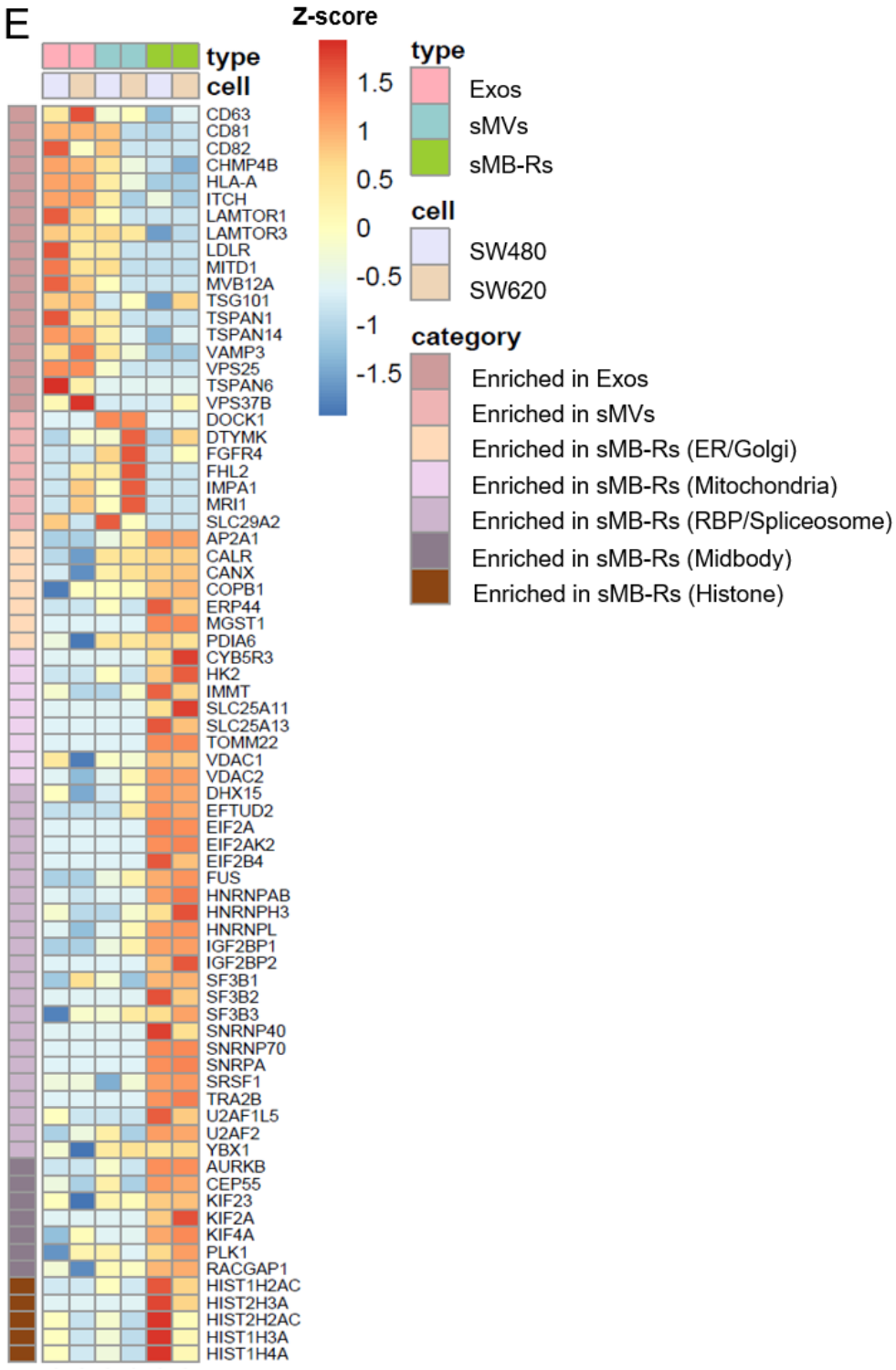

E
type
cell
CD63
CD81
CD82
CHMP4B
HLA-A
ITCH
LAMTOR1
LAMTOR3
LDLR
MITD1
MVB12A
TSG101
TSPAN1
TSPAN14
VAMP3
VPS25
TSPAN6
VPS37B
DOCK1
DTYMK
FGFR4
FHL2
IMPA1
MRI1
SLC29A2
AP2A1
CALR
CANX
COPB1
ERP44
MGST1
PDIA6
CYB5R3
HK2
IMMT
SLC25A11
SLC25A13
TOMM22
VDAC1
VDAC2
DHX15
EFTUD2
EIF2A
EIF2AK2
EIF2B4
FUS
HNRNPAB
HNRNPH3
HNRNPL
IGF2BP1
IGF2BP2
SF3B1
SF3B2
SF3B3
SNRNP40
SNRNP70
SNRPA
SRSF1
TRA2B
U2AF1L5
U2AF2
YBX1
AURKB
CEP55
KIF23
KIF2A
KIF4A
PLK1
RACGAP1
HIST1H2AC
HIST2H3A
HIST2H2AC
HIST1H3A
HIST1H4A
Z-score
1.5
1
0.5
0
-0.5
-1
-1.5
type
Exos
sMVs
sMB-Rs
cell
SW480
SW620
category
Enriched in Exos
Enriched in sMVs
Enriched in sMB-Rs (ER/Golgi)
Enriched in sMB-Rs (Mitochondria)
Enriched in sMB-Rs (RBP/Spliceosome)
Enriched in sMB-Rs (Midbody)
Enriched in sMB-Rs (Histone)

Table 3.1 Selectively-enriched proteins in Exos, sMVs and sMB-Rs secreted from SW480 and SW620 cells

EV class	Category	Gene name	Protein description	Protein Accession (UniProt)	Normalized LFQ intensity[a]					
					SW480-Exos	SW480-sMVs	SW480-sMB-Rs	SW620-Exos	SW620-sMVs	SW620-sMB-Rs
Exos	Exosomal biogenesis	CD9	Tetraspanin	A6NNI4	25,519,078	15,249,895	4,228,553	8,346,331	3,882,327	2,630,692
		SDCBP	Syntenin-1	O00560	5,704,720	2,535,526	267,311	3,124,564	64,040	74,096
		SDCBP2*	Syntenin-2	Q9H190	29,811	-	-	137,412	-	-
		CD81	Tetraspanin	E9PJK1	2,906,182	2,072,424	721,091	2,243,535	968,283	493,968
		CD63	CD63 antigen	F8VV56	962,989	376,641	-	525,147	36,030	34,869
		ARRDC1	Arrestin domain-containing protein 1	Q8N5I2	646,143	360,359	9,284	178,366	47,319	26,081
		CD82	CD82 antigen	P27701	476,529	279,614	-	120,106	35,493	8,852
		CHMP4B	Charged multivesicular body protein 4b	Q9H444	135,214	115,238	19,315	127,888	30,932	38,442
		TSG101	Tumor susceptibility gene 101 protein	Q99816	132,686	113,310	50,004	256,676	100,816	72,311
		CHMP1A	Charged multivesicular body protein 1a	F8VUA2	32,297	12,682	4,571	39,378	-	-
	Vesicle-associated proteins	VPS28	Vacuolar protein sorting-associated protein 28 homolog	Q9UK41	250,149	108,190	36,386	336,014	71,659	77,160
		VPS37B	Vacuolar protein sorting-associated protein 37B	Q9H9H4	96,613	36,452	-	106,115	16,535	-
		VPS25	Vacuolar protein-sorting-associated protein 25	Q9BRG1	63,915	29,509	-	99,477	39,172	39,607
		VAMP3	Vesicle-associated membrane protein 3	Q15836	61,516	36,033	-	28,717	-	-
		VAMP5	Vesicle-associated membrane protein 5	O95183	55,443	26,760	-	45,797	-	-
	Membrane associated proteins	TSPAN1	Tetraspanin-1	O60635	128,571	73,976	-	28,886	-	-
		TSPAN14	Tetraspanin-14	Q8NG11	67,265	19,594	-	79,319	-	-
		TSPAN6	Tetraspanin-6	O43657	32,107	14,280	-	20,362	-	-
		CD99*	CD99 antigen	P14209	3,780	-	-	16,497	-	-
		NECTIN1*	Nectin cell adhesion molecule 1 (CD antigen CD111)	Q15223	2,277	-	-	7,040	-	-
sMVs	Membrane associated proteins	SLC29A2	Equilibrative nucleoside transporter 2	Q14542	12,057	19,309	-	-	4,976	-
		FGFR4	Fibroblast growth factor receptor 4	P22455	-	2,420	-	6,308	10,476	-
	Enzyme	DTYMK*	Thymidylate kinase	P23919	-	18,102	-	-	15,154	-

		IMPA1	Inositol monophosphatase 1	P29218	-	9,695	-	13,144	47,096	25,899
		MRI1	Methylthioribose-1-phosphate isomerase	Q9BV20	-	7,804	-	-	34,210	8,556
sMB-Rs	Histone	HIST1H1C	Histone H1.2	P16403	146,951	103,434	17,457,590	-	-	599,781
		HIST2H3PS2	Histone H3	Q5TEC6	172,367	99,191	7,126,765	30,482	19,164	741,236
		HIST1H2AC	Histone H2A type 1-C	Q93077	-	138,759	2,051,667	-	-	306,244
		HIST2H3A*	Histone H3.2	Q71DI3	-	-	1,149,510	-	-	58,295
	RNA granule protein	IGF2BP1	Insulin-like growth factor 2 mRNA-binding protein 1	Q9NZI8	-	7,184	50,988	-	9,292	82,346
		FUS	RNA-binding protein FUS	P35637	-	4,148	17,101	-	6,866	63,974
		TARDBP	TAR DNA-binding protein 43	Q13148	-	-	7,698	-	8,158	43,250
		IGF2BP2*	Insulin-like growth factor 2 mRNA-binding protein 2	F8W930	-	-	6,518	-	-	10,777
	Ribonucleoprotein/spliceosome	HNRNPA1	Heterogeneous nuclear ribonucleoprotein A1	F8W6I7	189,305	280,594	667,383	204,208	577,883	1,543,355
		HNRNPK	Heterogeneous nuclear ribonucleoprotein K	P61978	102,706	143,624	623,067	63,661	164,451	702,952
		SRSF7*	Serine/arginine-rich-splicing factor 7	A0A0B4J1Z1	-	-	329,624	-	-	368,416
		HNRNPD	Heterogeneous nuclear ribonucleoprotein D0	Q14103	26,432	48,492	211,402	38,889	103,100	283,403
		SRSF1	Serine/arginine-rich-splicing factor 1	J3KTL2	22,406	5,156	176,279	17,424	21,063	394,435
		HNRNPU	Heterogeneous nuclear ribonucleoprotein U	Q00839	6,567	6,574	162,877	-	3,206	89,323
		U2AF1L5	Splicing factor U2AF 35 kDa subunit-like protein	P0DN76	11,168	-	89,469	-	-	60,371
		U2AF2	Splicing factor U2AF 65 kDa subunit	K7ENG2	-	16,353	54,726	4,994	-	28,231
		LARP1*	La ribonucleoprotein domain family member 1	Q6PKG0	-	-	5,240	-	-	3,272
		SFPQ	Splicing factor, proline- and glutamine-rich	P23246	-	698	44,513	-	-	39,108
		SRSF7*	Serine/arginine-rich-splicing factor 7	A0A0B4J1Z1	-	-	299,849	-	-	245,611
		SF3B1	Splicing factor 3B subunit 1	O75533	3,472	7,735	33,276	5,462	-	52,498
		SRSF10*	Serine/arginine-rich-splicing factor 10	Q5JRI1	-	-	20,123	-	-	47,477
		SF3B3	Splicing factor 3B subunit 3	Q15393	870	1,516	18,003	1,698	10,518	90,269
		SF3B2*	Splicing factor 3B subunit 2	E9PPJ0	-	-	6,861	-	-	2,550
	Organelle-associated protein	HSPA5	Endoplasmic reticulum luminal Ca(2+)-binding protein grp78	P11021	17,406	190,663	1,089,414	32,950	305,576	1,269,215
		CYB5B	Cytochrome b5 type B	J3KNF8	341,891	406,353	976,467	-	156,024	614,858
		VDAC1	Voltage-dependent anion-selective channel protein 1	P21796	113,357	204,643	943,804	-	81,780	514,641

		CANX	Calnexin	P27824	16,262	57,326	343,148	5,218	108,205	457,050
		VDAC2	Voltage-dependent anion-selective channel protein 2	P45880	7,296	11,766	189,210	-	29,286	319,459
		ERP44	Endoplasmic reticulum resident protein 44	Q9BS26	-	17,941	77,026	-	-	51,309
		TOMM22*	Mitochondrial import receptor subunit TOM22 homolog	Q9NS69	-	-	53,531	-	-	52,962
	Midbody	RACGAP1	Rac GTPase-activating protein 1	Q9H0H5	18,088	85,367	1,689,821	2,708	38,445	2,506,751
		KIF23	Kinesin-like protein KIF23	Q02241	11,504	38,477	1,062,271	91	14,435	1,733,403
		CEP55	Centrosomal protein of 55 kDa	Q53EZ4	9,834	12,795	97,651	-	-	47,455
		AURKB	Aurora kinase B (Fragment)	J3KT86	-	4,807	52,033	-	-	54,161
		KIF4A	Chromosome-associated kinesin KIF4A	O95239	-	793	40,044	2,015	2,394	189,978
		PLK1	Polo-like kinase 1	P53350	429	3,604	21,875	2,589	2,893	253,333
		KIF2A*	Kinesin-like protein KIF2A	O00139	-	-	11,375	-	-	84,425

a = LFQ (label free precursor intensity) normalized with protein length

** = uniquely identified proteins*

- = undetected in samples

3.3.4 Highly-enriched cancer associated cargo proteins in SW480-/SW620-EV classes that modulate CRC progression

It is well recognized that Exos secreted from human metastatic colorectal cancer cells harbor metastatic factors and signaling pathway components that engage in cross-talk between tumour and stromal cells in the tumour microenvironment[59]. To gain insights into how the proteome of individual EV classes might impact on CRC progression, I performed a differential protein enrichment analysis between SW480-EVs (combined) and SW620-EVs (combined) (see Methods, Section 3.2.6).

This analysis revealed 230 cancer-associated proteins that are highly enriched in combined SW480-EVs (such as CD44, STAT1, MDK, TGM2, EGFR, FAS, CLDN7) and 264 cancer-associated proteins in SW620-EVs such as RICTOR, MACC1, PRKACA, TGFBR2, MET (**Fig. 3.4A**). A list of highly-enriched protein identifications in this analysis is provided in **Appendix Table 3.9**). Next, I performed KEGG pathway analysis on these cancer-associated cargo proteins to gain insights into possible functional roles.

KEGG pathway analysis on SW480-EV highly-enriched cancer-associated proteins (230 proteins) revealed involvement of 7 pathways including ‘Ribosome’ (hsa03010), ‘Cell adhesion molecule’ (hsa04514), ‘Proteosome’ (hsa03050) and ‘Phagosome’ (hsa04145) (**Fig. 3.4B**). In contrast, KEGG pathway analysis on proteins enriched in SW620-EVs (264 proteins) showed 16 pathways that include ‘MAPK signaling pathway’

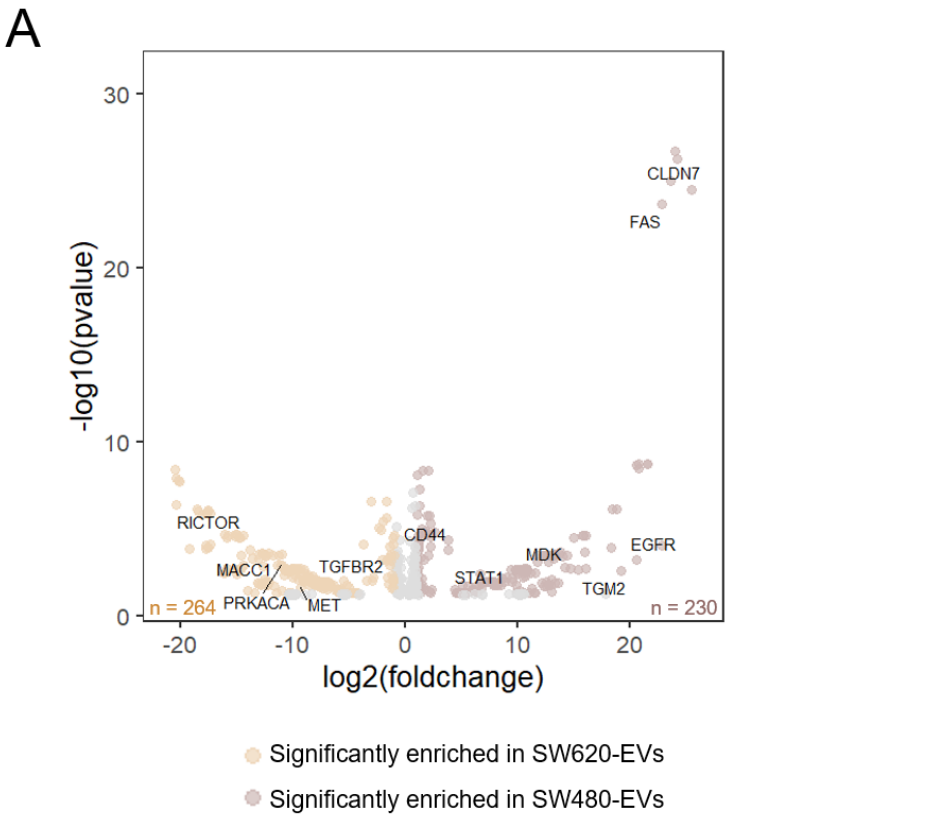
(hsa04010), ‘Proteoglycans in cancer’ (hsa05205), ‘Insulin signaling pathway’ (hsa04931) and ‘ErbB signaling pathway’ (hsa04012) (**Fig. 3.4B**).

I next asked whether the EV classes exhibit different biological functions in cancer progression. Proteomic analysis showed that many cancer progression-associated proteins are highly or uniquely sorted in the separate EV classes. For instance, protein-related to genetic stability such as PARP1 was highly-enriched in SW480-sMB-Rs and histone deacetylase (HDAC1) and its substrate MSH6 (DNA mismatch repair protein) were uniquely detected in SW480-sMB-Rs (**Table 3.2**). Several receptors and transporters are highly enriched in Exos and sMVs (CD44, EGFR, FAS in SW480-Exos/sMVs and MET, FGFR4 in SW620-Exos/sMVs) (**Table 3.2**). Interestingly, chemokine receptor (CXCR4) was uniquely detected in SW480-Exos and TGFBR2, AXL and ABCB1 were uniquely detected in SW620-Exos (**Table 3.2**). Signal transduction proteins such as CLDN7 was highly enriched in Exos and sMVs derived from SW480 cells (**Table 3.2**). Phospholipase D1 (PLD1) was only detected in SW480-Exos and STAT1, GYS1 and CTNNBL1 were uniquely detected in SW480-sMB-Rs (**Table 3.2**). Signal transduction proteins such as PLD2, SMAD5 and TP53RK were uniquely detected in SW620-Exos and LAMTOR3 is highly-enriched in SW620-Exos (**Table 3.2**). PRKACA, MACC1 and RICTOR were highly enriched in Exos, sMVs and sMB-Rs from SW620 compared to Exos, sMVs and sMB-Rs from SW480. Midkine (MDK) growth factors was highly-enriched in SW480-sMB-Rs and growth/differentiation factor-15 (GDF15) was uniquely detected in SW620-sMB-Rs (**Table 3.2**).

Interestingly, ECM remodeling proteins such as tissue transglutaminase-2 (TGM2) and matrix metalloproteinase MMP14 are enriched in SW480-EVs with selective enrichment in sMB-Rs and sMVs, respectively (**Table 3.2**). ADAM15 and tenascin (TNC) were uniquely identified in SW620-Exos and SW620-sMB-Rs, respectively.

In summary, proteomic analysis showed selective enrichment of cancer-associated proteins in different EV classes from primary CRC (SW480) and metastatic CRC (SW620) cell lines. It is interesting to speculate that the different cancer-associated proteins that selectively traffic to the three respective EV classes might act concomitantly in tumour progression and metastasis. However, further biological experiments are necessary to define the precise biological role of the individual EV classes.

Figure 3.4 Identification of cancer progression-related proteins and KEGG pathways in EVs derived from SW480 and SW620 cells. (A) Differential protein enrichment analysis of highly-enriched (log2 fold change > 1, pvalue < 0.05) cancer-associated cargo proteins in SW480-EVs (230 proteins) and SW620-EVs (264 proteins). **(B)** KEGG pathway analysis (ranked by pvalue) of highly-enriched cancer-associated proteins found in SW480-EVs and SW620-EVs.


A
30
20
10
0
-log10(pvalue)
CLDN7
FAS
RICTOR
CD44
MDK
EGFR
MACC1
TGFBR2
STAT1
TGM2
PRKACA
MET
n = 264
n = 230
-20
-10
0
10
20
log2(foldchange)
Significantly enriched in SW620-EVs
Significantly enriched in SW480-EVs

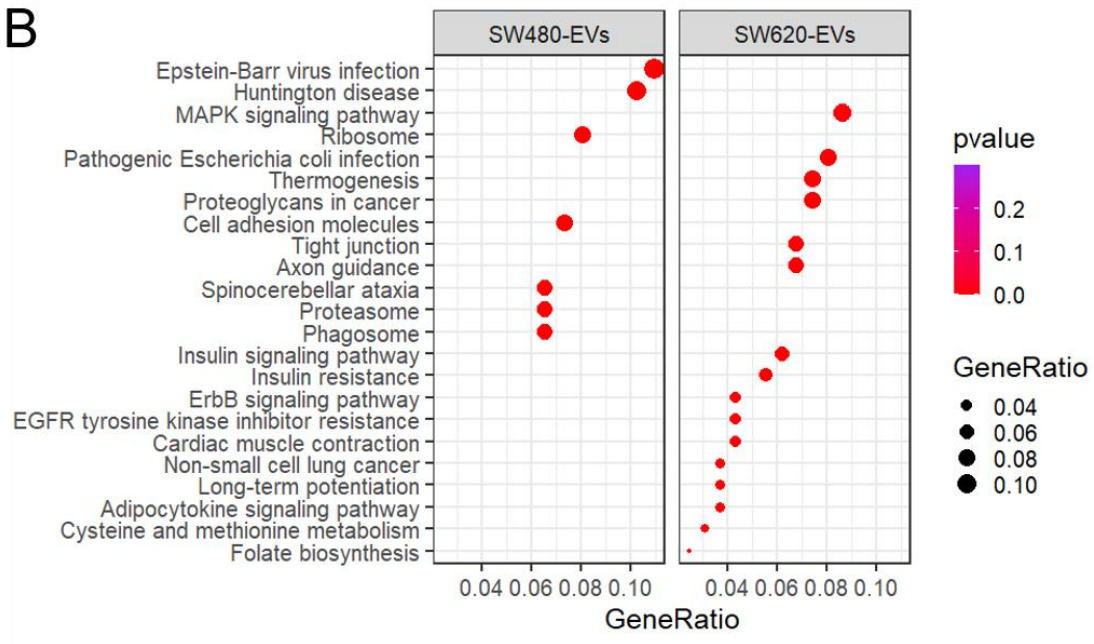

B
SW480-EVs
SW620-EVs
Epstein-Barr virus infection
Huntington disease
MAPK signaling pathway
Ribosome
Pathogenic Escherichia coli infection
Thermogenesis
Proteoglycans in cancer
Cell adhesion molecules
Tight junction
Axon guidance
Spinocerebellar ataxia
Proteasome
Phagosome
Insulin signaling pathway
Insulin resistance
ErbB signaling pathway
EGFR tyrosine kinase inhibitor resistance
Cardiac muscle contraction
Non-small cell lung cancer
Long-term potentiation
Adipocytokine signaling pathway
Cysteine and methionine metabolism
Folate biosynthesis
0.04 0.06 0.08 0.10
0.04 0.06 0.08 0.10
GeneRatio
pvalue
0.2
0.1
0.0
GeneRatio
0.04
0.06
0.08
0.10

Table 3.2 Cancer progression-associated proteins in Exos, sMVs and sMB-Rs secreted from SW480 and SW620 cells

Protein category	Gene name	Protein description	Protein Accession (UniProt)	Normalized LFQ intensity[a]					
				SW480-Exos	SW480-sMVs	SW480-sMB-Rs	SW620-Exos	SW620-sMVs	SW620-sMB-Rs
Genetic instability	PARP1	Poly [ADP-ribose] polymerase 1	P09874	383	2,201	85,855	584	342	3,273
	HDAC2*	Histone deacetylase 2	Q92769	-	-	24,887	-	-	-
	MSH6*	DNA mismatch repair protein Msh6	P52701	-	-	1,504	-	-	-
Receptor/transporter	CD44	CD44 antigen	P16070	470,072	409,137	149,257	57,617	36,076	8,596
	EGFR	Epidermal growth factor receptor	P00533	388,014	384,771	63,350	521	368	430
	FAS	Tumor necrosis factor receptor superfamily member 6	P25445	52,275	39,426	20,300	-	-	-
	MET	Hepatocyte growth factor receptor	P08581	11,471	13,078	-	118,556	62,139	8,963
	CXCR4*	C-X-C chemokine receptor type 4	P61073	2,781	-	-	-	-	-
	FGFR4	Fibroblast growth factor receptor 4	P22455	-	2,420	-	6,308	10,476	-
	TGFBR2*	TGF-beta receptor type-2	P37173	-	-	-	10,279	-	-
	AXL*	Tyrosine-protein kinase receptor	M0R0W6	-	-	-	1,436	-	-
	ABCB1*	Multidrug resistance protein 1	P08183	-	-	-	338	-	-
Signal transduction	CLDN7	Claudin (Fragment)	K7EP40	707,695	967,891	265,199	-	-	-
	LAMTOR3	Regulator complex protein LAMTOR3	Q9UHA4	163,164	-	-	663,677	-	106,844
	PRKACA	cAMP-dependent protein kinase catalytic subunit alpha	P17612	18,481	17,668	13,614	46,396	66,027	77,088
	MACC1	Metastasis-associated in colon cancer protein 1	Q6ZN28	9,017	4,772	5,786	35,719	40,552	27,503
	PLD1*	Phospholipase D1	Q13393	6,330	-	-	-	-	-
	RICTOR	Rapamycin-insensitive companion of mTOR	Q6R327	1,703	2,677	-	19,264	14,593	4,178
	CTNNBL1*	Beta-catenin-like protein 1	Q8WYA6	-	-	1,850	-	-	-
	STAT1*	Signal transducer and activator of transcription 1-alpha/beta	P42224	-	-	10,910	-	-	-
	GYS1*	Glycogen synthase	P13807	-	-	8,829	-	-	-
	PLD2*	Phospholipase D2 (Fragment)	I3L1F3	-	-	-	18,942	-	-
	SMAD5*	Mothers against decapentaplegic homolog 5	Q99717	-	-	-	5,114	-	-
	TP53RK*	TP53-regulating kinase	Q96S44	-	-	-	813	-	-

Growth factor	MDK	Midkine (Fragment)	E9PPJ5	19,889	-	413,613	-	-	-
	GDF15*	Growth/differentiation factor 15	Q99988	-	-	-	-	-	10,301
ECM remodeling enzyme/ECM protein	TGM2	Transglutaminase-2	P21980	652,639	660,063	1,411,989	1,075	584	16,397
	MMP14	Matrix metalloproteinase-14	P50281	6,401	21,885	5,236	4,129	-	-
	ADAM15*	Disintegrin and metalloproteinase domain-containing protein 15	Q13444	-	-	-	1,958	-	-
	TNC*	Tenascin	F5H7V9	-	-	-	-	-	6,919

a = LFQ (label free precursor intensity) normalized with protein length

** = uniquely identified proteins*

- = undetected in samples

3.4 Discussion and conclusions

In this chapter, I performed a detailed comparative proteome analysis of three EV classes (Exos, sMVs and sMB-Rs) secreted from isogenic human primary and metastatic colorectal cancer cells - SW480 cells (surrogate of adenocarcinoma) and SW620 (surrogate of metastatic colon cancer)[128].

Firstly, Exos, sMVs and sMB-Rs were purified in high yield from SW480/ SW620 CM using a combination of differential centrifugation and isopycnic buoyant density (iodixanol/ OptiPrep™) centrifugation (**Fig. 3.1A**). The yields of SW480-/ SW620-derived Exos (from 100K-10K pellet at buoyant density = 1.10 g/mL), sMVs (from 10K pellet at buoyant density = 1.10 g/mL), and sMB-Rs (from 10K pellet at buoyant density = 1.15 g/mL) from 180 mL CM were 868-987 µg, 749-827 µg, and 463-660 µg, respectively. TEM revealed sMVs and sMB-Rs are more ellipsoid in shape and heterogenous in size (100-500 nm) compared to Exos, which displayed a smaller size distribution range (50-200 nm) (**Fig. 3.1F**). Nanoparticle tracking analysis showed mean particle diameters of 186.2±3.4 nm/183.1±2.4, 350±28.4 nm/ 337.1±9.3 nm and 404.9±13.6 nm/401.2±1.7 nm for SW480/SW620-Exos, sMVs, and sMB-Rs, respectively (**Fig. 3.1G**). The diameters of Exos, and sMVs are similar to those reported in previous studies[59, 134, 163, 204, 205]. Western blot analysis revealed ALIX, TSG101, CD63, CD9 and CD81 were more enriched in Exos than in sMVs (**Fig. 3.1H**), a finding consistent with previous studies[163, 164, 204]. The

midbody centraspindlin complex components KIF23/MKLP1 and RACGAP1[96, 99] are selectively enriched in sMB-Rs (**Fig. 3.1 H**).

An interesting finding was that Exos, sMVs and sMB-Rs are molecularly distinct. Interestingly, the proteome of sMB-Rs is dissimilar to Exos and sMVs as indicated in PCA and clustering plots (**Fig. 3.2C** and **D**). Differential protein enrichment analysis revealed 80, 14 and 492 proteins are selectively enriched (using a cutoff, >1 or <-1 log2 fold change and pvalue < 0.05) in Exos, sMVs and sMB-Rs, respectively (**Fig. 3.3A, B** and **C**).

In Exos, 80 selectively-enriched proteins were identified. These include exosomal protein markers[161, 163, 206] such as CD63, CD81, TSG101 (**Fig. 3.3E**). KEGG pathway analysis of these selectively-enriched proteins showed that they are involved in pathways such as 'Endocytosis' (hsa04144) and 'SNARE interactions in vesicular transport' (hsa04130) and 'Cell adhesion molecules' (hsa04514) (**Appendix Fig. 3.2A-F**). This finding further supports the notion that Exos biogenesis involves endosome dynamics and trafficking, as previously reported[79, 207].

When compared to Exos and sMB-Rs, sMVs were found to contain 14 selectively-enriched proteins (**Fig. 3.3B**, **Appendix Table 3.6**) and GO analysis revealed that they are involved in vacuolar membranes and nuclear envelopes (**Fig. 3.3D**). Interestingly, sMVs did not share any of KEGG pathways identified Exos and sMB-Rs. sMVs contain enriched proteins such as DOCK1, DTYMK, FGFR4, FHL2, IMPA2, MRI1 and SLC29A2 not observed in Exos and sMB-Rs (**Fig. 3.3E**). Interestingly, proteins purported to be

involved in sMV biogenesis such as ARRDC1[208], ARF1[194], ARF6[195], RHOA[209] and ANXA5[210] are not only enriched in sMVs but also in Exos derived from SW480 and SW620 cells in this current study. This paradox raises the specter of EV purity – for example, possible cross contamination with other EV types is highly probable, especially in the case of sMVs and Exos which overlap in size distribution and have similar biophysical properties such as buoyant density.

Concerning the 492 selectively-enriched proteins observed in sMB-Rs (compared to Exos and sMVs, **Fig. 3.3C**), GO analysis revealed terms such as 'small nuclear ribonucleoprotein (RNP) complex', 'spliceosomal complex', and organelles such as 'mitochondrial membrane part', 'Golgi-associated vesicle membrane', 'endoplasmic reticulum membrane' and 'nuclear membrane' (**Fig. 3.3D** and **E**). KIF23/ MKLP1, CEP55, and RACGAP1, which are core structural proteins of midbody/midbody remnants were highly enriched in sMB-Rs, as well as key enzymes involved in midbody formation (AURKB, PLK1)[99] (**Fig. 3.3E**). KEGG pathway analysis of the 492 proteins that selectively traffic to sMB-Rs showed pathways such as 'RNA transport' (hsa03013), 'Spliceosome' (hsa03040), 'Protein processing in endoplasmic reticulum' (hsa04141) and 'Citrate cycle (TCA cycle) and mitochondrial enzymes' (hsa00020) (**Appendix Fig. 3.2G-N**).

The observed enrichment of RNA granule and mitochondrial proteins in sMB-Rs is intriguing and raises the question of contamination of my sMB-R preparations. The isolation and purification methods used in this thesis are

similar to RNA granule and mitochondrion isolation methods[211, 212]. However, TEM images of sMB-Rs derived from both SW480 and SW620 cells (**Fig. 3.1F**) did not show any evidence of mitochondria (see additional TEM images of sMB-Rs in **Appendix Fig. 3.4**) Besides, the size distribution range of intact mitochondria is 500-1200 nm[213] and should have been clearly noticeable using TEM analysis. Interestingly, Skop and colleagues reported the identification of mitochondrial (26%), nuclear (16%) and ribosomal (13%) proteins in the midbody proteome[214] and RNA localization, nuclear transport, RNA splicing, mitochondrion organization as top fold enrichment GO annotation profiles of midbody proteome and interactome using PANTHER[215]. Furthermore, 25/492 selectively-enriched proteins in my sMB-R preparation (RACGAP1, KIF4A, KIF23, CEP55, PLK1, for example) co-identified with proteins listed in the MiCroKITS v.4.0 database[216] (a database of proteins temporally and spatially localized in distinct subcellular positions including midbody, centrosome, kinetochore, telomere, and mitotic spindle during cytokinesis (cell division/ mitosis) (http://microkit.biocuckoo.org, **Appendix Table 3.7**). These proteins, for instance, KIF4A and KIF23 are important for MB formation and cytokinesis[217, 218]. Down regulation of KIF23 by si-RNA decreased the proliferation of glioma cells[219] and abrogated midbody formation and completion of cytokinesis[220]. In another study by Addi *et al*., enriched vesicular traffic transport and protein-translation related proteins were reported in the 'Flemmingsome' using STRING functional association network[103]. I also compared my sMB-R proteome with the 'Flemmingsome' proteome from Addi

and colleagues. Interestingly, 271/492 enriched proteins in my purified sMB-Rs are found in the 'Flemmingsome' proteome. These identified proteins include RNA granule proteins (FUS, IGF2BP1, TARDBP), ribonucleoproteins and splicing factors (HNRNPs, SFPQ, SF3B3) endoplasmic reticulum proteins (CALR, CANX), mitochondrial proteins (VDAC1, VDAC2, SLC25A3, SLC25A5, SLC25A6), histones (HIST1H1C, HIST2H3A) and midbody proteins (KIF2A, KIF4A, KIF23, RACGAP1, PLK1, CEP55) (**Appendix Table 3.7**), indicating similarity between my sMB-R proteome and that of the 'Flemmingsome'. While there is accumulating evidence of the presence mitochondrial proteins[221], RNA binding proteins (RBPs)[222] in EVs such as Exos, the comparative proteomic analysis in this chapter reveals that mitochondrial and RBPs are also enriched in sMB-Rs. Clearly, further studies are required to discern whether RNA granules, mitochondrial proteins and other cellular organelle proteins are physiologically relevant components of sMB-Rs.

Lastly, cancer progression-associated proteins are uniquely and highly sorted into EVs[59, 164, 222]). In addition, receptors such as CXCR4 and TGFBR2/AXL are uniquely detected in SW480-Exos and SW620-Exos, respectively. Interestingly, EGFR and MET receptors are highly enriched in SW480- and SW620-Exos, respectively. These receptors have been shown to promote cancer progression and found in Exos from other cancer cells, for example - EGFR in non-small cell lung cancer and gastric cancer-derived Exos[223, 224] and MET in melanoma-derived Exos[197]. A possible reason for

receptor enrichment in Exos is that receptors such as EGFR, HER2, ERBB3 and ERBB4 bind to their cognate ligands and are then internalized into the intercellular early endosome which further develops to late endosome and multivesicular body (MVB), respectively[225]. MVBs that contain receptor-intact intraluminal vesicles (exosomes) can fuse either with lysosome (leading to proteolytic degradation) or traffic to the plasma membrane whereupon Exos are released into the extracellular milieu [226]. In sMB-Rs, I identified HDAC2 and its substrate MSH6[227] uniquely detected in SW480-sMB-Rs (**Table 3.2**). Deacetylation of DNA mismatch repair proteins (MSHs) by HDAC1 leads to destabilization of MSHs resulting in genetic instability[228]. Furthermore, cancer progression-proteins such as STAT1[229]/TGM2[230]/MDK[231] and PRKACA[232] are highly enriched in sMB-Rs from SW480 and SW620 cells, respectively (**Table 3.2**). While the functionalities of these proteins in other EV functions are not clear, Shafiq et al. have indicated TGM2 derived from a mixture of sMVs and sMB-Rs, enhances anchorage-independent growth (Anoikis resistance) in MDCK cells which is a key cancer hallmark[233], while TGM2 derived from Exos has been shown to induce metastatic niche formation[234].

In summary, the proteome of sMB-Rs is distinct from Exos and sMVs; the sMB-R proteome is high enriched with mitochondrial proteins (membrane proteins and enzymes), RNA granule proteins, splicing factors, ribonucleoproteins, histone subunits, translation initiation factors and components of midbodies. Exos are highly-enriched in tetraspanins/glycoproteins (TSPAN1, TSPAN14, CD63, CD81, CD82) and

ESCRT components (TSG101, CHMP1A, CHMP4B). sMVs are highly-enriched in metabolic enzymes (DTYMK, IMPA1 and MRI) and membrane-associated proteins (SLC29A2, FGFR4). This chapter provides, for the first time, an in-depth comparative proteomic analysis of three EV classes (Exos, sMVs and sMB-Rs) which were purified simultaneously from two CRC cell types (SW480 and SW620 cells). This comparative analysis provides a better understanding of the protein cargos in three EV classes (Exos, sMVs, and sMB-Rs). I believe this chapter paves the way to advancing the characterization of EV classes and in doing so may impact on our understanding of intercellular communication.

3.5 Future perspectives

There are some questions arising from this chapter that have not been fully addressed due to time constraints and unavailability of reagents and new technologies. To improve this chapter, I would like to validate several enriched proteins shown in **Fig. 3.3E** using Western blot analysis. This would enable me to identify stereotypic protein markers for characterization of EV classes, especially in body fluids. I would also like to perform extensive purification of sMB-Rs using an immunoprecipitation technique with anti-KIF23 and anti-RACGAP1 to capture sMB-Rs and perform mass spectrometry on immunocaptured sMB-Rs. This could answer the vexed question of whether mitochondria, ER and RNA granules co-purify with or are integral components of sMB-Rs. For example, I could use anti-TOMM22 and anti-FUS/IGF2BP2 antibodies for immunoprecipitation to confirm whether mitochondria and RNA granules are integral cargo of sMB-Rs. Ultimately, it would be important to investigate whether different EV classes derived from CRC cell lines (and other cancer types) elicit different biological functions using a range of recipient cells.

Chapter 4

Comparative transcriptomic analysis of Exos, sMVs and sMB-Rs released from the isogenic colorectal cell lines SW480 and SW620

4.1 Introduction

It is well established that EVs contain genetic material such as DNA fragments and a broad spectrum of RNA species including protein-coding (mRNA), non-protein coding (small and long non protein-coding RNAs), and pseudogene transcripts[79]. Because RNA species are diverse, they can regulate cellular activities at many levels including RNA metabolism, RNA transcription and protein translation[235]. It is well recognized that EV-derived RNAs can be horizontally transferred to recipient cells to mediate biochemical and phenotypic changes in recipient cells[236]. EV encapsulated RNAs are gaining much attention as body-fluid-based diagnostic biomarkers of disease[237] due to their resistance against degradation [238]. Importantly, RNA profiles of circulating EVs in human patients have been shown to reflect (mirror) the RNA profile of disease biopsy material[57]. Therefore, the study of EV RNA profiles not only paves a way to understand the role of EV-derived RNAs on intercellular communication but it also provides a source of potential biomarkers of disease.

Previously, my laboratory reported EV-derived RNA analyses, including miRNA, mRNA, lncRNA, pseudogene transcripts and fusion genes in Exos and sMVs derived from the human colorectal cancer LIM1863[57, 58]. We also reported Exo-derived microRNA (miRNAs) signatures from the isogenic human colorectal cancer cell lines SW480 (adenocarcinoma) and SW620 (lymph-node metastatic cancer)) for the purpose of studying the role of

miRNAs on cancer progression, as well as potential metastatic cancer detection[112]. In my previous chapter (**Chapter 3**), I provided an insight into proteomic analysis of Exos, sMVs and sMB-Rs simultaneously isolated from the culture media (CM) of SW480-/ SW620 cells. In this chapter, I plan to use next-generation RNA sequencing in conjunction with bioinformatics tools to undertake a comparative study of the transcriptome profiles of Exos, sMVs and sMB-Rs to ascertain similarities and differences of RNA signatures in EV classes. In this chapter, I also plan to interrogate the fusion gene expression of SW480- and SW620-EV classes in order to gain insights into the possible role of EV-encapsulated fusion genes in CRC cancer progression.

4.2 Methods

4.2.1 Cell culture and large-scale purification of Exos, sMVs and sMB-Rs.

SW480 and SW620 cell lines were cultured as described in **Chapter 2**, Sections 2.2. For each biological replicate, large-scale purification of Exos, sMVs and sMB-Rs was from 180 mL culture media (CM) acquired over 18 days (30 mL CM harvested each day, 6 days for each biological replicate). The purification method for Exos, sMVs and sMB-Rs is described in **Chapter 3** (Section 3.2.1).

4.2.2 Total RNA isolation and quality control

Total RNA from SW480-/ SW620-derived sMVs, sMB-Rs, Exos and parental cell lysates was extracted by TRIzol™ reagent (Invitrogen). Total RNA from each sample was incubated with 1 mL TRIzol™ reagent (Invitrogen) for 5 min at 25 °C then 0.2 mL chloroform was added. Samples were vortexed vigorously for 15 sec) and further incubated at RT for 3 min. Total RNA was harvested from the aqueous phase by centrifugation at 12,000 x *g*, 4 °C for 15 min, mixed with 5-10 µg of glycogen (20 mg/mL aqueous glycogen from Invitrogen) and isopropyl alcohol (0.5 mL isopropyl alcohol/1 mL aqueous phase) and incubated at 25 °C for 10 min. Total RNA was pelleted by centrifugation at 12,000 x *g*, 4 °C for 15 min and then. washed once with 75% (v/v) aqueous ethanol, air-dried for 5 min and re-dissolved in RNase-free

water. The quality and composition of RNA samples were evaluated using an Agilent 2100 Bioanalyzer according to the manufacturer's instructions (Agilent Technologies, USA, 2100 Expert Software User's Guide). Isolated RNA samples were then transported to Beijing Genomics Institute (BGI, China) for RNA quantification and quality control (RNA Bioanalyzer), cDNA construction library and next-generation RNA sequencing (Illumina HiSeq4000 system).

4.2.3 cDNA library construction and strand-specific transcriptome sequencing

RNA libraries of EV subpopulations were constructed using Illumina, HiSeq 3000/4000 PE Cluster Kit and sequenced by HiSeq4000 system (Illumina), according to the manufacturer's protocol. Briefly, rRNA in total RNA sample was depleted by Ribo-Zero depletion kit (Illumina). Total RNA was fragmented and transcribed to be single strand DNA. The second-strand cDNA was then synthesized using dUTP, dATP, dGTP and dCTP as nucleotides. Adenine nucleotide was conjugated with first- and second-strand cDNA at 3' end using PCR technique. Adaptors were then conjugated at both 3' and 5' ends of cDNA. cDNA was enriched using PCR technique with dTTP, dATP, dGTP and dCTP as substrates. After that the cDNA library was sequenced using the Illumina HiSeq4000 system and RNA sequence of all samples was outputted as raw FASTQ files (paired end reads). Three biological replicates were performed for each sample.

4.2.4 Transcriptome sequencing analysis

Raw reads were cleaned by removing adaptor sequences, low quality reads containing >50% bases with quality [QA] ≤15 and reads with >2% undefined nucleotides[57]. Clean reads were aligned using HISAT[239] (v.0.1.6 beta, http://www.ccb.jhu.edu/software/hisat/index.shtml) and annotated using StringTie[240] (v.1.3.6, https://ccb.jhu.edu/software/stringtie/) with Human Ensembl GRCh38 v.96 (https://asia.ensembl.org/Homo_sapiens/Info/Index). Aligned reads were counted using featureCount (Rsubread[241], v.4.0, http://bioinf.wehi.edu.au/featureCounts/) operated in R. Raw read counts were converted to fragments per kilobase of transcript per million mapped reads (FPKM) followed formula in **Appendix Table 4.1**. To increase reliability of down-stream data analysis and RT-qPCR sensitivity, less than or equal 1.5 FPKM-transcripts in all samples were removed. Transcripts were classified into four categories: i) protein-coding, ii) long non-coding, iii) pseudogene transcripts and iv) other including short non-coding from Ensembl database using BiomaRT[242] (v.3.11, https://www.bioconductor.org/packages/release/bioc/html/biomaRt.html).
(The 3 digits such as 201, 203 are identifiers of transcript indicating transcript biotypes in ENSEMBL database[243])

4.2.5 Fusion gene discovery

Forward and reverse clean reads of samples (sMVs, sMB-Rs, Exos and cell lysate) were aligned with Human Ensembl GRCh38 v.96 using ChimeraScan[244] (v.0.4.5, http://chimerascan.googlecode.com). Lowly-expressed fusion genes (total fragment <10) were removed. Previously annotated fusion genes were identified with The Cancer Genome Atlas (TCGA) RNA-Seq analysis[245].

4.2.6 Differential transcript expression analysis

Differential transcript expression analysis was performed based on raw read counts for each sample using DESeq2[246] package (v.3.11, https://bioconductor.org/packages/release/bioc/html/DESeq2.html). Briefly, transcript expression was normalized using DESeq method[246]. Pvalues were calculated using Benjamini-Hochberg method[200] and normalized transcript expression was presented as log2.

EV class comparison consists of 3 specific comparisons: i) SW480-/ SW620-Exos vs SW480-/ SW620-sMVs, ii) SW480-/ SW620-Exos vs SW480-/ SW620-sMB-Rs, and iii) SW480-/SW620-sMVs vs SW480-/ SW620-sMB-Rs.

For cancer-progression-related transcript identification, SW480-combined EV classes were compared with SW620-combined EV classes.

Log_2 fold change and p-value were used to identified highly-enriched transcripts/fusion genes (Log_2 fold change <-1 or >1, p-value < 0.05).

4.2.7 Gene Ontology (GO) analysis

GO analysis was performed based on highly-enriched transcripts in each comparison (from Section 4.2.6) using clusterProfiler[201] (v.3.11, https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html) in R with stringent parameters (number of permutation: 1,000, minimum gene size: 20, maximum gene size: 500 and pvalue <0.05). Gene ontology (cellular component) of RNA-binding proteins was identified using gProfiler-g:GOst functional profiling (https://biit.cs.ut.ee/gprofiler/gost)[247].

4.2.8 Prediction of RNA-protein interactions

All human RNA-binding proteins were obtained from The Encyclopedia of RNA Interactomes (ENCORI) database[248] using curl (command line tool, v.7.73.0, https://curl.haxx.se/docs/manpage.html) operated in Ubuntu for Windows 10.

4.2.9 Data visualization

Gene Ontology and RNA-protein interactions were visualized using Cytoscape (v.3.7.1, https://cytoscape.org/). Principle component analysis, dot, box, ridge and volcano plots were visualized using ggplot (v.3.3.2, https://ggplot2.tidyverse.org/) in R. Heatmaps were visualized using pheatmap (v.1.0.12, https://www.rdocumentation.org/packages/pheatmap/versions/1.0.12) in R.

Venn diagram was generated using a web-based tool (http://www.interactivenn.net/)[202].

4.3 Results

4.3.1 Generation of transcript profiles of Exos, sMVs and sMB-Rs secreted from SW480 and SW620 cells

Total RNA was isolated from SW480-/SW620-derived Exos, sMVs and sMB-Rs – see **Chapter 3**, Section 3.2.1 for purification methods. RNA yields were 150.67±75.18 ng, 166.00±69.55 ng, 167.00±79.54 ng, 116±52.72 ng, 162.33±66.67 ng and 116.33±82.62 ng for SW480-Exos, SW480-sMVs, SW480-sMB-Rs, SW620-Exos, SW620-sMVs and SW620-sMB-Rs, respectively (RNA quality data shown in **Appendix Fig. 4.1**). RNA sequencing data were subjected to the transcriptomic analysis pipeline shown in **Fig. 4.1A**. RNA Sequencing analysis detected 92,289 transcripts in total for the 6 EV samples (SW480-Exos, SW480-sMVs, SW480-sMB-Rs, SW620-Exos, SW620-sMVs and SW620-sMB-Rs) of which 16,347 transcripts expressed FPKM >1.5 at least in one biological replicate; 5,954, 5,890, 5,808, 5,721, 5,814 and 5,933 transcripts exhibited a FPKM value more than 1.5 (in all three biological replicates) in SW480-Exos, SW480-sMVs, SW480-sMB-Rs, SW620-Exos, SW620-sMVs and SW620-sMB-Rs, respectively (**Fig. 4.1B** and **C,** a list of top 50 transcript identifications shown in **Appendix Table 4.2**).

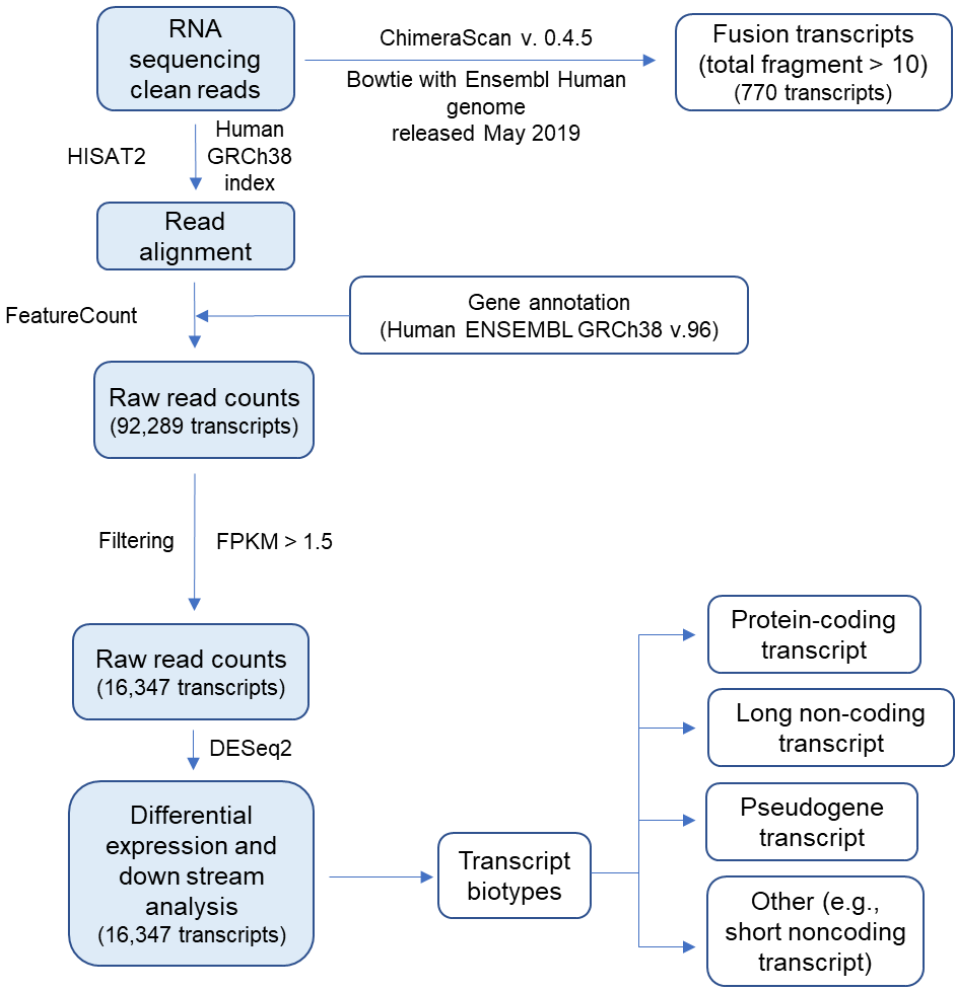
A three-way Venn diagram of transcripts identified in SW480-Exos, -sMVs and -sMB-Rs reveals 4971 commonly identified transcripts and 398, 220, and 509 uniquely identified transcripts in SW480-Exos, sMVs, and sMB-Rs, respectively (**Fig. 4.1B**). In SW620-derived EVs, 4577 transcripts were

commonly identified while 433, 298, and 894 transcripts were uniquely identified in SW620-Exos, sMVs, and sMB-Rs, respectively (**Fig. 4.1C**). To identify unique transcripts for each EV class, I focused on transcripts that were detected in both SW480- and SW620-EVs. In the case of sMB-Rs, 509 and 984 unique transcripts were found in SW480-/SW620-sMB-Rs, respectively (**Fig. 4.1 B, C**) and of these 134 transcripts were selectively enriched (i.e., found in both SW480 and SW620-sMB-Rs); of these and the top five average FPKM transcripts are MT-TG-201 (Mt-tRNA), MT-TH-201 (Mt-tRNA), MT-TR-201 (Mt-tRNA), MIR4479-201 (miRNA) and FBRSL1-203 (lncRNA) (for a list of selectively-enriched sMB-R transcripts see **Appendix Table 4.3**). In the case of Exos, 398 and 433 transcripts were uniquely identified in SW480-/SW620- Exos, respectively, with 34 unique transcripts found to be present in both (of these, the top 5 average FPKM transcripts are SNORD99-201 (snoRNA), HSP90AA1-205 (retained intron), HMGCS1-207 (retained intron), MIR7111-201 (miRNA) and AC092681.1-201 (pseudogene) (see full list of selectively-enriched Exos transcripts in **Appendix Table 4.3**). For sMVs, 220 and 298 transcripts were uniquely identified in SW480-/SW620-sMVs, respectively with 9 transcripts common to both (the top 5 average FPKM transcripts were RF00019.242-201 (misc RNA), AC018752.1-201 (lncRNA), RGS12-219 (lncRNA), FKRP-220 (lncRNA) and OR2A1-AS1-201 (lncRNA) – see list in **Appendix Table 4.3**). Principal component analysis (PCA) demonstrated closer correlation of transcript expression for Exos and sMVs compared to sMB-Rs from the same cell type. (**Fig. 4.1D**). Inspection of **Fig.**

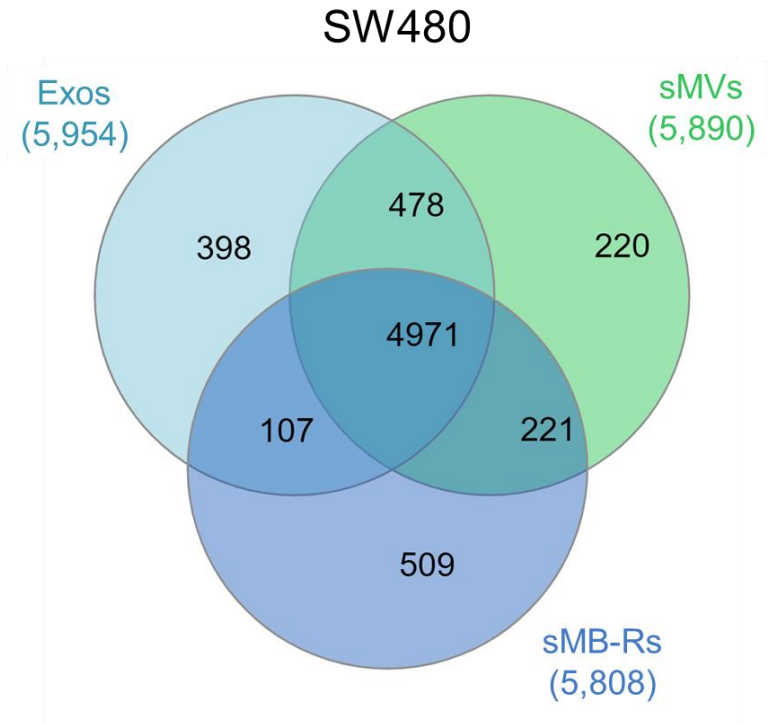
4.1E shows that the most abundant transcript biotypes in all 6 EV samples are protein-coding transcripts (mRNA) followed by long non-coding transcripts, other (e.g., short non-coding transcripts), and pseudogene transcripts (a list of transcript biotype classifications is given in **Appendix Table 4.4**).

Figure 4.1 Transcriptomic analysis pipeline and transcript profiling of Exos, sMVs and sMB-Rs derived from SW480 and SW620 cells. (A) HISAT2 identified 16,347 transcripts (FPKM>1.5) across all samples (Exos, sMVs and sMB-Rs) using Human GRCh38 index and gene annotation (Human ENSEMBL GRCh38, v.96). Differential transcript expression was performed using DESeq2 package with cutoff Log2 fold change >1 or <-1 and pvalue < 0.05 for highly-enriched transcript in each comparison (see Section 4.2.6). Transcripts were further annotated into four subclasses: protein-coding, long non-coding, pseudogene and other (e.g., short non-coding). Fusion genes were identified using ChimeraScan v.0.4.5. and those with total fragment > 10 were considered for further analysis (FPKM = Fragments Per Kilobase of transcript per Million mapped reads). **(B)** A three-way Venn diagram of transcripts identified in SW480-Exos, -sMVs and -sMB-Rs reveals 4971 transcripts were commonly identified, while 398, 220, and 509 transcripts were uniquely identified in SW480-Exos, sMVs, and sMB-Rs, respectively. **(C)** A three-way Venn diagram of transcripts identified in SW620-Exos, -sMVs and -sMB-Rs reveals 4577 transcripts were commonly identified, while 433, 298, and 894 transcripts were uniquely identified in SW620-Exos, sMVs, and sMB-Rs, respectively. **(D)** Principal component analysis (PCA) of total transcriptome of Exos, sMVs and sMB-Rs derived from SW480 and SW620 cells. **(E)** Number of protein-coding, long non-coding, pseudogene and other transcripts identified in Exos, sMVs and sMB-Rs derived from SW480 and SW620 cells (FPKM >1.5 in 3 biological replicates).

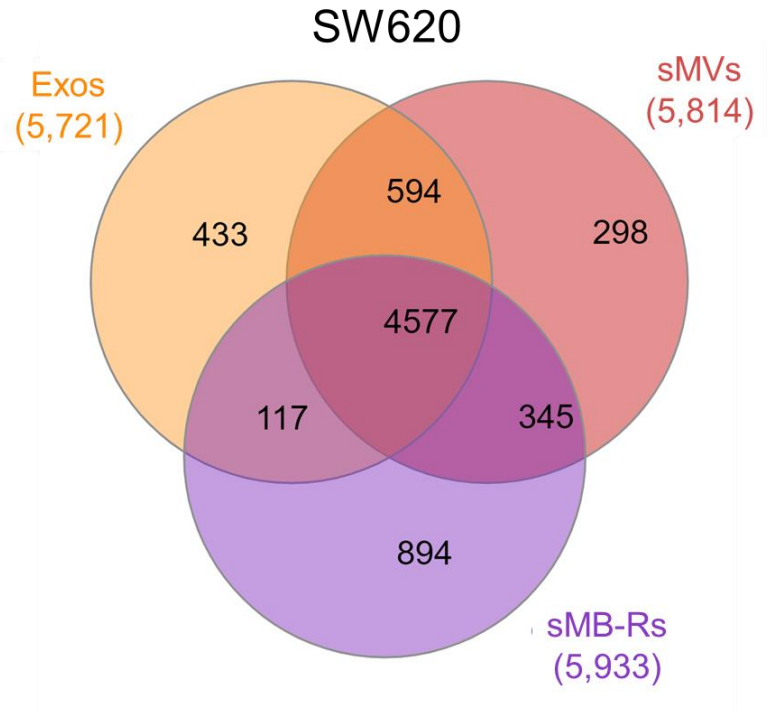
A


RNA sequencing clean reads
ChimeraScan v. 0.4.5
Bowtie with Ensembl Human genome released May 2019
Fusion transcripts (total fragment > 10) (770 transcripts)
HISAT2
Human GRCh38 index
Read alignment
FeatureCount
Gene annotation (Human ENSEMBL GRCh38 v.96)
Raw read counts (92,289 transcripts)
Filtering
FPKM > 1.5
Raw read counts (16,347 transcripts)
DESeq2
Differential expression and down stream analysis (16,347 transcripts)
Transcript biotypes
Protein-coding transcript
Long non-coding transcript
Pseudogene transcript
Other (e.g., short noncoding transcript)

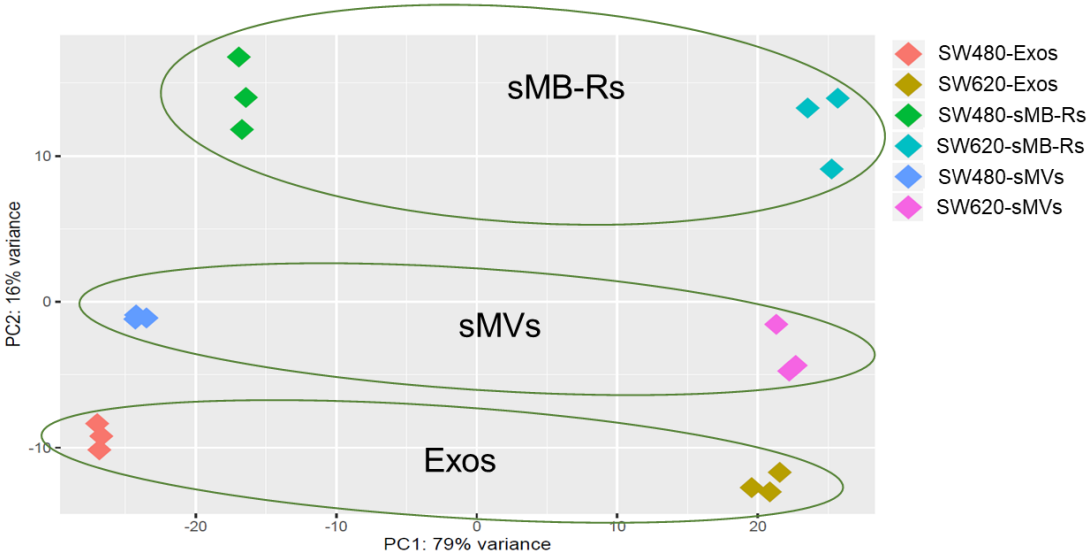
B


SW480
Exos
(5,954)
sMVs
(5,890)
sMB-Rs
(5,808)
398
478
220
4971
107
221
509

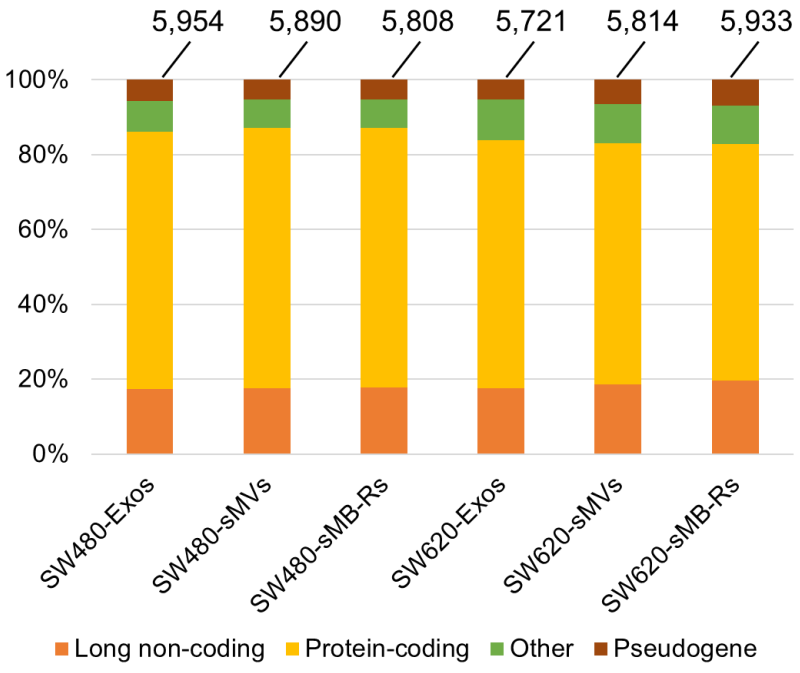
C


SW620
Exos
(5,721)
sMVs
(5,814)
sMB-Rs
(5,933)
433
594
298
4577
117
345
894

D


sMB-Rs
sMVs
Exos
PC2: 16% variance
PC1: 79% variance
10
0
-10
-20
-10
0
10
20
SW480-Exos
SW620-Exos
SW480-sMB-Rs
SW620-sMB-Rs
SW480-sMVs
SW620-sMVs

E


5,954
5,890
5,808
5,721
5,814
5,933
100%
80%
60%
40%
20%
0%
SW480-Exos
SW480-sMVs
SW480-sMB-Rs
SW620-Exos
SW620-sMVs
SW620-sMB-Rs
Long non-coding
Protein-coding
Other
Pseudogene

4.3.2 Differential transcript expression and Gene Ontology (GO) analysis of Exos, sMVs and sMB-Rs

To gain a greater insight into the transcriptomes of the three EV classes (Exos, sMVs, and sMB-Rs), differential transcript expression was performed in three different comparisons, i) SW480-/ SW620-Exos and SW480-/ SW620-sMVs, ii) SW480-/ SW620-Exos and SW480-/ SW620-sMB-Rs, and iii) SW480-/ SW620-sMVs and SW480-/ SW620-sMB-Rs. Transcripts with log2fold change >1 or <-1 and pvalue < 0.05 were considered as highly-enriched transcripts (see Section 4.2.6). GO analysis utilized log2fold change of transcripts from differential expression analysis for functional annotation.

Differential transcript expression analysis revealed 136 and 737 transcripts highly-enriched in Exos compared to sMVs and sMB-Rs (**Fig. 4.2A, B;** a list of top 100 enriched transcripts is given in **Appendix Tables 4.5** and **4.6**, respectively). Interestingly, exosome-related transcripts such as HSP90AA1-207, KRT19-205 and ANXA2-223 are highly-enriched in Exos compared to sMVs (**Fig. 2A**) while several histone subunit transcripts such as HIST1H3E-202, HIST1H4J-201, HIST1H2BG-201, HIST1H4E-201, HIST1H4B-201 and HIST1H4L-201, as well as ribosomal protein-coding transcripts such as RPS3-205, RPS-212 and MRPL27-207 are enriched in Exos compared with sMB-Rs (**Fig. 4.2B**). Selectively-enriched transcripts in Exos (i.e., transcripts observed in both comparisons) are highlighted with a red star in the Volcano plots. Prominent amongst these are BICD2-201 (BICD cargo adaptor 2, mRNA) and AC109326.1-201 (a novel transcript). GO

analysis of highly-enriched transcripts in Exos compared to sMVs and sMB-Rs revealed DNA-binding GO Terms such as ‘DNA package complex’, ‘protein-DNA complex’, ‘nuclear nucleosome’ and other GO terms such as ‘large ribosomal subunit’, and ‘cytosolic part’ (**Appendix Fig. 4.2C, D**, and **Appendix Tables 4.7, 4.8**).

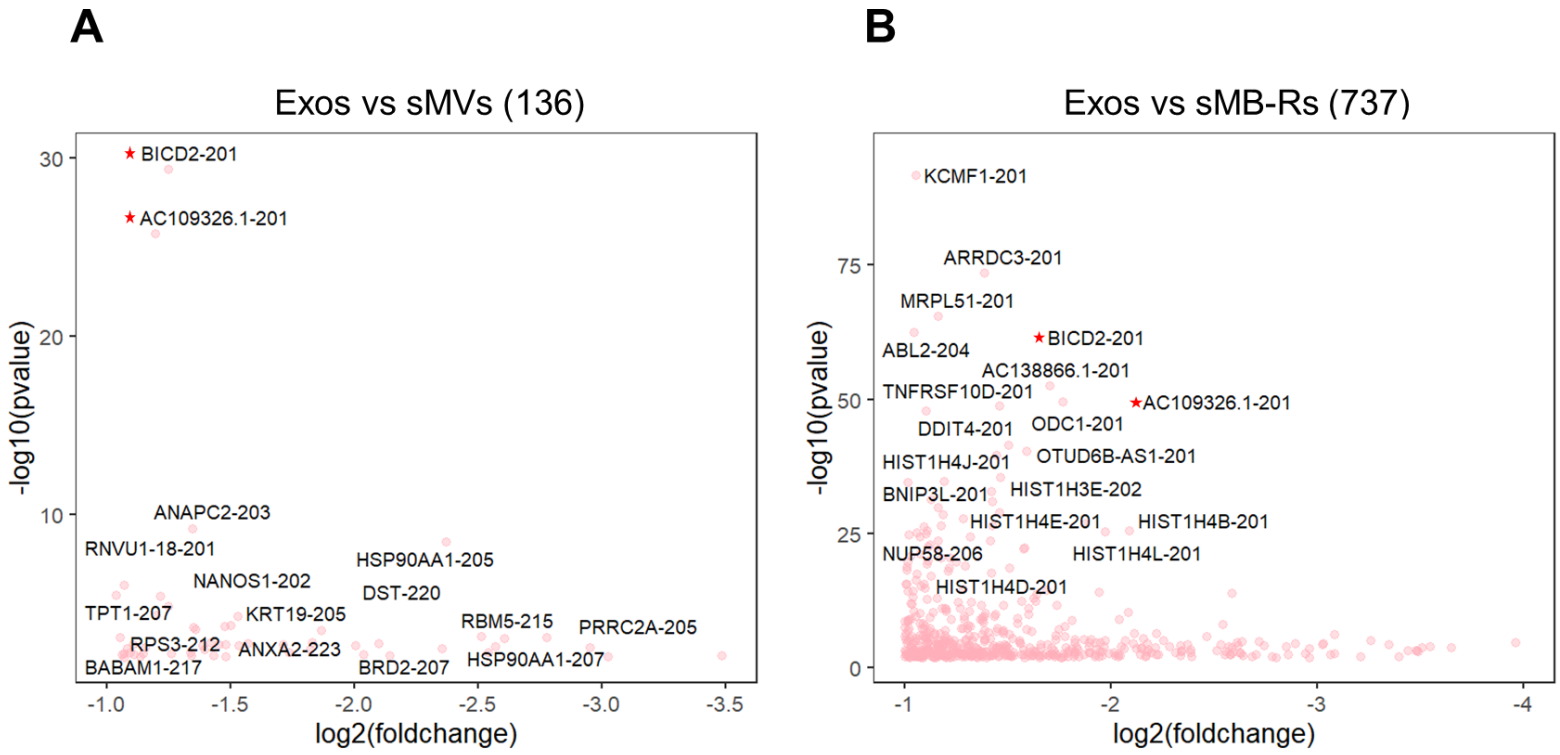
Differential transcript expression analysis of sMVs revealed 528 and 226 highly-enriched transcripts in sMVs compared to Exos and sMB-Rs, respectively (**Fig. 4.2C, D,** see a list of top 100 enriched transcripts in **Appendix Tables 4.5, 4.9**). Mitochondrial transcripts (MT-CO1, MT-CO3-201, MT-CYB-201, MT-ND4-201 and MT-ND5-201) and membrane-related transcripts (OPRD1-201 and AVL9-204) are highly-enriched in sMVs compared to Exos (**Fig. 4.2C**). GTF2IP7-202 (transcription factor pseudogene), LAMC2-201 (laminin gamma2) and histone protein-coding transcripts such as HIST1H4B-201 and HIST1H4L-201 when compared with sMB-Rs (**Fig. 4.2D**). GO analysis revealed enriched GO terms related to mitochondrial- (‘respiratory chain complex’, ‘respiratory chain’) and membrane-related (‘postsynapse, membrane protein complex’, ‘leading edge membrane’) in sMVs compared to Exos (**Appendix Fig. 4.2C** and **Appendix Table 4.8**) and enriched GO terms such as ‘nucleosome’ and ‘protein-DNA complex’ when compared to sMB-Rs (**Appendix Fig. 4.2D**, **Appendix table 4.10**).

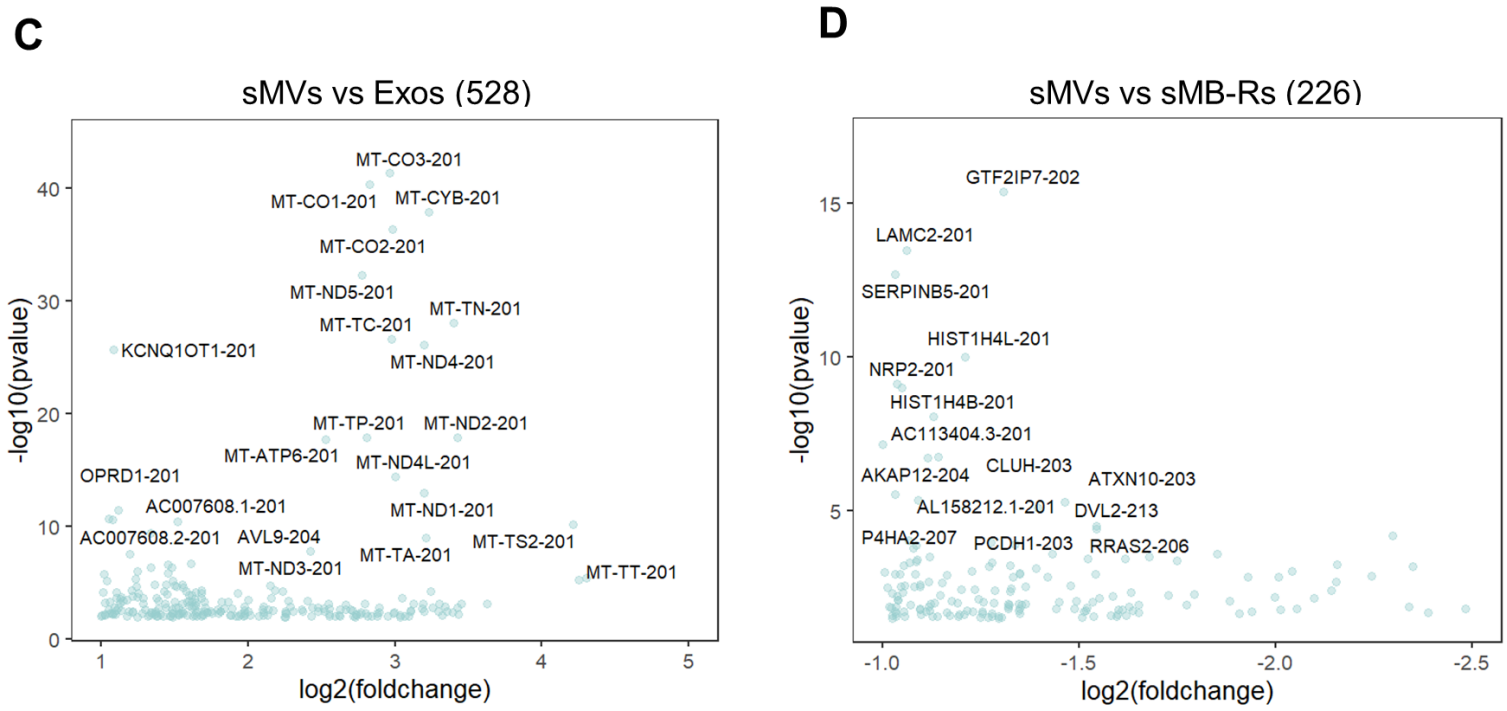
Differential transcript expression of sMB-Rs revealed 2,276 and 1,541 transcripts highly-enriched in sMB-Rs compared to Exos and sMVs,

respectively (**Fig. 4.2E, F** – a list of top 100 enriched transcripts is given in **Appendix Table 4.6, 4.9**). Selectively-enriched transcripts in sMB-Rs (i.e., transcripts observed in both datasets) include mitochondrial (MT-CO1, MT-CO2, MT-CO3, MT-ND-5), ribonucleoprotein complex- (NEAT1-202, SCAF1-201, LARP1-204), zinc finger- (ZNF703-201, ZFPM1-201) and signaling-related transcripts (TGFB1-201) and are highlighted with red stars in the Volcano plots (**Fig 4.2E, F**). GO analysis of transcripts highly enriched in sMB-Rs compared to Exos and sMVs include GO terms such as mitochondrial ('respiratory chain complex', 'mitochondrial inner membrane'), membrane-related ('synapse part', 'postsynaptic specialization'), (**Appendix Fig. 4.2E, F, and Appendix Tables 4.8, 4.10**). Interestingly, GO terms such as 'cytoplasmic stress granule' and 'transcription factor complex' were enriched in sMB-Rs compared to Exos (**Appendix Fig. 4.2F).**

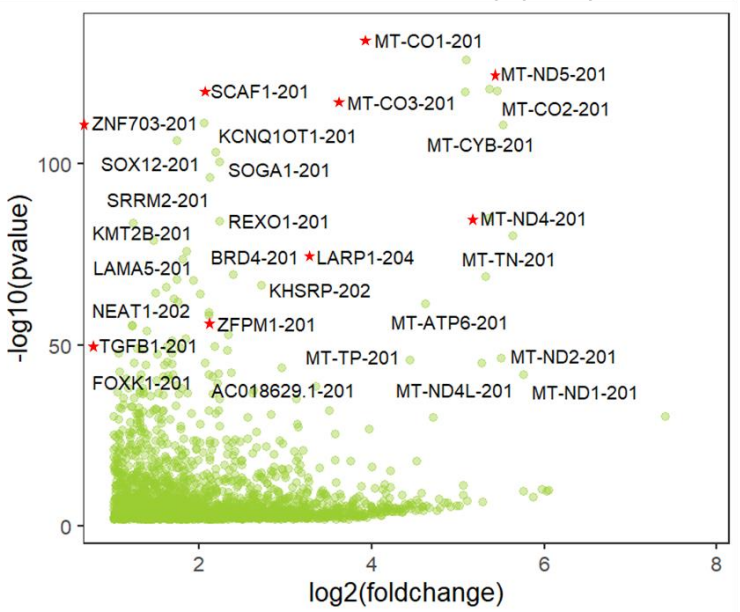
Expression levels of highly-enriched transcripts in EV class comparisons (**Fig 4.2A-F**) were compared in a heatmap to identify transcript signatures in each EV class. The heatmap illustrates that sMB-Rs are enriched in mitochondrial (MT-transcripts) and ribonucleoprotein complex/splicing factor-related transcripts (LARP1-204, ZNF703-201, SCAF1-201), regardless of cell types (**Fig. 4.2G**). Exos are enriched in histone transcripts (HIST1H-transcripts) and exosome biogenesis-related transcripts such as BICD2-203, ARRDC3-201 and CAV2-201, regardless of cell types (**Fig. 4.2G**).

Figure 4.2 Differential transcript expression analysis and Gene Ontology (GO) analysis of Exos, sMVs and sMB-Rs derived from SW480 and SW620 cells. (A) Differential transcript expression analysis revealed 136 transcripts highly-enriched in Exos compared with sMVs (log2 fold change <-1, pvalue <0.05) and **(B)** 737 transcripts highly-enriched in Exos compared with sMB-Rs (log2 foldchange <-1, pvalue <0.05). **(C)** Differential transcript expression analysis revealed 528 transcripts highly-enriched in sMVs compared with Exos (log2 fold change >1, pvalue <0.05) and **(D)** 226 transcripts highly-enriched in sMVs compared with sMB-Rs (log2 foldchange <-1, pvalue <0.05). **(E)** Differential transcript expression analysis revealed 2,276 transcripts highly-enriched in sMB-Rs compared with Exos (log2 fold change >1, pvalue <0.05) and **(F)** 1,541 transcripts highly-enriched in sMB-Rs compared with sMVs (log2 foldchange >1, pvalue <0.05). **(G)** Heat map illustration of highly-enriched transcripts in SW480-/SW620-derived Exos, sMVs and sMB-Rs (scale shown is normalized counts subtracted by mean and divided by standard deviation), red stars show selectively-enriched transcripts in each EV classes.

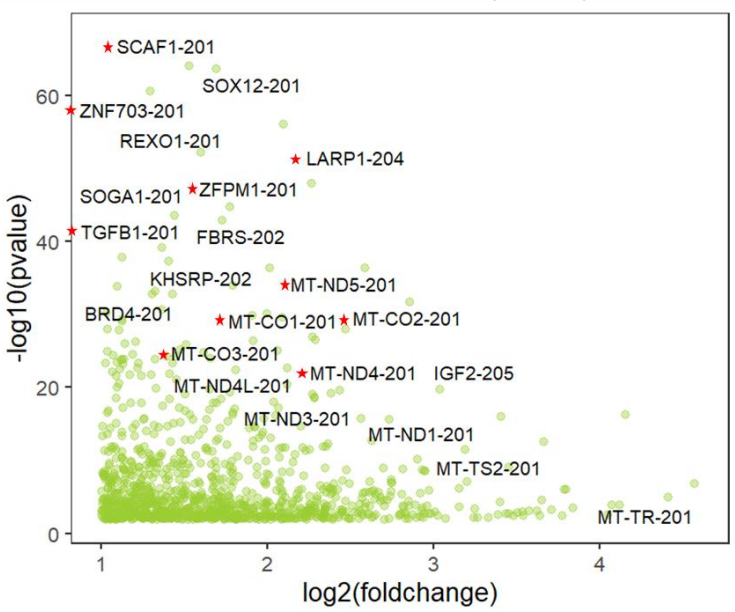

A
Exos vs sMVs (136)
-log10(pvalue)
log2(foldchange)
BICD2-201
AC109326.1-201
ANAPC2-203
RNVU1-18-201
NANOS1-202
TPT1-207
KRT19-205
RPS3-212
BABAM1-217
ANXA2-223
HSP90AA1-205
DST-220
BRD2-207
RBM5-215
HSP90AA1-207
PRRC2A-205
B
Exos vs sMB-Rs (737)
-log10(pvalue)
log2(foldchange)
KCMF1-201
ARRDC3-201
MRPL51-201
ABL2-204
BICD2-201
AC138866.1-201
TNFRSF10D-201
AC109326.1-201
DDIT4-201
ODC1-201
HIST1H4J-201
OTUD6B-AS1-201
BNIP3L-201
HIST1H3E-202
HIST1H4E-201
HIST1H4B-201
NUP58-206
HIST1H4L-201
HIST1H4D-201


C
sMVs vs Exos (528)
-log10(pvalue)
log2(foldchange)
MT-CO3-201
MT-CO1-201
MT-CYB-201
MT-CO2-201
MT-ND5-201
MT-TN-201
MT-TC-201
KCNQ1OT1-201
MT-ND4-201
MT-TP-201
MT-ND2-201
MT-ATP6-201
MT-ND4L-201
OPRD1-201
AC007608.1-201
MT-ND1-201
AC007608.2-201
AVL9-204
MT-TS2-201
MT-TA-201
MT-ND3-201
MT-TT-201
D
sMVs vs sMB-Rs (226)
-log10(pvalue)
log2(foldchange)
GTF2IP7-202
LAMC2-201
SERPINB5-201
HIST1H4L-201
NRP2-201
HIST1H4B-201
AC113404.3-201
AKAP12-204
CLUH-203
ATXN10-203
AL158212.1-201
DVL2-213
P4HA2-207
PCDH1-203
RRAS2-206

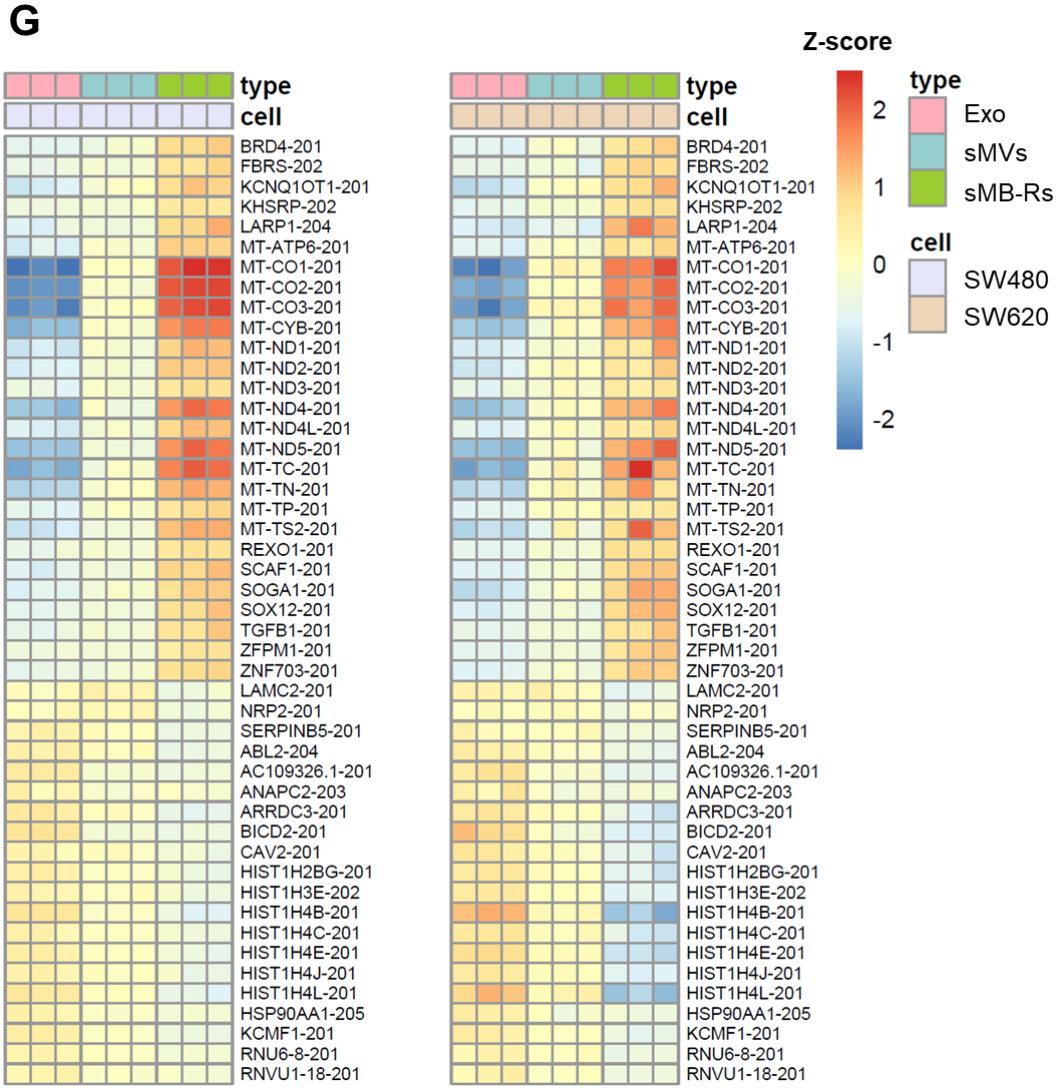
E


sMB-Rs vs Exos (2,276)
-log10(pvalue)
log2(foldchange)
MT-CO1-201
MT-ND5-201
SCAF1-201
MT-CO3-201
MT-CO2-201
ZNF703-201
KCNQ1OT1-201
MT-CYB-201
SOX12-201
SOGA1-201
SRRM2-201
REXO1-201
MT-ND4-201
KMT2B-201
BRD4-201
LARP1-204
MT-TN-201
LAMA5-201
KHSRP-202
NEAT1-202
ZFPM1-201
MT-ATP6-201
TGFB1-201
MT-TP-201
MT-ND2-201
FOXK1-201
AC018629.1-201
MT-ND4L-201
MT-ND1-201

F


sMB-Rs vs sMVs (1,541)
-log10(pvalue)
log2(foldchange)
SCAF1-201
SOX12-201
ZNF703-201
REXO1-201
LARP1-204
SOGA1-201
ZFPM1-201
TGFB1-201
FBRS-202
KHSRP-202
MT-ND5-201
BRD4-201
MT-CO1-201
MT-CO2-201
MT-CO3-201
MT-ND4-201
IGF2-205
MT-ND4L-201
MT-ND3-201
MT-ND1-201
MT-TS2-201
MT-TR-201

G


type
cell
BRD4-201
FBRS-202
KCNQ1OT1-201
KHSRP-202
LARP1-204
MT-ATP6-201
MT-CO1-201
MT-CO2-201
MT-CO3-201
MT-CYB-201
MT-ND1-201
MT-ND2-201
MT-ND3-201
MT-ND4-201
MT-ND4L-201
MT-ND5-201
MT-TC-201
MT-TN-201
MT-TP-201
MT-TS2-201
REXO1-201
SCAF1-201
SOGA1-201
SOX12-201
TGFB1-201
ZFPM1-201
ZNF703-201
LAMC2-201
NRP2-201
SERPINB5-201
ABL2-204
AC109326.1-201
ANAPC2-203
ARRDC3-201
BICD2-201
CAV2-201
HIST1H2BG-201
HIST1H3E-202
HIST1H4B-201
HIST1H4C-201
HIST1H4E-201
HIST1H4J-201
HIST1H4L-201
HSP90AA1-205
KCMF1-201
RNU6-8-201
RNVU1-18-201
type
cell
BRD4-201
FBRS-202
KCNQ1OT1-201
KHSRP-202
LARP1-204
MT-ATP6-201
MT-CO1-201
MT-CO2-201
MT-CO3-201
MT-CYB-201
MT-ND1-201
MT-ND2-201
MT-ND3-201
MT-ND4-201
MT-ND4L-201
MT-ND5-201
MT-TC-201
MT-TN-201
MT-TP-201
MT-TS2-201
REXO1-201
SCAF1-201
SOGA1-201
SOX12-201
TGFB1-201
ZFPM1-201
ZNF703-201
LAMC2-201
NRP2-201
SERPINB5-201
ABL2-204
AC109326.1-201
ANAPC2-203
ARRDC3-201
BICD2-201
CAV2-201
HIST1H2BG-201
HIST1H3E-202
HIST1H4B-201
HIST1H4C-201
HIST1H4E-201
HIST1H4J-201
HIST1H4L-201
HSP90AA1-205
KCMF1-201
RNU6-8-201
RNVU1-18-201
Z-score
2
1
0
-1
-2
type
Exo
sMVs
sMB-Rs
cell
SW480
SW620

4.3.3 LncRNA and pseudogene transcripts and their association with RNA-binding proteins in Exos, sMVs and sMB-Rs

Genomic DNA is transcribed into a board spectrum of RNAs including non-protein coding RNAs and pseudogenes which function in both normal physiological and pathophysiological conditions[249]. Accumulating evidence shows that non-coding RNAs and pseudogenes can be secreted into the extracellular space and, more recently, be detected in extracellular vesicles[236].

LncRNA and pseudogene transcripts are detected in Exos, sMVs and sMB-Rs derived from SW480 and SW620 cells as shown in **Fig. 4.3A**. Differential transcript expression analysis revealed that highly-enriched transcripts in sMB-Rs include KCNQ1OT1-201 (lncRNA), GABPB1-AS1-202 (lncRNA), NEAT1-202 (lncRNA), AC007608.2-201 (pseudogene) (yellow star-labelled in **Fig. 4.3A**). AC131392.1-201 (pseudogene), ANAPC2-203 (lncRNA) and GUSBP9-201 (pseudogene) were highly-enriched in Exos compared to sMVs and sMB-Rs (yellow star-labelled in **Fig. 4.3A**).

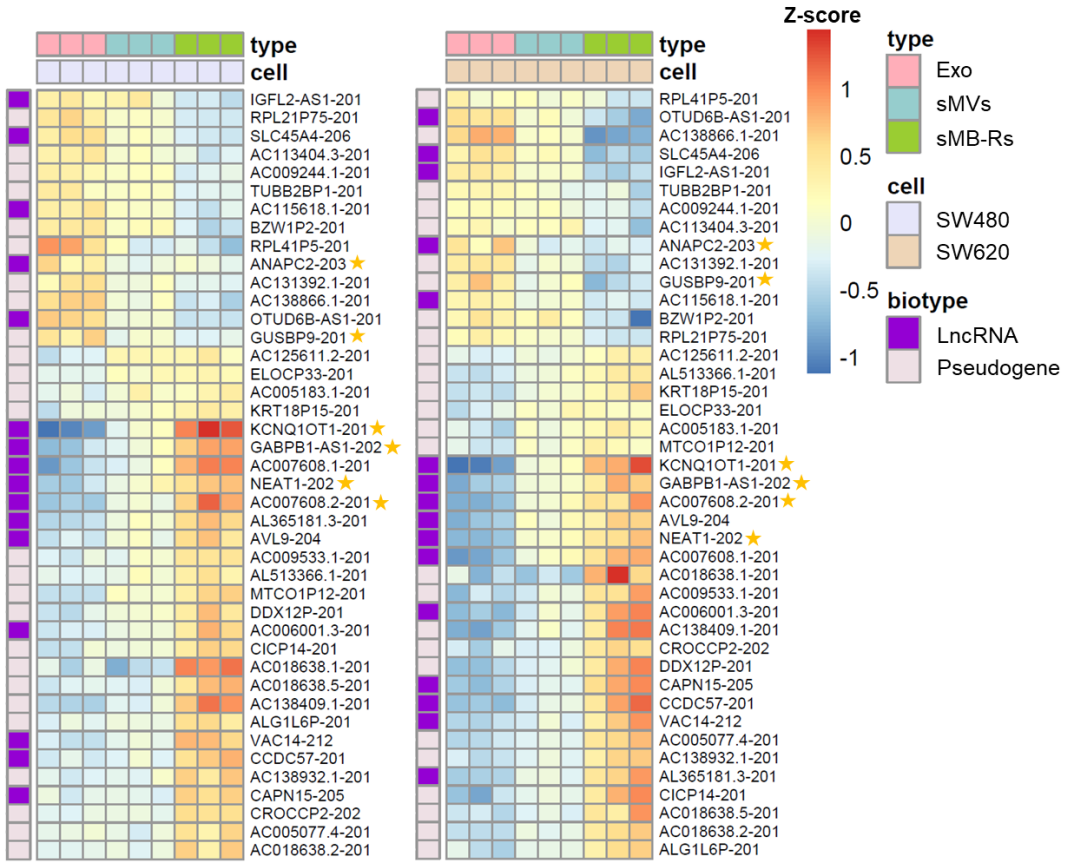
LncRNA and pseudogene transcripts have been shown to be associated with RNA-binding proteins (RBPs), RNA metabolism and RNA interference processes[250, 251]. To explore this relationship the lncRNA and pseudogene transcripts identified in all EV samples were interrogated to find their potential RBP targets using Encyclopedia of RNA Interactomes database (ENCORI)[248]. The ENCORI-RBP target database revealed that 275, 265, 276, 290, 284, 337 lncRNA transcripts and 92, 112, 109, 124, 116, 141 pseudogene

transcripts in SW480-Exos, SW620-Exos, SW480-sMVs, SW620-sMVs, SW480- sMB-Rs and SW620- sMB-Rs, respectively, were shown to be associated with 27 RBPs listed in the ENCORI-RBP target database (**Appendix Fig. 4.3**). Next, the 27 identified RBP partners (see **Appendix Table 4.11**) were functionally annotated using a g:GOst functional profiling from gProfiler[247]. GO analysis revealed that the 27 RBPs involve ribonucleoprotein complex/spliceosomal complex (HNRNP families, U2AF2), ribonucleoprotein granule (ELAVL1, FUS, IGF2BP families) and nuclear speckles (ALYREF, METTL3) (**Appendix Table 4.12**).

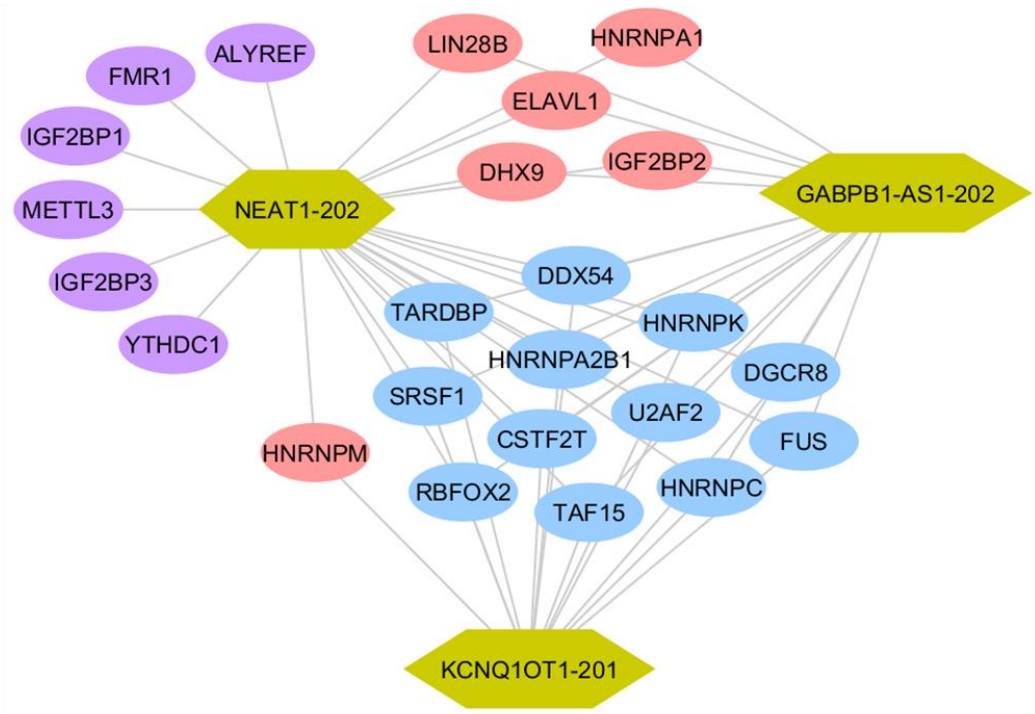
Interestingly, 19/27 RBPs were co-identified at the proteome level (See **Appendix Table 4.11**) and highly-enriched lncRNA and pseudogene transcripts are predicted to bind with the RBPs (**Table 4.1**). In addition, it can be seen in an interaction network between lncRNA transcripts and RBPs that highly-enriched lncRNA transcripts in sMB-Rs such as NEAT1-202 (involved in RNA retention, nuclear speckles and metabolism[252]), KCNQ1OT1-201 and GABPB1-AS1-202 can bind to RNA/stress granule markers (FUS, TARDBP, IGF2BP2, TAF15, ELAVL1) and ribonucleoproteins (HNRNP families) (**Fig. 4.3B**).

Figure 4.3 LncRNA and pseudogene transcripts and their association with RNA-binding proteins (RBPs) in Exos, sMVs and sMB-Rs derived from SW480 and SW620 cells. (A) Heat map illustration of highly-enriched lncRNA and pseudogene transcripts in SW480-/SW620-derived Exos, sMVs and sMB-Rs (scale shown is normalized counts subtracted by mean and divided by standard deviation). **(B)** Binary interaction networks of RBPs (data from **Chapter 3**) and three highly-enriched lncRNA transcripts identified in sMB-Rs (NEAT1-202, KCNQ1OT1-201 and GABPB1-AS1-202).

A


type
cell
IGFL2-AS1-201
RPL21P75-201
SLC45A4-206
AC113404.3-201
AC009244.1-201
TUBB2BP1-201
AC115618.1-201
BZW1P2-201
RPL41P5-201
ANAPC2-203
AC131392.1-201
AC138866.1-201
OTUD6B-AS1-201
GUSBP9-201
AC125611.2-201
ELOCP33-201
AC005183.1-201
KRT18P15-201
KCNQ1OT1-201
GABPB1-AS1-202
AC007608.1-201
NEAT1-202
AC007608.2-201
AL365181.3-201
AVL9-204
AC009533.1-201
AL513366.1-201
MTCO1P12-201
DDX12P-201
AC006001.3-201
CICP14-201
AC018638.1-201
AC018638.5-201
AC138409.1-201
ALG1L6P-201
VAC14-212
CCDC57-201
AC138932.1-201
CAPN15-205
CROCCP2-202
AC005077.4-201
AC018638.2-201
type
cell
RPL41P5-201
OTUD6B-AS1-201
AC138866.1-201
SLC45A4-206
IGFL2-AS1-201
TUBB2BP1-201
AC009244.1-201
AC113404.3-201
ANAPC2-203
AC131392.1-201
GUSBP9-201
AC115618.1-201
BZW1P2-201
RPL21P75-201
AC125611.2-201
AL513366.1-201
KRT18P15-201
ELOCP33-201
AC005183.1-201
MTCO1P12-201
KCNQ1OT1-201
GABPB1-AS1-202
AC007608.2-201
AVL9-204
NEAT1-202
AC007608.1-201
AC018638.1-201
AC009533.1-201
AC006001.3-201
AC138409.1-201
CROCCP2-202
DDX12P-201
CAPN15-205
CCDC57-201
VAC14-212
AC005077.4-201
AC138932.1-201
AL365181.3-201
CICP14-201
AC018638.5-201
AC018638.2-201
ALG1L6P-201
Z-score
1
0.5
0
-0.5
-1
type
Exo
sMVs
sMB-Rs
cell
SW480
SW620
biotype
LncRNA
Pseudogene

B


ALYREF
FMR1
IGF2BP1
METTL3
IGF2BP3
YTHDC1
NEAT1-202
LIN28B
HNRNPA1
ELAVL1
DHX9
IGF2BP2
GABPB1-AS1-202
DDX54
TARDBP
HNRNPK
HNRNPA2B1
DGCR8
SRSF1
U2AF2
CSTF2T
FUS
HNRNPM
RBFOX2
HNRNPC
TAF15
KCNQ1OT1-201

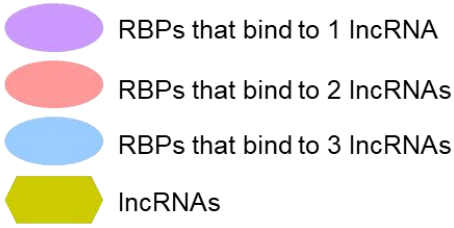

RBPs that bind to 1 lncRNA
RBPs that bind to 2 lncRNAs
RBPs that bind to 3 lncRNAs
lncRNAs

Table 4.1 LncRNA and pseudogene transcripts associated with RBPs

RNA-binding proteins*	Number of lncRNA and pseudogene transcripts identified in each SW480/ SW620 EV class											
	Exos/sMVs[a]		Exos/sMB-Rs[b]		sMVs/Exos[c]		sMVs/sMB-Rs[d]		sMB-Rs/Exos[e]		sMB-Rs/sMVs[f]	
	lncRNA	Pseudogene	lncRNA	Pseudogene	lncRNA	Pseudogene	lncRNA	Pseudogene	lncRNA	Pseudogene	lncRNA	Pseudogene
HNRNPA1	0	0	5	2	12	1	3	1	59	12	31	6
HNRNPA2B1	0	0	2	1	7	2	2	0	33	10	17	6
HNRNPC	0	1	4	5	24	4	1	1	95	16	50	8
HNRNPK	1	0	2	0	7	1	2	0	34	5	21	4
HNRNPM	0	0	4	1	8	0	1	1	29	1	17	1
SRSF1	0	1	5	7	11	10	3	2	56	27	32	14
U2AF2	0	1	9	8	28	8	4	3	110	24	61	15
DDX3X	0	0	0	0	0	0	0	0	1	0	1	0
DHX9	0	0	3	0	3	0	0	0	18	1	12	1
ELAVL1	0	1	5	2	4	2	3	1	23	2	9	1
FUS	0	0	4	0	2	1	1	0	16	1	9	0
IGF2BP1	0	1	3	1	6	1	2	0	28	5	13	3
IGF2BP2	0	1	8	7	9	4	3	2	54	20	26	13
IGF2BP3	0	0	3	0	4	3	2	0	20	6	12	4
TARDBP	0	0	6	0	11	0	3	0	49	3	25	3
ALYREF	0	0	0	0	0	0	0	0	4	0	1	0
CSTF2	1	0	7	2	10	3	2	1	53	11	25	8
TAF15	0	1	16	4	16	7	7	1	81	23	49	12
DDX54	1	1	3	2	6	3	1	1	39	17	19	8

*=19/27 RBPs co-identified in proteomic data (**Appendix Table 4.11**) and ENCORI-RBP target database

a =Highly-enriched transcripts in Exos compared to sMVs (Fig. 4.2A)

b =Highly-enriched transcripts in Exos compared to sMB-Rs (Fig. 4.2B)

c = Highly-enriched transcripts in sMVs compared to Exos (Fig. 4.2C)

d = Highly-enriched transcripts in sMV compared to sMB-Rs (Fig 4.2D)

e = Highly-enriched transcripts in sMB-Rs compared to Exos (Fig. 4.2E)

f = Highly-enriched transcripts in sMB-Rs compared to sMVs (Fig 4.2F)

4.3.4 Identification of fusion genes in Exos, sMVs, sMB-Rs and cell lysate derived from SW480 and SW620 cells

In disease biomarker research, EVs are served as molecular targets for liquid biopsies[57]. Fusion genes (a hybrid gene from two previously different genes) have been detected in CRC-derived Exos (LIM1863 cells)[57] and proposed as potential cancer biomarkers. In this study, I used an isogenic human colorectal cancer model that consists of adenocarcinoma (SW480) and lymph-node metastatic cancer (SW620) and questioned - i) if SW480- and SW620-EV classes contain fusion genes, -ii) if EVs are a more enriched source of fusion genes than parental cells, and -iii) if fusion genes could be utilized as potential markers to study cancer progression. To address the questions, fusion genes were first detected in parental cells (SW480 and SW620 cell lysates) and their EV classes using ChimeraScan (v.0.4.5)[244]. Enriched fusion genes (total fragments >10) were considered for further analysis.

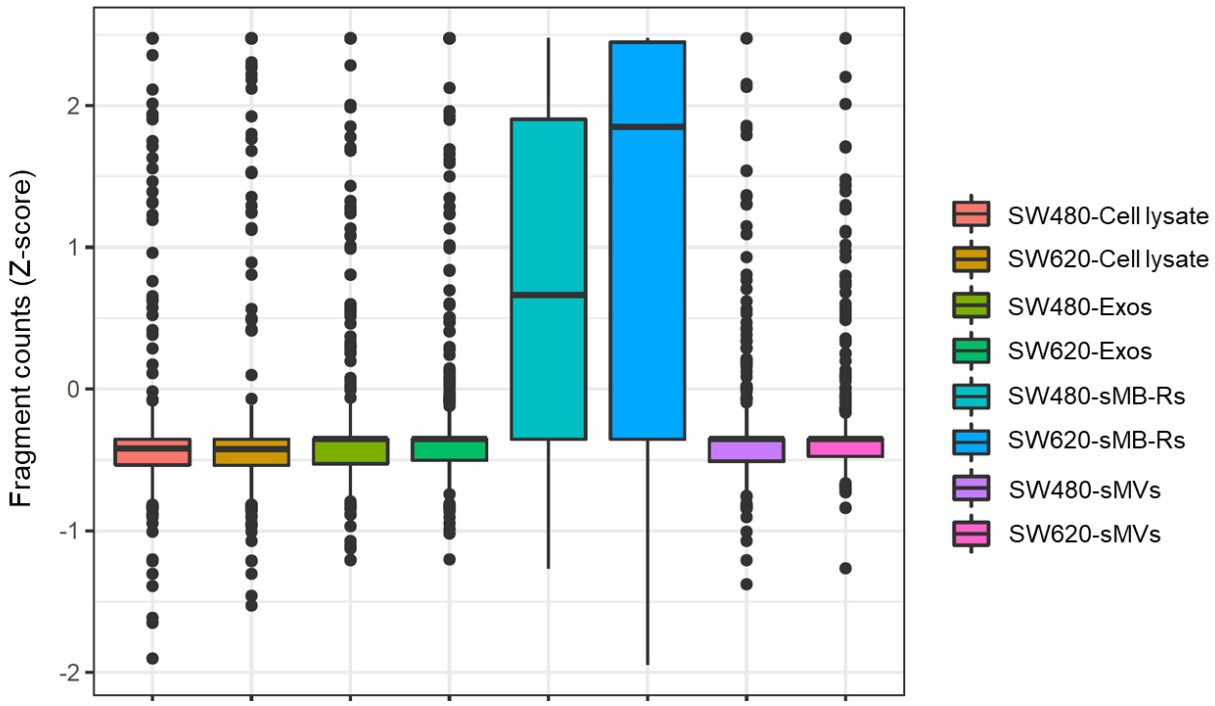
ChimeraScan (see Section 4.2.5) identified 770 fusion genes (total fragments >10) across parental cell lysates, Exos, sMVs and sMB-Rs derived from SW480 and SW620 cells (a list of top 100 enriched fusion genes is given in **Appendix table 4.13**). Interestingly, only 10/770 fusion genes have been identified in the TCGA fusion gene database[245] (AKAP13-PDE8A, ARL17A-KANSL1, C15orf57-CBX3, CTNNBIP1-CLSTN1, KANSL1-ARL17A, LOC100130000-LINC00623, LOC541471-ANAPC1, MX1-FNBP1, NSD1-

ZNF346 and PRRG4-QSER1). The majority of 770 fusion genes are highly enriched in sMB-Rs (for both SW480 and SW620 cell types) compared to cell lysates, Exos and sMVs (**Fig. 4.4A**). For example, sMB-Rs contained 496 (SW480-sMB-Rs) and 577 (SW620-sMB-Rs) fusion genes compared to their corresponding Exos (127 and 184), sMVs (145 and 193 fusion genes) and cell lysate (62 and 55 fusion genes) (**Fig. 4.4B, C**).

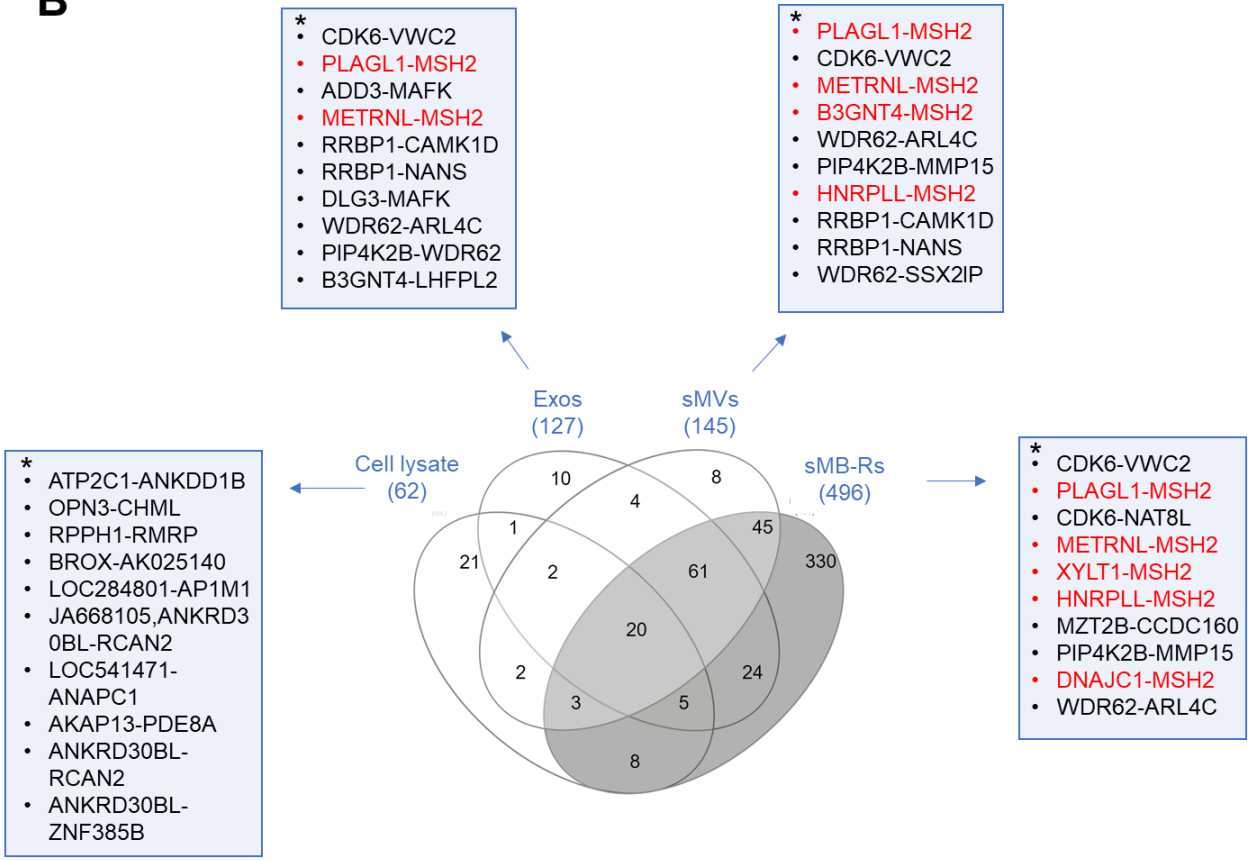
Tumor suppressor gene (MSH2)[253] was found fused to several genes (fusion genes) in all EV classes. Several MSH2 fusion genes such as PLAGL1-MSH2, METRNL-MSH2 and HNRPLL-MSH2 are the top 10 most abundant fusion genes observed in Exos, sMVs, sMB-Rs but not detectable in corresponding cell lysates (red labelled, **Fig. 4.4B, C**). MSH2 fusion genes are most abundant in sMB-Rs (relative to Exos and sMVs) (**Fig. 4.4D**). Moreover, sMB-Rs contain oncogenic genes[254] such as RAS (RRAS-BMP8B, C16orf52-DIRAS1) and CDK (CDK6-VWC2, HNRNPU-CDK6 and CDK6-STX4) (**Fig. 4.4E**). Intriguingly, RAS fusion genes were highly enriched in SW620-sMB-Rs in contrast to the majority of CDK fusion genes which are detected in SW480-sMB-Rs (**Fig. 4.4E**). The novel fusion genes CDK6-ATIP1B2 and ADD3-PARN are uniquely detected in SW480-EVs (but not in SW480 cell lysate) with selective enrichment in SW480-sMB-Rs (**Fig. 4.4F**). In contrast, NSD1-ZNF346 and CPS1-ATXN10 (TCGA-computationally annotated)[245] fusion genes (see Section 4.2.5) are uniquely detected in SW620-EVs (and not in SW620 cell lysate) with selective enrichment in SW620-sMB-Rs (**Fig. 4.4F**).

Figure 4.4 Identification of fusion genes in Exos, sMVs, sMB-Rs and cell lysate derived from SW480 and SW620 cells. (A) Enrichment level of fusion genes of Exos, sMVs, sMB-Rs and cell lysate derived from SW480 and SW620 cells. **(B)** A four-way Venn diagram of fusion genes identified in SW480-Exos, -sMVs, -sMB-Rs and cell lysate reveals 20 fusion genes were commonly identified, while 10, 8, 330 and 21 fusion genes were uniquely identified in SW480-Exos, sMVs, sMB-Rs and cell lysate, respectively (*Top 10 fusion genes). **(C)** A four-way Venn diagram of fusion genes identified in SW620-Exos, -sMVs, -sMB-Rs and cell lysate reveals 15 fusion genes were commonly identified, while 21, 4, 369 and 26 fusion genes were uniquely identified in SW620-Exos, sMVs, sMB-Rs and cell lysate, respectively (*Top 10 fusion genes). **(D)** Heat map illustration of MSH2 (tumor suppressor) fusion genes in SW480-/SW620-derived Exos, sMVs, sMB-Rs and cell lysate (scale shown is average total fragment of each fusion gene in log2). **(E)** Heat map illustration of RAS and CDK (oncogenic) fusion genes in SW480-/SW620-derived Exos, sMVs, sMB-Rs and cell lysate (scale shown is total fragment of each fusion gene in log2). **(F)** Number of selected fusion gene fragments of CDK6-ATP1B2, ADD3-PARN, NSD1-ZNF346, and CPS1-ATXN10 identified in SW480-/SW620-EV classes and their corresponding cell lysates.

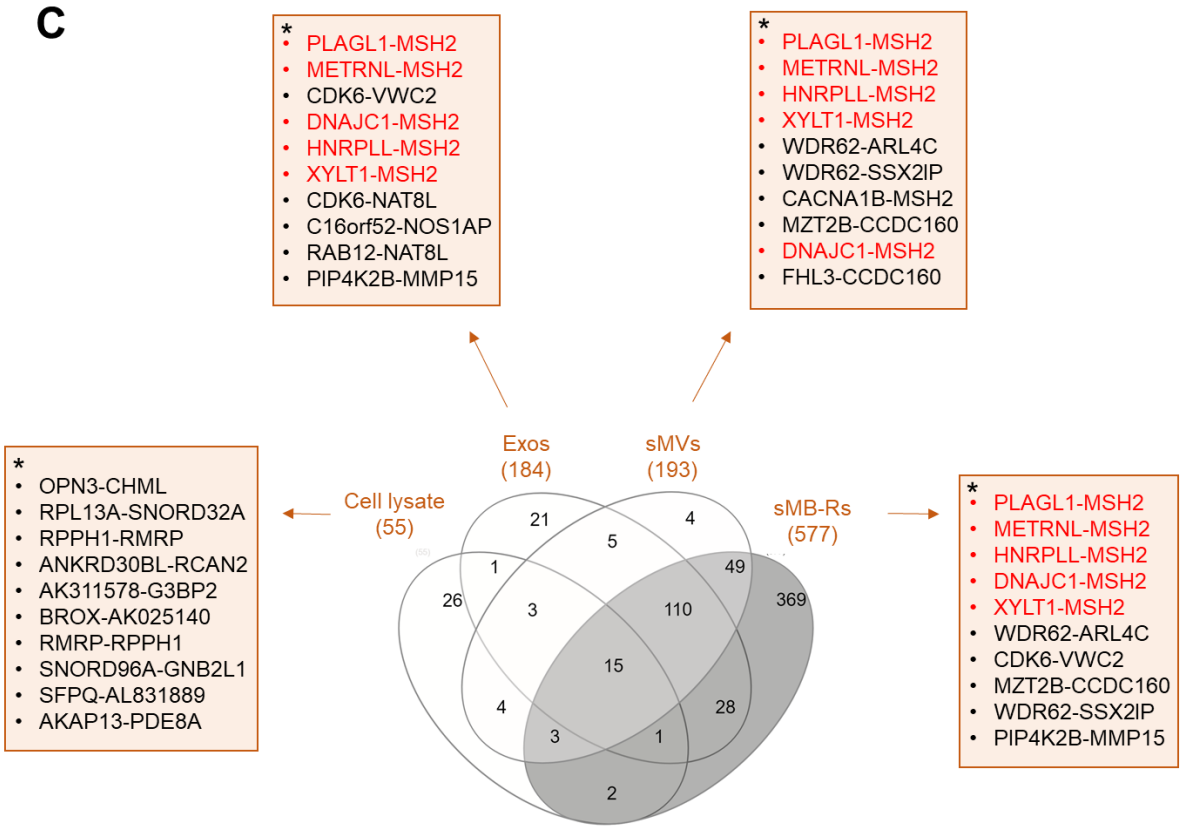
A


Fragment counts (Z-score)
2
1
0
-1
-2
SW480-Cell lysate
SW620-Cell lysate
SW480-Exos
SW620-Exos
SW480-sMB-Rs
SW620-sMB-Rs
SW480-sMVs
SW620-sMVs

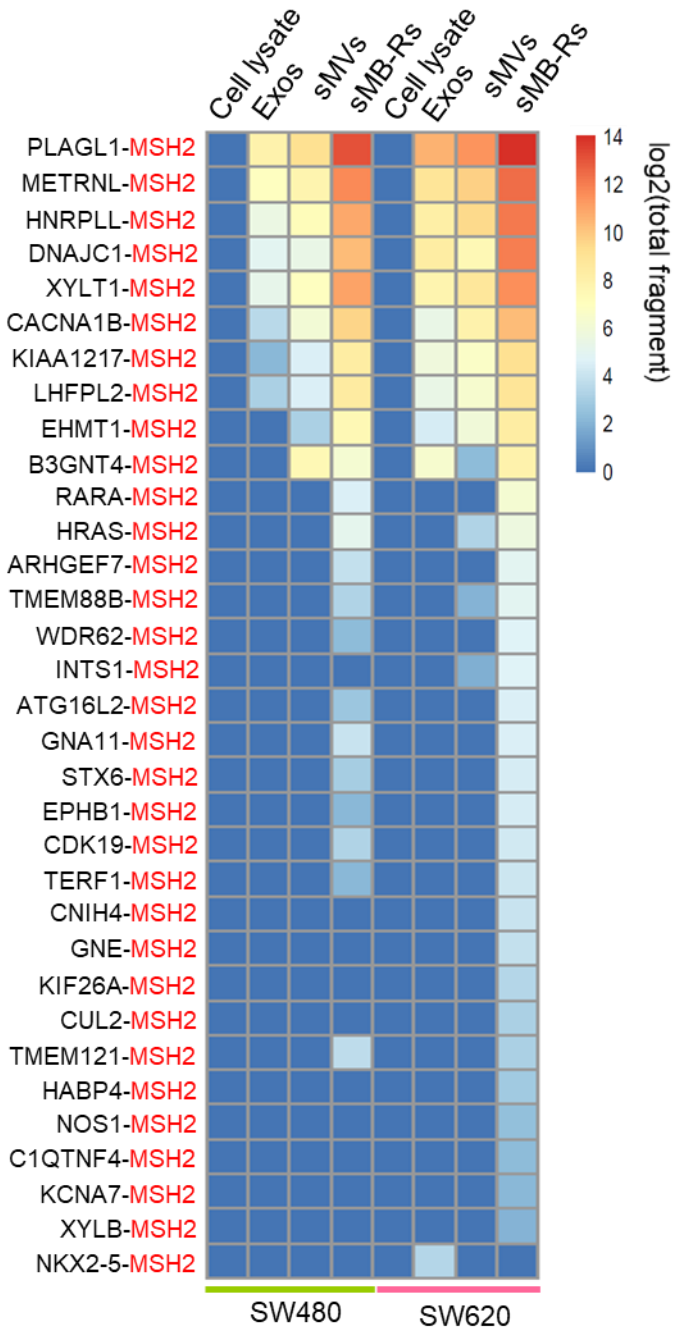
B

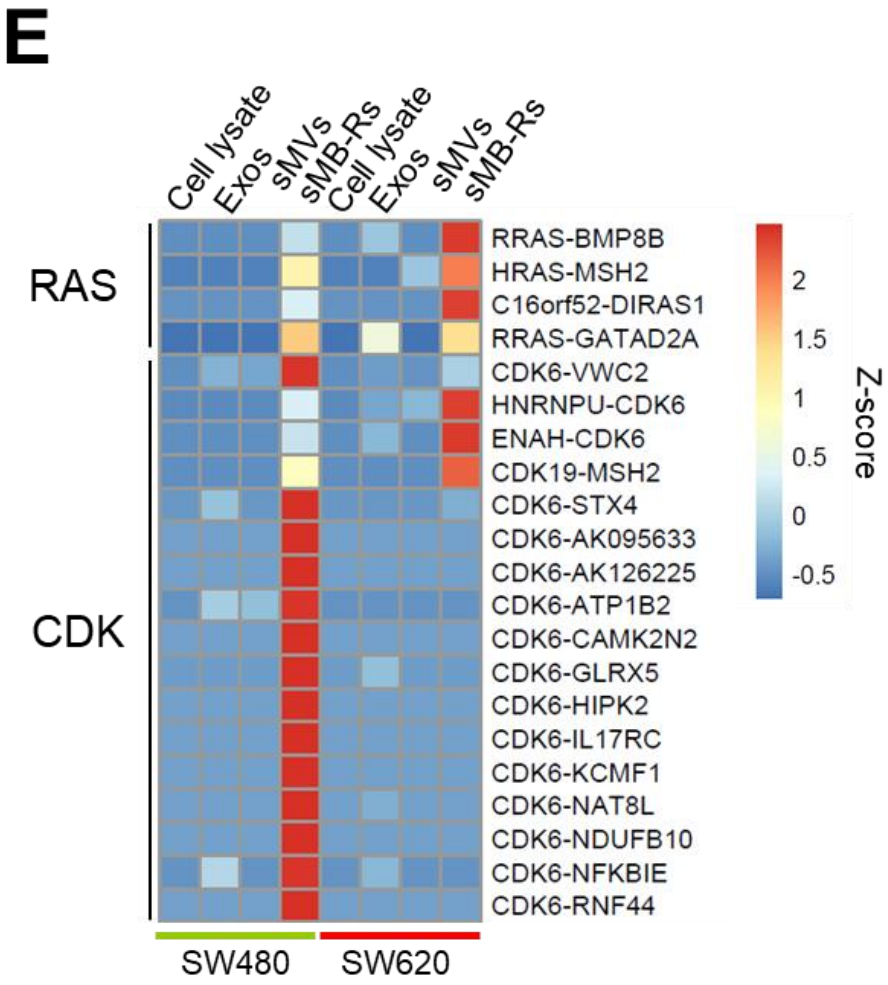

*
• CDK6-VWC2
• PLAGL1-MSH2
• ADD3-MAFK
• METRNL-MSH2
• RRBP1-CAMK1D
• RRBP1-NANS
• DLG3-MAFK
• WDR62-ARL4C
• PIP4K2B-WDR62
• B3GNT4-LHFPL2
*
• PLAGL1-MSH2
• CDK6-VWC2
• METRNL-MSH2
• B3GNT4-MSH2
• WDR62-ARL4C
• PIP4K2B-MMP15
• HNRPLL-MSH2
• RRBP1-CAMK1D
• RRBP1-NANS
• WDR62-SSX2IP
Exos (127)
sMVs (145)
Cell lysate (62)
sMB-Rs (496)
10
8
4
1
45
21
2
61
330
20
2
24
3
5
8
*
• ATP2C1-ANKDD1B
• OPN3-CHML
• RPPH1-RMRP
• BROX-AK025140
• LOC284801-AP1M1
• JA668105,ANKRD30BL-RCAN2
• LOC541471-ANAPC1
• AKAP13-PDE8A
• ANKRD30BL-RCAN2
• ANKRD30BL-ZNF385B
*
• CDK6-VWC2
• PLAGL1-MSH2
• CDK6-NAT8L
• METRNL-MSH2
• XYLT1-MSH2
• HNRPLL-MSH2
• MZT2B-CCDC160
• PIP4K2B-MMP15
• DNAJC1-MSH2
• WDR62-ARL4C

C

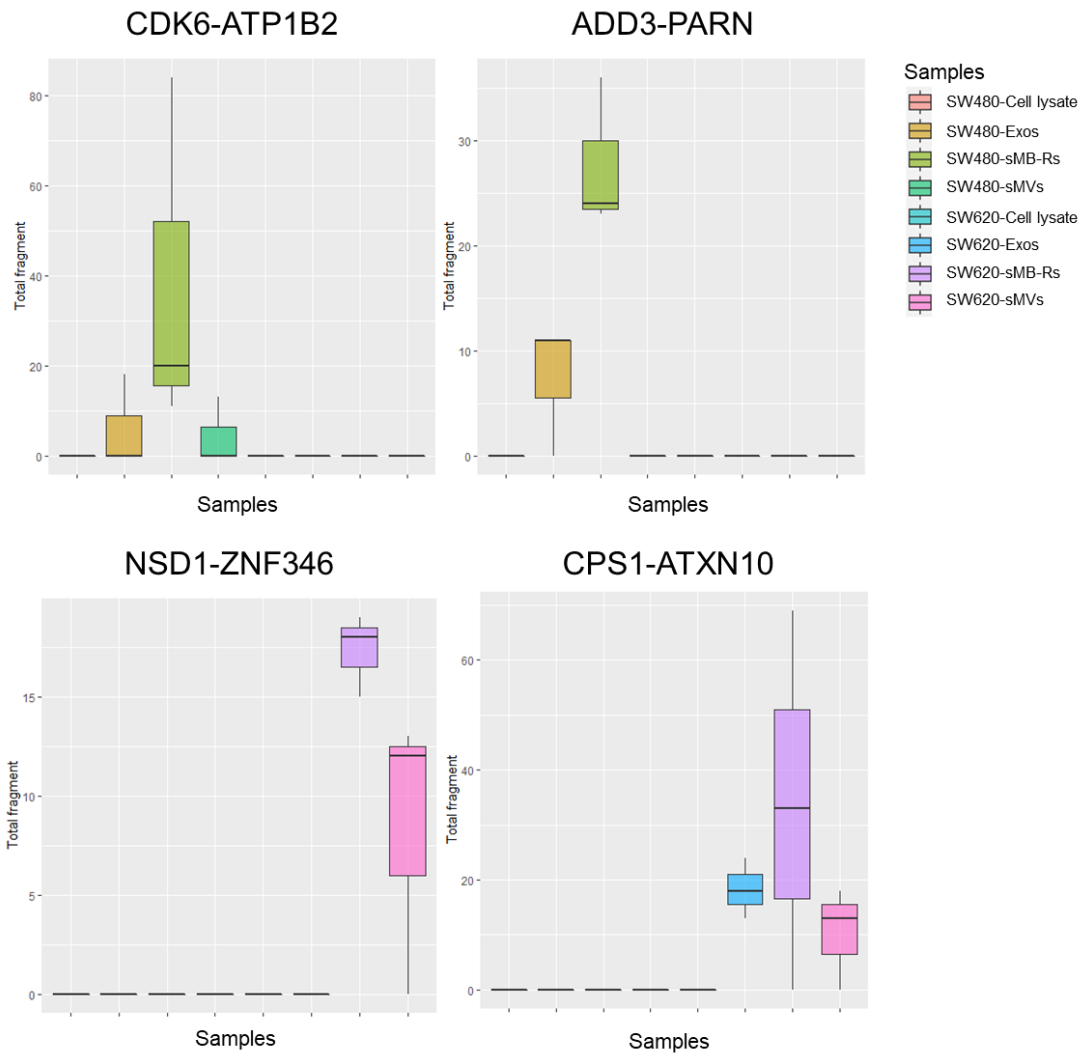

*
• PLAGL1-MSH2
• METRNL-MSH2
• CDK6-VWC2
• DNAJC1-MSH2
• HNRPLL-MSH2
• XYLT1-MSH2
• CDK6-NAT8L
• C16orf52-NOS1AP
• RAB12-NAT8L
• PIP4K2B-MMP15
*
• PLAGL1-MSH2
• METRNL-MSH2
• HNRPLL-MSH2
• XYLT1-MSH2
• WDR62-ARL4C
• WDR62-SSX2IP
• CACNA1B-MSH2
• MZT2B-CCDC160
• DNAJC1-MSH2
• FHL3-CCDC160
*
• OPN3-CHML
• RPL13A-SNORD32A
• RPPH1-RMRP
• ANKRD30BL-RCAN2
• AK311578-G3BP2
• BROX-AK025140
• RMRP-RPPH1
• SNORD96A-GNB2L1
• SFPQ-AL831889
• AKAP13-PDE8A
Exos (184)
sMVs (193)
Cell lysate (55)
sMB-Rs (577)
21
4
5
1
49
26
3
110
369
15
4
28
3
1
2
*
• PLAGL1-MSH2
• METRNL-MSH2
• HNRPLL-MSH2
• DNAJC1-MSH2
• XYLT1-MSH2
• WDR62-ARL4C
• CDK6-VWC2
• MZT2B-CCDC160
• WDR62-SSX2IP
• PIP4K2B-MMP15

D


Cell lysate
Exos
sMVs
sMB-Rs
Cell lysate
Exos
sMVs
sMB-Rs
PLAGL1-MSH2
METRNL-MSH2
HNRPLL-MSH2
DNAJC1-MSH2
XYLT1-MSH2
CACNA1B-MSH2
KIAA1217-MSH2
LHFPL2-MSH2
EHMT1-MSH2
B3GNT4-MSH2
RARA-MSH2
HRAS-MSH2
ARHGEF7-MSH2
TMEM88B-MSH2
WDR62-MSH2
INTS1-MSH2
ATG16L2-MSH2
GNA11-MSH2
STX6-MSH2
EPHB1-MSH2
CDK19-MSH2
TERF1-MSH2
CNIH4-MSH2
GNE-MSH2
KIF26A-MSH2
CUL2-MSH2
TMEM121-MSH2
HABP4-MSH2
NOS1-MSH2
C1QTNF4-MSH2
KCNA7-MSH2
XYLB-MSH2
NKX2-5-MSH2
SW480
SW620
14
12
10
8
6
4
2
0
log2(total fragment)


E
Cell lysate
Exos
sMVs
sMB-Rs
Cell lysate
Exos
sMVs
sMB-Rs
RAS
CDK
RRAS-BMP8B
HRAS-MSH2
C16orf52-DIRAS1
RRAS-GATAD2A
CDK6-VWC2
HNRNPU-CDK6
ENAH-CDK6
CDK19-MSH2
CDK6-STX4
CDK6-AK095633
CDK6-AK126225
CDK6-ATP1B2
CDK6-CAMK2N2
CDK6-GLRX5
CDK6-HIPK2
CDK6-IL17RC
CDK6-KCMF1
CDK6-NAT8L
CDK6-NDUFB10
CDK6-NFKBIE
CDK6-RNF44
2
1.5
1
0.5
0
-0.5
Z-score
SW480
SW620

F


CDK6-ATP1B2
ADD3-PARN
NSD1-ZNF346
CPS1-ATXN10
Total fragment
Samples
Samples
SW480-Cell lysate
SW480-Exos
SW480-sMB-Rs
SW480-sMVs
SW620-Cell lysate
SW620-Exos
SW620-sMB-Rs
SW620-sMVs

4.3.5 Highly-enriched cancer associated transcripts in SW480-/SW620-derived EV classes that modulate CRC progression

It has been shown that EVs derived from cancer cells mediate the tumor microenvironment to initiate and maintain hallmarks of cancer such as sustaining cell division and growth, evading immune cells, resisting cell death, reprograming neighboring cells, and acquiring genome stability[255]. Because EV classes in this study were secreted from human CRC adenocarcinoma (SW480 cells) and lymph-node metastasis (SW620 cells)) I questioned whether their transcript profiles might yield insights into cancer progression. Differential transcript expression analysis was performed using DESeq2[246] with log2fold change and p-value indicating transcript expression level and degree of statistical tests, respectively. Transcripts with log2 foldchange >1 (enriched in SW620-EVs) and <-1 (enriched in SW480-EVs) and pvalue <0.05 were considered as highly-enriched transcripts and Reactome pathway analysis was employed for functional annotation (see Section 4.2.6).

Differential transcript expression analysis revealed 1,759 and 3,923 transcripts were highly-enriched in combined SW480-EVs and combined SW620-EVs, respectively (**Fig. 4.5A**). Highly-enriched transcripts in SW480-EVs included CCND1-201 (protein coding, cyclin D1), Wnt-signaling transcripts such as WNT9A201, WNT6-201, WNT10A-201 (protein-coding, Wnt family member 9A, 6, 10A), receptors such as EGFR-201 (protein-coding, epithelial growth factor receptor), NOTCH2-201 (protein-coding, NOTCH receptor2). Metallopeptidases/ inhibitors such as ADAM19-201 (protein-

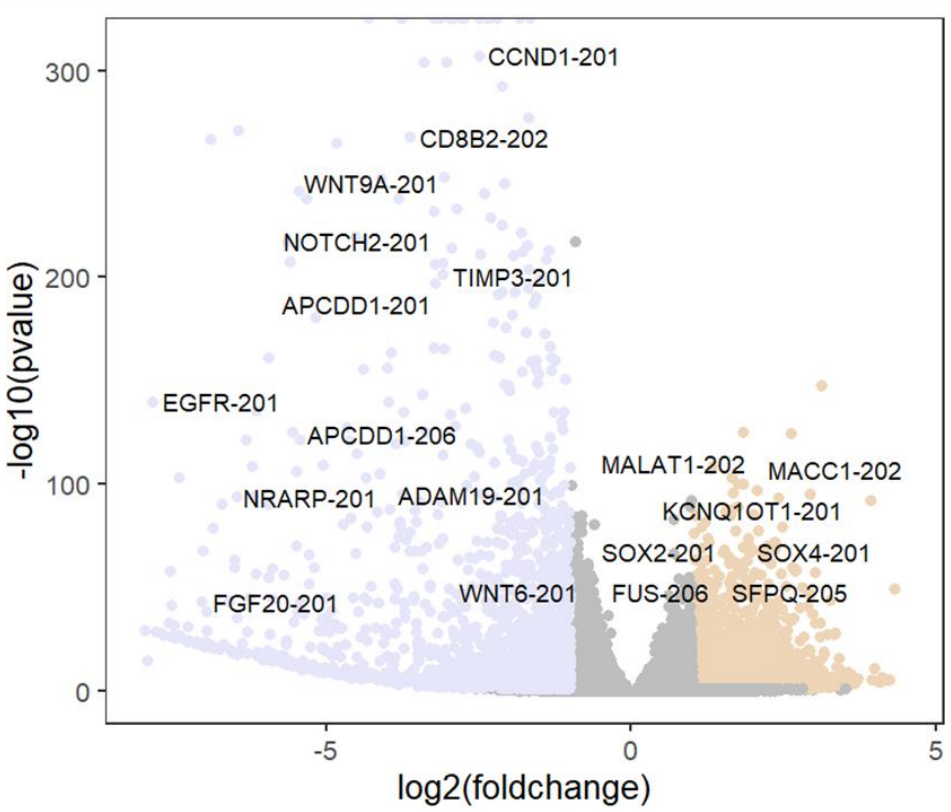
coding, ADAM metallopeptidase domain 19) and TIMP3-201 (protein-coding, TIMP metallopeptidase inhibitor 3) and growth factors such as FGF9-201, FGF20-201 (protein-coding, fibroblast growth factor 9, 20) (**Fig. 4.5A, B,** a list top 100 enriched transcripts in SW480- and SW620-EVs is given in **Appendix Table 4.14**). Reactome pathway analysis showed pathways related to 'Signaling by EGFR in cancer', 'Signaling by non-receptor tyrosine kinase', 'IGF1R signaling cascade' and 'PI3K/AKT signaling in cancer' were enriched in SW480-EVs (**Fig. 4.5C, Appendix Table 4.15**), whereas highly-enriched in transcripts in SW620 EVs included lncRNA (MALAT1-202, MANCR-201, NONO-211, SFPQ-205), transcription factors (SOX2-201, 4-201 (protein-coding, SRY-box 2, 4)), receptors (GPR183-201 (protein-coding, G protein-coupled receptor 183), ITPR2-202, (protein-coding, inositol 1,4,5-trisphosphate receptor type 2)) and signaling molecules such as metastasis-associated in colon cancer 1 also known as MET transcriptional regulator (MACC1-202, protein-coding), MAP2K6-206 (protein-coding, mitogen-activated protein kinase-kinase 6), TDGF-201 (protein-coding, teratocarcinoma-derived growth factor 1) and MARCKSL-201 (protein-coding, MARCKS-like 1) (**Fig. 4.5A, B**). Reactome pathway analysis showed that SW620-EVs are enriched in pathways related to RNA metabolism (e.g., 'Eukaryotic translation initiation', 'Translation initiation complex formation', 'mRNA splicing' and 'tRNA processing'), as well as 'Oncogenic MAPK pathway', 'PTEN regulation' and 'Nucleotide-binding domain, leucine rich

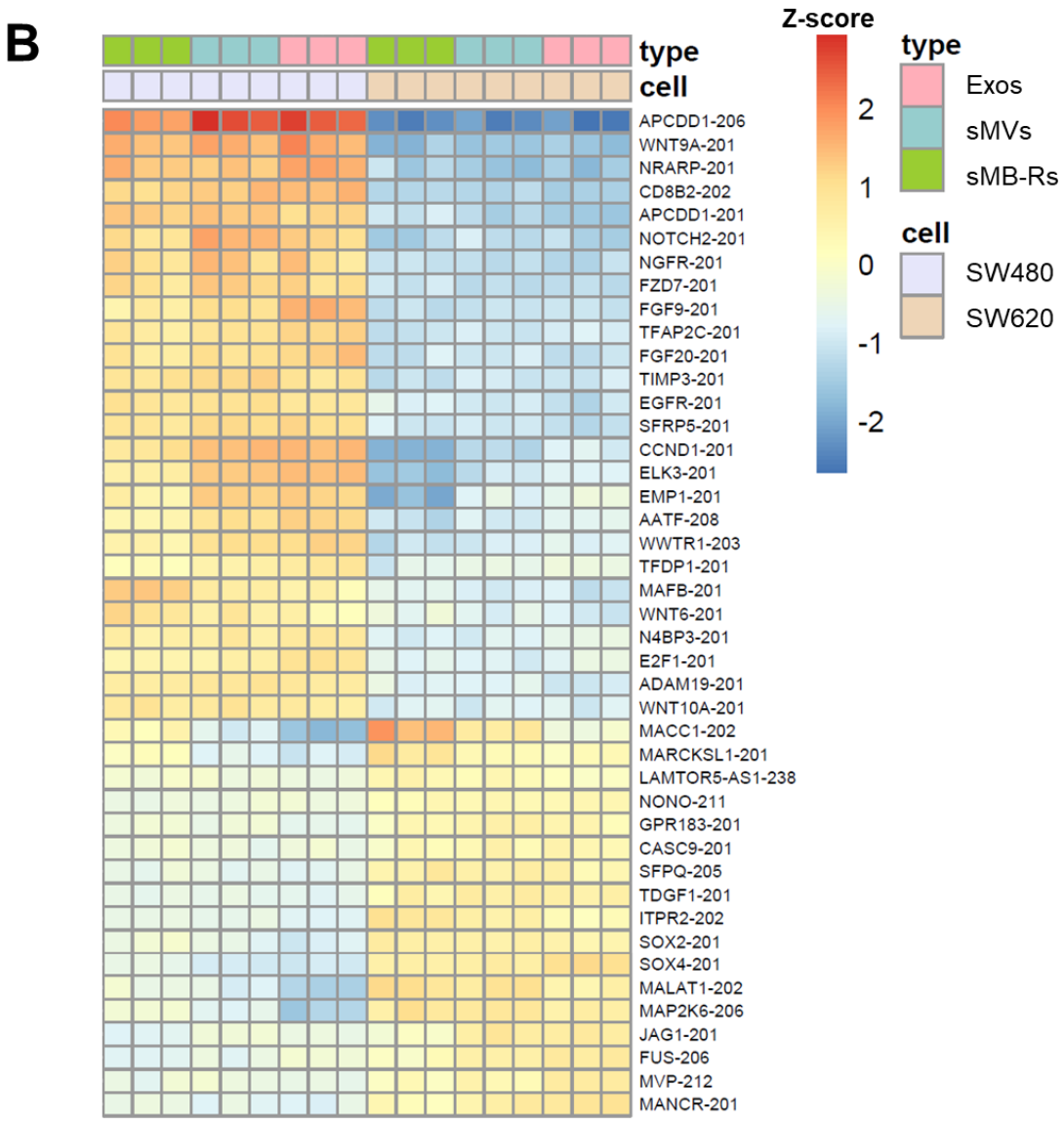
repeat containing receptor (NLR) signaling pathways' (**Fig. 4.5C**, **Appendix Table 4.15**).

These analyses show that SW480-EVs and SW620-EVs contain highly-enriched transcripts related to signaling molecules, receptors, growth factors, and transcription factors that might modulate the tumor microenvironment and play a role in CRC progression.

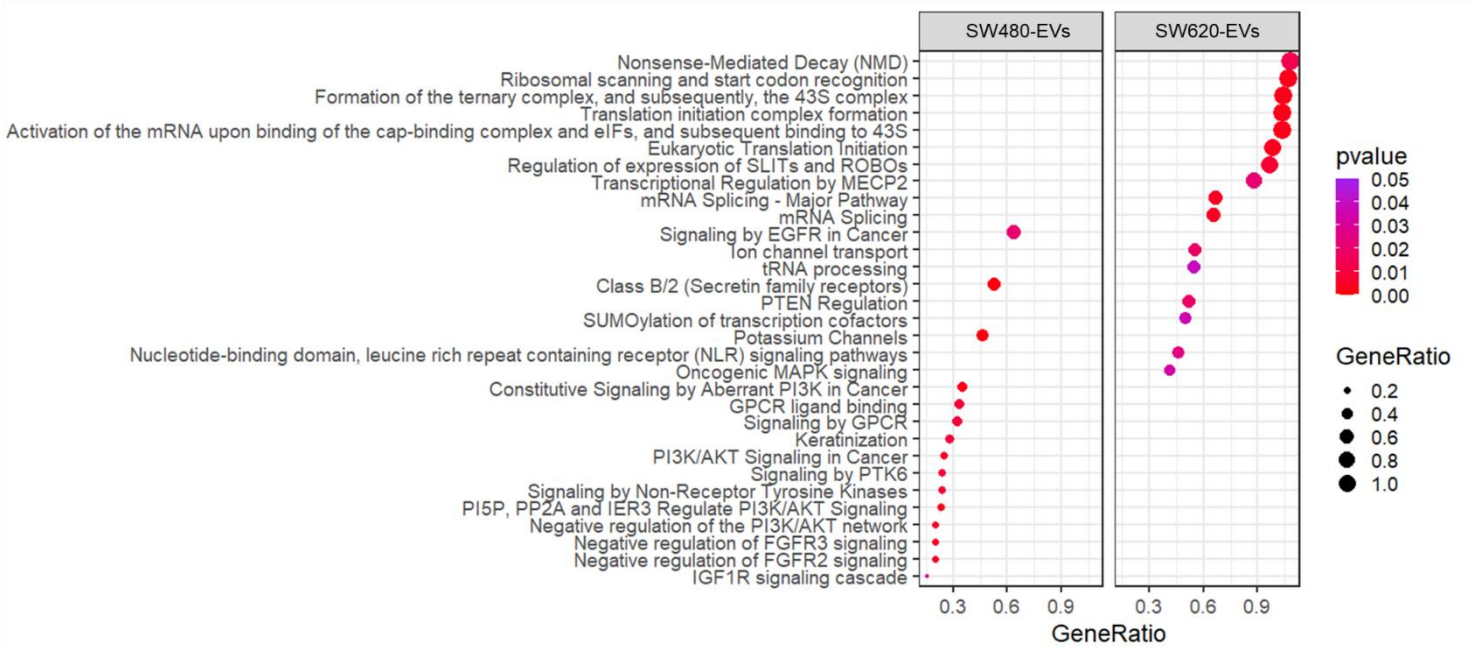
Figure 4.5 Identification of cancer progression-related transcripts and Reactome pathway analysis in EVs derived from SW480 and SW620 cells. (A) Differential transcript expression analysis of highly-enriched (log2 fold change <-1, pvalue < 0.05) cancer-associated transcripts in SW480-EVs (1,759 transcripts) and SW620-EVs (3,923 transcripts). **(B)** Heat map illustration of cancer-associated transcripts in SW480-/SW620-derived Exos, sMVs and sMB-Rs (scale shown is normalized counts subtracted by mean and divided by standard deviation). **(C)** Reactome pathway analysis of highly-enriched transcripts in SW480-EVs (n=1,759) and SW620-EVs (n=3,923) (cutoff: pvalue <0.05).

A


CCND1-201
CD8B2-202
WNT9A-201
NOTCH2-201
TIMP3-201
APCDD1-201
EGFR-201
APCDD1-206
MALAT1-202
MACC1-202
NRARP-201
ADAM19-201
KCNQ1OT1-201
SOX2-201
SOX4-201
FGF20-201
WNT6-201
FUS-206
SFPQ-205
300
200
100
0
-5
0
5
-log10(pvalue)
log2(foldchange)


B
type
cell
APCDD1-206
WNT9A-201
NRARP-201
CD8B2-202
APCDD1-201
NOTCH2-201
NGFR-201
FZD7-201
FGF9-201
TFAP2C-201
FGF20-201
TIMP3-201
EGFR-201
SFRP5-201
CCND1-201
ELK3-201
EMP1-201
AATF-208
WWTR1-203
TFDP1-201
MAFB-201
WNT6-201
N4BP3-201
E2F1-201
ADAM19-201
WNT10A-201
MACC1-202
MARCKSL1-201
LAMTOR5-AS1-238
NONO-211
GPR183-201
CASC9-201
SFPQ-205
TDGF1-201
ITPR2-202
SOX2-201
SOX4-201
MALAT1-202
MAP2K6-206
JAG1-201
FUS-206
MVP-212
MANCR-201
Z-score
2
1
0
-1
-2
type
Exos
sMVs
sMB-Rs
cell
SW480
SW620

C


SW480-EVs
SW620-EVs
Nonsense-Mediated Decay (NMD)
Ribosomal scanning and start codon recognition
Formation of the ternary complex, and subsequently, the 43S complex
Translation initiation complex formation
Activation of the mRNA upon binding of the cap-binding complex and eIFs, and subsequent binding to 43S
Eukaryotic Translation Initiation
Regulation of expression of SLITs and ROBOs
Transcriptional Regulation by MECP2
mRNA Splicing - Major Pathway
mRNA Splicing
Signaling by EGFR in Cancer
Ion channel transport
tRNA processing
Class B/2 (Secretin family receptors)
PTEN Regulation
SUMOylation of transcription cofactors
Potassium Channels
Nucleotide-binding domain, leucine rich repeat containing receptor (NLR) signaling pathways
Oncogenic MAPK signaling
Constitutive Signaling by Aberrant PI3K in Cancer
GPCR ligand binding
Signaling by GPCR
Keratinization
PI3K/AKT Signaling in Cancer
Signaling by PTK6
Signaling by Non-Receptor Tyrosine Kinases
PI5P, PP2A and IER3 Regulate PI3K/AKT Signaling
Negative regulation of the PI3K/AKT network
Negative regulation of FGFR3 signaling
Negative regulation of FGFR2 signaling
IGF1R signaling cascade
0.3
0.6
0.9
0.3
0.6
0.9
GeneRatio
pvalue
0.05
0.04
0.03
0.02
0.01
0.00
GeneRatio
0.2
0.4
0.6
0.8
1.0

4.4 Discussion and conclusions

In this chapter, I provided an in-depth transcriptomic analysis of EV classes (Exos, sMVs and sMB-Rs) derived from the isogenic human colorectal cancer SW480/SW620 cell line model (SW480 cells, surrogate of adenocarcinoma, and SW620 cells, surrogate of CRC metastasis)[128]. Inspection of transcript data indicates that the three EV classes have distinct transcriptome profiles (**Fig. 4.1D**) and the most abundant transcripts are (in order) protein-coding (mRNA), long non-coding, other (such as short non-coding) and pseudogene transcripts (**Fig. 4.1E**). This finding mirrors our observation that the three EV classes have distinct proteome profiles (**Chapter 3**, **Fig. 3.2C**). A major finding is that each of the EV classes, as well as containing transcripts in common, contain selectively-enriched transcripts (i.e., only seen in one EV class) – for example, sMB-Rs contain 134 uniquely identified transcript, Exos 34 transcripts and sMVs 9 transcripts (**Appendix Table 4.3**). It is not clear whether these differences translate into different functionalities of the EV classes. Clearly, further studies are required to address functional differences between the EV classes.

Differential transcription analysis was performed for each of the three EV classes. First, Exos displayed 136 and 737 transcripts highly-enriched compared to sMVs and sMB-Rs, respectively (**Fig. 4.2A, B**). These transcripts potentially translate into proteins previously shown to be Exo-cargo related to

Exos. For example, highly-enriched transcripts in Exos include Exo-related transcripts such as HSP90AA1-205 (heat shock protein 90 alpha family class A member 1)[256], ANXA2-223 (Annexin A2)[257] and BICD2-201 (BICD cargo adaptor 2). BICD protein interacts with the dynein–dynactin motor complex for the intracellular transport of small membranous organelles[258] and release of Exos[259]. An interesting finding was the presence of a novel transcript AC109326.1-201 highly-enriched in Exos compared to sMVs and sMB-Rs. Interestingly, several histone subunit transcripts are highly-enriched in Exos compared to sMB-Rs with GO terms such as 'DNA packaging complex', 'protein-DNA complex' and 'Nuclear nucleosome' were enriched in Exo-transcriptome (**Appendix Fig. 4.2A, B**). Whilst this finding suggests Exos might be involved in genomic DNA (gDNA) sorting/replication, histone subunits (essential for gDNA replication[260]) were not found to be enriched in my Exos proteome (**Chapter 3**, **Fig. 3.3E**). My data cannot distinguish intact DNA from DNA fragments. (It is noteworthy that DNA fragments containing mutant oncogenes have been recently reported in plasma[261]) Clearly, further studies are warranted to clarify this paradox.

Another key finding of this study was the enriched presence of mitochondrial, ribonucleoprotein, and signaling-related transcripts in sMB-Rs compared to Exos and sMVs (**Fig. 4.2E, F**) These transcripts include the mitochondrial transcripts (MT-CO1, MT-CO2, MT-CO3, MT-ND-5)[262], ribonucleoprotein-complex transcripts (NEAT1-202[263], LARP1-204[264]) and signaling-related transcripts (TGFB1-201[265]). GO analysis revealed enriched

mitochondrial-related GO terms ‘respiratory chain complex’, ‘mitochondrial inner membrane’, membrane-related GO terms (‘synapse part’, ‘postsynaptic specialization’) (**Appendix Fig. 4.2E, F)**. Importantly, ‘cytoplasmic stress granule’ and ‘transcription factor complex’ GO terms are significantly enriched in sMB-Rs compared to Exos (**Appendix Fig. 4.2E**). Enrichment of mitochondrial and ribonucleoprotein transcripts in sMB-Rs (relative to Exos and sMVs) poses the question whether mitochondria and ribonucleoprotein complexes might have co-purified along with sMB-Rs. Previous studies have shown that mitochondria and ribonucleoprotein complexes/RNA granules can be released into the extracellular space either in a free-form[52, 266] or embedded within EVs[267]. Mitochondria are known to sediment at 7,000-12,000 x *g* (10 min)[212, 268] and 18,000 x *g* (15 min) for RNA granules[211] which is similar to that required for microvesicle isolation[87, 164]. Proteomic analysis (**Chapter 3**) also demonstrated abundant mitochondrial markers such as TOMM22 and VDAC1/2 in sMB-Rs. Interestingly, sMB-Rs not only contain ribonucleoprotein-complex transcripts but also their ribonucleoprotein complexes protein partners such as splicing factors, and translation initiation factors (see **Chapter 3**, **Fig. 3.3E**). This observation suggests a tight relationship of RNA and their cognate protein complexes in sMB-Rs. Following this observation, I next questioned whether enriched transcripts in sMB-Rs might bind to RBPs. Target RBPs for the highly-enriched transcripts in sMB-Rs were identified (27 in total) in the ENCORI database[248]. Functional annotation of these 27 RBPs (using gProfiler[247]) suggested GO terms such as ‘ribonucleoprotein complex’/

'spliceosomal complex' (HNRNP families, U2AF2), 'ribonucleoprotein granule' (ELAVL1, FUS, IGF2BP families) and 'nuclear speckles' (ALYREF, METTL3) (**Appendix Table 4.12**). Intriguingly, 19/27 putative RBPs (**Table 4.1**) are found in my proteome data (**Appendix Table 4.11**). For example, NEAT1-202 that is associated with RNA retention and RNA granules[252, 269] is predicted to bind to RNA/stress granule protein markers (FUS, TARDBP, IGF2BP2, TAF15, ELAVL1) and ribonucleoproteins (HNRNP families) (all of these RBPs were identified in my proteomic data, **Chapter 3**) (**Fig. 4.3B**). The co-existence of these RBPs and their cognate transcripts suggests they might occur as pre-formed complexes in SW480/SW620-sMB-Rs. It is interesting to speculate that pre-formed RNA/protein complexes in sMB-Rs might be implicated in RNA sorting[222, 270]. Whilst this concept has been proposed as an RNA-sorting mechanism in Exos[222, 270], it has not been previously reported for larger EVs such as sMB-Rs.

Highly-enriched mitochondrial transcripts (MT-CO1-201, MT-CO2-201, MT-CYB-201) were found to be prominent in sMVs when compared with Exos, but to a lesser extent when compared with sMB-Rs (**Fig. 4.2E, F**). However, sMVs showed highly enriched histone subunit transcripts when compared with sMB-Rs, but not enriched when compared to Exos (**Fig. 4.2F**). In contrast to Exos and sMB-Rs, transcript levels in sMVs were of moderate abundance **(Fig. 4.2G)**.

EV-containing fusion genes are gaining much attention as potential targets in biomarker discovery[271]. In my study, 770 fusion genes were detected

across all samples (cell lysate, Exos, sMVs and sMB-Rs). It is interesting that fusion gene enrichment levels much higher in sMB-Rs compared to other Exos, sMVs and cell lysates (**Fig. 4.4A**). I detected the tumour suppression gene MSH2 linked to many other genes. The fusion gene analysis showed 33 novel MSH2 fusion genes were highly detected in EV classes (Exos, sMVs, sMB-Rs, but not in SW480/SW620 cell lysates) (**Fig. 4.4D**). MSH2 is a key mammalian DNA mismatch repair (MMR) gene and mutations or deficiencies in mammalian MSH2 gene result in microsatellite instability (MSI^{+})[272]. SW480 and SW620 cells have been characterized as microsatellite stable (MSS)[130] and it is interesting that most MSH2 fusion genes were not detected in SW480- and SW620-cell lysate but were highly abundant in both SW480-EVs and SW620-EVs. It is not clear why (or how) low-abundance MSH-2 fusion genes, presumably at almost undetectable levels in the cytoplasm, selectively traffic to EVs where they are found at abundant levels. If MSH-2 fusion genes are considered to be toxic to genome stability, their selective enrichment in EVs might offer a mechanism for their removal from the cell. Clearly, this hypothesis warrants further experimentation. Another key finding is the selective enrichment of the novel fusion genes CDK6-ATIP1B2 and ADD3-PARN in SW480-sMB-Rs but not in SW480 cell lysates (**Fig. 4.4F**). Similarly, NSD1-ZNF346 and CPS1-ATXN10 (TCGA-computationally annotated[245]) fusion genes were uniquely detected in SW620-EVs (and not in SW620 cell lysate) with high enrichment in SW620-sMB-Rs (**Fig. 4.4F**). NSD1-ZNF346 fusion gene has been detected in high-grade serous carcinoma[273]. Thus, NSD1-

ZNF346 fusion gene could be a biomarker for aggressive cancer, but is not specific for metastatic colorectal cancer. More clinical EV samples are required to increase specificity of fusion gene-base biomarkers.

EV-derived RNAs have been shown to mediate the tumor microenvironment and maintain hallmarks of cancer[79]. Because EV classes in this study were isolated from different cell origins (colorectal adenocarcinoma (SW480) and lymph-node metastasis cancer cell (SW620)) that have distinct morphology, tumorigenic properties[274] and clinical cancer stages[128], I opined whether there is a difference of transcriptome between combined SW480-EVs and combined SW620-EVs. Differential transcript expression analysis revealed 1,759 and 3,923 transcripts were highly-enriched in SW480-EVs and SW620-EVs, respectively (**Fig. 4.5A**). The enriched transcripts in SW480-EVs that are important signaling pathways for cancer proliferation and growth, include signaling related to Wnt[275] (WNT9A201, 6-201, 10A-201), EGFR[276] (EGFR-201), NOTCH[277] (NOTCH2-201) and extracellular remodeling mediators such as ADAM19-201 and TIMP3-201[278]. In this study, I found that my transcriptomic data (for example, EGFR) correlates with enriched levels of EGFR protein found in SW480-Exos[59], SW480-sMVs[164] and SW480-sMB-Rs (**Chapter 3**). Similarly, I found a correlation of enriched levels of NOTCH2 protein and transcripts in SW480-sMVs[164] and SW480-sMB-Rs (**Chapter 3**) compared to the same EV classes derived from SW620. Interestingly, SW620-EVs were enriched in lncRNA (MALAT1-202[250]), transcription factors (SOX2-201[279]) and signaling transcript molecules such as metastasis-associated in

colon cancer 1/ MET transcriptional regulator (MACC1-202[280]). This finding is consistent with our observation that the MACC1 protein is also enriched in SW620-Exos, -sMVs and -sMB-Rs (**Table 3.2**), compared to the same EV classes derived from SW480. Reactome pathway analysis demonstrated that SW620-EVs are enriched in pathways related to translation processes and signaling pathways - for example, 'Oncogenic MAPK pathway' (**Fig. 4.5C**) which is consistent with previously reported transcriptome analyses of primary colorectal adenocarcinomas and metastatic colorectal cancer[281] where metastatic cancers have been shown to display aberrant RNA translation processes[282, 283].

In summary, the transcriptome of sMB-Rs is dissimilar to Exos and sMVs with high enrichment of mitochondrial transcripts, lncRNA/pseudogene transcripts that are predicted to bind to ribonucleoproteins, splicing factors and RNA/stress granule proteins. The transcriptome of sMB-Rs is highly enriched with many fusion genes, some of which have not been hitherto described in the literature (CDK6-ATIP1B2, ADD3-PARN). Exos contain transcripts related to release and biogenesis of Exos such as BICD2-201, ARRDC3-201 and CAV2-201. Unlike the other two EV classes, sMVs do not contain a distinct transcript signature. Importantly, the expression of cancer progression-related transcripts found in EV classes derived from SW480 and SW620 cells reflects their parental cell types and positively correlate with protein abundance levels (described in **Chapter 3, Table 3.2**). Finally, this chapter provides, for the first time, a comprehensive transcriptomic analysis of three different EV classes

(Exos, sMVs and sMB-Rs) secreted from human CRC cells SW480 and SW620. My findings, I believe, provide a better understanding of RNA compositions in EV classes and how they might impact on EV functionality in cancer progression. This study also offers several potential EV-based RNA/ fusion gene candidates for cancer biomarkers.

4.5 Future perspectives

There are some questions arising from my thesis that have not been fully addressed due to time constraints and unavailability of reagents and new technologies. To improve this chapter, I would like to perform qRT-PCR to validate enriched transcripts/fusion genes in EV classes. I would also like to perform an immunoprecipitation technique using anti-TOMM22 and FUS/IGF2BP2 to detect free-mitochondria and free-RNA granules to investigate whether mitochondria and RNA granules are precipitated during sMB-Rs isolation for deeper characterization of sMB-Rs. Another approach to characterize EV classes is to investigate unannotated transcripts derived from EV classes. I have established a method for identifying unannotated transcripts from RNA sequencing data (the published article is attached below). Due to time (and budget) constraints I could not perform these experiments. I believe such experiments might provide an in-depth understanding of EV class/ subtype heterogeneity which, in turn, might allow the development of EV isolation kits for increasing of sensitivity and clinical utility of EV-based biomarker.

Wittaya Suwakulsiri, Maoshan Chen, David W. Greening, Rong Xu and Richard J, Simpson. Analysis of Annotated and Unannotated Long Noncoding RNAs from Exosome Subtypes Using Next-Generation RNA Sequencing. *Methods in Molecular Biology*, Haiming Cao (Editor), vol. 2254. PMID:

33326077. https://doi.org/10.1007/978-1-0716-1158-6_12 (**Appendix Article 3**).

Chapter 5

Summary and future directions

5.1 Background context of my thesis

At the commencement of my PhD there were at least two well-defined classes of EVs described in the literature – namely, exosomes (Exos) and shed microvesicles (sMVs)[68, 79]. These classes were identified primarily on the basis of their mechanism of biogenesis. Exos biogenesis involves the inward budding of the plasma membrane (PM) to form the early endosome, which matures to form a late endosome. During this process the limiting membrane of the late endosome buds inwards and pinches off to form small (50-200 nm) membrane-enclosed vesicles referred to as intraluminal vesicles (ILVs) within late endosomes – now referred to as multivesicular bodies (MVB). MVBs are integrated into the endosomal recycling system and traffic to and fuse with the PM thereupon releasing their ILV cargo (now termed Exos) into the extracellular space. By contrast, sMVs (also known as microparticles and ectosomes) form by direct budding from the PM. Despite differences in their mechanism of biogenesis and membrane of origin – the MVB limiting membrane in the case of Exos and PM for sMVs - Exos and sMVs are functionally similar following release into the extracellular space but molecularly distinct at the proteome level[163]. During the initial phase of my PhD, I was involved in work underway in the Simpson lab that led to the identification of a novel third class of EV referred to as shed midbody remnants (sMB-Rs). sMB-Rs arise during the final stages of cell division whereby newly-formed daughter cells remain connected by a thin intercellular bridge containing the midbody (MB), a microtubule-rich organelle responsible for

cytokinetic abscission. Following cell division, the MB is asymmetrically inherited by one daughter cell where it persists as a midbody remnant (MB-R) (for a summary of our current understanding of the biogenesis of the three EV classes, see **Fig. 1.7**). Accumulating evidence shows sMB-Rs are secreted into the extracellular medium and engulfed by neighbouring non-sister cells[70]. This latter finding led me to focus on the systematic comparative analysis of the protein and RNA cargo in the three EV classes (Exos, sMVs, and sMB-Rs) as the overarching aim of my thesis. As a biological model for my thesis, I focused on the release of the three EV classes from the isogenic human CRC cell lines SW480 (derived from primary colorectal cancer tumor) and SW620 (derived from lymph node-metastatic colorectal cancer tumor from the same patient).

The following is a summary of the major findings of my thesis:

5.2 Comparative proteome analysis of Exos, sMVs, and sMB-Rs released from colorectal cancer SW480 and SW620 cells

In **Chapter 2**, I focused on the characterizing the proteome of sMVs isolated and purified from the isogenic human CRC cell lines (SW480 cells, surrogate of adenocarcinoma, and SW620 cells, surrogate of CRC metastasis[128]). The purification strategy involved a combination of differential ultracentrifugation and iodixanol density-based separation. Purified sMVs were obtained at 10,000 x *g* centrifugation and buoyant density 1.10 g/ mL. I first determined biophysical properties of sMVs using transmission electron microscopy (TEM) and nanoparticle tracking analysis (NTA). TEM revealed sMVs display round-like membranous vesicle structures 100-500 nm in size, and NTA showed heterogenous particle mean diameters of 350±28.4 nm for SW480-sMVs and 337.1±9.3 nm for SW620-sMVs. In addition, Western blot analysis shows that sMVs are $ALIX^-$, $TSG101^-$, $CD9^-$, and $CD63^-$ in comparison with Exos. Label-free quantitative mass spectrometry identified 834 common proteins in both SW480- and SW620-sMVs. GO analysis of the 834 proteins identifies enriched following processes – 'cell adhesion/adhesion molecules' (PKP2/3, CLDND1, OCLN, CTN families), 'signal transduction' (KRAS, NRAS, RHOA, MAPK1/2K1), 'RNA binding proteins' (e.g., ribonuclear proteins HNRNPK/Q/E2, YBX-1), 'translation-associated proteins' (ribosomal proteins, translation initiation factors, aminoacyl tRNA ligases), and 'transport vesicle' (VPS29/35, SNAP23). Interestingly, cancer progression-related proteins found in SW480- and SW620-sMV reflects their parental cell types which is

consistent with cancer progression-related proteins identified in SW480- and SW620-Exos[59, 134]. For example, EGFR, CD44, CLDN7 are enriched in SW480-sMVs while MET, PRKCA, MACC1, CLDN1 are abundant in SW620-sMVs.

In **Chapter 3**, I isolated, purified and characterized sMB-Rs using the same starting materials as in **Chapter 2**. sMB-Rs were purified from 10,000 x *g* centrifugation with high buoyant density 1.15 g/mL. Using the same materials, I also isolated and purified Exos (from supernatant of 10,000 x g centrifugation) at 100,000 x *g* centrifugation (buoyant density 1.10 g/mL) to perform comparative proteomic analysis of the three EV classes. TEM revealed sMB-Rs are ellipsoid in shape and heterogenous in size (100-500 nm) and NTA showed mean particle diameters of 404.9±13.6 nm/401.2±1.7 nm for SW480/SW620-sMB-Rs. Western blot analysis revealed ALIX, TSG101, CD63, CD9 and CD81 were more enriched in Exos than in sMVs (**Fig. 3.1H**), a finding consistent with previous studies[204]. The midbody centraspindlin complex components KIF23/MKLP1 and RACGAP1[96] are selectively enriched in sMB-Rs compared to Exos and sMVs. Proteome of sMB-Rs is very dissimilar to Exos and sMVs. Using differential protein enrichment analysis and KEGG pathway analysis, I identified selectively-enriched proteins in Exos, sMVs and sMB-Rs. Exos are selectively-enriched in exosomal protein markers (CD63, CD81, CHMP4B (ALIX), TSG101)[161, 163, 206] and enriched KEGG pathways such as ‘Endocytosis’ (hsa04144) and ‘SNARE interactions in vesicular transport’ (hsa04130) and ‘Cell adhesion

molecules' (hsa04514). sMVs are selectively-enriched in metabolic enzymes (DTYMK, IMPA1 and MRI) and membrane-associated proteins (SLC29A2). sMB-Rs are selectively-enriched in centraspindlin and midbody components[96] (KIF23/MKLP1, RACGAP1, PLK1, AURKB), mitochondrial proteins (TOM22, VDAC1, VDAC2), proteins in ribonucleoprotein complex/ splicing complex/ RNA granules (HNRNPs, SF3s, FUS, IGF2BP1) and enriched KEGG pathways such as 'RNA transport' (hsa03013), 'Spliceosome' (hsa03040), 'Protein processing in endoplasmic reticulum' (hsa04141) and 'Citrate cycle (TCA cycle) and mitochondrial enzymes' (hsa00020). Furthermore, I observed that cancer progression-related proteins are selective sorted in EV classes. Several receptors and transporters are highly enriched in Exos and sMVs (CD44, EGFR, FAS in SW480-Exos/sMVs and MET, FGFR4 in SW620-Exos/sMVs). Signaling molecules such as STAT1, GYS1 and CTNNBL1 were highly-enriched in SW480-sMB-Rs and GDF15, ADAM15, TNC were highly-enriched in SW620-sMB-Rs.

In summary, **Chapters 2** and **3** of my thesis provide, for the first time, an in-depth comparative proteomic analysis of three EV classes (Exos, sMVs and sMB-Rs) which were isolated and purified from SW480 and SW620 cancer cells. This comparative analysis provides a comprehensive understanding of the protein cargos in three EV classes (Exos, sMVs, and sMB-Rs) from CRC SW480 and SW620 cell lines.

5.3 Comparative transcriptomic analysis of Exos, sMVs and sMB-Rs released from colorectal cancer SW480 and SW620 cells

In **Chapter 4**, I undertook a comparative transcriptomic analysis of Exos, sMVs and sMB-Rs from the same material I used for the proteome studies described in **Chapters 2** and **3**. Total RNA was isolated from the same materials as in **Chapter 3** and sequenced using RNA-next generation sequencing technique (Illumina HiSeq 4000). The most abundant transcripts in Exos, sMVs and sMB-Rs are (in order) protein-coding (mRNA), long non-coding, other (such as short non-coding) and pseudogene transcripts. Interestingly, the transcriptome of sMB-Rs is distinct from Exos and sMVs (observed in EVs from both SW480 and SW620 cells). Differential transcript expression analysis and Gene Ontology (GO) analysis revealed that sMB-Rs are highly-enriched in mitochondrial transcripts (MT-CO1-201, MT-CO2-201, MT-CO3-201, MT-CYB-201, MT-ND1-201), transcripts that are translated to transcription factor/ signaling proteins (SOX12-201, TGFB1-201), lncRNA/pseudogene transcripts (NEAT1-201, KCNQ1OT1-201, GABPB1-AS1-202) that are predicted to bind to ribonucleoproteins (HNRNPs), splicing factors (SRSF1, U2AF2) and RNA/stress granule proteins (FUS, TAF15) that were detected in proteome of sMB-Rs (**Supplementary Fig. 3.4**). Interestingly, sMB-Rs also contain higher number of fusion genes compared to Exos, sMVs and cell lysates derived from SW480 and SW620. I identified fusion genes from tumour suppressor gene (MSH2) such as PLAGL1-MSH2, METRNL-MSH2 and oncogenes (CDK, RAS) such as RRAS-BMP8B and

CDK-STX4 selectively enriched in sMB-Rs. Moreover, NSD1-ZNF346 and CPS1-ATXN10 are only detected in SW620-EVs, with selective enrichment in sMB-Rs. This finding suggests sMB-Rs-derived fusion genes should be considered as liquid biopsy targets for metastatic cancer detection. Exos contain transcripts that are translated to proteins related to release and biogenesis of Exos such as BICD2-201, ARRDC3-201 and CAV2-201. Unlike the other two EV classes, sMVs do not contain a distinct transcript signature. Importantly, the expression of cancer progression-related transcripts found in Exos, sMVs and sMB-Rs derived from SW480 and SW620 cell lines positively correlate with known proteins (EGFR, CD44, FAS, MET, MACC1, PRCAKA) found in their parental cell types (**Table 3.2**). Exos, sMVs and sMB-Rs secreted from SW620 cells are enriched in lncRNA transcripts that have been shown to promote cancer progression such as MALAT1-202[250, 284] and MANCR-201[285].

In conclusion, my **Chapter 4** is the first report of a comprehensive transcriptomic analysis of three different EV classes (Exos, sMVs and sMB-Rs) secreted from human CRC SW480 and SW620 cells. These findings provide a better understanding of RNA compositions in EV classes and their possible roles in cancer progression. This study also provides several potential EV-based RNA/ fusion gene candidates for cancer biomarkers.

5.4 Conclusions and future directions

Cells release at least three EV classes that vary in size, cargo components and mechanism of biogenesis. One outstanding question in the EV field is whether EV classes show different functionalities upon uptake by a recipient cell(s). To understand their precise functionality, EV classes need to be well-characterized with respect to their protein and RNA compositions. Hence, to tackle this important question, and to enable comparison between laboratories, comprehensive biochemical and biophysical characterization of EV classes is essential.

This thesis demonstrates a novel comparative proteomic and transcriptomic analysis of three EV classes (Exos, sMVs and sMB-Rs) isolated and purified from an isogenic human colorectal cancer cells SW480 (adenocarcinoma) and SW620 (lymph node-metastatic cancer) using a combination of differential ultracentrifugation and iodixanol density-based separation. sMB-Rs from both SW480 and SW620 cells show highly distinct proteomic and transcriptomic profiles from Exos and sMVs (**Chapter 3 and 4**). At the protein level (**Chapter 3**), midbody components, mitochondrial proteins, histone subunits and RNA-binding proteins (RBPs) such as ribonucleoproteins, RNA stress/granule proteins, splicing factors and translation initiation factors are selectively-enriched in sMB-Rs (**Fig. 5.1**). Interestingly, the transcriptome of sMB-Rs reveals highly-enriched mitochondrial transcripts, lncRNA/pseudogene transcripts (**Chapter 4**) that are

predicted to bind to RBPs (identified in **Chapter 3)** and novel fusion genes. Exosomal proteins such as CD63, CD81, CHMP4B (ALIX), TSG101 as well as transcripts that translate into proteins involved in exosome biogenesis/ release (ARRDC3-201, BICD2-201, CAV2-201) (**Chapter 4**) are selectively-enriched in Exos (**Fig. 5.1**). Metabolic enzymes (DTYMK, IMPA1 and MRI) and membrane-associated proteins (SLC29A2) are selectively-enriched in sMVs (**Fig. 5.1**).

Expression of cancer progression-related proteins and transcripts in Exos, sMVs and sMB-Rs (**Chapters 3** and **4**) to a large extent positively correlate with the two -omic profiles. For example, EGFR protein and transcript are enriched in SW480-EV classes and MET, MACC1 proteins and transcripts are enriched in SW620-EV classes (**Fig. 5.2**). Moreover, SW620-EV classes contain enriched non-coding transcripts that are known to promote cancer progression (**Fig. 5.2**).

There are outstanding questions and incomplete confirmation of transcript findings arising from my findings in this thesis that have not been fully answered due to time constraints, unavailability of reagents and new technologies. For example, - 1) orthogonal confirmation of proteins/ transcripts identified in EV classes is required, - 2) the question of EV purity needs further examination, - 3) the possibility of additional RNA species in my EV samples and incomplete annotation of RNA species needs to be explored further, and - 5) there is a pressing need to further investigate the tumorigenic potential of the different EV classes derived from CRC cell lines (and other cancer types).

To address the first question, I would like to perform Western blot analysis on highly-enriched proteins as well as qRT-PCR on highly-enriched transcripts to confirm their expression levels in EV classes. To answer the second question, I would like to perform extensive purification of sMB-Rs using an immunoprecipitation technique with anti-KIF23 and anti-RACGAP1 to capture sMB-Rs and perform mass spectrometry on immunocaptured sMB-Rs. This approach could answer whether mitochondria and RNA stress/ granules are co-purified or integral components of sMB-Rs. To address the third question, I would like to perform small RNA analysis (e.g., miRNA, snRNA, snoRNA). I have also published an RNA analysis method for the identification of unannotated transcripts[286] and I would like to apply this method directly to the RNA sequencing data in **Chapter 4**. I would also like to identify fusion proteins (from fusion genes observed in **Chapter 4**) and attempt to identify neoantigens (unidentified proteins in UniProt database) using my transcriptome data of Exos, sMVs, and sMB-Rs. To answer the last question, I would like to perform a tumour xenograft experiment by injecting SW480- and SW620-EV classes into mice and monitor tumour growth[196].

Collectively, I believe my thesis findings pave the way to advancing the characterization of EV classes and - in doing so - may impact on our understanding of intercellular communication, cancer progression and EV-based liquid biopsy in clinical use.

Figure 5.1 Schemata of highly-enriched proteins and transcripts in Exos, sMVs and sMB-Rs derived from SW480 and SW620 cells. Exos contain proteins involved in exosome biogenesis/ release/ trafficking such as CD63, CD81, CD82, TSG101, VPS25, TSPAN1 and TSPAN14 as well as transcripts that are translated into proteins in exosome biogenesis/ release/ trafficking such as BICD2-201, CAV2-201 and ARRCD3-201. sMVs are enriched in metabolic enzymes (DTYMK, IMPA1 and MRI), membrane-associated proteins (SLC29A2) as well as long non-coding transcripts, AC018752.1-201 (novel) and RGS12-219. sMB-Rs are enriched in midbody components, histone subunits and RNA-binding proteins (RBPs) such as ribonucleoproteins, RNA stress/granule proteins, splicing factors and translation initiation factors as well as mitochondrial proteins/transcripts. sMB-Rs also contain long non-coding transcripts (NEAT1-202, GABPB1-AS1-202) that are predicted to bind to enriched RBPs

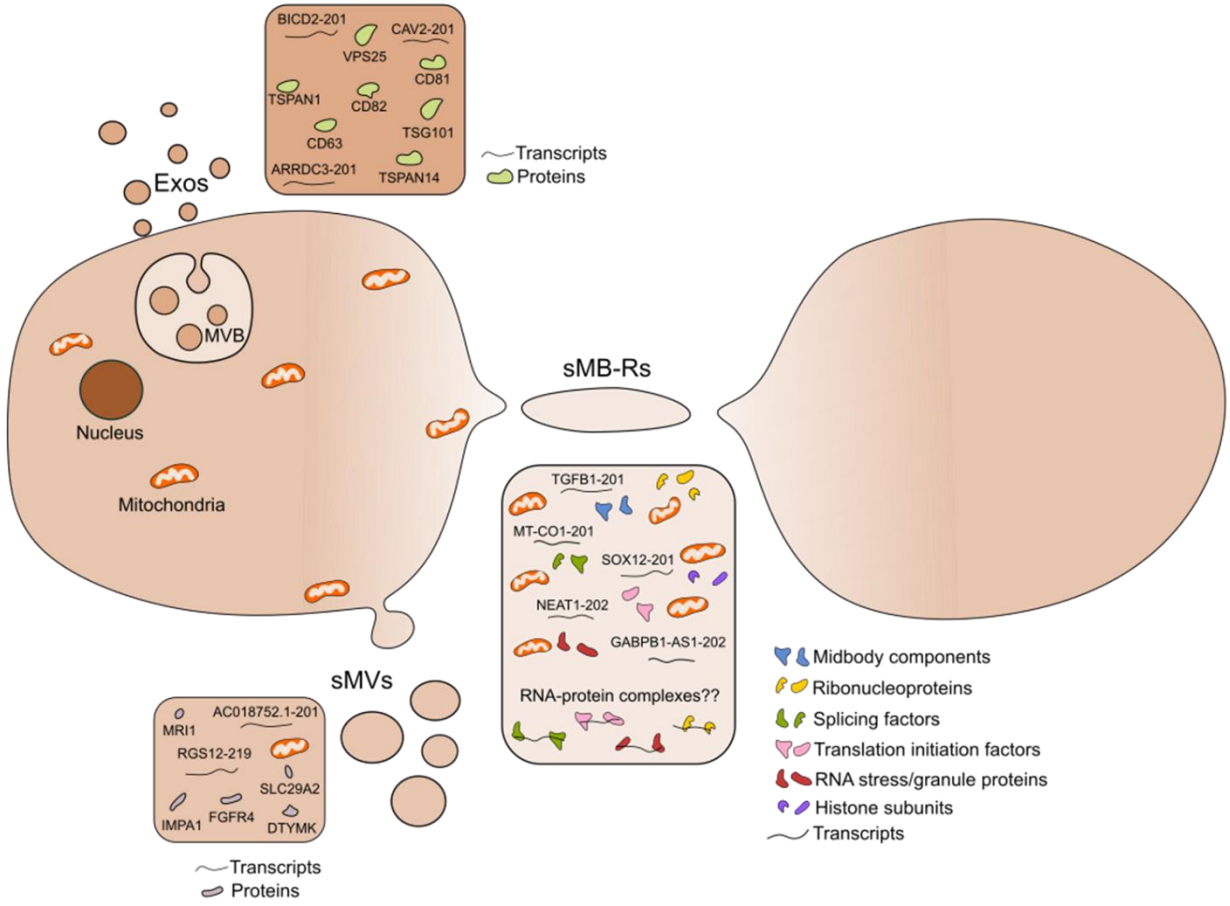
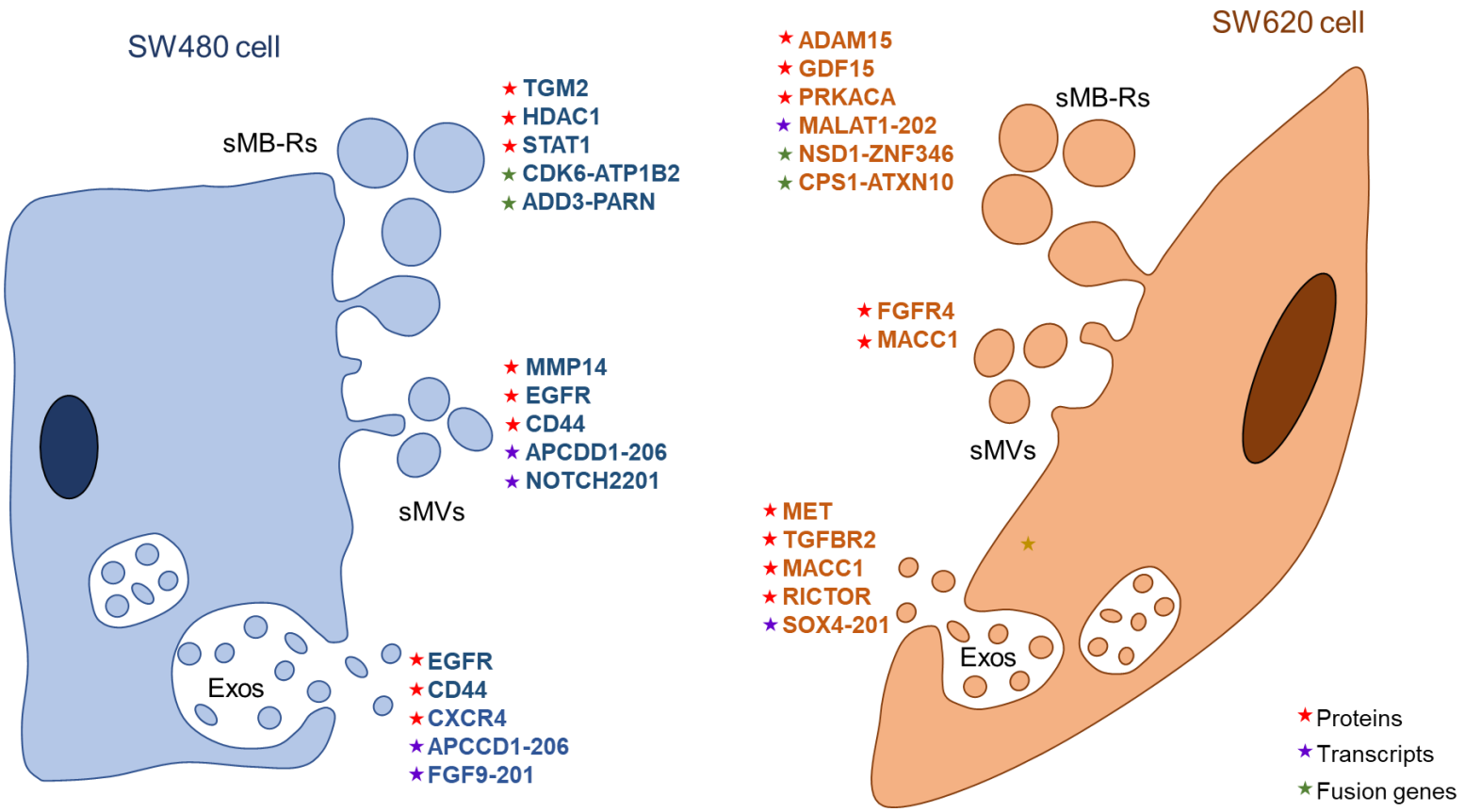

BICD2-201
CAV2-201
VPS25
CD81
TSPAN1
CD82
TSG101
CD63
ARRDC3-201
TSPAN14
Transcripts
Proteins
Exos
MVB
Nucleus
Mitochondria
sMB-Rs
TGFB1-201
MT-CO1-201
SOX12-201
NEAT1-202
GABPB1-AS1-202
RNA-protein complexes??
sMVs
AC018752.1-201
MRI1
RGS12-219
SLC29A2
IMPA1
FGFR4
DTYMK
Transcripts
Proteins
Midbody components
Ribonucleoproteins
Splicing factors
Translation initiation factors
RNA stress/granule proteins
Histone subunits
Transcripts

Figure 5.2 Schemata of selected cancer progression-related proteins, transcripts and fusion genes in Exos, sMVs and sMB-Rs derived from SW480 and SW620 cells. Exos from SW480 contain receptors/proteins such as EGFR, CD44, CXCR4 and transcripts such as APCCD1-206 and FGF9-201. SW480-sMVs contain MMP14 protein and similar abundance of EGFR and CD44 proteins, APCDD1-206 transcript to SW480-Exos. sMB-Rs from SW480 contain TGM2, HDAC1, STAT1 and novel fusion genes (CDK6-ATP1B2, ADD3-PARN). In SW620, Exos contain receptors/proteins such as MET, TGFBR2, MACC1 and RICTOR as well as SOX4-201 transcript. SW620-sMVs contain FGFR4 and similar abundance of MACC1 proteins to SW620-Exos. sMB-Rs from SW620 contain proteins such as ADAM15, GDF15, PRKACA and long non-coding transcript (MALAT1-202). Furthermore, fusion genes such as NSD1-ZNF346 and CPS1-ATXN10 are enriched in SW620-sMB-Rs.


SW480 cell
sMB-Rs
TGM2
HDAC1
STAT1
CDK6-ATP1B2
ADD3-PARN
MMP14
EGFR
CD44
APCDD1-206
NOTCH2201
sMVs
Exos
EGFR
CD44
CXCR4
APCCD1-206
FGF9-201
SW620 cell
ADAM15
GDF15
PRKACA
MALAT1-202
NSD1-ZNF346
CPS1-ATXN10
sMB-Rs
FGFR4
MACC1
sMVs
MET
TGFBR2
MACC1
RICTOR
SOX4-201
Exos
Proteins
Transcripts
Fusion genes

References

1. Kindler, H.L. & Shulman, K.L. Metastatic colorectal cancer. *Curr Treat Options Oncol* **2**, 459-471 (2001).
2. Board, R.E. & Valle, J.W. Metastatic colorectal cancer: current systemic treatment options. *Drugs* **67**, 1851-1867 (2007).
3. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **71**, 209-249 (2021).
4. Markowitz, S.D. & Bertagnolli, M.M. Molecular basis of colorectal cancer. *New England Journal of Medicine* **361**, 2449-2460 (2009).
5. Markowitz, S.D., Dawson, D.M., Willis, J. & Willson, J.K. Focus on colon cancer. *Cancer cell* **1**, 233-236 (2002).
6. Fearon, E.R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759-767 (1990).
7. Grady, W.M. Genomic instability and colon cancer. *Cancer and Metastasis Reviews* **23**, 11-27 (2004).
8. Lengauer, C., Kinzler, K.W. & Vogelstein, B. Genetic instability in colorectal cancers. *Nature* **386**, 623-627 (1997).
9. Pino, M.S. & Chung, D.C. The chromosomal instability pathway in colon cancer. *Gastroenterology* **138**, 2059-2072 (2010).
10. Ganem, N.J., Godinho, S.A. & Pellman, D. A mechanism linking extra centrosomes to chromosomal instability. *Nature* **460**, 278-282 (2009).

11. Lengauer, C., Kinzler, K.W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396**, 643-649 (1998).

12. Okugawa, Y., Grady, W.M. & Goel, A. Epigenetic alterations in colorectal cancer: emerging biomarkers. *Gastroenterology* **149**, 1204-1225. e1212 (2015).

13. Fodde, R., Smits, R. & Clevers, H. APC, signal transduction and genetic instability in colorectal cancer. *Nature Reviews Cancer* **1**, 55-67 (2001).

14. Strate, L.L. & Syngal, S. Hereditary colorectal cancer syndromes. *Cancer Causes Control* **16**, 201-213 (2005).

15. Haraldsdottir, S. *et al.* Comprehensive population-wide analysis of Lynch syndrome in Iceland reveals founder mutations in MSH6 and PMS2. *Nat Commun* **8**, 14755 (2017).

16. Boland, C.R. & Goel, A. Microsatellite instability in colorectal cancer. *Gastroenterology* **138**, 2073-2087 e2073 (2010).

17. Leach, F.S. *et al.* Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell* **75**, 1215-1225 (1993).

18. Papadopoulos, N. *et al.* Mutation of a mutL homolog in hereditary colon cancer. *Science* **263**, 1625-1629 (1994).

19. Kinzler, K.W. *et al.* Identification of FAP locus genes from chromosome 5q21. *Science* **253**, 661-665 (1991).

20. Nishisho, I. *et al.* Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. *Science* **253**, 665-669 (1991).

21. Jia, Y. & Guo, M. Epigenetic changes in colorectal cancer. *Chinese journal of cancer* **32**, 21 (2013).

22. Hermsen, M. *et al.* Colorectal adenoma to carcinoma progression follows multiple pathways of chromosomal instability. *Gastroenterology* **123**, 1109-1119 (2002).

23. Aoki, K. & Taketo, M.M. Adenomatous polyposis coli (APC): a multi-functional tumor suppressor gene. *J Cell Sci* **120**, 3327-3335 (2007).

24. Lee, S.K., Hwang, J.H. & Choi, K.Y. Interaction of the Wnt/beta-catenin and RAS-ERK pathways involving co-stabilization of both beta-catenin and RAS plays important roles in the colorectal tumorigenesis. *Adv Biol Regul* **68**, 46-54 (2018).

25. Boland, C.R. & Goel, A. Microsatellite instability in colorectal cancer. *Gastroenterology* **138**, 2073-2087. e2073 (2010).

26. Yamamoto, H., Sawai, H. & Perucho, M. Frameshift somatic mutations in gastrointestinal cancer of the microsatellite mutator phenotype. *Cancer Res* **57**, 4420-4426 (1997).

27. Grady, W.M. & Markowitz, S. Genomic instability and colorectal cancer. *Curr Opin Gastroenterol* **16**, 62-67 (2000).

28. Kane, M.F. *et al.* Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. *Cancer Res* **57**, 808-811 (1997).

29. Samowitz, W.S. *et al.* Evaluation of a large, population-based sample supports a CpG island methylator phenotype in colon cancer. *Gastroenterology* **129**, 837-845 (2005).

30. Rajagopalan, H. *et al.* Tumorigenesis: RAF/RAS oncogenes and mismatch-repair status. *Nature* **418**, 934-934 (2002).

31. Parsons, D.W. *et al.* Colorectal cancer: mutations in a signalling pathway. *Nature* **436**, 792-792 (2005).

32. Jabbari, K. & Bernardi, G. Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. *Gene* **333**, 143-149 (2004).

33. Lao, V.V. & Grady, W.M. Epigenetics and colorectal cancer. *Nat Rev Gastroenterol Hepatol* **8**, 686-700 (2011).

34. Toyota, M. *et al.* CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A* **96**, 8681-8686 (1999).

35. Ogino, S. *et al.* CpG island methylator phenotype (CIMP) of colorectal cancer is best characterised by quantitative DNA methylation analysis and prospective cohort studies. *Gut* **55**, 1000-1006 (2006).

36. Weisenberger, D.J. *et al.* CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet* **38**, 787-793 (2006).

37. Cheng, Y.W. *et al.* CpG island methylator phenotype associates with low-degree chromosomal abnormalities in colorectal cancer. *Clin Cancer Res* **14**, 6005-6013 (2008).

38. Issa, J.P. Colon cancer: it's CIN or CIMP. *Clin Cancer Res* **14**, 5939-5940 (2008).

39. Whiteside, T.L. The tumor microenvironment and its role in promoting tumor growth. *Oncogene* **27**, 5904-5912 (2008).

40. Fodde, R., Smits, R. & Clevers, H. APC, signal transduction and genetic instability in colorectal cancer. *Nat Rev Cancer* **1**, 55-67 (2001).

41. Purnak, T., Ozaslan, E. & Efe, C. Molecular basis of colorectal cancer. *N Engl J Med* **362**, 1246; author reply 1246-1247 (2010).

42. Whiteside, T. The tumor microenvironment and its role in promoting tumor growth. *Oncogene* **27**, 5904-5912 (2008).

43. Mbeunkui, F. & Johann Jr, D.J. Cancer and the tumor microenvironment: a review of an essential relationship. *Cancer chemotherapy and pharmacology* **63**, 571-582 (2009).

44. Dvorak, H.F. Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing. *N Engl J Med* **315**, 1650-1659 (1986).

45. Negrini, S., Gorgoulis, V.G. & Halazonetis, T.D. Genomic instability—an evolving hallmark of cancer. *Nature reviews Molecular cell biology* **11**, 220-228 (2010).

46. Joyce, J.A. & Pollard, J.W. Microenvironmental regulation of metastasis. *Nature Reviews Cancer* **9**, 239-252 (2009).

47. Albini, A. & Sporn, M.B. The tumour microenvironment as a target for chemoprevention. *Nature Reviews Cancer* **7**, 139-147 (2007).

48. Krtolica, A., Parrinello, S., Lockett, S., Desprez, P.-Y. & Campisi, J. Senescent fibroblasts promote epithelial cell growth and tumorigenesis: a link between cancer and aging. *Proceedings of the National Academy of Sciences* **98**, 12072-12077 (2001).

49. Gallucci, S. & Matzinger, P. Danger signals: SOS to the immune system. *Current opinion in immunology* **13**, 114-119 (2001).

50. Mantovani, A., Sozzani, S., Locati, M., Allavena, P. & Sica, A. Macrophage polarization: tumor-associated macrophages as a paradigm for polarized M2 mononuclear phagocytes. *Trends in immunology* **23**, 549-555 (2002).

51. Jin, M.Z. & Jin, W.L. The updated landscape of tumor microenvironment and drug repurposing. *Signal Transduct Target Ther* **5**, 166 (2020).

52. Arroyo, J.D. *et al.* Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proc Natl Acad Sci U S A* **108**, 5003-5008 (2011).

53. Gu, D., Ao, X., Yang, Y., Chen, Z. & Xu, X. Soluble immune checkpoints in cancer: production, function and biological significance. *J Immunother Cancer* **6**, 132 (2018).

54. Mittelbrunn, M. & Sanchez-Madrid, F. Intercellular communication: diverse structures for exchange of genetic information. *Nat Rev Mol Cell Biol* **13**, 328-335 (2012).

55. Haglund, K., Nezis, I.P. & Stenmark, H. Structure and functions of stable intercellular bridges formed by incomplete cytokinesis during development. *Commun Integr Biol* **4**, 1-9 (2011).

56. Valadi, H. *et al.* Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat Cell Biol* **9**, 654-659 (2007).

57. Chen, M. *et al.* Transcriptome and long noncoding RNA sequencing of three extracellular vesicle subtypes released from the human colon cancer LIM1863 cell line. *Sci Rep* **6**, 38397 (2016).

58. Ji, H. *et al.* Deep sequencing of RNA from three different extracellular vesicle (EV) subtypes released from the human LIM1863 colon cancer cell line uncovers distinct miRNA-enrichment signatures. *PLoS One* **9**, e110314 (2014).

59. Ji, H. *et al.* Proteome profiling of exosomes derived from human primary and metastatic colorectal cancer cells reveal differential expression of key metastatic factors and signal transduction components. *Proteomics* **13**, 1672-1686 (2013).

60. Santi, A. *et al.* Cancer associated fibroblasts transfer lipids and proteins to cancer cells through cargo vesicles supporting tumor growth. *Biochim Biophys Acta* **1853**, 3211-3223 (2015).

61. Raposo, G. & Stoorvogel, W. Extracellular vesicles: exosomes, microvesicles, and friends. *The Journal of cell biology* **200**, 373-383 (2013).

62. Kamerkar, S. *et al.* Exosomes facilitate therapeutic targeting of oncogenic KRAS in pancreatic cancer. *Nature* **546**, 498-503 (2017).

63. Zheng, P. *et al.* Exosomal transfer of tumor-associated macrophage-derived miR-21 confers cisplatin resistance in gastric cancer cells. *J Exp Clin Cancer Res* **36**, 53 (2017).

64. Cocucci, E. & Meldolesi, J. Ectosomes and exosomes: shedding the confusion between extracellular vesicles. *Trends in cell biology* **25**, 364-372 (2015).

65. Shi, T., Gao, G. & Cao, Y. Long Noncoding RNAs as Novel Biomarkers Have a Promising Future in Cancer Diagnostics. *Dis Markers* **2016**, 9085195 (2016).

66. Taylor, D.D. & Gercel-Taylor, C. MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer. *Gynecol Oncol* **110**, 13-21 (2008).

67. Nilsson, J. *et al.* Prostate cancer-derived urine exosomes: a novel approach to biomarkers for prostate cancer. *Br J Cancer* **100**, 1603-1607 (2009).

68. Xu, R., Greening, D.W., Zhu, H.J., Takahashi, N. & Simpson, R.J. Extracellular vesicle isolation and characterization: toward clinical application. *J Clin Invest* **126**, 1152-1162 (2016).

69. Joyce, J.A. & Pollard, J.W. Microenvironmental regulation of metastasis. *Nat Rev Cancer* **9**, 239-252 (2009).

70. Rai, A. *et al.* Secreted midbody remnants are a class of extracellular vesicles molecularly distinct from exosomes and microparticles. *Commun Biol* **4**, 400 (2021).

71. Christ, L., Raiborg, C., Wenzel, E.M., Campsteijn, C. & Stenmark, H. Cellular Functions and Molecular Mechanisms of the ESCRT Membrane-Scission Machinery. *Trends Biochem Sci* **42**, 42-56 (2017).

72. Colombo, M. *et al.* Analysis of ESCRT functions in exosome biogenesis, composition and secretion highlights the heterogeneity of extracellular vesicles. *J Cell Sci* **126**, 5553-5565 (2013).

73. van Niel, G. *et al.* The tetraspanin CD63 regulates ESCRT-independent and -dependent endosomal sorting during melanogenesis. *Dev Cell* **21**, 708-721 (2011).

74. Andreu, Z. & Yanez-Mo, M. Tetraspanins in extracellular vesicle formation and function. *Front Immunol* **5**, 442 (2014).

75. Stuffers, S., Sem Wegner, C., Stenmark, H. & Brech, A. Multivesicular endosome biogenesis in the absence of ESCRTs. *Traffic* **10**, 925-937 (2009).

76. Thery, C., Amigorena, S., Raposo, G. & Clayton, A. Isolation and characterization of exosomes from cell culture supernatants and biological fluids. *Curr Protoc Cell Biol* **Chapter 3**, Unit 3 22 (2006).

77. Tauro, B.J. *et al.* Comparison of ultracentrifugation, density gradient separation, and immunoaffinity capture methods for isolating human

colon cancer cell line LIM1863-derived exosomes. *Methods* **56**, 293-304 (2012).

78. Zhang, H. & Lyden, D. Asymmetric-flow field-flow fractionation technology for exomere and small extracellular vesicle separation and characterization. *Nat Protoc* **14**, 1027-1053 (2019).

79. Xu, R. *et al.* Extracellular vesicles in cancer - implications for future improvements in cancer care. *Nat Rev Clin Oncol* **15**, 617-638 (2018).

80. Kharaziha, P., Ceder, S., Li, Q. & Panaretakis, T. Tumor cell-derived exosomes: a message in a bottle. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer* **1826**, 103-111 (2012).

81. Akers, J.C., Gonda, D., Kim, R., Carter, B.S. & Chen, C.C. Biogenesis of extracellular vesicles (EV): exosomes, microvesicles, retrovirus-like vesicles, and apoptotic bodies. *J Neurooncol* **113**, 1-11 (2013).

82. Vietri, M., Radulovic, M. & Stenmark, H. The many functions of ESCRTs. *Nat Rev Mol Cell Biol* **21**, 25-42 (2020).

83. Andrews, N.W., Almeida, P.E. & Corrotte, M. Damage control: cellular mechanisms of plasma membrane repair. *Trends in cell biology* **24**, 734-742 (2014).

84. Andrews, N.W., Almeida, P.E. & Corrotte, M. Damage control: cellular mechanisms of plasma membrane repair. *Trends Cell Biol* **24**, 734-742 (2014).

85. Morello, M. *et al.* Large oncosomes mediate intercellular transfer of functional microRNA. *Cell Cycle* **12**, 3526-3536 (2013).

86. Bussche, L. *et al.* Microvesicle-mediated Wnt/beta-Catenin Signaling Promotes Interspecies Mammary Stem/Progenitor Cell Growth. *J Biol Chem* **291**, 24390-24405 (2016).

87. Menck, K. *et al.* Induction and transport of Wnt 5a during macrophage-induced malignant invasion is mediated by two types of extracellular vesicles. *Oncotarget* **4**, 2057-2066 (2013).

88. Wang, T. *et al.* Hypoxia-inducible factors and RAB22A mediate formation of microvesicles that stimulate breast cancer invasion and metastasis. *Proc Natl Acad Sci U S A* **111**, E3234-3242 (2014).

89. Bebawy, M. *et al.* Membrane microparticles mediate transfer of P-glycoprotein to drug sensitive cancer cells. *Leukemia* **23**, 1643-1649 (2009).

90. Van Doormaal, F., Kleinjan, A., Di Nisio, M., Büller, H. & Nieuwland, R. Cell-derived microvesicles and cancer. *Neth J Med* **67**, 266-273 (2009).

91. Iero, M. *et al.* Tumour-released exosomes and their implications in cancer immunity. *Cell Death & Differentiation* **15**, 80-88 (2008).

92. Mack, M. *et al.* Transfer of the chemokine receptor CCR5 between cells by membrane-derived microparticles: a mechanism for cellular human immunodeficiency virus 1 infection. *Nature medicine* **6**, 769-775 (2000).

93. del Conde, I., Shrimpton, C.N., Thiagarajan, P. & López, J.A. Tissue-factor–bearing microvesicles arise from lipid rafts and fuse with activated platelets to initiate coagulation. *Blood* **106**, 1604-1611 (2005).

94. Sanderson, M.P. *et al.* Generation of novel, secreted epidermal growth factor receptor (EGFR/ErbB1) isoforms via metalloprotease-dependent ectodomain shedding and exosome secretion. *Journal of cellular biochemistry* **103**, 1783-1797 (2008).

95. Kim, C.W. *et al.* Extracellular membrane vesicles from tumor cells promote angiogenesis via sphingomyelin. *Cancer Research* **62**, 6312-6317 (2002).

96. Peterman, E. *et al.* The post-abscission midbody is an intracellular signaling organelle that regulates cell proliferation. *Nat Commun* **10**, 3181 (2019).

97. Mullins, J.M. & McIntosh, J.R. Isolation and initial characterization of the mammalian midbody. *J Cell Biol* **94**, 654-661 (1982).

98. Schiel, J.A. *et al.* Endocytic membrane fusion and buckling-induced microtubule severing mediate cell abscission. *J Cell Sci* **124**, 1411-1424 (2011).

99. Glotzer, M. The 3Ms of central spindle assembly: microtubules, motors and MAPs. *Nat Rev Mol Cell Biol* **10**, 9-20 (2009).

100. Hu, C.K., Coughlin, M. & Mitchison, T.J. Midbody assembly and its regulation during cytokinesis. *Mol Biol Cell* **23**, 1024-1034 (2012).

101. Carlton, J.G., Jones, H. & Eggert, U.S. Membrane and organelle dynamics during cell division. *Nat Rev Mol Cell Biol* **21**, 151-166 (2020).

102. Mullins, J.M. & Biesele, J.J. Terminal phase of cytokinesis in D-98s cells. *J Cell Biol* **73**, 672-684 (1977).

103. Addi, C. *et al.* The Flemmingsome reveals an ESCRT-to-membrane coupling via ALIX/syntenin/syndecan-4 required for completion of cytokinesis. *Nature Communications* **11**, 1-15 (2020).

104. Skop, A.R., Liu, H., Yates, J., 3rd, Meyer, B.J. & Heald, R. Dissection of the mammalian midbody proteome reveals conserved cytokinesis mechanisms. *Science* **305**, 61-66 (2004).

105. Heijnen, H.F., Schiel, A.E., Fijnheer, R., Geuze, H.J. & Sixma, J.J. Activated platelets release two types of membrane vesicles: microvesicles by surface shedding and exosomes derived from exocytosis of multivesicular bodies and alpha-granules. *Blood* **94**, 3791-3799 (1999).

106. Clark, D.J. *et al.* Redefining the Breast Cancer Exosome Proteome by Tandem Mass Tag Quantitative Proteomics and Multivariate Cluster Analysis. *Anal Chem* **87**, 10462-10469 (2015).

107. Greening, D.W. & Simpson, R.J. Understanding extracellular vesicle diversity - current status. *Expert Rev Proteomics* **15**, 887-910 (2018).

108. Skotland, T., Sandvig, K. & Llorente, A. Lipids in exosomes: Current knowledge and the way forward. *Prog Lipid Res* **66**, 30-41 (2017).

109. Record, M., Silvente-Poirot, S., Poirot, M. & Wakelam, M.J.O. Extracellular vesicles: lipids as key components of their biogenesis and functions. *J Lipid Res* **59**, 1316-1324 (2018).

110. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347-355 (2016).

111. Tauro, B.J. *et al.* Two distinct populations of exosomes are released from LIM1863 colon carcinoma cell-derived organoids. *Mol Cell Proteomics* **12**, 587-598 (2013).

112. Chen, M. *et al.* Distinct shed microvesicle and exosome microRNA signatures reveal diagnostic markers for colorectal cancer. *PLoS One* **14**, e0210003 (2019).

113. Liang, H. *et al.* MicroRNA-223 delivered by platelet-derived microvesicles promotes lung cancer cell invasion via targeting tumor suppressor EPB41L3. *Mol Cancer* **14**, 58 (2015).

114. Pitt, J.M., Kroemer, G. & Zitvogel, L. Extracellular vesicles: masters of intercellular communication and potential clinical interventions. *J Clin Invest* **126**, 1139-1143 (2016).

115. Wiklander, O.P.B., Brennan, M.A., Lotvall, J., Breakefield, X.O. & El Andaloussi, S. Advances in therapeutic applications of extracellular vesicles. *Sci Transl Med* **11** (2019).

116. Goricar, K., Dolzan, V. & Lenassi, M. Extracellular Vesicles: A Novel Tool Facilitating Personalized Medicine and Pharmacogenomics in Oncology. *Front Pharmacol* **12**, 671298 (2021).

117. Mathivanan, S. & Simpson, R.J. ExoCarta: A compendium of exosomal proteins and RNA. *Proteomics* **9**, 4997-5000 (2009).

118. Simpson, R.J., Kalra, H. & Mathivanan, S. ExoCarta as a resource for exosomal research. *J Extracell Vesicles* **1** (2012).

119. Kalra, H. *et al.* Vesiclepedia: a compendium for extracellular vesicles with continuous community annotation. *PLoS Biol* **10**, e1001450 (2012).

120. Kim, D.K. *et al.* EVpedia: a community web portal for extracellular vesicles research. *Bioinformatics* **31**, 933-939 (2015).

121. Liu, T. *et al.* EVmiRNA: a database of miRNA profiling in extracellular vesicles. *Nucleic Acids Res* **47**, D89-D93 (2019).

122. Zhang, Y., Liu, Y., Liu, H. & Tang, W.H. Exosomes: biogenesis, biologic function and clinical potential. *Cell Biosci* **9**, 19 (2019).

123. Bobrie, A., Colombo, M., Raposo, G. & Thery, C. Exosome secretion: molecular mechanisms and roles in immune responses. *Traffic* **12**, 1659-1668 (2011).

124. Xu, R. *et al.* Surfaceome of Exosomes Secreted from the Colorectal Cancer Cell Line SW480: Peripheral and Integral Membrane Proteins Analyzed by Proteolysis and TX114. *Proteomics* **19**, e1700453 (2019).

125. Cvjetkovic, A. *et al.* Detailed Analysis of Protein Topology of Extracellular Vesicles-Evidence of Unconventional Membrane Protein Orientation. *Sci Rep* **6**, 36338 (2016).

126. Wu, D. *et al.* Profiling surface proteins on individual exosomes using a proximity barcoding assay. *Nat Commun* **10**, 3854 (2019).

127. Lane, R.E., Korbie, D., Hill, M.M. & Trau, M. Extracellular vesicles as circulating cancer biomarkers: opportunities and challenges. *Clin Transl Med* **7**, 14 (2018).

128. Hewitt, R.E. *et al.* Validation of a model of colon cancer progression. *J Pathol* **192**, 446-454 (2000).

129. Leibovitz, A. *et al.* Classification of human colorectal adenocarcinoma cell lines. *Cancer Res* **36**, 4562-4569 (1976).

130. Ahmed, D. *et al.* Epigenetic and genetic features of 24 colon cancer cell lines. *Oncogenesis* **2**, e71 (2013).

131. Ghosh, D. *et al.* Identification of key players for colorectal cancer metastasis by iTRAQ quantitative proteomics profiling of isogenic SW480 and SW620 cell lines. *J Proteome Res* **10**, 4373-4387 (2011).

132. El-Athman, R., Fuhr, L. & Relogio, A. A Systems-Level Analysis Reveals Circadian Regulation of Splicing in Colorectal Cancer. *EBioMedicine* **33**, 68-81 (2018).

133. Choi, D.S. *et al.* Proteomic analysis of microvesicles derived from human colorectal cancer cells. *J Proteome Res* **6**, 4646-4655 (2007).

134. Choi, D.S. *et al.* Quantitative proteomics of extracellular vesicles derived from human primary and metastatic colorectal cancer cells. *J Extracell Vesicles* **1** (2012).

135. Nakurte, I. *et al.* Colorectal Cancer Cell Line SW480 and SW620 Released Extravascular Vesicles: Focus on Hypoxia-induced Surface Proteome Changes. *Anticancer Res* **38**, 6133-6138 (2018).

136. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **68**, 394-424 (2018).

137. Haggar, F.A. & Boushey, R.P. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clin Colon Rectal Surg* **22**, 191-197 (2009).

138. Breivik, J. & Gaudernack, G. Carcinogenesis and natural selection: a new perspective to the genetics and epigenetics of colorectal cancer. *Adv Cancer Res* **76**, 187-212 (1999).

139. Tomlinson, I. & Bodmer, W. Selection, the mutation rate and cancer: ensuring that the tail does not wag the dog. *Nat Med* **5**, 11-12 (1999).

140. Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **171**, 1029-1041 e1021 (2017).

141. Bignell, G.R. *et al.* Signatures of mutation and selection in the cancer genome. *Nature* **463**, 893-898 (2010).

142. Rivlin, N., Brosh, R., Oren, M. & Rotter, V. Mutations in the p53 Tumor Suppressor Gene: Important Milestones at the Various Steps of Tumorigenesis. *Genes Cancer* **2**, 466-474 (2011).

143. Hobbs, G.A., Der, C.J. & Rossman, K.L. RAS isoforms and mutations in cancer at a glance. *J Cell Sci* **129**, 1287-1292 (2016).

144. Kwong, L.N. & Dove, W.F. APC and its modifiers in colon cancer. *Adv Exp Med Biol* **656**, 85-106 (2009).

145. Jass, J.R. Familial colorectal cancer: pathology and molecular characteristics. *Lancet Oncol* **1**, 220-226 (2000).

146. Jass, J.R., Stewart, S.M., Stewart, J. & Lane, M.R. Hereditary non-polyposis colorectal cancer--morphologies, genes and mutations. *Mutat Res* **310**, 125-133 (1994).

147. Davies, H. *et al.* Mutations of the BRAF gene in human cancer. *Nature* **417**, 949-954 (2002).

148. Fearnhead, N.S., Britton, M.P. & Bodmer, W.F. The ABC of APC. *Hum Mol Genet* **10**, 721-733 (2001).

149. Ma, H. *et al.* Pathology and genetics of hereditary colorectal cancer. *Pathology* **50**, 49-59 (2018).

150. Marmol, I., Sanchez-de-Diego, C., Pradilla Dieste, A., Cerrada, E. & Rodriguez Yoldi, M.J. Colorectal Carcinoma: A General Overview and Future Perspectives in Colorectal Cancer. *Int J Mol Sci* **18** (2017).

151. Watson, I.R., Takahashi, K., Futreal, P.A. & Chin, L. Emerging patterns of somatic mutations in cancer. *Nat Rev Genet* **14**, 703-718 (2013).

152. Rowan, A.J. *et al.* APC mutations in sporadic colorectal tumors: A mutational "hotspot" and interdependence of the "two hits". *Proc Natl Acad Sci U S A* **97**, 3352-3357 (2000).

153. Mueller-Koch, Y. *et al.* Hereditary non-polyposis colorectal cancer: clinical and molecular evidence for a new entity of hereditary colorectal cancer. *Gut* **54**, 1733-1740 (2005).

154. Peltomaki, P. Deficient DNA mismatch repair: a common etiologic factor for colon cancer. *Hum Mol Genet* **10**, 735-740 (2001).

155. Abbas, T., Keaton, M.A. & Dutta, A. Genomic instability in cancer. *Cold Spring Harb Perspect Biol* **5**, a012914 (2013).

156. Hawkes, N. Cancer survival data emphasise importance of early diagnosis. *BMJ* **364**, l408 (2019).

157. Moghimi-Dehkordi, B. & Safaee, A. An overview of colorectal cancer survival rates and prognosis in Asia. *World J Gastrointest Oncol* **4**, 71-75 (2012).

158. Etzioni, R. *et al.* The case for early detection. *Nat Rev Cancer* **3**, 243-252 (2003).

159. Duggan, M.A., Anderson, W.F., Altekruse, S., Penberthy, L. & Sherman, M.E. The Surveillance, Epidemiology, and End Results (SEER) Program and Pathology: Toward Strengthening the Critical Relationship. *Am J Surg Pathol* **40**, e94-e102 (2016).

160. Maacha, S. *et al.* Extracellular vesicles-mediated intercellular communication: roles in the tumor microenvironment and anti-cancer drug resistance. *Mol Cancer* **18**, 55 (2019).

161. Willms, E., Cabanas, C., Mager, I., Wood, M.J.A. & Vader, P. Extracellular Vesicle Heterogeneity: Subpopulations, Isolation Techniques, and Diverse Functions in Cancer Progression. *Front Immunol* **9**, 738 (2018).

162. Mathivanan, S. *et al.* Proteomics analysis of A33 immunoaffinity-purified exosomes released from the human colon tumor cell line

LIM1215 reveals a tissue-specific protein signature. *Mol Cell Proteomics* **9**, 197-208 (2010).

163. Xu, R., Greening, D.W., Rai, A., Ji, H. & Simpson, R.J. Highly-purified exosomes and shed microvesicles isolated from the human colon cancer cell line LIM1863 by sequential centrifugal ultrafiltration are biochemically and functionally distinct. *Methods* **87**, 11-25 (2015).

164. Suwakulsiri, W. *et al.* Proteomic profiling reveals key cancer progression modulators in shed microvesicles released from isogenic human primary and metastatic colorectal cancer cell lines. *Biochim Biophys Acta Proteins Proteom* **1867**, 140171 (2019).

165. Rai, A. *et al.* Exosomes Derived from Human Primary and Metastatic Colorectal Cancer Cells Contribute to Functional Heterogeneity of Activated Fibroblasts by Reprogramming Their Proteome. *Proteomics* **19**, e1800148 (2019).

166. Rai, A., Greening, D.W., Xu, R., Suwakulsiri, W. & Simpson, R.J. Exosomes Derived from the Human Primary Colorectal Cancer Cell Line SW480 Orchestrate Fibroblast-Led Cancer Invasion. *Proteomics* **20**, e2000016 (2020).

167. Alhomrani, M. *et al.* The Human Amnion Epithelial Cell Secretome Decreases Hepatic Fibrosis in Mice with Chronic Liver Fibrosis. *Front Pharmacol* **8**, 748 (2017).

168. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372 (2008).

169. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).

170. Dennis, G., Jr. *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**, P3 (2003).

171. Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47**, D607-D613 (2019).

172. O'Brien, L.E. *et al.* Rac1 orientates epithelial apical polarity through effects on basolateral laminin assembly. *Nat Cell Biol* **3**, 831-838 (2001).

173. Harburger, D.S. & Calderwood, D.A. Integrin signalling at a glance. *J Cell Sci* **122**, 159-163 (2009).

174. van der Horst, G. *et al.* Targeting of alpha-v integrins reduces malignancy of bladder carcinoma. *PLoS One* **9**, e108464 (2014).

175. Li, Y. *et al.* Pleiotropic regulation of macrophage polarization and tumorigenesis by formyl peptide receptor-2. *Oncogene* **30**, 3887-3899 (2011).

176. Khau, T. *et al.* Annexin-1 signals mitogen-stimulated breast tumor cell proliferation by activation of the formyl peptide receptors (FPRs) 1 and 2. *FASEB J* **25**, 483-496 (2011).

177. Gonzalez-Mariscal, L., Tapia, R. & Chamorro, D. Crosstalk of tight junction components with signaling pathways. *Biochim Biophys Acta* **1778**, 729-756 (2008).

178. Sigismund, S., Avanzato, D. & Lanzetti, L. Emerging functions of the EGFR in cancer. *Mol Oncol* **12**, 3-20 (2018).

179. Louderbough, J.M. & Schroeder, J.A. Understanding the dual nature of CD44 in breast cancer progression. *Mol Cancer Res* **9**, 1573-1586 (2011).

180. Rodriguez-Vita, J., Tetzlaff, F. & Fischer, A. Notch controls endothelial cells. *Oncoscience* **4**, 45-46 (2017).

181. Huang, J. *et al.* TGF-beta1 Promotes Hepatocellular Carcinoma Invasion and Metastasis via ERK Pathway-Mediated FGFR4 Expression. *Cell Physiol Biochem* **45**, 1690-1699 (2018).

182. Sun, Y. *et al.* Cancer-associated fibroblasts secrete FGF-1 to promote ovarian proliferation, migration, and invasion through the activation of FGF-1/FGFR4 signaling. *Tumour Biol* **39**, 1010428317712592 (2017).

183. Stein, U. *et al.* MACC1, a newly identified key regulator of HGF-MET signaling, predicts colon cancer metastasis. *Nat Med* **15**, 59-67 (2009).

184. Jacinto, E. *et al.* Mammalian TOR complex 2 controls the actin cytoskeleton and is rapamycin insensitive. *Nat Cell Biol* **6**, 1122-1128 (2004).

185. Angliker, N. & Ruegg, M.A. In vivo evidence for mTORC2-mediated actin cytoskeleton rearrangement in neurons. *Bioarchitecture* **3**, 113-118 (2013).

186. Sit, S.T. & Manser, E. Rho GTPases and their role in organizing the actin cytoskeleton. *J Cell Sci* **124**, 679-683 (2011).

187. Gong, B. *et al.* MiR-21/RASA1 axis affects malignancy of colon cancer cells via RAS pathways. *World J Gastroenterol* **21**, 1488-1497 (2015).

188. Sun, D. *et al.* C/EBP-beta-activated microRNA-223 promotes tumour growth through targeting RASA1 in human colorectal cancer. *Br J Cancer* **112**, 1491-1500 (2015).

189. Peinado, H. *et al.* Pre-metastatic niches: organ-specific homes for metastases. *Nat Rev Cancer* **17**, 302-317 (2017).

190. Crawford, N.P., Colliver, D.W. & Galandiuk, S. Tumor markers and colorectal cancer: utility in management. *J Surg Oncol* **84**, 239-248 (2003).

191. Zhang, Y. *et al.* Co-expression of TIM-3 and CEACAM1 promotes T cell exhaustion in colorectal cancer patients. *Int Immunopharmacol* **43**, 210-218 (2017).

192. Liang, F., Wang, Y., Shi, L. & Zhang, J. Association of Ezrin expression with the progression and prognosis of gastrointestinal cancer: a meta-analysis. *Oncotarget* **8**, 93186-93195 (2017).

193. Gupta, B.K. *et al.* Increased expression and aberrant localization of mucin 13 in metastatic colon cancer. *J Histochem Cytochem* **60**, 822-831 (2012).

194. Schlienger, S., Campbell, S. & Claing, A. ARF1 regulates the Rho/MLC pathway to control EGF-dependent breast cancer cell invasion. *Mol Biol Cell* **25**, 17-29 (2014).

195. Muralidharan-Chari, V. *et al.* ARF6-regulated shedding of tumor cell-derived plasma membrane microvesicles. *Curr Biol* **19**, 1875-1885 (2009).

196. Gopal, S.K. *et al.* YBX1/YB-1 induces partial EMT and tumourigenicity through secretion of angiogenic factors into the extracellular microenvironment. *Oncotarget* **6**, 13718-13730 (2015).

197. Peinado, H. *et al.* Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. *Nat Med* **18**, 883-891 (2012).

198. Capalbo, L. *et al.* The midbody interactome reveals unexpected roles for PP1 phosphatases in cytokinesis. *Nat Commun* **10**, 4513 (2019).

199. Robinson, M.D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**, R25 (2010).

200. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57**, 289-300 (1995).

201. Yu, G., Wang, L.G., Han, Y. & He, Q.Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284-287 (2012).

202. Heberle, H., Meirelles, G.V., da Silva, F.R., Telles, G.P. & Minghim, R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* **16**, 169 (2015).

203. Luo, W. & Brouwer, C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* **29**, 1830-1831 (2013).

204. Kowal, J. *et al.* Proteomic comparison defines novel markers to characterize heterogeneous populations of extracellular vesicle subtypes. *Proc Natl Acad Sci U S A* **113**, E968-977 (2016).

205. Elad, N., Abramovitch, S., Sabanay, H. & Medalia, O. Microtubule organization in the final stages of cytokinesis as revealed by cryo-electron tomography. *J Cell Sci* **124**, 207-215 (2011).

206. Doyle, L.M. & Wang, M.Z. Overview of Extracellular Vesicles, Their Origin, Composition, Purpose, and Methods for Exosome Isolation and Analysis. *Cells* **8** (2019).

207. Hessvik, N.P. & Llorente, A. Current knowledge on exosome biogenesis and release. *Cell Mol Life Sci* **75**, 193-208 (2018).

208. Nabhan, J.F., Hu, R., Oh, R.S., Cohen, S.N. & Lu, Q. Formation and release of arrestin domain-containing protein 1-mediated microvesicles (ARMMs) at plasma membrane by recruitment of TSG101 protein. *Proc Natl Acad Sci U S A* **109**, 4146-4151 (2012).

209. Li, B., Antonyak, M.A., Zhang, J. & Cerione, R.A. RhoA triggers a specific signaling pathway that generates transforming microvesicles in cancer cells. *Oncogene* **31**, 4740-4749 (2012).

210. Jeppesen, D.K. *et al.* Reassessment of Exosome Composition. *Cell* **177**, 428-445 e418 (2019).

211. Wheeler, J.R., Jain, S., Khong, A. & Parker, R. Isolation of yeast and mammalian stress granule cores. *Methods* **126**, 12-17 (2017).

212. Djafarzadeh, S. & Jakob, S.M. Isolation of Intact Mitochondria from Skeletal Muscle by Differential Centrifugation for High-resolution Respirometry Measurements. *J Vis Exp* (2017).

213. Mannella, C.A. Structure and dynamics of the mitochondrial inner membrane cristae. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1763**, 542-548 (2006).

214. Skop, A.R., Liu, H., Yates, J., Meyer, B.J. & Heald, R. Dissection of the mammalian midbody proteome reveals conserved cytokinesis mechanisms. *Science* **305**, 61-66 (2004).

215. Capalbo, L. *et al.* The midbody interactome reveals unexpected roles for PP1 phosphatases in cytokinesis. *Nature communications* **10**, 1-17 (2019).

216. Huang, Z. *et al.* MiCroKiTS 4.0: a database of midbody, centrosome, kinetochore, telomere and spindle. *Nucleic Acids Res* **43**, D328-334 (2015).

217. Kurasawa, Y., Earnshaw, W.C., Mochizuki, Y., Dohmae, N. & Todokoro, K. Essential roles of KIF4 and its binding partner PRC1 in organized central spindle midzone formation. *EMBO J* **23**, 3237-3248 (2004).

218. Guse, A., Mishima, M. & Glotzer, M. Phosphorylation of ZEN-4/MKLP1 by aurora B regulates completion of cytokinesis. *Curr Biol* **15**, 778-786 (2005).

219. Takahashi, S. *et al.* Downregulation of KIF23 suppresses glioma proliferation. *J Neurooncol* **106**, 519-529 (2012).

220. Zhu, C., Bossy-Wetzel, E. & Jiang, W. Recruitment of MKLP1 to the spindle midzone/midbody by INCENP is essential for midbody formation and completion of cytokinesis in human cells. *Biochem J* **389**, 373-381 (2005).

221. Hough, K.P. *et al.* Exosomal transfer of mitochondria from airway myeloid-derived regulatory cells to T cells. *Redox Biol* **18**, 54-64 (2018).

222. Statello, L. *et al.* Identification of RNA-binding proteins in exosomes capable of interacting with different types of RNA: RBP-facilitated transport of RNAs into exosomes. *PLoS One* **13**, e0195969 (2018).

223. Castellanos-Rizaldos, E. *et al.* Exosome-based detection of activating and resistance EGFR mutations from plasma of non-small cell lung cancer patients. *Oncotarget* **10**, 2911-2920 (2019).

224. Zhang, H. *et al.* Exosome-delivered EGFR regulates liver microenvironment to promote gastric cancer liver metastasis. *Nat Commun* **8**, 15016 (2017).

225. Ceresa, B.P. & Schmid, S.L. Regulation of signal transduction by endocytosis. *Curr Opin Cell Biol* **12**, 204-210 (2000).

226. Keller, S., Sanderson, M.P., Stoeck, A. & Altevogt, P. Exosomes: from biogenesis and secretion to biological function. *Immunol Lett* **107**, 102-108 (2006).

227. Nalawansha, D.A., Zhang, Y., Herath, K. & Pflum, M.K.H. HDAC1 Substrate Profiling Using Proteomics-Based Substrate Trapping. *ACS Chem Biol* **13**, 3315-3324 (2018).

228. Li, S., Shi, B., Liu, X. & An, H.X. Acetylation and Deacetylation of DNA Repair Proteins in Cancers. *Front Oncol* **10**, 573502 (2020).

229. Wang, S., Darini, C., Desaubry, L. & Koromilas, A.E. STAT1 Promotes KRAS Colon Tumor Growth and Susceptibility to Pharmacological Inhibition of Translation Initiation Factor eIF4A. *Mol Cancer Ther* **15**, 3055-3063 (2016).

230. Antonyak, M.A. *et al.* Cancer cell-derived microvesicles induce transformation by transferring tissue transglutaminase and fibronectin to recipient cells. *Proc Natl Acad Sci U S A* **108**, 4852-4857 (2011).

231. Filippou, P.S., Karagiannis, G.S. & Constantinidou, A. Midkine (MDK) growth factor: a key player in cancer progression and a promising therapeutic target. *Oncogene* **39**, 2040-2054 (2020).

232. Berthon, A.S., Szarek, E. & Stratakis, C.A. PRKACA: the catalytic subunit of protein kinase A and adrenocortical tumors. *Front Cell Dev Biol* **3**, 26 (2015).

233. Shafiq, A. *et al.* Transglutaminase-2, RNA-binding proteins and mitochondrial proteins selectively traffic to MDCK cell-derived microvesicles following H-Ras-induced epithelial-mesenchymal transition. *Proteomics*, e2000221 (2021).

234. Shinde, A. *et al.* Transglutaminase-2 facilitates extracellular vesicle-mediated establishment of the metastatic niche. *Oncogenesis* **9**, 16 (2020).

235. Hu, W. *et al.* Comprehensive landscape of extracellular vesicle-derived RNAs in cancer initiation, progression, metastasis and cancer immunology. *Mol Cancer* **19**, 102 (2020).

236. Dragomir, M., Chen, B. & Calin, G.A. Exosomal lncRNAs as new players in cell-to-cell communication. *Transl Cancer Res* **7**, S243-S252 (2018).

237. Pang, B. *et al.* Extracellular vesicles: the next generation of biomarkers for liquid biopsy-based prostate cancer diagnosis. *Theranostics* **10**, 2309-2326 (2020).

238. Ge, Q. *et al.* miRNA in plasma exosome is stable under different storage conditions. *Molecules* **19**, 1568-1575 (2014).

239. Kim, D., Langmead, B. & Salzberg, S.L. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**, 357-360 (2015).

240. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290-295 (2015).

241. Liao, Y., Smyth, G.K. & Shi, W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res* **47**, e47 (2019).

242. Durinck, S. *et al.* BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439-3440 (2005).

243. Hubbard, T. *et al.* The Ensembl genome database project. *Nucleic Acids Res* **30**, 38-41 (2002).

244. Iyer, M.K., Chinnaiyan, A.M. & Maher, C.A. ChimeraScan: a tool for identifying chimeric transcription in sequencing data. *Bioinformatics* **27**, 2903-2904 (2011).

245. Lee, M. *et al.* ChimerDB 3.0: an enhanced database for fusion genes from cancer transcriptome and literature data mining. *Nucleic Acids Res* **45**, D784-D789 (2017).

246. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

247. Reimand, J. *et al.* g:Profiler-a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res* **44**, W83-89 (2016).

248. Li, J.H., Liu, S., Zhou, H., Qu, L.H. & Yang, J.H. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res* **42**, D92-97 (2014).

249. Sahraeian, S.M.E. *et al.* Gaining comprehensive biological insight into the transcriptome by performing a broad-spectrum RNA-seq analysis. *Nat Commun* **8**, 59 (2017).

250. Ji, Q. *et al.* MALAT1 regulates the transcriptional and translational levels of proto-oncogene RUNX2 in colorectal cancer metastasis. *Cell Death Dis* **10**, 378 (2019).

251. Oliveira-Mateos, C. *et al.* The transcribed pseudogene RPSAP52 enhances the oncofetal HMGA2-IGF2BP2-RAS axis through LIN28B-dependent and independent let-7 inhibition. *Nat Commun* **10**, 3979 (2019).

252. Lin, Y., Schmidt, B.F., Bruchez, M.P. & McManus, C.J. Structural analyses of NEAT1 lncRNAs suggest long-range RNA interactions that may contribute to paraspeckle architecture. *Nucleic Acids Res* **46**, 3742-3752 (2018).

253. Dowty, J.G. *et al.* Cancer risks for MLH1 and MSH2 mutation carriers. *Hum Mutat* **34**, 490-497 (2013).

254. Choi, Y. *et al.* Integrative analysis of oncogenic fusion genes and their functional impact in colorectal cancer. *Br J Cancer* **119**, 230-240 (2018).

255. Xavier, C.P.R. *et al.* The Role of Extracellular Vesicles in the Hallmarks of Cancer and Drug Resistance. *Cells* **9** (2020).

256. McCready, J., Sims, J.D., Chan, D. & Jay, D.G. Secretion of extracellular hsp90alpha via exosomes increases cancer cell motility: a role for plasminogen activation. *BMC Cancer* **10**, 294 (2010).

257. Hagiwara, K., Katsuda, T., Gailhouste, L., Kosaka, N. & Ochiya, T. Commitment of Annexin A2 in recruitment of microRNAs into extracellular vesicles. *FEBS Lett* **589**, 4071-4078 (2015).

258. Hoogenraad, C.C. *et al.* Mammalian Golgi-associated Bicaudal-D2 functions in the dynein-dynactin pathway by interacting with these complexes. *EMBO J* **20**, 4041-4054 (2001).

259. Kimura, N., Inoue, M., Okabayashi, S., Ono, F. & Negishi, T. Dynein dysfunction induces endocytic pathology accompanied by an increase in Rab GTPases: a potential mechanism underlying age-dependent endocytic dysfunction. *J Biol Chem* **284**, 31291-31302 (2009).

260. Bar-Ziv, R., Voichek, Y. & Barkai, N. Chromatin dynamics during DNA replication. *Genome Res* **26**, 1245-1256 (2016).

261. Seki, Y. *et al.* Circulating cell-free plasma tumour DNA shows a higher incidence of EGFR mutations in patients with extrathoracic disease progression. *ESMO Open* **3**, e000292 (2018).

262. D'Souza, A.R. & Minczuk, M. Mitochondrial transcription and translation: overview. *Essays Biochem* **62**, 309-320 (2018).

263. Taiana, E. *et al.* LncRNA NEAT1 in Paraspeckles: A Structural Scaffold for Cellular DNA Damage Response Systems? *Noncoding RNA* **6** (2020).

264. Maraia, R.J., Mattijssen, S., Cruz-Gallardo, I. & Conte, M.R. The La and related RNA-binding proteins (LARPs): structures, functions, and evolving perspectives. *Wiley Interdiscip Rev RNA* **8** (2017).

265. Caja, F. & Vannucci, L. TGFbeta: A player on multiple fronts in the tumor microenvironment. *J Immunotoxicol* **12**, 300-307 (2015).

266. Al Amir Dache, Z. *et al.* Blood contains circulating cell-free respiratory competent mitochondria. *FASEB J* **34**, 3616-3630 (2020).

267. Puhm, F. *et al.* Mitochondria Are a Subset of Extracellular Vesicles Released by Activated Monocytes and Induce Type I IFN and TNF Responses in Endothelial Cells. *Circ Res* **125**, 43-52 (2019).

268. Franko, A. *et al.* Efficient isolation of pure and functional mitochondria from mouse tissues using automated tissue disruption and enrichment with anti-TOM22 magnetic beads. *PLoS One* **8**, e82392 (2013).

269. Nakagawa, S., Yamazaki, T. & Hirose, T. Molecular dissection of nuclear paraspeckles: towards understanding the emerging world of the RNP milieu. *Open Biol* **8** (2018).

270. Yokoi, A. *et al.* Mechanisms of nuclear content loading to exosomes. *Sci Adv* **5**, eaax8849 (2019).

271. Kim, R.N., Moon, H.G., Han, W. & Noh, D.Y. Perspective Insight into Future Potential Fusion Gene Transcript Biomarker Candidates in Breast Cancer. *Int J Mol Sci* **19** (2018).

272. Wang, S.M. *et al.* Clinical significance of MLH1/MSH2 for stage II/III sporadic colorectal cancer. *World J Gastrointest Oncol* **11**, 1065-1080 (2019).

273. Agostini, A. *et al.* Identification of novel cyclin gene fusion transcripts in endometrioid ovarian carcinomas. *Int J Cancer* **143**, 1379-1387 (2018).

274. Bergmann-Leitner, E.S. & Abrams, S.I. Differential role of Fas/Fas ligand interactions in cytolysis of primary and metastatic colon carcinoma cell lines by human antigen-specific CD8+ CTL. *J Immunol* **164**, 4941-4954 (2000).

275. Schatoff, E.M., Leach, B.I. & Dow, L.E. Wnt Signaling and Colorectal Cancer. *Curr Colorectal Cancer Rep* **13**, 101-110 (2017).

276. Pabla, B., Bissonnette, M. & Konda, V.J. Colon cancer and the epidermal growth factor receptor: Current treatment paradigms, the importance of diet, and the role of chemoprevention. *World J Clin Oncol* **6**, 133-141 (2015).

277. Reedijk, M. *et al.* Activation of Notch signaling in human colon adenocarcinoma. *Int J Oncol* **33**, 1223-1229 (2008).

278. Rooprai, H.K. *et al.* Comparative gene expression profiling of ADAMs, MMPs, TIMPs, EMMPRIN, EGF-R and VEGFA in low grade meningioma. *Int J Oncol* **49**, 2309-2318 (2016).

279. Zheng, J. *et al.* Sox2 modulates motility and enhances progression of colorectal cancer via the Rho-ROCK signaling pathway. *Oncotarget* **8**, 98635-98645 (2017).

280. Arlt, F. & Stein, U. Colon cancer metastasis: MACC1 and Met as metastatic pacemakers. *Int J Biochem Cell Biol* **41**, 2356-2359 (2009).

281. Kamal, Y., Schmit, S.L., Hoehn, H.J., Amos, C.I. & Frost, H.R. Transcriptomic Differences between Primary Colorectal Adenocarcinomas and Distant Metastases Reveal Metastatic Colorectal Cancer Subtypes. *Cancer Res* **79**, 4227-4241 (2019).

282. Truitt, M.L. & Ruggero, D. New frontiers in translational control of the cancer genome. *Nat Rev Cancer* **16**, 288-304 (2016).

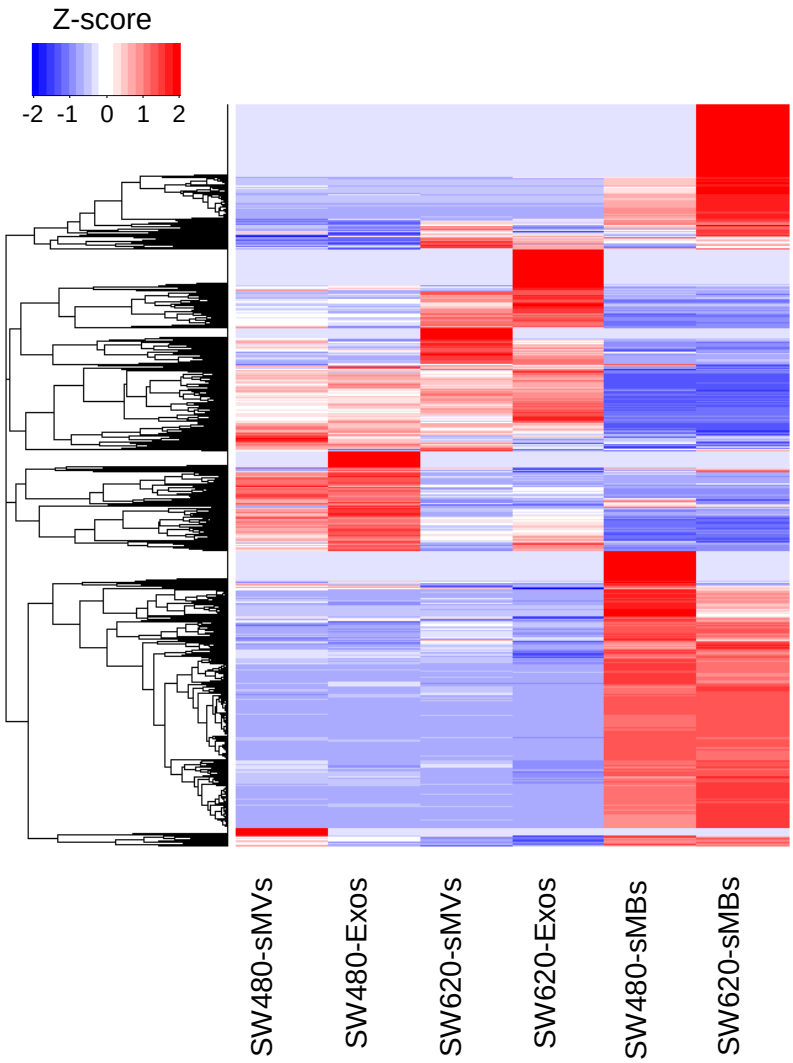
283. Leibovitch, M. & Topisirovic, I. Dysregulation of mRNA translation and energy metabolism in cancer. *Adv Biol Regul* **67**, 30-39 (2018).

284. Sun, Y. & Ma, L. New Insights into Long Non-Coding RNA MALAT1 in Cancer and Metastasis. *Cancers (Basel)* **11** (2019).

285. Wen, S. *et al.* Downregulation of MANCR inhibits cancer cell proliferation in mantle cell lymphoma possibly by interacting with RUNX2. *Acta Biochim Biophys Sin (Shanghai)* **51**, 1142-1147 (2019).

286. Suwakulsiri, W., Chen, M., Greening, D.W., Xu, R. & Simpson, R.J. Analysis of Annotated and Unannotated Long Noncoding RNAs from Exosome Subtypes Using Next-Generation RNA Sequencing. *Methods Mol Biol* **2254**, 195-218 (2021).

Appendices



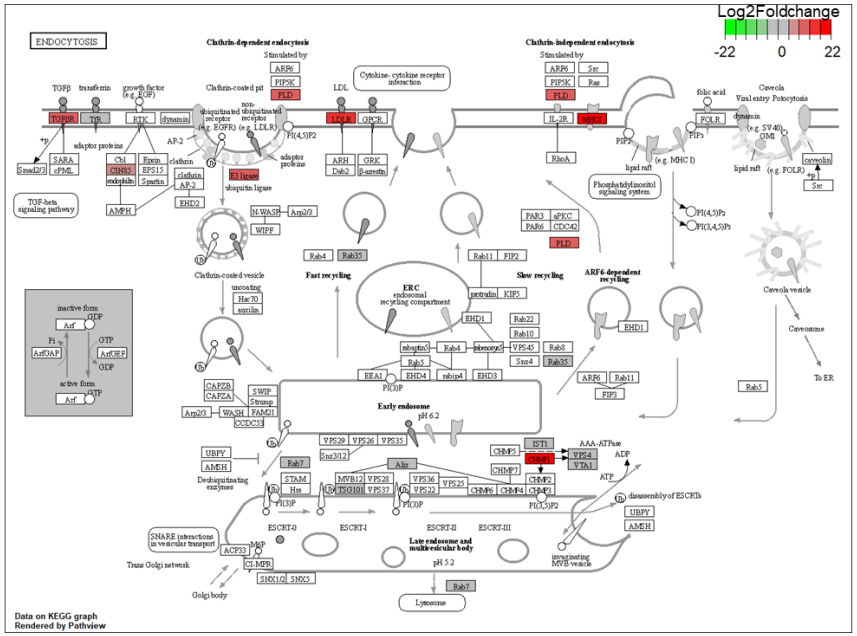
Appendix Figure 3.1 Heatmap of total proteome in Exos, sMVs and sMB-Rs derived from SW480 and SW620 cells

Appendix Figure 3.2 KEGG pathway analysis of highly-enriched proteins found in Exos (95 and 409 proteins, see Fig. 3.3A) and sMB-Rs (604 and 533 proteins, see Fig. 3.3C). (A) Proteins from 'Endocytosis (hsa04144)' highly-enriched in Exos (red) compared to sMVs (green). **(B)** Proteins from 'Endocytosis (hsa04144)' highly-enriched in Exos (red) compared to sMB-Rs (green). **(C)** Proteins from 'SNARE interactions in vesicular transport (hsa04130)' highly-enriched in Exos (red) compared to sMVs (green). **(D)** Proteins from 'SNARE interactions in vesicular transport (hsa04130)' highly-enriched in Exos (red) compared to sMB-Rs (green). **(E)** Proteins from 'Cell adhesion molecules (hsa04514)' highly-enriched in Exos (red) compared to sMVs (green). **(F)** Proteins from 'Cell adhesion molecules (hsa04514)' highly-enriched in Exos (red) compared to sMB-Rs (green). **(G)** Proteins from 'RNA transport (hsa03013)' highly-enriched in sMB-Rs (red) compared to Exos (green). **(H)** Proteins from 'RNA transport (hsa03013)' highly-enriched in sMB-Rs (red) compared to sMVs (green). **(I)** Proteins from 'Spliceosome (hsa03040)' highly-enriched in sMB-Rs (red) compared to Exos (green). **(J)** Proteins from 'Spliceosome (hsa03040)' highly-enriched in sMB-Rs (red) compared to sMVs (green). **(K)** Proteins from 'Protein processing in endoplasmic reticulum (hsa04141)' highly-enriched in sMB-Rs (red) compared to Exos (green). **(L)** Proteins from 'Protein processing in endoplasmic reticulum (hsa04141)' highly-enriched in sMB-Rs (red) compared to sMVs (green). **(M)** Proteins from 'Citrate cycle, TCA cycle (hsa00020)' highly-enriched in sMB-Rs (red) compared to Exos (green). **(N)** Proteins from 'Citrate

cycle, TCA cycle (hsa00020)' highly-enriched in sMB-Rs (red) compared to sMVs (green).

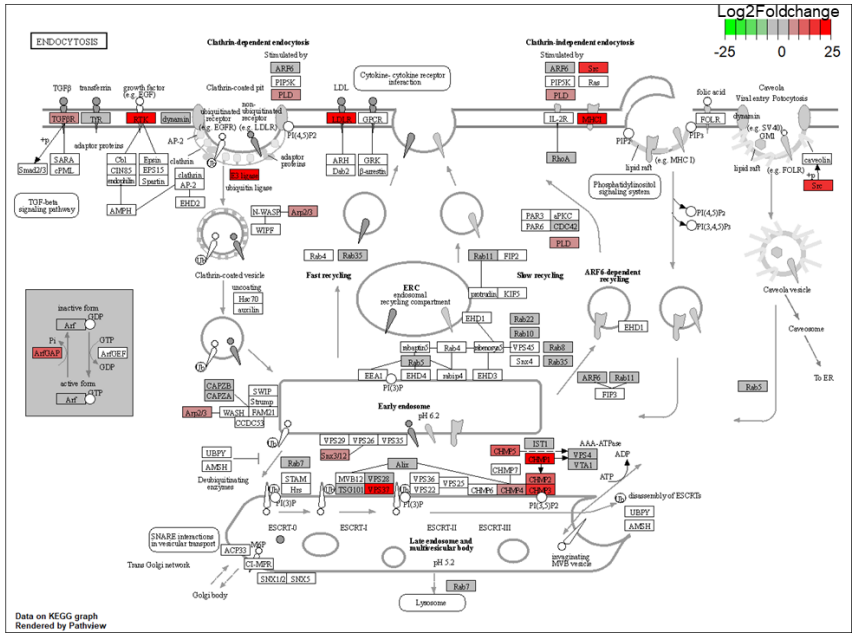
A

Exos vs sMVs



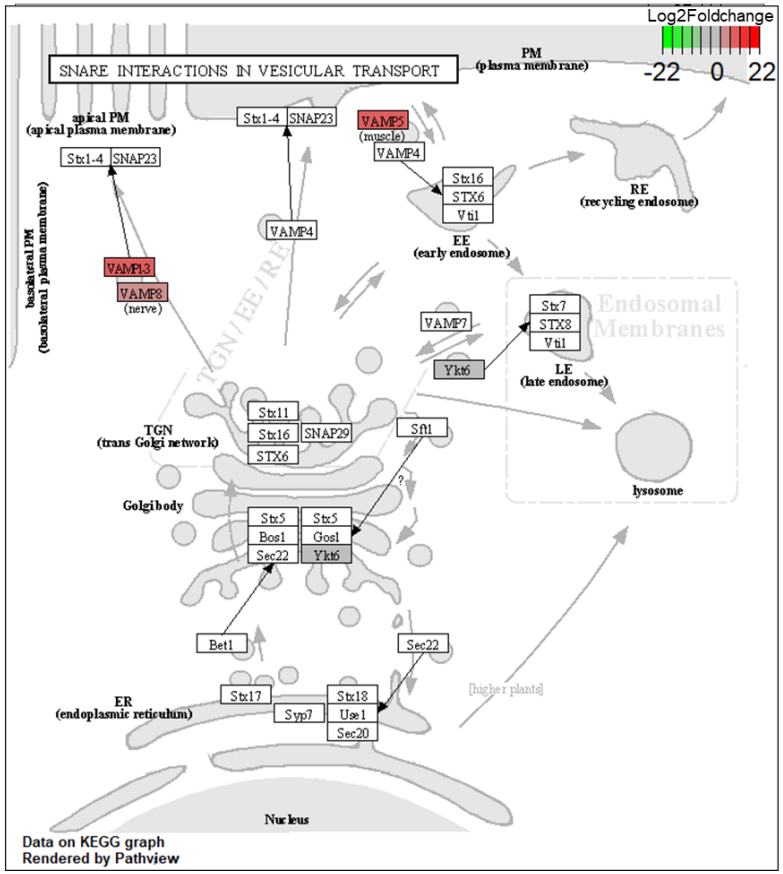
B

Exos vs sMB-Rs



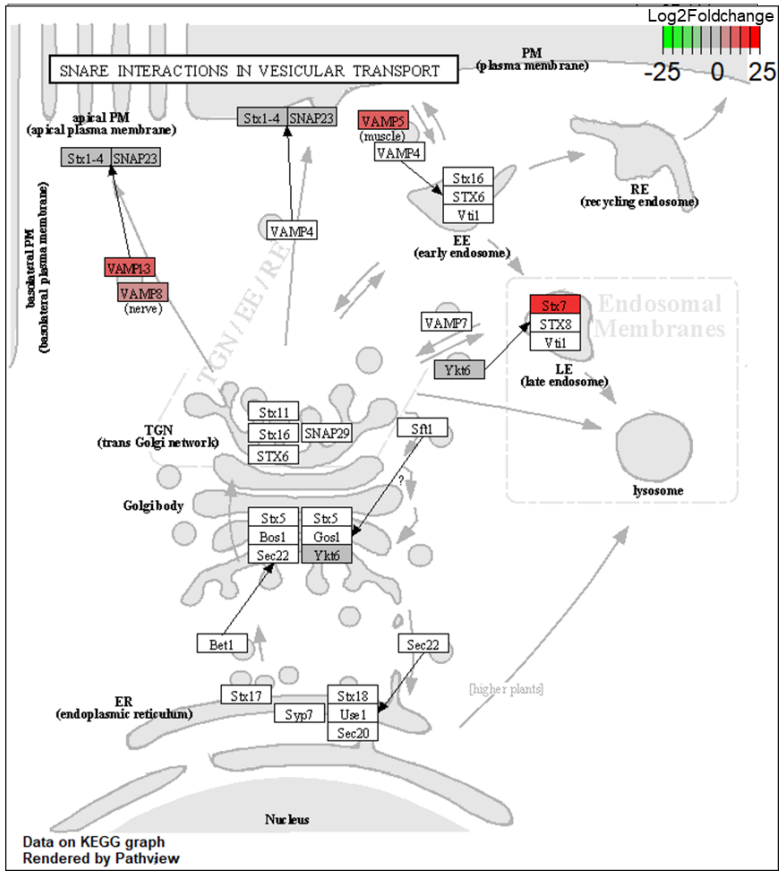
C

Exos vs sMVs



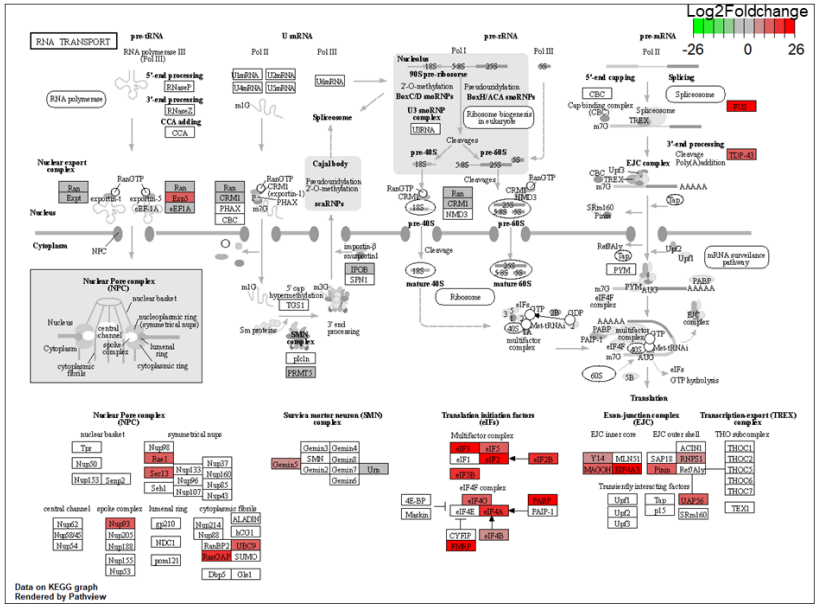
D

Exos vs sMVBRs



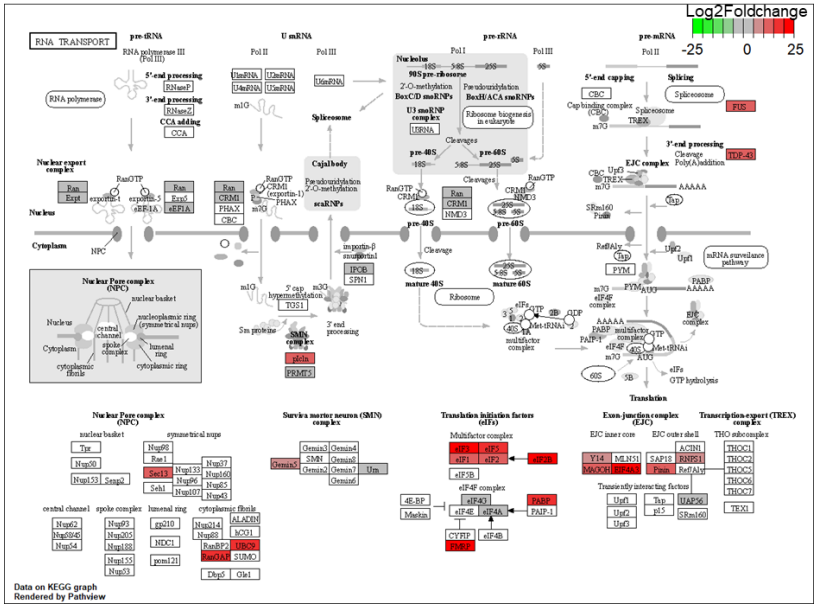
G

sMB-Rs vs Exos



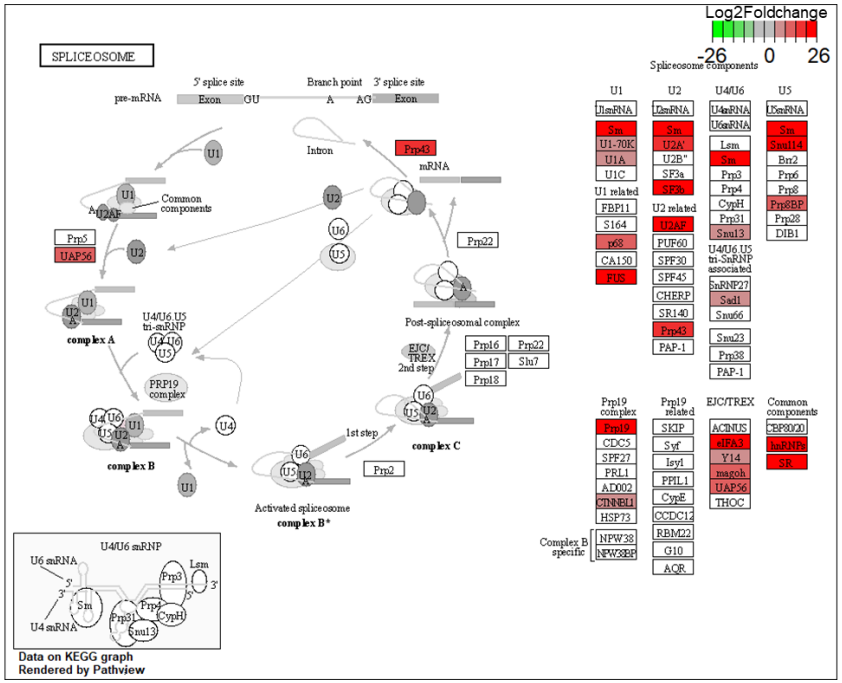
H

sMB-Rs vs sMVs



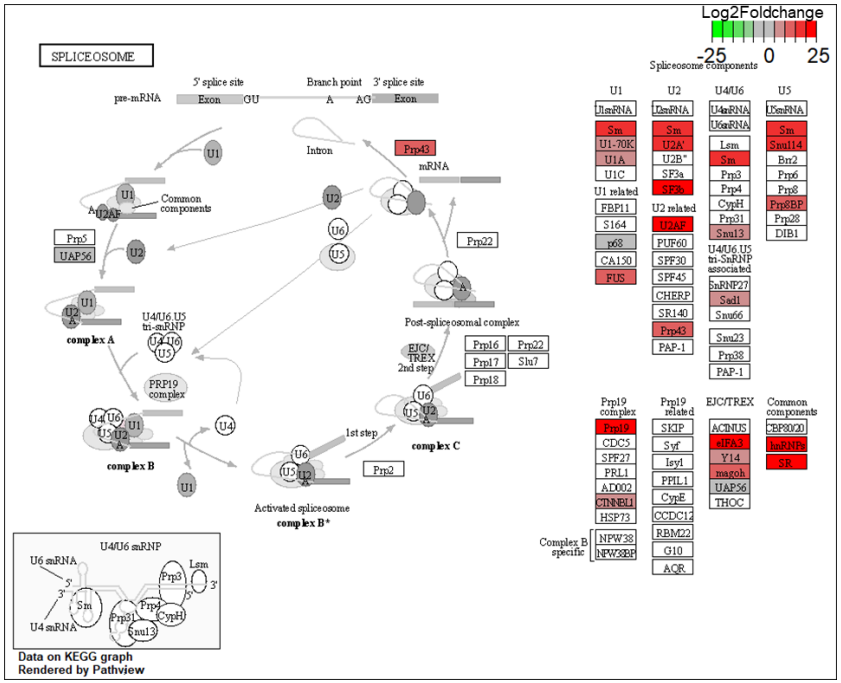
I

sMB-Rs vs Exos



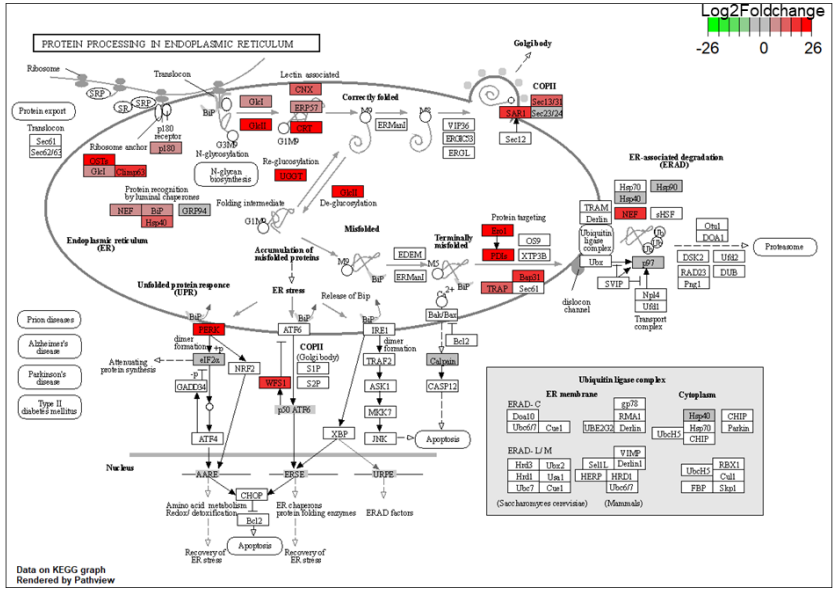
J

sMB-Rs vs sMVs



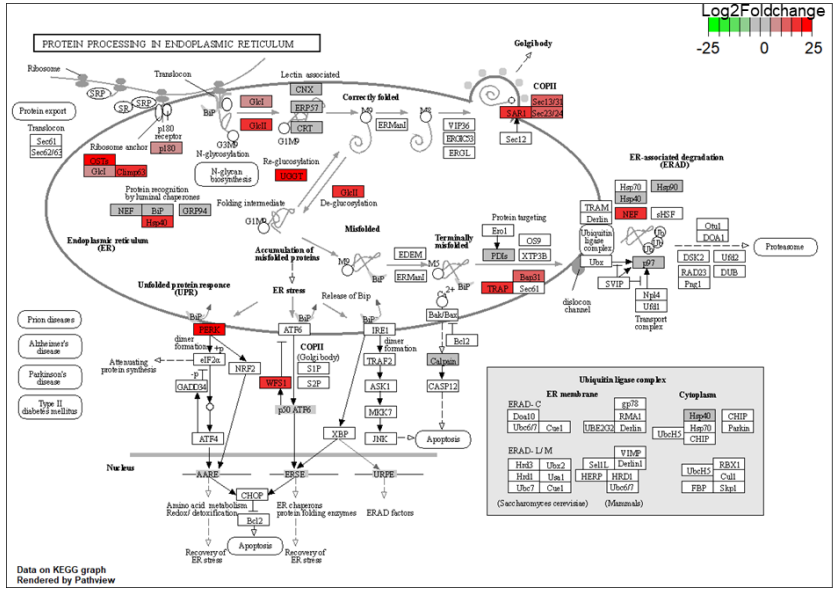
K

sMB-Rs vs Exos



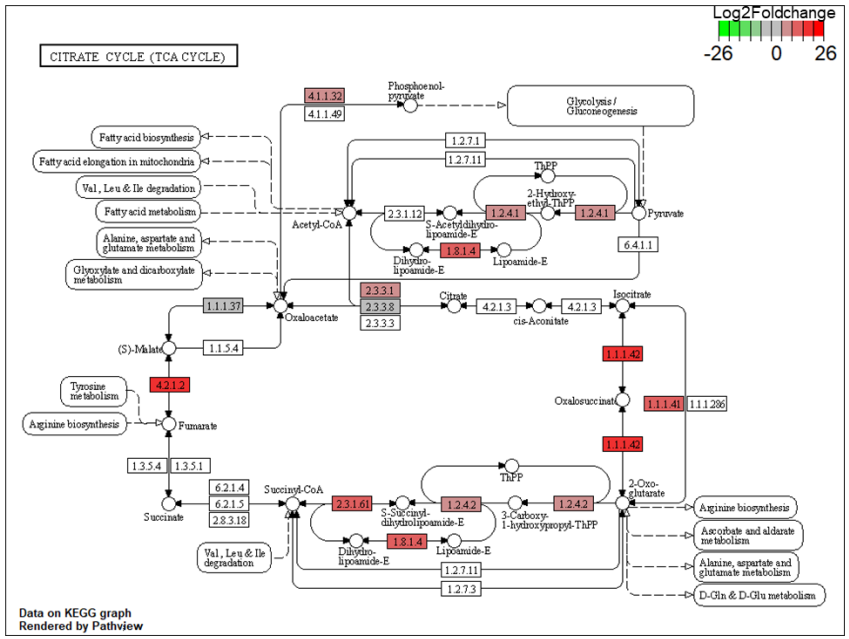
L

sMB-Rs vs sMVs



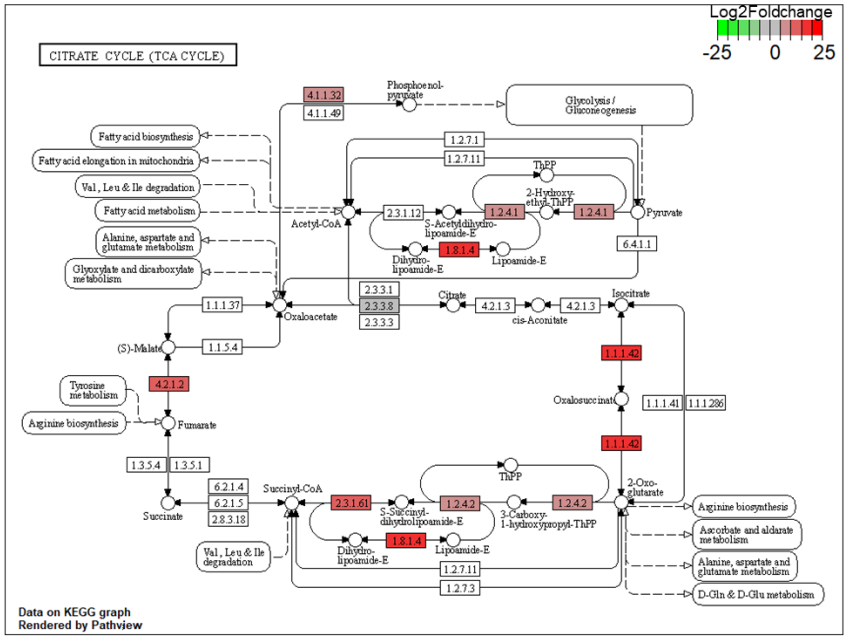
M

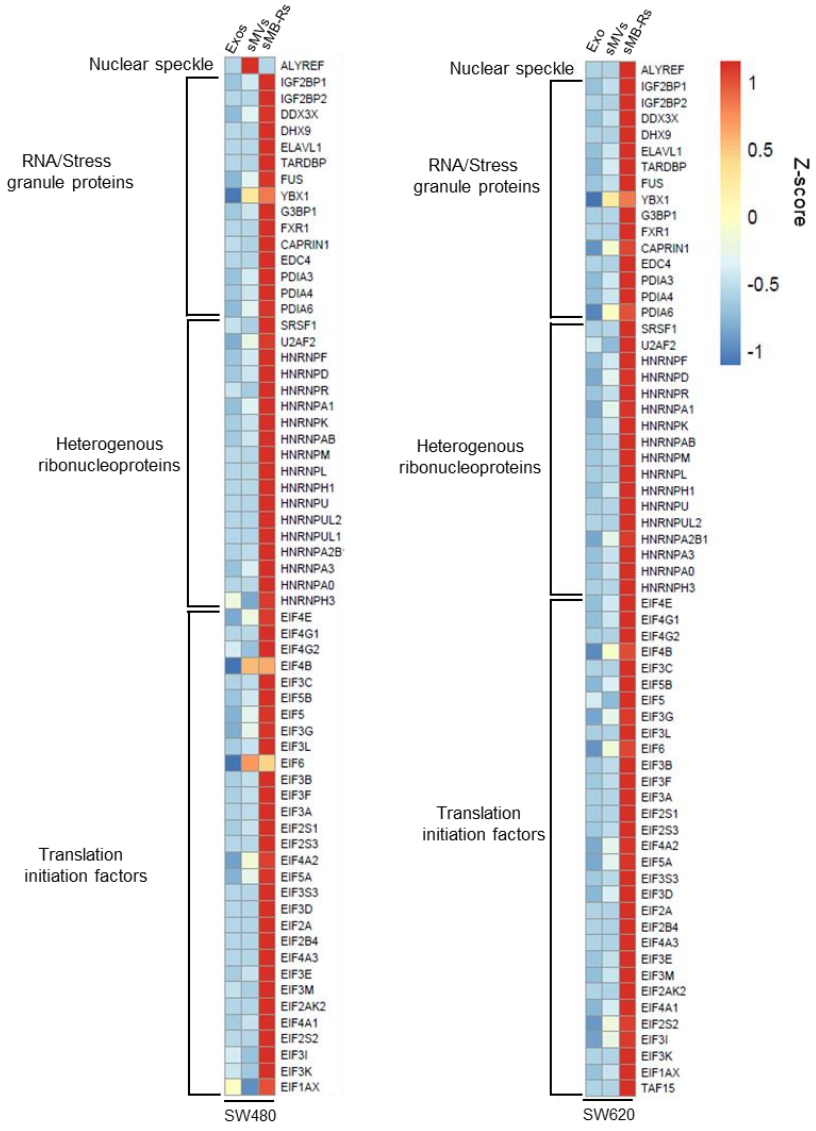
sMB-Rs vs Exos



N

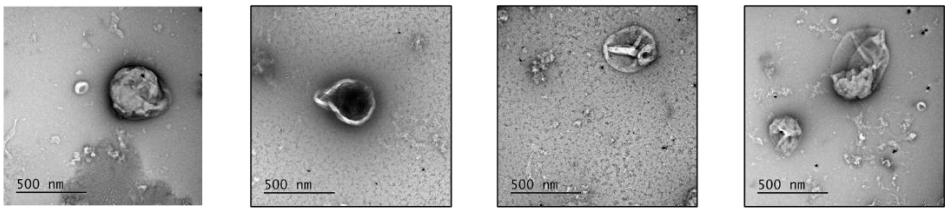
sMB-Rs vs sMVs



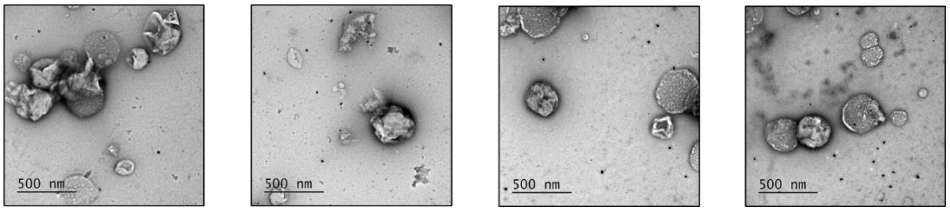


Appendix Figure 3.3 Heat map illustration of RBPs related to nuclear speckles, RNA/stress granule, heterogenous ribonucleoprotein and translation initiation factor in Exos, sMVs and sMB-Rs (scale shown is average normalized LFQ subtracted by mean and divided by standard deviation).

SW480-sMB-Rs



SW620-sMB-Rs



Appendix Figure 3.4 Transmission electron microscopic analysis of SW480-/ SW620-sMB-Rs

Appendix Table 3.1 A list of top 50 protein identifications in Exos, sMVs and sMB-Rs by label-free mass spectrometry

Protein Access (UniProt)	Protein description	Gene name	SW480-sMVs	SW480-sMB-Rs	SW480-Exos	SW620-sMVs	SW620-sMB-Rs	SW620-Exos
P63261	Actin, cytoplasmic 2	ACTG1	166,002,667	95,824,889	142,833,778	229,384,889	148,960,889	214,093,333
P62805	Histone H4	HIST1H4A	5,767,961	196,896,440	9,062,524	2,145,728	16,966,958	2,701,715
P23528	Cofilin-1	CFL1	27,737,952	18,215,863	28,390,763	42,574,096	28,529,317	46,175,100
P07355	Annexin A2	ANXA2	34,900,688	20,446,411	35,390,364	20,333,628	14,143,363	20,622,222
P57053	Histone H2B type F-S	H2BFS	3,047,196	114,148,148	4,667,169	927,741	8,294,841	1,182,899
P04406	Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	20,682,587	22,173,234	19,539,701	21,484,677	23,460,100	18,476,517
P68431	Histone H3.1	HIST1H3A	2,696,250	108,375,000	3,897,377	788,556	8,037,623	1,000,868
P62937	Peptidyl-prolyl cis-trans isomerase A	PPIA	21,257,172	16,707,879	20,368,485	21,006,263	18,313,333	21,481,414
Q16777	Histone H2A type 2-C	HIST2H2AC	4,296,098	75,644,444	4,962,817	1,399,612	7,236,770	1,651,137
O43707	Alpha-actinin-4	ACTN4	20,774,241	16,234,175	14,954,629	17,530,187	7,857,007	15,085,254
P07737	Profilin-1	PFN1	13,839,524	8,151,667	11,499,762	21,401,429	16,778,333	19,806,190
P06733	Alpha-enolase	ENO1	12,075,422	8,487,174	9,055,453	14,618,126	10,129,877	12,636,559
O00299	Chloride intracellular channel protein 1	CLIC1	11,737,068	5,647,856	10,000,138	13,516,874	8,136,791	17,348,686
P68032	Actin, alpha cardiac muscle 1	ACTC1	6,696,640	6,172,060	5,577,100	16,710,433	8,826,525	22,160,831
P62258	14-3-3 protein epsilon	YWHAE	13,076,732	5,973,987	13,134,379	12,456,863	8,490,980	12,952,288
P26038	Moesin	MSN	12,395,898	3,767,995	13,289,832	13,761,352	4,445,812	16,688,388
A6NNI4	Tetraspanin	CD9	15,249,895	4,228,553	25,519,078	3,882,327	2,630,692	8,346,331
P00338	L-lactate dehydrogenase A chain	LDHA	11,751,406	9,026,606	11,482,831	9,638,353	8,376,908	7,754,518
P68104	Elongation factor 1-alpha 1	EEF1A1	6,543,290	13,059,668	6,182,900	7,315,296	13,124,242	5,502,092
P62979	Ubiquitin-40S ribosomal protein S27a	RPS27A	10,395,085	3,252,607	12,921,581	4,939,038	3,266,859	13,191,239
Q15758	Neutral amino acid transporter B	SLC1A5	13,396,057	6,186,137	14,277,203	4,498,275	2,552,742	5,480,345
P00558	Phosphoglycerate kinase 1	PGK1	6,268,026	6,762,670	4,868,585	9,884,412	9,492,886	8,505,755
P68371	Tubulin beta-4B chain	TUBB4B	5,176,929	12,511,086	4,032,959	5,905,843	11,636,704	3,839,476
P11166	Solute carrier family 2, facilitated glucose transporter member 1	SLC2A1	13,977,100	5,386,382	14,467,886	3,341,260	1,500,488	4,347,764
P68363	Tubulin alpha-1B chain	TUBA1B	4,877,162	11,203,917	4,348,854	5,127,864	12,848,189	3,583,740
P14618	Pyruvate kinase PKM	PKM	6,238,104	8,382,109	5,923,854	6,894,162	8,819,900	4,963,842

P63104	14-3-3 protein zeta/delta	YWHAZ	8,235,782	3,809,646	7,623,673	8,209,524	5,197,918	7,999,864
P60174	Triosephosphate isomerase	TPI1	8,235,315	3,938,345	7,716,783	6,723,427	4,712,821	5,633,916
P09211	Glutathione S-transferase P	GSTP1	6,422,381	4,478,222	5,902,540	6,657,143	5,956,349	5,063,667
P60903	Protein S100-A10	S100A10	7,103,952	4,901,134	6,644,674	5,752,612	3,346,667	5,830,000
P04075	Fructose-bisphosphate aldolase A	ALDOA	7,125,366	5,249,725	5,884,249	5,582,418	4,491,667	4,432,509
P62873	Guanine nucleotide-binding protein G	GNB1	4,158,824	2,022,000	4,400,686	7,343,235	4,133,912	7,464,020
P07900	Heat shock protein HSP 90-alpha	HSP90AA1	3,989,891	8,487,705	3,348,862	3,436,475	6,613,570	2,372,723
Q06830	Peroxiredoxin-1	PRDX1	5,459,950	5,235,176	4,503,082	4,644,556	4,926,281	3,167,806
F8WCF6	Actin-related protein 2/3 complex subunit 4	ARPC4-TTLL3	4,264,365	1,799,042	4,800,239	6,123,941	3,866,059	6,547,514
P12814	Alpha-actinin-1	ACTN1	5,204,559	4,127,877	3,893,012	5,780,755	2,593,647	5,028,027
P63000	Ras-related C3 botulinum toxin substrate 1	RAC1	5,259,948	1,704,028	5,775,625	4,569,010	2,618,993	5,910,243
P09455	Retinol-binding protein 1	RBP1	6,959,407	5,953,284	7,455,901	1,942,025	1,694,642	1,805,556
P60660	Myosin light polypeptide 6	MYL6	3,832,583	3,318,896	3,746,932	5,275,254	4,266,887	5,269,647
P09382	Galectin-1	LGALS1	4,229,531	5,707,802	3,141,778	4,344,198	3,827,309	3,645,407
F5GZS6	4F2 cell-surface antigen heavy chain	SLC3A2	6,830,161	3,703,172	6,308,625	3,072,454	1,787,813	3,193,434
P04899	Guanine nucleotide-binding protein G	GNAI2	4,227,700	1,924,620	4,156,714	5,538,779	3,207,831	5,823,568
P11142	Heat shock cognate 71 kDa protein	HSPA8	4,217,441	4,091,589	4,678,277	3,257,688	4,046,182	3,452,116
P26447	Protein S100-A4	S100A4	2,820,396	2,322,145	2,642,145	4,292,937	4,580,429	5,247,789
H0Y7A7	Calmodulin-2	CALM2	5,330,570	4,909,608	4,250,642	2,006,934	1,903,458	2,234,581
P18085	ADP-ribosylation factor 4	ARF4	3,758,870	2,760,444	4,409,185	3,117,611	2,492,074	3,593,426
P54709	Sodium/potassium-transporting ATPase subunit beta-3	ATP1B3	4,309,677	3,236,977	4,230,824	2,818,722	2,001,016	2,902,951
Q9Y696	Chloride intracellular channel protein 4	CLIC4	4,927,997	2,136,113	4,945,455	2,485,349	1,428,248	3,557,918
P07195	L-lactate dehydrogenase B chain	LDHB	3,740,918	2,863,593	3,259,691	3,338,224	3,391,218	2,601,377
P61586	Transforming protein RhoA	RHOA	3,196,235	1,945,786	3,592,435	3,421,157	2,241,537	4,281,554

Appendix Table 3.2 Uniquely identified proteins in Exos from SW480 and SW620 cells by label-free mass spectrometry

Protein Access. (UniProt)	Protein description	Gene name	SW480-sMVs	SW480-sMB-Rs	SW480-Exos	SW620-sMVs	SW620-sMB-Rs	SW620-Exos
Q6IAA8	Regulator complex protein LAMTOR1	LAMTOR1	-	-	51,765	-	-	15,325
Q9H190	Syntenin-2	SDCBP2	-	-	29,811	-	-	137,412
O60869	Endothelial differentiation-related factor 1	EDF1	-	-	24,439	-	-	27,839
Q9NRX5	Serine incorporator 1	SERINC1	-	-	7,458	-	-	18,016
O75317	Ubiquitin carboxyl-terminal hydrolase 12	USP12	-	-	5,846	-	-	14,222
P14209	CD99 antigen	CD99	-	-	3,780	-	-	16,497
Q15223	Nectin-1	NECTIN1	-	-	2,277	-	-	7,040
P49023	Paxillin	PXN	-	-	3,872	-	-	2,864
E9PS74	Solute carrier family 43 member 3	SLC43A3	-	-	3,147	-	-	20,888

Appendix Table 3.3 Uniquely identified proteins in sMVs from SW480 and SW620 cells by label-free mass spectrometry

Protein Access. (UniProt)	Protein description	Gene name	SW480-sMVs	SW480-sMB-Rs	SW480-Exos	SW620-sMVs	SW620-sMB-Rs	SW620-Exos
P23919	Thymidylate kinase	DTYMK	18,102	-	-	15,154	-	-

Appendix Table 3.4 Uniquely identified proteins in sMB-Rs from SW480 and SW620 cells by label-free mass spectrometry

Protein Access. (UniProt)	Protein description	Gene name	SW480-sMVs	SW480-sMB-Rs	SW480-Exos	SW620-sMVs	SW620-sMB-Rs	SW620-Exos
Q71DI3	Histone H3.2	HIST2H3A;	-	1,149,510	-	-	58,295	-
A0A0B4J1Z1	Serine/arginine-rich-splicing factor 7	SRSF7	-	299,849	-	-	245,611	-
Q08945	FACT complex subunit SSRP1	SSRP1	-	81,476	-	-	22,185	-
Q9NS69	Mitochondrial import receptor subunit TOM22 homolog	TOMM22	-	53,531	-	-	52,962	-
Q14978	Nucleolar and coiled-body phosphoprotein 1	NOLC1	-	85,150	-	-	15,812	-
O43809	Cleavage and polyadenylation specificity factor subunit 5	NUDT21	-	41,260	-	-	51,701	-
I3L1P8	Mitochondrial 2-oxoglutarate/malate carrier protein	SLC25A11	-	20,017	-	-	53,020	-
P09661	U2 small nuclear ribonucleoprotein A'	SNRPA1	-	56,284	-	-	12,791	-
Q00577	Transcriptional activator protein Pur-alpha	PURA	-	50,421	-	-	31,812	-
P46060	Ran GTPase-activating protein 1	RANGAP1	-	50,824	-	-	21,213	-
P61009	Signal peptidase complex subunit 3	SPCS3	-	15,445	-	-	17,500	-
B4DJ81	NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial	NDUFS1	-	14,969	-	-	14,089	-
O75569	Interferon-inducible double-stranded RNA-dependent protein kinase activator A	PRKRA	-	45,428	-	-	53,810	-
O43765	Small glutamine-rich tetratricopeptide repeat-containing protein alpha	SGTA	-	32,484	-	-	8,734	-
P08621	U1 small nuclear ribonucleoprotein 70 kDa	SNRNP70	-	12,217	-	-	11,273	-
P24539	ATP synthase F	ATP5F1	-	11,257	-	-	21,721	-
O60832	H/ACA ribonucleoprotein complex subunit 4	DKC1	-	20,807	-	-	5,123	-
P33992	DNA replication licensing factor MCM5	MCM5	-	24,443	-	-	13,108	-
P26583	High mobility group protein B2	HMGB2	-	21,689	-	-	14,704	-
P55735	Protein SEC13 homolog	SEC13	-	10,735	-	-	42,561	-
Q9NZ01	Very-long-chain enoyl-CoA reductase	TECR	-	34,148	-	-	48,025	-

Q9BY44	Eukaryotic translation initiation factor 2A	EIF2A	-	28,077	-	-	19,342	-
G8JLD5	Dynamin-1-like protein	DNM1L	-	32,110	-	-	28,519	-
G3V1C3	Apoptosis inhibitor 5	API5	-	24,996	-	-	16,134	-
Q00059	Transcription factor A, mitochondrial	TFAM	-	9,404	-	-	39,998	-
Q13501	Sequestosome-1	SQSTM 1	-	18,171	-	-	14,288	-
Q96DI7	U5 small nuclear ribonucleoprotein 40 kDa protein	SNRNP 40	-	18,407	-	-	10,785	-
B4DXZ6	Fragile X mental retardation syndrome-related protein 1	FXR1	-	32,618	-	-	15,218	-
P16435	NADPH--cytochrome P450 reductase	POR	-	22,702	-	-	18,011	-
P18754	Regulator of chromosome condensation	RCC1	-	7,994	-	-	8,557	-
Q1KMD3	Heterogeneous nuclear ribonucleoprotein U-like protein 2	HNRNP UL2	-	18,744	-	-	9,982	-
Q9NVJ2	ADP-ribosylation factor-like protein 8B	ARL8B	-	7,470	-	-	8,815	-
Q14258	E3 ubiquitin/ISG15 ligase TRIM25	TRIM25	-	23,562	-	-	27,226	-
Q92688	Acidic leucine-rich nuclear phosphoprotein 32 family member B	ANP32 B	-	21,388	-	-	20,475	-
A0A087W TA5	Translation initiation factor eIF-2B subunit delta	EIF2B4	-	20,829	-	-	12,745	-
P32322	Pyrroline-5-carboxylate reductase 1, mitochondrial	PYCR1	-	18,470	-	-	65,404	-
Q8NBP7	Proprotein convertase subtilisin/kexin type 9	PCSK9	-	23,475	-	-	21,675	-
Q9Y383	Putative RNA-binding protein Luc7-like 2	LUC7L2	-	10,891	-	-	3,501	-
Q9UBT2	SUMO-activating enzyme subunit 2	UBA2	-	21,216	-	-	6,990	-
Q9UJS0	Calcium-binding mitochondrial carrier protein Aralar2	SLC25 A13	-	19,624	-	-	12,016	-
O60568	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3	PLOD3	-	20,569	-	-	21,197	-
Q9P035	Very-long-chain	HACD3	-	13,040	-	-	40,484	-
K7EIG1	Clustered mitochondria protein homolog	CLUH	-	16,203	-	-	15,360	-
Q8WXF1	Paraspeckle component 1	PSPC1	-	22,385	-	-	3,245	-
Q96HS1	Serine/threonine-protein phosphatase PGAM5, mitochondrial	PGAM5	-	18,403	-	-	22,274	-
O00139	Kinesin-like protein KIF2A	KIF2A	-	11,375	-	-	84,425	-
P33993	DNA replication licensing factor MCM7	MCM7	-	10,345	-	-	2,588	-
Q9GZP4	PITH domain-containing protein 1	PITHD1	-	4,698	-	-	7,115	-
B0QYK0	RNA-binding protein EWS	EWSR1	-	13,098	-	-	15,524	-
P78347	General transcription factor II-I	GTF2I	-	10,453	-	-	6,093	-
Q93009	Ubiquitin carboxyl-terminal hydrolase 7	USP7	-	4,336	-	-	7,254	-

Q5SQP8	C-terminal-binding protein 2	CTBP2	-	4,256	-	-	5,674	-
F8W726	Ubiquitin-associated protein 2-like	UBAP2L	-	7,090	-	-	4,249	-
Q07065	Cytoskeleton-associated protein 4	CKAP4	-	10,001	-	-	7,235	-
P36957	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	DLST	-	6,826	-	-	9,612	-
P38919	Eukaryotic initiation factor 4A-III	EIF4A3	-	15,600	-	-	25,195	-
P48735	Isocitrate dehydrogenase [NADP], mitochondrial	IDH2	-	3,602	-	-	61,780	-
P40939	Trifunctional enzyme subunit alpha, mitochondrial	HADHA	-	14,158	-	-	6,525	-
A0A0G2JL47	Large proline-rich protein BAG6	BAG6	-	6,654	-	-	8,935	-
Q14694	Ubiquitin carboxyl-terminal hydrolase 10	USP10	-	7,734	-	-	2,657	-
P46821	Microtubule-associated protein 1B	MAP1B	-	8,265	-	-	7,986	-
P49736	DNA replication licensing factor MCM2	MCM2	-	6,531	-	-	2,540	-
Q6P2E9	Enhancer of mRNA-decapping protein 4	EDC4	-	2,779	-	-	30,513	-
P19525	Interferon-induced, double-stranded RNA-activated protein kinase	EIF2AK2	-	8,783	-	-	13,238	-
Q14980	Nuclear mitotic apparatus protein 1	NUMA1	-	5,500	-	-	2,367	-
C9J2Y9	DNA-directed RNA polymerase subunit beta	POLR2B	-	5,749	-	-	6,420	-
Q8TCJ2	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3B	STT3B	-	4,284	-	-	11,463	-
Q6PKG0	La-related protein 1	LARP1	-	5,240	-	-	3,272	-
P42167	Lamina-associated polypeptide 2, isoforms beta/gamma	TMPO	-	1,801	-	-	5,318	-
Q9NYU2	UDP-glucose:glycoprotein glucosyltransferase 1	UGGT1	-	5,329	-	-	5,183	-
E9PPJ0	Splicing factor 3B subunit 2	SF3B2	-	7,321	-	-	1,700	-
P98160	Basement membrane-specific heparan sulfate proteoglycan core protein	HSPG2	-	4,065	-	-	3,117	-
Q12931	Heat shock protein 75 kDa, mitochondrial	TRAP1	-	16,906	-	-	28,223	-
Q13423	NAD	NNT	-	2,904	-	-	4,009	-
P25205	DNA replication licensing factor MCM3	MCM3	-	2,806	-	-	3,102	-
P48681	Nestin	NES	-	1,000	-	-	547	-
J3QQY2	Calcium load-activated calcium channel	TMCO1	-	91,099	-	-	100,179	-
P00387	NADH-cytochrome b5 reductase 3	CYB5R3	-	26,505	-	-	68,053	-
H0YMF9	WD repeat-containing protein 61	WDR61	-	22,356	-	-	64,369	-
Q5JRI1	Serine/arginine-rich-splicing factor 10	SRSF10	-	35,437	-	-	31,651	-
P62995	Transformer-2 protein homolog beta	TRA2B	-	11,578	-	-	31,037	-

F8VU90	Peptidylprolyl isomerase	FKBP11	-	20,954	-	-	54,630	-
Q9BTV4	Transmembrane protein 43	TMEM43	-	6,598	-	-	16,865	-
P09622	Dihydrolipoyl dehydrogenase, mitochondrial	DLD	-	9,747	-	-	16,663	-
Q9UBS4	DnaJ homolog subfamily B member 11	DNAJB11	-	19,713	-	-	19,809	-
Q9H307	Pinin	PNN	-	11,811	-	-	8,376	-
E9PHA6	DNA mismatch repair protein	MSH2	-	5,524	-	-	6,602	-
O76024	Wolframin	WFS1	-	3,320	-	-	13,877	-
F8W930	Insulin-like growth factor 2 mRNA-binding protein 2	IGF2BP2	-	6,518	-	-	10,777	-
E9PDE8	Heat shock 70 kDa protein 4L	HSPA4L	-	3,508	-	-	10,579	-
P20700	Lamin-B1	LMNB1	-	11,899	-	-	2,083	-
Q13085	Acetyl-CoA carboxylase 1	ACACA	-	638	-	-	1,118	-
B5MBZ0	Echinoderm microtubule-associated protein-like 4	EML4	-	6,904	-	-	5,233	-
P07099	Epoxide hydrolase 1	EPHX1	-	12,373	-	-	7,585	-
P10620	Microsomal glutathione S-transferase 1	MGST1	-	18,671	-	-	23,301	-
C9JDR0	Sterol-4-alpha-carboxylate 3-dehydrogenase, decarboxylating	NSDHL	-	22,313	-	-	10,030	-
P08559	Pyruvate dehydrogenase E1 component subunit alpha, somatic form, mitochondrial	PDHA1	-	6,317	-	-	9,010	-
P40937	Replication factor C subunit 5	RFC5	-	6,000	-	-	5,785	-
Q9P2E9	Ribosome-binding protein 1	RRBP1	-	2,240	-	-	4,977	-
P09012	U1 small nuclear ribonucleoprotein A	SNRPA	-	19,567	-	-	44,080	-

Appendix Table 3.5 Selectively enriched proteins in Exos by label-free mass spectrometry

Protein Access. (UniProt)	Protein description	Gene name	SW480-sMVs	SW480-sMB-Rs	SW480-Exos	SW620-sMVs	SW620-sMB-Rs	SW620-Exos
F8VV56	CD63 antigen	CD63	376,641	-	962,989	36,030	34,869	525,147
Q9P2B2	Prostaglandin F2 receptor negative regulator	PTGFRN	486,366	2,039	726,223	4,783	3,921	117,066
Q8N5I2	Arrestin domain-containing protein 1	ARRDC1	360,359	9,284	646,143	47,319	26,081	178,366
Q96QD8	Sodium-coupled neutral amino acid transporter 2	SLC38A2	253,005	51,865	304,664	144,874	58,070	520,520
A8MXQ1	Pituitary tumor-transforming gene 1 protein-interacting protein	PTTG1IP	83,305	-	238,340	-	11,434	154,878
Q9BV40	Vesicle-associated membrane protein 8	VAMP8	-	-	152,817	-	-	-
O14672	Disintegrin and metalloproteinase domain-containing protein 10	ADAM10	114,652	13,356	215,793	92,112	37,559	208,133
Q9Y237	Peptidyl-prolyl cis-trans isomerase NIMA-interacting 4	PIN4	-	-	136,985	-	-	-
P50151	Guanine nucleotide-binding protein G	GNG10	107,638	-	186,266	-	-	143,527
Q8WWI5	Choline transporter-like protein 1	SLC44A1	126,794	18,411	213,871	65,948	15,754	125,536
P53990	IST1 homolog	IST1	134,635	24,427	194,853	128,125	75,573	434,185
Q9BW04	Specifically androgen-regulated gene protein	SARG	121,128	11,364	173,658	63,472	31,508	179,250
Q8WV92	MIT domain-containing protein 1	MITD1	46,652	-	98,508	-	-	47,245
O60635	Tetraspanin-1	TSPAN1	73,976	-	128,571	-	-	28,886
P12830	Cadherin-1	CDH1	62,362	11,294	114,132	118,224	66,644	302,060
A0A087WXP3	Integrin beta	ITGB6	45,911	-	82,426	24,403	-	97,763
P10316	HLA class I histocompatibility antigen, A-69 alpha chain	HLA-A	53,592	-	84,784	-	-	24,680
O95164	Ubiquitin-like protein 3	UBL3	27,321	-	96,436	-	-	17,907
A8MXZ4	G-protein-coupled receptor family C group 5 member C	GPRC5C	58,462	-	89,055	12,279	10,851	39,553
Q2TAP0	Golgin subfamily A member 7B	GOLGA7B	-	-	65,484	-	-	-
O95183	Vesicle-associated membrane protein 5	VAMP5	26,760	-	55,443	-	-	45,797
Q15836	Vesicle-associated membrane protein 3	VAMP3	36,033	-	61,516	-	-	28,717
Q16790	Carbonic anhydrase 9	CA9	38,867	-	71,195	46,938	18,351	134,448
P02790	Hemopexin	HPX	23,522	-	72,818	-	-	54,098

Q8NG11	Tetraspanin-14	TSPAN14	19,594	-	67,265	-	-	79,319
Q9UN37	Vacuolar protein sorting-associated protein 4A	VPS4A	45,385	22,284	79,558	38,460	79,049	140,528
A2A2V1	Major prion protein	PRNP	17,791	16,256	51,925	-	-	31,930
Q9C0H2	Protein tweety homolog 3	TTYH3	36,233	-	55,723	5,912	-	58,236
Q8IYI6	Exocyst complex component 8	EXOC8	40,023	-	55,332	6,178	11,114	27,730
O43657	Tetraspanin-6	TSPAN6	14,280	-	32,107	-	-	20,362
Q6IAA8	Ragulator complex protein LAMTOR1	LAMTOR1	-	-	51,765	-	-	15,325
E7ESP4	Integrin alpha-2	ITGA2	15,447	-	31,554	-	-	21,056
F8VUA2	Charged multivesicular body protein 1a	CHMP1A	12,682	4,571	32,297	-	-	39,378
O60869	Endothelial differentiation-related factor 1	EDF1	-	-	24,439	-	-	27,839
Q9NV70	Exocyst complex component 1	EXOC1	23,806	-	33,232	-	-	25,750
F8WCD0	E3 ubiquitin-protein ligase RNF149	RNF149	17,540	-	31,483	-	-	3,494
Q92692	Nectin-2	NECTIN2	13,297	-	28,389	4,421	-	23,885
P78536	Disintegrin and metalloproteinase domain-containing protein 17	ADAM17	23,543	-	31,449	4,464	7,933	20,680
Q5W0Z9	Probable palmitoyltransferase ZDHHC20	ZDHHC20	25,786	-	32,013	-	-	6,226
J3KMZ9	Low-density lipoprotein receptor	LDLR	17,310	6,430	27,370	-	-	24,474
X6RM00	ELKS/Rab6-interacting/CAST family member 1	ERC1	15,343	-	19,814	-	-	3,747
Q96J02	E3 ubiquitin-protein ligase Itchy homolog	ITCH	17,663	-	25,592	5,671	-	19,219
Q9NRX5	Serine incorporator 1	SERINC1	-	-	7,458	-	-	18,016
F5GYQ1	V-type proton ATPase subunit	ATP6V0D1	-	-	12,794	-	-	-
P56199	Integrin alpha-1	ITGA1	13,793	-	17,580	-	-	7,095
O75317	Ubiquitin carboxyl-terminal hydrolase 12	USP12	-	-	5,846	-	-	14,222
Q08345	Epithelial discoidin domain-containing receptor 1	DDR1	10,457	-	15,504	-	-	3,545
Q9Y3R5	Protein dopey-2	DOPEY2	14,251	-	15,000	-	-	7,264
O75128	Protein cordon-bleu	COBL	8,522	-	12,286	2,983	-	14,020
P14209	CD99 antigen	CD99	-	-	3,780	-	-	16,497
Q8IZW8	Tensin-4	TNS4	6,964	-	11,198	-	-	3,594
E7EW84	Exocyst complex component	EXOC6	7,267	-	10,238	1,861	5,898	9,383
O75110	Probable phospholipid-transporting ATPase IIA	ATP9A	7,940	-	10,324	-	-	1,046
O94887	FERM, RhoGEF and pleckstrin domain-containing protein 2	FARP2	5,292	-	7,453	2,362	-	8,849

Q99569	Plakophilin-4	PKP4	5,730	-	7,428	2,306	4,186	9,037
O75054	Immunoglobulin superfamily member 3	IGSF3	8,090	-	10,714	-	-	2,549
Q15223	Nectin-1	NECTIN1	-	-	2,277	-	-	7,040
P46531	Neurogenic locus notch homolog protein 1	NOTCH1	4,585	-	7,441	-	-	5,043
Q13393	Phospholipase D1	PLD1	-	-	6,330	-	-	-
E7ENQ1	Mitogen-activated protein kinase kinase kinase kinase 4	MAP4K4	2,691	-	5,502	-	-	7,336
Q7Z403	Transmembrane channel-like protein 6	TMC6	5,550	-	5,999	5,012	-	14,125
Q7Z3U7	Protein MON2 homolog	MON2	4,569	-	3,729	-	-	1,854
O75882	Attractin	ATRN	1,928	-	2,477	596	-	1,578
P36894	Bone morphogenetic protein receptor type-1A	BMPR1A	-	-	-	-	-	3,959
Q5VT25	Serine/threonine-protein kinase MRCK alpha	CDC42BPA	-	-	-	3,087	-	9,814
Q9UKY7	Protein CDV3 homolog	CDV3	-	-	-	-	-	17,736
Q8NCR9	Clarin-3	CLRN3	-	-	-	48,226	-	147,665
Q5VW36	Focadhesin	FOCAD	-	-	-	52,686	-	107,242
Q9Y4H2	Insulin receptor substrate 2	IRS2	-	-	-	-	-	3,167
P78504	Protein jagged-1	JAG1	-	-	511	3,023	718	7,738
Q9UHA4	Ragulator complex protein LAMTOR3	LAMTOR3	-	-	163,164	-	106,844	663,677
Q9NS86	LanC-like protein 2	LANCL2	-	-	-	-	-	14,567
O15344	E3 ubiquitin-protein ligase Midline-1	MID1	-	-	-	3,148	-	12,889
Q8IZ21	Phosphatase and actin regulator 4	PHACTR4	-	-	-	2,092	-	6,546
Q86UU1	Pleckstrin homology-like domain family B member 1	PHLDB1	-	-	9,121	-	-	-
P49023	Paxillin	PXN	-	-	3,872	-	-	2,864
Q8TEB7	E3 ubiquitin-protein ligase RNF128	RNF128	-	-	-	-	-	28,543
E9PS74	Solute carrier family 43 member 3	SLC43A3	-	-	3,147	-	-	20,888
P37173	TGF-beta receptor type-2	TGFBR2	-	-	-	-	-	10,279

Appendix Table 3.6 Selectively enriched proteins in sMVs by label-free mass spectrometry

Protein Access. (UniProt)	Protein description	Gene name	SW480-sMVs	SW480-sMB-Rs	SW480-Exos	SW620-sMVs	SW620-sMB-Rs	SW620-Exos
Q14542	Equilibrative nucleoside transporter 2	SLC29A2	19,309	-	12,057	4,976	-	-
M0R0V2	Deoxyhypusine synthase	DHPS	20,175	-	-	-	-	-
Q14185	Dedicator of cytokinesis protein 1	DOCK1	789	-	-	3,676	-	2,539
P23919	Thymidylate kinase	DTYMK	18,102	-	-	15,154	-	-
Q14156	Protein EFR3 homolog A	EFR3A	12,970	-	-	-	-	-
P22413	Ectonucleotide pyrophosphatase/phosphodiesterase family member 1	ENPP1	-	-	-	1,656	-	-
B1AK53	Espin	ESPN	-	-	-	21,304	-	6,374
P22455	Fibroblast growth factor receptor 4	FGFR4	2,420	-	-	10,476	-	6,308
J3KNW4	Four and a half LIM domains protein 2	FHL2	4,279	-	-	11,574	-	4,147
P29218	Inositol monophosphatase 1	IMPA1	9,695	-	-	47,096	25,899	13,144
Q9BV20	Methylthioribose-1-phosphate isomerase	MRI1	7,804	-	-	34,210	8,556	-
A0A024R412	Neuropilin	NRP2	2,170	-	-	-	-	-
K7ELV2	Nucleoporin SEH1	SEH1L	-	-	-	17,180	-	-
Q99805	Transmembrane 9 superfamily member 2	TM9SF2	19,139	-	-	-	-	-

Appendix Table 3.7 Selectively enriched proteins in sMB-Rs by label-free mass spectrometry

Protein Access.	Protein description	Gene name	SW480-sMVs	SW480-sMB-Rs_3	SW480-Exos	SW620-sMVs	SW620-sMB-Rs	SW620-Exos	A	B	C
P16403	Histone H1.2	HIST1H1C	103,434	17,457,590	146,951	-	599,781	-	Y		
P62906	60S ribosomal protein L10a	RPL10A	568,065	3,202,273	568,725	331,439	1,426,886	207,402	Y		Y
P08238	Heat shock protein HSP 90-beta	HSP90AB1	1,293,633	3,059,576	1,122,274	1,744,475	3,706,538	1,191,137	Y		Y
P21980	Protein-glutamine gamma-glutamyltransferase 2	TGM2	660,063	1,411,989	652,639	584	16,397	1,075			
P62249	40S ribosomal protein S16	RPS16	785,993	2,650,890	829,087	509,555	2,329,406	258,237	Y		Y
P62277	40S ribosomal protein S13	RPS13	863,857	2,577,417	1,021,766	622,208	2,195,106	313,883	Y		Y
P62241	40S ribosomal protein S8	RPS8	491,255	2,507,821	462,771	347,603	1,710,431	144,005	Y		Y
P05783	Keratin, type I cytoskeletal 18	KRT18	523,798	2,564,496	352,001	437,674	1,167,496	174,021	Y		Y
P46781	40S ribosomal protein S9	RPS9	769,536	2,390,481	704,863	450,442	1,854,306	174,338	Y		Y
P23396	40S ribosomal protein S3	RPS3	722,840	2,271,879	610,604	646,927	2,502,483	337,219	Y		Y
P62701	40S ribosomal protein S4, X isoform	RPS4X	539,957	1,914,588	469,923	392,015	1,651,711	107,000	Y		
Q02878	60S ribosomal protein L6	RPL6	381,319	1,872,813	370,706	284,990	1,459,326	72,529	Y		
P62424	60S ribosomal protein L7a	RPL7A	317,617	1,903,333	307,254	206,179	1,136,628	8,133	Y		Y
P62280	40S ribosomal protein S11	RPS11	539,454	1,912,342	515,618	306,772	1,531,424	133,535	Y		Y
J3QQ67	60S ribosomal protein L18	RPL18	422,228	1,746,456	388,668	281,514	1,368,070	101,454			
P62269	40S ribosomal protein S18	RPS18	602,373	1,882,368	675,969	447,820	1,698,461	199,888	Y		Y
P18124	60S ribosomal protein L7	RPL7	462,989	1,899,032	500,444	287,970	1,519,641	81,042	Y		Y
P36578	60S ribosomal protein L4	RPL4	346,073	1,771,717	337,244	272,982	1,585,566	42,598	Y		
E7EPB3	60S ribosomal protein L14	RPL14	360,089	1,516,667	370,124	274,554	1,133,648	116,640			
P62081	40S ribosomal protein S7	RPS7	550,414	1,785,412	534,277	426,880	1,445,773	124,813	Y		Y
P26641	Elongation factor 1-gamma	EEF1G	659,115	1,709,039	472,021	626,568	1,628,589	310,099	Y		Y
Q9H0H5	Rac GTPase-activating protein 1	RACGAP1	85,367	1,689,821	18,088	38,445	2,506,751	2,708	Y	Y	Y

P15880	40S ribosomal protein S2	RPS2	347,379	1,673,231	350,185	261,849	1,108,594	99,564	Y		
P05388	60S acidic ribosomal protein P0	RPLP0	371,148	1,591,146	306,855	287,083	939,546	137,479	Y		Y
P61247	40S ribosomal protein S3a	RPS3A	422,285	1,609,949	355,357	320,995	1,214,025	84,442	Y		Y
C9J9K3	40S ribosomal protein SA	RPSA	539,100	1,159,493	418,537	414,867	1,441,381	96,750			
P61353	60S ribosomal protein L27	RPL27	380,400	1,333,064	432,223	307,066	1,240,039	122,294	Y		
P63244	Receptor of activated protein C kinase 1	RACK1	579,737	1,454,784	513,113	486,151	1,559,159	296,460	Y	Y	Y
P35579	Myosin-9	MYH9	620,891	1,464,065	443,148	1,781,616	2,131,820	1,138,571	Y		Y
P62854	40S ribosomal protein S26	RPS26	293,812	1,249,536	295,901	210,316	1,062,171	72,316	Y		Y
P60866	40S ribosomal protein S20	RPS20	445,829	1,182,997	366,137	339,950	956,913	170,146	Y		Y
P05387	60S acidic ribosomal protein P2	RPLP2	211,619	1,226,319	132,037	49,858	751,544	7,590	Y		
Q02543	60S ribosomal protein L18a	RPL18A	211,754	1,022,670	196,104	147,045	936,167	7,808	Y		Y
Q71DI3	Histone H3.2	HIST2H3A;	-	1,149,510	-	-	58,295	-	Y		
P11021	78 kDa glucose-regulated protein	HSPA5	190,663	1,089,414	17,406	305,576	1,269,215	32,950	Y	Y	Y
P40429	60S ribosomal protein L13a	RPL13A	218,652	1,107,077	249,435	195,878	944,530	33,709	Y		
J3QRI7	60S ribosomal protein L26	RPL26	179,729	707,352	194,657	70,268	676,172	18,508			
P62753	40S ribosomal protein S6	RPS6	214,886	1,004,137	216,201	126,940	747,285	54,878	Y		
E7ETK0	40S ribosomal protein S24	RPS24	229,557	842,163	217,064	112,160	591,746	-			
J3KNF8	Cytochrome b5 type B	CYB5B	406,353	976,467	341,891	156,024	614,858	-			
P62266	40S ribosomal protein S23	RPS23	151,946	1,050,443	149,044	127,841	911,434	16,019	Y		Y
Q02241	Kinesin-like protein KIF23	KIF23	38,477	1,062,271	11,504	14,435	1,733,403	91	Y	Y	Y
P17096	High mobility group protein HMG-I/HMG-Y	HMGA1	211,339	2,907,975	286,505	-	388,305	16,640			
X1WI28	60S ribosomal protein L10	RPL10	269,643	1,102,867	219,938	172,772	985,268	18,280			
P39023	60S ribosomal protein L3	RPL3	143,221	985,451	156,693	129,429	859,462	17,405	Y		Y
P26373	60S ribosomal protein L13	RPL13	145,657	933,397	209,327	104,292	881,719	27,472	Y		
P49368	T-complex protein 1 subunit gamma	CCT3	365,450	960,734	356,807	530,263	1,267,627	300,349	Y		
K7EM56	40S ribosomal protein S15	RPS15	188,732	534,842	143,652	158,074	595,442	-			
P08727	Keratin, type I cytoskeletal 19	KRT19	452,650	1,951,017	310,194	566,808	1,616,708	322,533			Y
S4R435	RPS10-NUDT3 readthrough	RPS10-NUDT3	168,933	433,522	195,380	223,938	582,722	115,189			

P21796	Voltage-dependent anion-selective channel protein 1	VDAC1	204,643	943,804	113,357	81,780	514,641	-	Y		Y
F8W6I7	Heterogeneous nuclear ribonucleoprotein A1	HNRNPA1	280,594	667,383	189,305	577,883	1,543,355	204,208			
P78371	T-complex protein 1 subunit beta	CCT2	484,187	961,857	433,271	515,259	1,389,134	341,034	Y		Y
P61313	60S ribosomal protein L15	RPL15	115,544	944,641	142,737	94,503	871,874	15,611	Y		
P32969	60S ribosomal protein L9	RPL9	221,608	919,497	181,557	153,139	645,453	24,651	Y		
P55060	Exportin-2	CSE1L	418,819	983,598	306,646	259,327	614,418	211,490	Y		Y
P24534	Elongation factor 1-beta	EEF1B2	415,698	803,081	260,801	224,498	705,332	83,867	Y		Y
P12268	Inosine-5'-monophosphate dehydrogenase 2	IMPDH2	256,481	790,084	168,623	178,899	441,783	95,078	Y		
P46776	60S ribosomal protein L27a	RPL27A	95,471	698,009	97,426	70,318	757,676	27,235	Y		
P62917	60S ribosomal protein L8	RPL8	156,206	811,933	182,802	124,528	734,900	21,002	Y		Y
Q9Y265	RuvB-like 1	RUVBL1	291,075	736,389	190,980	168,327	353,596	117,105	Y		
C9J4Z3	60S ribosomal protein L37a	RPL37A	158,855	700,172	144,089	135,692	616,052	-			
P25398	40S ribosomal protein S12	RPS12	269,595	759,667	212,629	294,240	738,505	50,244	Y		Y
P50991	T-complex protein 1 subunit delta	CCT4	331,596	716,784	304,873	379,443	949,920	256,877	Y		Y
Q9Y3U8	60S ribosomal protein L36	RPL36	105,807	684,384	108,878	77,853	491,844	-			Y
P55072	Transitional endoplasmic reticulum ATPase	VCP	290,993	712,469	207,523	254,103	483,867	161,381	Y	Y	Y
P08708	40S ribosomal protein S17	RPS17	223,175	842,099	104,797	179,573	895,686	23,360			
J3QR09	Ribosomal protein L19	RPL19	86,444	345,611	92,739	65,314	388,107	-			
P17987	T-complex protein 1 subunit alpha	TCP1	287,218	706,427	279,281	367,272	986,517	228,243	Y		Y
P05386	60S acidic ribosomal protein P1	RPLP1	185,120	1,077,836	98,772	129,918	546,529	-			
P30101	Protein disulfide-isomerase A3	PDIA3	113,931	642,944	8,874	231,934	1,355,162	14,622	Y		Y
P41091	Eukaryotic translation initiation factor 2 subunit 3	EIF2S3	171,683	665,706	158,602	64,373	332,267	44,785	Y		Y
P61978	Heterogeneous nuclear ribonucleoprotein K	HNRNPK	143,624	623,067	102,706	164,451	702,952	63,661	Y		
C9JXB8	60S ribosomal protein L24	RPL24	65,956	1,033,747	109,799	70,167	700,096	-			
Q5TEC6	Histone H3	HIST2H3PS2	99,191	7,126,765	172,367	19,164	741,236	30,482			
M0R0F0	40S ribosomal protein S5	RPS5	165,125	635,383	47,798	138,983	720,042	15,072			
P12956	X-ray repair cross-complementing protein 6	XRCC6	11,915	613,552	6,172	6,941	360,476	2,525	Y		Y
Q99832	T-complex protein 1 subunit eta	CCT7	299,472	589,411	281,363	358,828	900,773	213,913	Y		

Q969Q0	60S ribosomal protein L36a-like	RPL36AL	-	204,758	30,302	18,916	171,670	27,925			
D3YTB1	60S ribosomal protein L32	RPL32	42,329	817,644	72,627	54,182	696,953	4,851			Y
Q04837	Single-stranded DNA-binding protein, mitochondrial	SSBP1	3,703	506,414	12,497	10,980	152,144	-	Y		
P46777	60S ribosomal protein L5	RPL5	145,052	569,113	85,318	154,519	672,746	24,584	Y		Y
E9PK01	Elongation factor 1-delta	EEF1D	210,978	510,600	151,461	194,538	593,114	81,167			Y
E7EU96	Casein kinase II subunit alpha	CSNK2A1	84,287	277,074	91,387	91,203	253,760	85,964			
Q58FF8	Putative heat shock protein HSP 90-beta 2	HSP90AB2P	212,973	568,031	246,521	287,507	533,578	211,812	Y		
P42677	40S ribosomal protein S27	RPS27	95,217	576,639	70,033	81,910	632,730	-	Y		
P35268	60S ribosomal protein L22	RPL22	122,753	536,214	135,531	72,987	465,813	-	Y		
P14625	Endoplasmin	HSP90B1	101,881	548,742	28,364	241,706	926,704	69,509	Y	Y	Y
P27797	Calreticulin	CALR	176,320	539,720	8,473	185,432	821,918	-	Y		
P53999	Activated RNA polymerase II transcriptional coactivator p15	SUB1	-	527,764	38,554	-	118,339	-	Y		
Q9UQ80	Proliferation-associated protein 2G4	PA2G4	151,772	526,396	115,882	249,790	766,345	38,690	Y		Y
E9PPJ5	Midkine	MDK	-	413,613	19,889	-	-	-			
Q15084	Protein disulfide-isomerase A6	PDIA6	153,578	513,803	46,412	83,459	171,365	-	Y		Y
Q15366	Poly	PCBP2	157,973	466,055	135,501	289,742	521,416	199,063	Y		
P14868	Aspartate--tRNA ligase, cytoplasmic	DARS	159,532	447,166	142,420	117,438	489,707	62,228	Y		Y
O60506	Heterogeneous nuclear ribonucleoprotein Q	SYNCRIP	123,210	472,204	84,283	90,039	418,694	16,295	Y		
O75367	Core histone macro-H2A.1	H2AFY	21,600	2,872,939	57,283	1,250	142,802	-			Y
P13010	X-ray repair cross-complementing protein 5	XRCC5	6,737	423,224	3,829	3,476	249,199	5,046	Y		Y
Q5SRQ6	Casein kinase II subunit beta	CSNK2B	23,956	120,167	35,469	21,809	114,544	19,768			
Q9NY12	H/ACA ribonucleoprotein complex subunit 1	GAR1	-	57,069	16,275	-	-	-			
F8VPD4	CAD protein	CAD	63,349	156,986	53,313	52,336	135,848	46,561			Y
Q9P016	Thymocyte nuclear protein 1	THYN1	28,566	41,291	15,258	-	24,296	-			
P27824	Calnexin	CANX	57,326	343,148	16,262	108,205	457,050	5,218	Y		Y
F8VVA7	Coatomer subunit zeta-1	COPZ1	31,159	184,985	47,495	64,259	218,394	36,995			
K7EK07	Histone H3	H3F3B	-	1,878,409	14,707	-	68,889	-			
P07237	Protein disulfide-isomerase	P4HB	150,632	418,038	14,720	191,118	912,513	28,377	Y		Y
Q9UNM6	26S proteasome non-ATPase regulatory subunit 13	PSMD13	74,512	361,356	59,257	37,397	206,027	25,916	Y		Y

P12236	ADP/ATP translocase 3	SLC25A6	67,501	385,593	31,210	69,209	632,416	-	Y		Y
Q9Y230	RuvB-like 2	RUVBL2	159,199	386,587	118,179	116,379	213,859	65,024	Y	Y	Y
P62314	Small nuclear ribonucleoprotein Sm D1	SNRPD1	22,665	279,930	41,814	31,225	287,876	5,127			
O00231	26S proteasome non-ATPase regulatory subunit 11	PSMD11	101,988	374,534	100,611	52,956	251,392	7,485	Y		Y
P49207	60S ribosomal protein L34	RPL34	14,456	332,895	45,727	-	450,607	-	Y		
P62195	26S proteasome regulatory subunit 8	PSMC5	56,177	290,534	55,794	41,255	173,802	8,385	Y		Y
E9PKG1	Protein arginine N-methyltransferase 1	PRMT1	23,982	197,596	21,326	10,721	46,847	11,554			
Q9NRW3	DNA dC->dU-editing enzyme APOBEC-3C	APOBEC3C	-	41,718	12,433	-	18,248	-			
R4GNH3	26S proteasome regulatory subunit 6A	PSMC3	39,887	208,103	31,730	29,980	130,528	10,843			
Q14204	Cytoplasmic dynein 1 heavy chain 1	DYNC1H1	146,378	302,360	137,097	112,677	285,019	106,195	Y	Y	Y
O96019	Actin-like protein 6A	ACTL6A	-	51,605	12,249	-	10,042	-			Y
O00264	Membrane-associated progesterone receptor component 1	PGRMC1	43,450	123,660	36,734	32,293	116,585	-			
P06576	ATP synthase subunit beta, mitochondrial	ATP5B	8,607	256,377	-	8,996	243,938	-	Y		
Q14697	Neutral alpha-glucosidase AB	GANAB	34,927	279,936	3,152	120,007	465,823	14,015	Y		Y
P62899	60S ribosomal protein L31	RPL31	61,912	300,512	78,813	17,009	294,264	-	Y		
P26599	Polypyrimidine tract-binding protein 1	PTBP1	41,558	281,695	17,180	83,917	263,935	9,840	Y		
Q15008	26S proteasome non-ATPase regulatory subunit 6	PSMD6	83,643	264,844	84,300	36,674	162,326	27,163	Y		Y
P47914	60S ribosomal protein L29	RPL29	9,861	261,694	25,814	9,602	123,442	-	Y		
H3BR35	Eukaryotic peptide chain release factor GTP-binding subunit ERF3A	GSPT1	36,347	103,898	29,695	23,397	107,199	26,506			
P41252	Isoleucine--tRNA ligase, cytoplasmic	IARS	53,010	268,072	39,994	39,179	254,464	16,032	Y		
Q13200	26S proteasome non-ATPase regulatory subunit 2	PSMD2	83,182	270,422	76,335	33,366	154,916	28,377	Y		Y
P62333	26S proteasome regulatory subunit 10B	PSMC6	64,132	233,434	53,002	60,686	181,284	12,702	Y		
Q15582	Transforming growth factor-beta-induced protein ig-h3	TGFBI	6,738	259,009	16,707	-	63,270	-			Y
P04843	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1	RPN1	31,334	222,729	-	71,314	408,473	4,030	Y		
P68366	Tubulin alpha-4A chain	TUBA4A	75,880	220,150	67,534	94,528	247,894	54,592	Y	Y	Y
O43324	Eukaryotic translation elongation factor 1 epsilon-1	EEF1E1	20,352	229,425	62,140	39,889	203,851	-	Y		
P54136	Arginine--tRNA ligase, cytoplasmic	RARS	51,955	228,914	42,702	46,033	220,409	12,424	Y		Y
P43686	26S proteasome regulatory subunit 6B	PSMC4	29,682	238,173	23,113	21,754	112,918	12,013	Y		Y
F5H5D3	Tubulin alpha chain	TUBA1C	61,972	162,566	30,162	77,933	162,168	34,408			
Q9UKM9	RNA-binding protein Raly	RALY	-	28,622	7,024	-	38,510	-			

P50454	Serpin H1	SERPINH1	12,906	207,114	16,809	33,991	201,575	3,377	Y		
P56192	Methionine--tRNA ligase, cytoplasmic	MARS	30,482	196,685	19,216	25,314	271,041	1,701	Y		Y
O14980	Exportin-1	XPO1	88,014	208,382	60,360	65,228	130,990	54,019	Y		Y
Q14103	Heterogeneous nuclear ribonucleoprotein D0	HNRNPD	48,492	211,402	26,432	103,100	283,403	38,889	Y		Y
P45880	Voltage-dependent anion-selective channel protein 2	VDAC2	11,766	189,210	7,296	29,286	319,459	-	Y		Y
O00487	26S proteasome non-ATPase regulatory subunit 14	PSMD14	70,708	195,994	51,776	31,948	120,357	7,605	Y		
J3KTL2	Serine/arginine-rich-splicing factor 1	SRSF1	3,438	172,316	14,937	23,318	280,418	17,424			Y
Q9BVK6	Transmembrane emp24 domain-containing protein 9	TMED9	9,177	47,009	5,957	7,923	43,650	4,507			
P54577	Tyrosine--tRNA ligase, cytoplasmic	YARS	85,614	195,997	74,940	105,869	245,631	69,538	Y		Y
P07814	Bifunctional glutamate/proline--tRNA ligase	EPRS	20,948	195,893	11,518	19,047	193,640	3,753	Y		Y
J3QLE5	Small nuclear ribonucleoprotein-associated protein N	SNRPN	12,286	38,286	13,197	9,448	67,126	6,568			
A6XND1	Insulin-like growth factor binding protein 3 isoform b	IGFBP3	-	128,745	10,382	-	-	-			
Q7KZF4	Staphylococcal nuclease domain-containing protein 1	SND1	71,789	174,604	53,123	86,638	235,872	66,038	Y		Y
Q53GQ0	Very-long-chain 3-oxoacyl-CoA reductase	HSD17B12	90,947	312,640	42,620	135,493	488,427	9,364			
Q9P2J5	Leucine--tRNA ligase, cytoplasmic	LARS	6,051	189,855	895	7,322	182,225	945	Y		Y
Q7L2H7	Eukaryotic translation initiation factor 3 subunit M	EIF3M	51,296	162,469	57,445	43,647	180,594	18,897	Y		Y
P20042	Eukaryotic translation initiation factor 2 subunit 2	EIF2S2	28,288	192,095	29,332	38,085	100,434	-	Y		
O00232	26S proteasome non-ATPase regulatory subunit 12	PSMD12	40,734	165,711	26,050	16,782	113,192	3,075	Y		Y
P55884	Eukaryotic translation initiation factor 3 subunit B	EIF3B	23,558	152,273	15,966	18,069	137,793	7,274	Y		
Q13838	Spliceosome RNA helicase DDX39B	DDX39B	8,902	161,217	5,552	17,607	184,607	1,598	Y		Y
P19338	Nucleolin	NCL	5,492	155,469	3,064	4,493	114,641	348	Y		
P39656	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit	DDOST	38,785	175,343	24,710	61,084	318,512	6,285	Y		Y
Q00839	Heterogeneous nuclear ribonucleoprotein U	HNRNPU	6,574	162,877	6,567	3,206	89,323	-	Y		Y
Q96AG4	Leucine-rich repeat-containing protein 59	LRRC59	25,103	163,391	21,958	33,262	135,242	27,287	Y		
P61289	Proteasome activator complex subunit 3	PSME3	-	143,341	10,864	-	75,012	-	Y		Y
O00303	Eukaryotic translation initiation factor 3 subunit F	EIF3F	26,758	151,831	18,505	21,511	154,811	11,841	Y		
Q9Y4L1	Hypoxia up-regulated protein 1	HYOU1	11,977	132,763	4,187	7,271	93,754	3,037	Y		Y
P60228	Eukaryotic translation initiation factor 3 subunit E	EIF3E	31,531	136,422	21,294	23,217	141,352	11,282	Y		
Q99733	Nucleosome assembly protein 1-like 4	NAP1L4	-	113,541	-	15,119	89,208	-	Y	Y	
P52292	Importin subunit alpha-1	KPNA2	69,069	126,187	59,820	35,365	105,001	21,927	Y		

Q9NQW6	Anillin	ANLN	35,600	105,559	19,572	13,734	32,927	8,349	Y	Y	
Q14152	Eukaryotic translation initiation factor 3 subunit A	EIF3A	11,994	119,428	8,309	5,906	149,100	1,758	Y		Y
Q8N163	Cell cycle and apoptosis regulator protein 2	CCAR2	-	5,829	2,914	-	14,960	-			
Q15046	Lysine--tRNA ligase	KARS	27,465	109,493	6,590	73,643	392,066	8,650	Y		Y
Q99460	26S proteasome non-ATPase regulatory subunit 1	PSMD1	18,589	119,129	11,757	7,991	58,802	12,194	Y		
Q15149	Plectin	PLEC	50,793	119,675	39,077	48,116	108,634	27,825	Y		Y
P53618	Coatomer subunit beta	COPB1	9,184	103,378	4,040	11,177	144,320	9,041	Y		Y
B4DY09	Interleukin enhancer-binding factor 2	ILF2	10,273	209,691	9,401	5,695	199,542	8,131			Y
O43390	Heterogeneous nuclear ribonucleoprotein R	HNRNPR	1,157	45,206	5,323	8,280	70,642	811			Y
P08670	Vimentin	VIM	5,445	128,945	6,717	6,746	67,150	1,687	Y	Y	Y
P53621	Coatomer subunit alpha	COPA	30,772	110,354	26,971	31,603	134,875	17,518	Y		Y
P17858	ATP-dependent 6-phosphofructokinase, liver type	PFKL	38,018	119,083	31,389	48,804	103,648	19,328	Y		Y
Q9UBQ5	Eukaryotic translation initiation factor 3 subunit K	EIF3K	23,664	116,616	33,609	-	79,171	-	Y		
Q99613	Eukaryotic translation initiation factor 3 subunit C	EIF3C	6,591	108,945	1,834	1,298	90,863	1,345	Y		
P61221	ATP-binding cassette sub-family E member 1	ABCE1	37,506	98,604	27,448	54,084	137,277	23,843	Y		Y
P20290	Transcription factor BTF3	BTF3	27,811	87,652	18,230	35,851	104,512	9,447	Y		
E7EX73	Eukaryotic translation initiation factor 4 gamma 1	EIF4G1	3,326	38,342	2,701	3,656	21,635	745			
Q53EZ4	Centrosomal protein of 55 kDa	CEP55	12,795	97,651	9,834	-	47,455	-	Y	Y	
J3KTA4	Probable ATP-dependent RNA helicase DDX5	DDX5	2,780	32,833	1,489	3,176	66,407	946			
O76094	Signal recognition particle subunit SRP72	SRP72	-	93,874	27,802	11,824	101,842	-	Y		
Q9NSD9	Phenylalanine--tRNA ligase beta subunit	FARSB	11,520	91,846	5,666	21,137	195,014	892	Y		
P05141	ADP/ATP translocase 2	SLC25A5	10,791	104,056	6,601	10,300	192,923	-	Y		Y
P00367	Glutamate dehydrogenase 1, mitochondrial	GLUD1	-	73,716	-	8,611	113,787	-	Y		
P05121	Plasminogen activator inhibitor 1	SERPINE1	3,090	677,728	2,632	927	728	-			
Q08945	FACT complex subunit SSRP1	SSRP1	-	81,476	-	-	22,185	-	Y	Y	Y
Q14978	Nucleolar and coiled-body phosphoprotein 1	NOLC1	-	85,150	-	-	15,812	-	Y		
P14866	Heterogeneous nuclear ribonucleoprotein L	HNRNPL	1,158	74,103	961	6,000	144,736	-	Y		Y
Q09028	Histone-binding protein RBBP4	RBBP4	3,977	79,252	-	17,167	54,536	6,256	Y		Y
P26196	Probable ATP-dependent RNA helicase DDX6	DDX6	25,362	79,261	22,633	21,501	87,502	21,565	Y		Y
Q9NR31	GTP-binding protein SAR1a	SAR1A	-	85,737	-	25,348	60,057	-	Y		

Q9Y224	UPF0568 protein C14orf166	C14orf166	15,081	78,792	31,523	12,480	59,134	-	Y		
O14744	Protein arginine N-methyltransferase 5	PRMT5	30,148	79,087	25,417	44,567	102,065	25,869	Y	Y	Y
P51572	B-cell receptor-associated protein 31	BCAP31	21,107	54,008	18,138	7,052	50,721	-	Y		
Q96A72	Protein mago nashi homolog 2	MAGOHB	12,573	77,288	24,096	-	77,847	-	Y		
P38646	Stress-70 protein, mitochondrial	HSPA9	3,789	79,901	9,143	10,867	115,531	1,211	Y		Y
P25705	ATP synthase subunit alpha, mitochondrial	ATP5A1	3,420	82,538	1,256	3,491	111,794	-	Y		Y
O95573	Long-chain-fatty-acid--CoA ligase 3	ACSL3	-	8,524	-	2,228	14,237	-			Y
P02545	Prelamin-A/C [Cleaved into: Lamin-A/C	LMNA	8,349	69,147	4,309	500	5,832	489	Y	Y	Y
O00468	Agrin [Cleaved into: Agrin N-terminal 110 kDa subunit; Agrin C-terminal 110 kDa subunit; Agrin C-terminal 90 kDa fragment	AGRN	-	58,791	-	-	-	-			Y
P47897	Glutamine--tRNA ligase	QARS	18,063	80,474	19,988	10,304	77,136	2,242	Y		
P27144	Adenylate kinase 4, mitochondrial	AK4	-	-	-	-	29,652	-			
A0A0B4J1Z1	Serine/arginine-rich-splicing factor 7	SRSF7	-	299,849	-	-	245,611	-			Y
O95782	AP-2 complex subunit alpha-1	AP2A1	1,567	18,865	-	5,810	15,118	-			Y
Q9Y678	Coatomer subunit gamma-1	COPG1	9,980	76,184	8,816	18,071	84,426	4,485	Y		Y
Q9HDC9	Adipocyte plasma membrane-associated protein	APMAP	25,700	83,425	10,178	24,351	52,166	-	Y		
V9GYM8	Rho guanine nucleotide exchange factor 2	ARHGEF2	-	-	-	-	4,398	-			
Q9NS69	Mitochondrial import receptor subunit TOM22 homolog	TOMM22	-	53,531	-	-	52,962	-			
P16615	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	ATP2A2	3,491	29,597	-	8,430	53,543	-			Y
P35606	Coatomer subunit beta'	COPB2	10,401	68,203	9,266	13,803	76,631	4,565	Y		Y
P36542	ATP synthase subunit gamma, mitochondrial	ATP5C1	-	-	-	-	33,871	-			Y
Q9Y3I0	tRNA-splicing ligase RtcB homolog	RTCB	20,828	71,823	17,999	24,390	51,745	4,039	Y		Y
Q9UHB9	Signal recognition particle subunit SRP68	SRP68	7,071	64,715	-	-	71,343	-	Y		Y
P48047	ATP synthase subunit O, mitochondrial	ATP5O	-	40,858	-	-	52,781	11,193			
G3V126	ATPase, H+ transporting, lysosomal 50/57kDa, V1 subunit H, isoform CRA_c	ATP6V1H	-	23,252	-	7,916	25,442	-			
J3KT86	Aurora kinase B	AURKB	4,807	52,033	-	-	54,161	-			
O43809	Cleavage and polyadenylation specificity factor subunit 5	NUDT21	-	41,260	-	-	51,701	-			
Q16576	Histone-binding protein RBBP7	RBBP7	-	68,943	5,629	21,148	49,368	3,474	Y		
P28070	Proteasome subunit beta type-4	PSMB4	25,794	82,399	49,920	24,831	46,273	-	Y		
P49915	GMP synthase [glutamine-hydrolyzing]	GMPS	10,494	60,516	11,114	7,700	20,756	-	Y		
Q12906	Interleukin enhancer-binding factor 3	ILF3	-	60,193	2,207	1,424	54,774	-	Y		Y

P48444	Coatomer subunit delta	ARCN1	15,728	65,529	12,374	17,298	58,783	-	Y		
E7ES10	Calpastatin	CAST	-	11,229	-	-	-	-			
Q8WUJ3	Cell migration-inducing and hyaluronan-binding protein	CEMIP	-	-	-	-	15,842	1,037			Y
F5GWX5	Chromodomain-helicase-DNA-binding protein 4	CHD4	-	1,357	-	-	-	-			
I3L1P8	Mitochondrial 2-oxoglutarate/malate carrier protein	SLC25A11	-	20,017	-	-	53,020	-			
P09661	U2 small nuclear ribonucleoprotein A'	SNRPA1	-	56,284	-	-	12,791	-			
Q9UMS4	Pre-mRNA-processing factor 19	PRPF19	3,224	64,511	-	-	52,239	-	Y		Y
P39060	Collagen alpha-1	COL18A1	833	59,768	-	-	12,196	-	Y		Y
Q00325	Phosphate carrier protein, mitochondrial	SLC25A3	3,212	57,783	-	2,637	57,574	-	Y		
E9PGT6	COP9 signalosome complex subunit 8	COPS8	-	74,505	20,308	11,188	58,154	-			
P10606	Cytochrome c oxidase subunit 5B, mitochondrial	COX5B	-	-	-	-	43,875	-			
P36551	Oxygen-dependent coproporphyrinogen-III oxidase, mitochondrial	CPOX	-	-	-	-	6,326	-			
P50416	Carnitine O-palmitoyltransferase 1, liver isoform	CPT1A	-	20,167	-	-	-	-			
Q9UBM7	7-dehydrocholesterol reductase	DHCR7	6,970	56,738	-	3,873	55,435	-	Y		
Q13363	C-terminal-binding protein 1	CTBP1	-	-	-	-	11,070	-			
Q00577	Transcriptional activator protein Pur-alpha	PURA	-	50,421	-	-	31,812	-			
Q8WYA6	Beta-catenin-like protein 1	CTNNBL1	-	1,850	-	-	-	-			
A0A1B0GW44	Cathepsin D	CTSD	2,175	14,098	3,476	-	2,346	-			
O15371	Eukaryotic translation initiation factor 3 subunit D	EIF3D	-	54,389	-	12,363	60,452	-	Y		Y
Q16850	Lanosterol 14-alpha demethylase	CYP51A1	-	-	-	-	10,211	-			
K7EQ02	DAZ-associated protein 1	DAZAP1	-	-	-	-	23,727	-			
P46060	Ran GTPase-activating protein 1	RANGAP1	-	50,824	-	-	21,213	-	Y		
P35659	Protein DEK	DEK	-	32,146	-	-	-	-			Y
Q13263	Transcription intermediary factor 1-beta	TRIM28	1,113	49,035	737	-	12,915	387	Y		
P27816	Microtubule-associated protein 4	MAP4	-	42,383	4,797	-	-	-	Y	Y	
P61009	Signal peptidase complex subunit 3	SPCS3	-	15,445	-	-	17,500	-			
B4DJ81	NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial	NDUFS1	-	14,969	-	-	14,089	-			
O75569	Interferon-inducible double-stranded RNA-dependent protein kinase activator A	PRKRA	-	45,428	-	-	53,810	-			
Q9NZI8	Insulin-like growth factor 2 mRNA-binding protein 1	IGF2BP1	7,184	50,988	-	9,292	82,346	-	Y		Y
P08621	U1 small nuclear ribonucleoprotein 70 kDa	SNRNP70	-	12,217	-	-	11,273	-			

Q5QPK2	Dolichol-phosphate mannosyltransferase subunit 1	DPM1	10,828	17,698	-	-	28,497	-			
A8MZF9	Developmentally-regulated GTP-binding protein 2	DRG2	-	-	-	-	16,593	-			
Q9BQA1	Methylosome protein 50	WDR77	-	48,872	10,688	21,596	74,476	-	Y		Y
Q13011	Delta	ECH1	-	32,141	-	12,909	40,901	-	Y		
Q16401	26S proteasome non-ATPase regulatory subunit 5	PSMD5	22,474	45,667	21,228	21,415	48,171	12,248	Y		
P24539	ATP synthase F	ATP5F1	-	11,257	-	-	21,721	-			
Q9Y2Q3	Glutathione S-transferase kappa 1	GSTK1	-	36,207	-	23,367	149,136	-	Y		Y
O60832	H/ACA ribonucleoprotein complex subunit 4	DKC1	-	20,807	-	-	5,123	-			Y
O95239	Chromosome-associated kinesin KIF4A	KIF4A	793	40,044	-	2,394	189,978	2,015	Y	Y	Y
O43143	Pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	DHX15	1,318	43,502	1,914	1,519	31,423	-	Y		Y
P23246	Splicing factor, proline- and glutamine-rich	SFPQ	698	44,513	-	-	39,108	-	Y		
B3KS98	Eukaryotic translation initiation factor 3 subunit H	EIF3S3	-	126,785	-	5,322	90,168	-			
P55735	Protein SEC13 homolog	SEC13	-	10,735	-	-	42,561	-			
Q86U42	Polyadenylate-binding protein 2	PABPN1	9,960	30,344	-	-	35,502	-	Y		
Q8NC51	Plasminogen activator inhibitor 1 RNA-binding protein	SERBP1	-	33,267	-	6,947	32,320	-	Y		
Q08211	ATP-dependent RNA helicase A	DHX9	874	41,100	1,709	1,391	60,434	1,277	Y		Y
Q9H3N1	Thioredoxin-related transmembrane protein 1	TMX1	-	36,504	-	-	44,208	7,477	Y		
Q9BS26	Endoplasmic reticulum resident protein 44	ERP44	17,941	77,026	-	-	51,309	-			
G8JLD5	Dynamin-1-like protein	DNM1L	-	32,110	-	-	28,519	-			
P22087	rRNA 2'-O-methyltransferase fibrillarin	FBL	9,468	28,531	-	-	18,446	-			
O43765	Small glutamine-rich tetratricopeptide repeat-containing protein alpha	SGTA	-	32,484	-	-	8,734	-	Y	Y	
G3V1C3	Apoptosis inhibitor 5	API5	-	24,996	-	-	16,134	-			
P55010	Eukaryotic translation initiation factor 5	EIF5	13,237	37,623	4,526	-	32,827	6,052	Y		Y
O43592	Exportin-T	XPOT	11,717	37,622	12,662	10,001	19,392	9,652	Y		
P31942	Heterogeneous nuclear ribonucleoprotein H3	HNRNPH3	-	15,412	5,047	4,385	155,996	-	Y		Y
Q00059	Transcription factor A, mitochondrial	TFAM	-	9,404	-	-	39,998	-			
P51116	Fragile X mental retardation syndrome-related protein 2	FXR2	-	-	-	-	6,757	-			
O60256	Phosphoribosyl pyrophosphate synthase-associated protein 2	PRPSAP2	5,208	33,399	9,738	5,486	34,182	-	Y		
P33992	DNA replication licensing factor MCM5	MCM5	-	24,443	-	-	13,108	-	Y		
P26583	High mobility group protein B2	HMGB2	-	21,689	-	-	14,704	-	Y		

K7ERP4	Glutathione peroxidase	GPX4	-	74,323	-	-	-	-			
Q9HAV7	GrpE protein homolog 1, mitochondrial	GRPEL1	-	26,471	-	14,669	40,670	-			
Q9NZ01	Very-long-chain enoyl-CoA reductase	TECR	-	34,148	-	-	48,025	-	Y		Y
Q9BY44	Eukaryotic translation initiation factor 2A	EIF2A	-	28,077	-	-	19,342	-	Y		Y
P13807	Glycogen [starch] synthase, muscle	GYS1	-	8,829	-	-	-	-			
P07305	Histone H1.0	H1F0	-	2,700,825	-	-	-	-			
Q92522	Histone H1x	H1FX	6,155	364,919	-	-	12,320	-			
Q71UI9	Histone H2A.V	H2AFV	23,702	6,235,313	45,073	7,430	181,183	20,409			
Q9P0M6	Core histone macro-H2A.2	H2AFY2	-	44,892	-	-	-	-			
Q96DI7	U5 small nuclear ribonucleoprotein 40 kDa protein	SNRNP40	-	18,407	-	-	10,785	-			
B4DXZ6	Fragile X mental retardation syndrome-related protein 1	FXR1	-	32,618	-	-	15,218	-			
Q92769	Histone deacetylase 2	HDAC2	-	24,887	-	-	-	-			Y
P30419	Glycylpeptide N-tetradecanoyltransferase 1	NMT1	10,568	36,138	4,512	-	35,922	-	Y		Y
Q93077	Histone H2A type 1-C	HIST1H2AC	138,759	2,051,667	-	-	306,244	-			
Q9Y5B9	FACT complex subunit SPT16	SUPT16H	-	45,566	2,141	-	7,528	-	Y		Y
E9PB90	Hexokinase-2	HK2	3,071	7,850	-	-	12,271	-			
Q8TCT9	Minor histocompatibility antigen H13	HM13	20,247	71,534	-	5,002	30,243	-			
P18754	Regulator of chromosome condensation	RCC1	-	7,994	-	-	8,557	-			Y
D6R9P3	Heterogeneous nuclear ribonucleoprotein A/B	HNRNPAB	12,105	28,756	9,931	10,951	77,165	4,427			
O14979	Heterogeneous nuclear ribonucleoprotein D-like	HNRNPDL	-	-	-	15,005	46,669	-			
P24752	Acetyl-CoA acetyltransferase, mitochondrial	ACAT1	-	30,029	-	6,834	56,051	-	Y		Y
Q13057	Bifunctional coenzyme A synthase	COASY	13,049	29,190	6,384	7,184	34,524	-	Y		
Q13501	Sequestosome-1	SQSTM1	-	18,171	-	-	14,288	-	Y		
Q1KMD3	Heterogeneous nuclear ribonucleoprotein U-like protein 2	HNRNPUL2	-	18,744	-	-	9,982	-			Y
Q99714	3-hydroxyacyl-CoA dehydrogenase type-2	HSD17B10	-	83,143	-	5,701	131,855	-			Y
Q9NVJ2	ADP-ribosylation factor-like protein 8B	ARL8B	-	7,470	-	-	8,815	-			
P13861	cAMP-dependent protein kinase type II-alpha regulatory subunit	PRKAR2A	-	26,666	-	17,236	25,560	-	Y		
Q92688	Acidic leucine-rich nuclear phosphoprotein 32 family member B	ANP32B	-	21,388	-	-	20,475	-			
P16435	NADPH--cytochrome P450 reductase	POR	-	22,702	-	-	18,011	-	Y		
A0A087WTA5	Translation initiation factor eIF-2B subunit delta	EIF2B4	-	20,829	-	-	12,745	-			

Q92616	eIF-2-alpha kinase activator GCN1	GCN1	2,300	23,503	1,083	7,902	54,563	2,445	Y		
Q7Z6Z7	E3 ubiquitin-protein ligase HUWE1	HUWE1	2,901	22,065	1,474	6,162	23,603	3,298	Y		Y
F8WDV0	Importin-11	IPO11	-	9,894	-	-	-	-			
O60341	Lysine-specific histone demethylase 1A	KDM1A	-	1,873	-	-	-	-			
O75534	Cold shock domain-containing protein E1	CSDE1	-	18,213	5,208	5,653	23,714	1,398	Y		
Q14258	E3 ubiquitin/ISG15 ligase TRIM25	TRIM25	-	23,562	-	-	27,226	-	Y		Y
P32322	Pyrroline-5-carboxylate reductase 1, mitochondrial	PYCR1	-	18,470	-	-	65,404	-			
P52272	Heterogeneous nuclear ribonucleoprotein M	HNRNPM	-	23,388	350	2,670	34,237	583	Y		Y
O60749	Sorting nexin-2	SNX2	-	22,464	-	-	-	-	Y		
Q8NBP7	Proprotein convertase subtilisin/kexin type 9	PCSK9	-	23,475	-	-	21,675	-			Y
P04181	Ornithine aminotransferase, mitochondrial	OAT	-	27,136	-	3,402	38,492	-	Y		Y
K7EJE8	Lon protease homolog, mitochondrial	LONP1	-	15,537	-	4,403	4,816	-			
P46977	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3A	STT3A	-	22,365	-	4,788	39,050	-	Y		
P53350	Serine/threonine-protein kinase PLK1	PLK1	3,604	21,875	429	2,893	253,333	2,589	Y	Y	Y
Q07666	KH domain-containing, RNA-binding, signal transduction-associated protein 1	KHDRBS1	1,095	15,580	-	595	13,160	-	Y		
Q9Y383	Putative RNA-binding protein Luc7-like 2	LUC7L2	-	10,891	-	-	3,501	-	Y		
Q9UJS0	Calcium-binding mitochondrial carrier protein Aralar2	SLC25A13	-	19,624	-	-	12,016	-			
A0A0R4J2E8	Matrin-3	MATR3	-	1,964	-	121	1,857	-			
Q9UBT2	SUMO-activating enzyme subunit 2	UBA2	-	21,216	-	-	6,990	-	Y		
Q9P035	Very-long-chain	HACD3	-	13,040	-	-	40,484	-			
K7EIG1	Clustered mitochondria protein homolog	CLUH	-	16,203	-	-	15,360	-			
Q15393	Splicing factor 3B subunit 3	SF3B3	1,519	18,226	580	8,954	73,369	1,698	Y		Y
Q8WXF1	Paraspeckle component 1	PSPC1	-	22,385	-	-	3,245	-			
Q14696	LDLR chaperone MESD	MESDC2	-	7,552	-	5,779	17,311	-			
Q96HS1	Serine/threonine-protein phosphatase PGAM5, mitochondrial	PGAM5	-	18,403	-	-	22,274	-			Y
Q13724	Mannosyl-oligosaccharide glucosidase	MOGS	-	-	-	-	10,587	-			
J3KSI4	Mannose-P-dolichol utilization defect 1 protein	MPDU1	-	97,751	-	20,338	129,369	-			
E9PE17	28S ribosomal protein S17, mitochondrial	MRPS17	-	-	-	-	10,172	-			
O60568	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3	PLOD3	-	20,569	-	-	21,197	-	Y		
B4DHE8	RNA-binding protein Musashi homolog 2	MSI2	-	-	-	-	19,402	-			

Q9Y6C9	Mitochondrial carrier homolog 2	MTCH2	-	-	-	-	38,853	-			Y
P35637	RNA-binding protein FUS	FUS	4,148	17,101	-	6,866	63,974	-	Y		
C9JW55	N-alpha-acetyltransferase 10	NAA10	-	44,907	-	-	-	-			
Q9UNH7	Sorting nexin-6	SNX6	-	13,111	-	4,648	8,939	-	Y		
Q9H0A0	RNA cytidine acetyltransferase	NAT10	-	5,050	-	-	-	-			
O95347	Structural maintenance of chromosomes protein 2	SMC2	-	18,589	-	-	-	-	Y		
Q9GZP4	PITH domain-containing protein 1	PITHD1	-	4,698	-	-	7,115	-			
P78527	DNA-dependent protein kinase catalytic subunit	PRKDC	2,670	17,022	2,047	2,101	6,118	3,531	Y	Y	Y
Q9BPW8	Protein NipSnap homolog 1	NIPSNAP1	-	-	-	5,424	58,285	-			
P42224	Signal transducer and activator of transcription 1-alpha/beta	STAT1	-	10,910	-	-	-	-	Y		
B0QYK0	RNA-binding protein EWS	EWSR1	-	13,098	-	-	15,524	-			
P78347	General transcription factor II-I	GTF2I	-	10,453	-	-	6,093	-			
Q9Y2X3	Nucleolar protein 58	NOP58	-	8,648	-	-	-	-			Y
Q6P988	Palmitoleoyl-protein carboxylesterase NOTUM	NOTUM	-	84,399	-	-	117,370	2,497			Y
Q93009	Ubiquitin carboxyl-terminal hydrolase 7	USP7	-	4,336	-	-	7,254	-			
Q5SQP8	C-terminal-binding protein 2	CTBP2	-	4,256	-	-	5,674	-			
F8W726	Ubiquitin-associated protein 2-like	UBAP2L	-	7,090	-	-	4,249	-			
P07954	Fumarate hydratase, mitochondrial	FH	12,980	5,184	-	-	28,239	-	Y		Y
A0A0D9SFS3	2-oxoglutarate dehydrogenase, mitochondrial	OGDH	-	3,042	-	-	-	-			
F8VUA7	Oxysterol-binding protein	OSBPL8	-	7,652	-	-	-	-			
P13674	Prolyl 4-hydroxylase subunit alpha-1	P4HA1	-	15,366	-	4,436	19,818	-			
B1ANR0	Polyadenylate-binding protein	PABPC4	3,011	24,423	-	2,317	29,791	-			
Q9NR30	Nucleolar RNA helicase 2	DDX21	-	11,663	-	-	14,160	454	Y		Y
P09874	Poly [ADP-ribose] polymerase 1	PARP1	2,201	85,855	383	342	3,273	584			
H0YML5	Phosphoenolpyruvate carboxykinase [GTP], mitochondrial	PCK2	-	-	-	-	6,965	-			
O00139	Kinesin-like protein KIF2A	KIF2A	-	11,375	-	-	84,425	-	Y		
P36957	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	DLST	-	6,826	-	-	9,612	-			
Q14566	DNA replication licensing factor MCM6	MCM6	963	12,471	-	-	-	-	Y		Y
Q8N1G4	Leucine-rich repeat-containing protein 47	LRRC47	4,675	16,375	-	6,952	25,130	2,506	Y		Y
Q5T4S7	E3 ubiquitin-protein ligase UBR4	UBR4	957	14,787	1,097	112	14,128	595	Y		Y

P33993	DNA replication licensing factor MCM7	MCM7	-	10,345	-	-	2,588	-	Y		
P60891	Ribose-phosphate pyrophosphokinase 1	PRPS1	8,128	19,149	4,491	13,528	40,058	-	Y		
A0A0G2JL47	Large proline-rich protein BAG6	BAG6	-	6,654	-	-	8,935	-			
Q14694	Ubiquitin carboxyl-terminal hydrolase 10	USP10	-	7,734	-	-	2,657	-			
P46821	Microtubule-associated protein 1B	MAP1B	-	8,265	-	-	7,986	-			
O00505	Importin subunit alpha-4	KPNA3	3,098	14,650	-	-	-	-	Y		
K7ELL7	Glucosidase 2 subunit beta	PRKCSH	7,524	71,687	-	32,770	127,347	-			
Q07065	Cytoskeleton-associated protein 4	CKAP4	-	10,001	-	-	7,235	-	Y		
Q96PK6	RNA-binding protein 14	RBM14	-	9,529	-	-	-	-	Y		
P38919	Eukaryotic initiation factor 4A-III	EIF4A3	-	15,600	-	-	25,195	-	Y		Y
P49748	Very long-chain specific acyl-CoA dehydrogenase, mitochondrial	ACADVL	4,870	12,422	-	-	47,413	-	Y		
O95831	Apoptosis-inducing factor 1, mitochondrial	AIFM1	-	7,241	-	3,008	38,880	-	Y		Y
P48735	Isocitrate dehydrogenase [NADP], mitochondrial	IDH2	-	3,602	-	-	61,780	-	Y		Y
Q6P2E9	Enhancer of mRNA-decapping protein 4	EDC4	-	2,779	-	-	30,513	-			Y
P40939	Trifunctional enzyme subunit alpha, mitochondrial	HADHA	-	14,158	-	-	6,525	-	Y		Y
Q14980	Nuclear mitotic apparatus protein 1	NUMA1	-	5,500	-	-	2,367	-		Y	Y
C9J2Y9	DNA-directed RNA polymerase subunit beta	POLR2B	-	5,749	-	-	6,420	-			
O00203	AP-3 complex subunit beta-1	AP3B1	2,179	11,829	1,582	-	15,357	-	Y		
Q15029	116 kDa U5 small nuclear ribonucleoprotein component	EFTUD2	-	12,996	-	1,281	4,100	-	Y		
Q14498	RNA-binding protein 39	RBM39	-	21,088	-	-	-	-			
Q9Y5S9	RNA-binding protein 8A	RBM8A	-	-	-	-	44,186	-			
Q8TCJ2	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3B	STT3B	-	4,284	-	-	11,463	-			
Q9P258	Protein RCC2	RCC2	540	57,367	-	13,003	69,073	-		Y	
Q14444	Caprin-1	CAPRIN1	1,331	9,346	1,562	2,579	6,117	-	Y		
H3BPG5	RNA binding protein S1, serine-rich domain, isoform CRA_c	RNPS1	-	-	-	-	30,282	-			
P27694	Replication protein A 70 kDa DNA-binding subunit	RPA1	-	26,469	-	-	-	-			Y
Q13045	Protein flightless-1 homolog	FLII	-	13,666	-	3,043	9,649	2,060	Y		Y
P49736	DNA replication licensing factor MCM2	MCM2	-	6,531	-	-	2,540	-	Y		
P42167	Lamina-associated polypeptide 2, isoforms beta/gamma	TMPO	-	1,801	-	-	5,318	-			
O76021	Ribosomal L1 domain-containing protein 1	RSL1D1	-	42,735	-	-	-	-			

M0QZS6	SUMO-activating enzyme subunit 1	SAE1	-	63,444	25,589	-	28,512	-			
O75475	PC4 and SFRS1-interacting protein	PSIP1	-	12,800	-	-	-	-	Y		
Q9NYU2	UDP-glucose:glycoprotein glucosyltransferase 1	UGGT1	-	5,329	-	-	5,183	-			
P42704	Leucine-rich PPR motif-containing protein, mitochondrial	LRPPRC	575	6,264	252	668	10,773	1,982	Y		Y
Q14563	Semaphorin-3A	SEMA3A	-	-	-	-	6,622	-			Y
P19525	Interferon-induced, double-stranded RNA-activated protein kinase	EIF2AK2	-	8,783	-	-	13,238	-	Y		
P07093	Glia-derived nexin	SERPINE2	-	35,099	-	-	93,379	5,368			
O75533	Splicing factor 3B subunit 1	SF3B1	6,760	35,059	2,315	2,225	44,004	5,462			Y
E9PPJ0	Splicing factor 3B subunit 2	SF3B2	-	7,321	-	-	1,700	-			
Q96AE4	Far upstream element-binding protein 1	FUBP1	-	6,238	-	719	8,062	-	Y		
P98160	Basement membrane-specific heparan sulfate proteoglycan core protein	HSPG2	-	4,065	-	-	3,117	-			Y
Q6PKG0	La-related protein 1	LARP1	-	5,240	-	-	3,272	-	Y		
P42285	Superkiller viralicidic activity 2-like 2	SKIV2L2	-	22,882	-	-	10,197	7,348			
P53007	Tricarboxylate transport protein, mitochondrial	SLC25A1	-	-	-	-	14,260	-			
Q16891	MICOS complex subunit MIC60	IMMT	-	7,757	1,164	1,565	4,611	-	Y		
Q13423	NAD	NNT	-	2,904	-	-	4,009	-			Y
Q8NBF2	NHL repeat-containing protein 2	NHLRC2	-	3,050	-	1,115	4,726	927	Y		
Q12931	Heat shock protein 75 kDa, mitochondrial	TRAP1	-	16,906	-	-	28,223	-	Y		
Q8TEQ6	Gem-associated protein 5	GEMIN5	-	4,360	-	-	-	-	Y		
O60264	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5	SMARCA5	6,515	3,033	-	-	4,683	-			
P25205	DNA replication licensing factor MCM3	MCM3	-	2,806	-	-	3,102	-	Y		
E9PD53	Structural maintenance of chromosomes protein	SMC4	-	8,965	-	-	-	-			
P48681	Nestin	NES	-	1,000	-	-	547	-	Y		
P35573	Glycogen debranching enzyme	AGL	-	-	-	-	3,476	-	Y		
J3QQY2	Calcium load-activated calcium channel	TMCO1	-	91,099	-	-	100,179	-			
O75947	ATP synthase subunit d, mitochondrial	ATP5H	-	-	-	-	45,913	-	Y		
B1AHD1	NHP2-like protein 1	SNU13	-	57,929	-	-	-	-			
Q8NBJ5	Procollagen galactosyltransferase 1	COLGALT1	-	-	-	-	5,340	-	Y		
O43237	Cytoplasmic dynein 1 light intermediate chain 2	DYNC1LI2	-	-	-	-	19,417	-	Y	Y	Y
H0YMF9	WD repeat-containing protein 61	WDR61	-	22,356	-	-	64,369	-			

Q5JRI1	Serine/arginine-rich-splicing factor 10	SRSF10	-	35,437	-	-	31,651	-			
P30084	Enoyl-CoA hydratase, mitochondrial	ECHS1	-	-	-	-	30,009	-	Y		
P84090	Enhancer of rudimentary homolog	ERH	-	-	-	-	75,490	-	Y		
Q9Y5M8	Signal recognition particle receptor subunit beta	SRPRB	4,646	29,089	15,428	-	43,731	-			Y
Q9BXP5	Serrate RNA effector molecule homolog	SRRT	-	15,145	-	-	-	-			
P62995	Transformer-2 protein homolog beta	TRA2B	-	11,578	-	-	31,037	-			
J3QL05	Serine/arginine-rich-splicing factor 2	SRSF2	-	-	-	-	136,751	-			
F8VU90	Peptidylprolyl isomerase	FKBP11	-	20,954	-	-	54,630	-			
Q13242	Serine/arginine-rich splicing factor 9	SRSF9	-	-	-	-	31,811	-			
P05455	Lupus La protein	SSB	3,956	70,007	1,825	-	38,650	2,220			
C9J3L8	Translocon-associated protein subunit alpha	SSR1	-	22,776	4,403	-	39,847	-			
Q9BTV4	Transmembrane protein 43	TMEM43	-	6,598	-	-	16,865	-			
Q9UN86	Ras GTPase-activating protein-binding protein 2	G3BP2	-	-	-	-	16,290	-	Y		
O95793	Double-stranded RNA-binding protein Staufen homolog 1	STAU1	-	-	-	-	14,123	-			
Q9UJZ1	Stomatin-like protein 2, mitochondrial	STOML2	10,405	62,299	11,502	-	54,927	-			Y
Q00341	Vigilin	HDLBP	-	6,883	2,228	-	12,642	-	Y		
Q92945	Far upstream element-binding protein 2	KHSRP	-	1,591	-	1,184	25,774	-	Y		
Q14764	Major vault protein	MVP	-	-	-	-	9,658	-	Y		Y
O75396	Vesicle-trafficking protein SEC22b	SEC22B	-	-	-	-	23,200	-	Y		
Q5T8U5	Surfeit 4	SURF4	8,741	57,508	-	-	47,174	-			
Q8NEM2	SHC SH2 domain-binding protein 1	SHCBP1	-	-	-	-	23,369	-	Y		
P00387	NADH-cytochrome b5 reductase 3	CYB5R3	-	26,505	-	-	68,053	-	Y		
P09622	Dihydrolipoyl dehydrogenase, mitochondrial	DLD	-	9,747	-	-	16,663	-	Y		
Q9H307	Pinin	PNN	-	11,811	-	-	8,376	-			
E9PHA6	DNA mismatch repair protein	MSH2	-	5,524	-	-	6,602	-			
E7EQ72	Transmembrane emp24 domain-containing protein 2	TMED2	-	-	-	-	28,216	-			
O76024	Wolframin	WFS1	-	3,320	-	-	13,877	-			
F8W930	Insulin-like growth factor 2 mRNA-binding protein 2	IGF2BP 2	-	6,518	-	-	10,777	-			
Q92804	TATA-binding protein-associated factor 2N	TAF15	-	-	-	-	78,063	-	Y		Y
F5H7V9	Tenascin	TNC	-	-	-	-	6,919	-			

E9PDE8	Heat shock 70 kDa protein 4L	HSPA4L	-	3,508	-	-	10,579	-			
Q13148	TAR DNA-binding protein 43	TARDBP	-	7,698	-	8,158	43,250	-	Y		
O94826	Mitochondrial import receptor subunit TOM70	TOMM70	-	11,780	-	-	-	-			
P11388	DNA topoisomerase 2-alpha	TOP2A	-	14,086	488	1,667	2,387	-			Y
Q9UBS4	DnaJ homolog subfamily B member 11	DNAJB11	-	19,713	-	-	19,809	-	Y		
O00463	TNF receptor-associated factor 5	TRAF5	-	7,792	-	5,746	133,382	1,817			
O96008	Mitochondrial import receptor subunit TOM40 homolog	TOMM40	-	-	-	-	26,187	-	Y		
B5MBZ0	Echinoderm microtubule-associated protein-like 4	EML4	-	6,904	-	-	5,233	-			
P0DN76	Splicing factor U2AF 35 kDa subunit-like protein	U2AF1L5	-	89,469	11,168	-	60,371	-			
K7ENG2	Splicing factor U2AF 65 kDa subunit	U2AF2	16,353	54,726	-	-	28,231	4,994			
P07099	Epoxide hydrolase 1	EPHX1	-	12,373	-	-	7,585	-			
P20700	Lamin-B1	LMNB1	-	11,899	-	-	2,083	-	Y	Y	
H3BPC4	SUMO-conjugating enzyme UBC9	UBE2I	36,332	168,170	61,076	-	368,567	105,650			
C9JDR0	Sterol-4-alpha-carboxylate 3-dehydrogenase, decarboxylating	NSDHL	-	22,313	-	-	10,030	-			
Q13085	Acetyl-CoA carboxylase 1	ACACA	-	638	-	-	1,118	-	Y		
H3BRG4	Cytochrome b-c1 complex subunit 2, mitochondrial	UQCRC2	-	-	-	-	21,433	-			
P08559	Pyruvate dehydrogenase E1 component subunit alpha, somatic form, mitochondrial	PDHA1	-	6,317	-	-	9,010	-			Y
Q53GS9	U4/U6.U5 tri-snRNP-associated protein 2	USP39	-	16,588	-	-	-	-			
P40937	Replication factor C subunit 5	RFC5	-	6,000	-	-	5,785	-			
P10620	Microsomal glutathione S-transferase 1	MGST1	-	18,671	-	-	23,301	-	Y		
Q9H269	Vacuolar protein sorting-associated protein 16 homolog	VPS16	-	-	-	-	7,475	-			
P31930	Cytochrome b-c1 complex subunit 1, mitochondrial	UQCRC1	-	-	-	-	21,440	-	Y		
Q9P2E9	Ribosome-binding protein 1	RRBP1	-	2,240	-	-	4,977	-	Y		
P09012	U1 small nuclear ribonucleoprotein A	SNRPA	-	19,567	-	-	44,080	-			
Q5T6H7	Xaa-Pro aminopeptidase 1	XPNPEP1	-	-	-	4,731	16,854	-			

A = Co-identified proteins with Flemmingsomes from Addi et al.

B = Co-identified proteins with MiCroKiTS-midbody database

C = Co-identified proteins with pulled down sMB-Rs[70].

Y = Co-identified proteins

Appendix Table 3.8 KEGG pathway analysis of Exos, sMVs and sMB-Rs

EV type	Compare with	ID	Description	pvalue	Count
Exos	sMVs	hsa04144	Endocytosis	9.74E-06	9
		hsa04520	Adherens junction	4.72E-05	5
		hsa05165	Human papillomavirus infection	8.41E-05	9
		hsa04130	SNARE interactions in vesicular transport	0.000837	3
		hsa04514	Cell adhesion molecules	0.001489	5
	sMB-Rs	hsa04144	Endocytosis	2.52E-21	41
		hsa04810	Regulation of actin cytoskeleton	5.44E-19	36
		hsa04520	Adherens junction	4.13E-18	22
		hsa04530	Tight junction	7.67E-17	30
		hsa05130	Pathogenic Escherichia coli infection	6.01E-15	30
		hsa05100	Bacterial invasion of epithelial cells	6.68E-15	20
		hsa04014	Ras signaling pathway	1.81E-11	28
		hsa05163	Human cytomegalovirus infection	4.91E-11	27
		hsa05132	Salmonella infection	9.93E-11	28
		hsa04670	Leukocyte transendothelial migration	1.5E-10	19
		hsa05205	Proteoglycans in cancer	1.9E-10	25
		hsa04015	Rap1 signaling pathway	3.21E-10	25
		hsa05135	Yersinia infection	3.76E-09	19
		hsa05165	Human papillomavirus infection	4E-09	30
		hsa04510	Focal adhesion	1.92E-08	22
		hsa04611	Platelet activation	3.08E-08	17
		hsa05131	Shigellosis	4.02E-08	24
		hsa04360	Axon guidance	7.49E-08	20
		hsa05412	Arrhythmogenic right ventricular cardiomyopathy	1.13E-07	13
		hsa04371	Apelin signaling pathway	1.38E-07	17
		hsa04062	Chemokine signaling pathway	2E-07	20
		hsa04666	Fc gamma R-mediated phagocytosis	2.77E-07	14
		hsa04730	Long-term depression	4.61E-07	11
		hsa04514	Cell adhesion molecules	4.69E-07	17
		hsa05170	Human immunodeficiency virus 1 infection	9.95E-07	20
		hsa04926	Relaxin signaling pathway	1.78E-06	15
		hsa04130	SNARE interactions in vesicular transport	1.92E-06	8
		hsa04145	Phagosome	3.04E-06	16
		hsa04540	Gap junction	3.72E-06	12
		hsa05203	Viral carcinogenesis	9.11E-06	18
		hsa04724	Glutamatergic synapse	1.1E-05	13
		hsa05120	Epithelial cell signaling in Helicobacter pylori infection	1.61E-05	10
		hsa04727	GABAergic synapse	2.47E-05	11
		hsa04928	Parathyroid hormone synthesis, secretion and action	2.6E-05	12
		hsa04725	Cholinergic synapse	4.94E-05	12
		hsa04713	Circadian entrainment	5.57E-05	11
		hsa04726	Serotonergic synapse	5.88E-05	12
		hsa04151	PI3K-Akt signaling pathway	7.84E-05	23
		hsa05210	Colorectal cancer	9.84E-05	10
		hsa04512	ECM-receptor interaction	0.00012	10
		hsa04072	Phospholipase D signaling pathway	0.000173	13

		hsa04912	GnRH signaling pathway	0.00019	10
		hsa05212	Pancreatic cancer	0.000191	9
		hsa05167	Kaposi sarcoma-associated herpesvirus infection	0.000217	15
		hsa04728	Dopaminergic synapse	0.000222	12
		hsa05414	Dilated cardiomyopathy	0.000248	10
		hsa04934	Cushing syndrome	0.000273	13
		hsa04390	Hippo signaling pathway	0.00031	13
		hsa04935	Growth hormone synthesis, secretion and action	0.000352	11
		hsa04010	MAPK signaling pathway	0.000359	19
		hsa05211	Renal cell carcinoma	0.000502	8
		hsa05410	Hypertrophic cardiomyopathy	0.000683	9
		hsa05032	Morphine addiction	0.000741	9
		hsa05160	Hepatitis C	0.001074	12
		hsa01521	EGFR tyrosine kinase inhibitor resistance	0.001245	8
		hsa01522	Endocrine resistance	0.001264	9
		hsa05231	Choline metabolism in cancer	0.001264	9
		hsa04071	Sphingolipid signaling pathway	0.001364	10
		hsa05034	Alcoholism	0.001608	13
		hsa04972	Pancreatic secretion	0.001677	9
		hsa04024	cAMP signaling pathway	0.002086	14
		hsa04650	Natural killer cell mediated cytotoxicity	0.002794	10
		hsa04921	Oxytocin signaling pathway	0.002942	11
		hsa04150	mTOR signaling pathway	0.003093	11
		hsa04915	Estrogen signaling pathway	0.004069	10
		hsa05213	Endometrial cancer	0.004532	6
		hsa05416	Viral myocarditis	0.005362	6
		hsa04022	cGMP-PKG signaling pathway	0.005448	11
		hsa04916	Melanogenesis	0.00582	8
		hsa05142	Chagas disease	0.006173	8
		hsa05146	Amoebiasis	0.006173	8
		hsa04962	Vasopressin-regulated water reabsorption	0.006308	5
		hsa04723	Retrograde endocannabinoid signaling	0.006643	10
		hsa04012	ErbB signaling pathway	0.007872	7
		hsa04927	Cortisol synthesis and secretion	0.007917	6
		hsa04720	Long-term potentiation	0.00915	6
		hsa04924	Renin secretion	0.010516	6
		hsa04217	Necroptosis	0.010777	10
		hsa04140	Autophagy - animal	0.011689	9
		hsa04152	AMPK signaling pathway	0.015699	8
		hsa04971	Gastric acid secretion	0.016431	6
		hsa05133	Pertussis	0.016431	6
		hsa04919	Thyroid hormone signaling pathway	0.016433	8
		hsa04925	Aldosterone synthesis and secretion	0.016475	7
		hsa04923	Regulation of lipolysis in adipocytes	0.018364	5
sMVs	Exo	hsa04141	Protein processing in endoplasmic reticulum	5.21E-09	13
		hsa00970	Aminoacyl-tRNA biosynthesis	2.46E-06	7
		hsa03013	RNA transport	5.66E-05	9
	sMB-Rs	hsa04810	Regulation of actin cytoskeleton	3.23E-19	33
		hsa04530	Tight junction	1.72E-18	29
		hsa05100	Bacterial invasion of epithelial cells	1.71E-15	19

		hsa04520	Adherens junction	5.84E-15	18
		hsa05130	Pathogenic Escherichia coli infection	1E-14	27
		hsa05205	Proteoglycans in cancer	1.13E-11	24
		hsa05163	Human cytomegalovirus infection	1.3E-11	25
		hsa04015	Rap1 signaling pathway	1.89E-11	24
		hsa05135	Yersinia infection	7.61E-10	18
		hsa04144	Endocytosis	8.72E-10	24
		hsa04014	Ras signaling pathway	9.18E-10	23
		hsa04670	Leukocyte transendothelial migration	2.63E-09	16
		hsa04360	Axon guidance	1.16E-08	19
		hsa04371	Apelin signaling pathway	3.9E-08	16
		hsa04510	Focal adhesion	6.42E-08	19
		hsa05131	Shigellosis	7.39E-08	21
		hsa04540	Gap junction	3.83E-07	12
		hsa05132	Salmonella infection	4.17E-07	20
		hsa04611	Platelet activation	4.38E-07	14
		hsa04730	Long-term depression	5.49E-07	10
		hsa04926	Relaxin signaling pathway	7.14E-07	14
		hsa05165	Human papillomavirus infection	7.18E-07	23
		hsa05412	Arrhythmogenic right ventricular cardiomyopathy	7.38E-07	11
		hsa04062	Chemokine signaling pathway	8.39E-07	17
		hsa04666	Fc gamma R-mediated phagocytosis	1.12E-06	12
		hsa04727	GABAergic synapse	3.22E-06	11
		hsa05170	Human immunodeficiency virus 1 infection	3.32E-06	17
		hsa04725	Cholinergic synapse	5.75E-06	12
		hsa04724	Glutamatergic synapse	6.3E-06	12
		hsa04726	Serotonergic synapse	6.9E-06	12
		hsa04713	Circadian entrainment	7.53E-06	11
		hsa04514	Cell adhesion molecules	2.03E-05	13
		hsa04151	PI3K-Akt signaling pathway	2.63E-05	21
		hsa04728	Dopaminergic synapse	2.83E-05	12
		hsa05034	Alcoholism	5.43E-05	14
		hsa04928	Parathyroid hormone synthesis, secretion and action	9.78E-05	10
		hsa05032	Morphine addiction	0.000152	9
		hsa04912	GnRH signaling pathway	0.00018	9
		hsa04010	MAPK signaling pathway	0.000219	17
		hsa04935	Growth hormone synthesis, secretion and action	0.000256	10
		hsa05167	Kaposi sarcoma-associated herpesvirus infection	0.00029	13
		hsa01521	EGFR tyrosine kinase inhibitor resistance	0.000303	8
		hsa04072	Phospholipase D signaling pathway	0.00037	11
		hsa04921	Oxytocin signaling pathway	0.00052	11
		hsa05210	Colorectal cancer	0.000541	8
		hsa04934	Cushing syndrome	0.000549	11
		hsa04390	Hippo signaling pathway	0.000612	11
		hsa04512	ECM-receptor interaction	0.000631	8
		hsa05120	Epithelial cell signaling in Helicobacter pylori infection	0.000781	7
		hsa04915	Estrogen signaling pathway	0.000833	10
		hsa04024	cAMP signaling pathway	0.000847	13
		hsa04022	cGMP-PKG signaling pathway	0.001023	11
		hsa04071	Sphingolipid signaling pathway	0.001118	9

		hsa05414	Dilated cardiomyopathy	0.001124	8
		hsa04919	Thyroid hormone signaling pathway	0.001258	9
		hsa04971	Gastric acid secretion	0.001277	7
		hsa05212	Pancreatic cancer	0.001277	7
		hsa01522	Endocrine resistance	0.001286	8
		hsa04723	Retrograde endocannabinoid signaling	0.001424	10
		hsa05213	Endometrial cancer	0.001556	6
		hsa04916	Melanogenesis	0.001563	8
		hsa05203	Viral carcinogenesis	0.001628	12
		hsa04972	Pancreatic secretion	0.001665	8
		hsa04145	Phagosome	0.00174	10
		hsa04012	ErbB signaling pathway	0.002447	7
		hsa05410	Hypertrophic cardiomyopathy	0.003383	7
		hsa04924	Renin secretion	0.003785	6
		hsa05211	Renal cell carcinoma	0.003785	6
		hsa05215	Prostate cancer	0.005124	7
		hsa04261	Adrenergic signaling in cardiomyocytes	0.005401	9
		hsa04925	Aldosterone synthesis and secretion	0.005419	7
		hsa05231	Choline metabolism in cancer	0.005419	7
		hsa04130	SNARE interactions in vesicular transport	0.005443	4
		hsa04961	Endocrine and other factor-regulated calcium reabsorption	0.00572	5
		hsa04150	mTOR signaling pathway	0.006669	9
		hsa05142	Chagas disease	0.006729	7
		hsa05160	Hepatitis C	0.007236	9
		hsa05206	MicroRNAs in cancer	0.007662	14
		hsa04650	Natural killer cell mediated cytotoxicity	0.007719	8
		hsa05416	Viral myocarditis	0.009638	5
		hsa05219	Bladder cancer	0.011776	4
		hsa04927	Cortisol synthesis and secretion	0.013367	5
		hsa04720	Long-term potentiation	0.015098	5
		hsa05226	Gastric cancer	0.016001	8
		hsa05230	Central carbon metabolism in cancer	0.017965	5
sMB-Rs	Exo	hsa03010	Ribosome	2.1E-34	57
		hsa03013	RNA transport	7.66E-17	42
		hsa03040	Spliceosome	1.42E-14	35
		hsa04141	Protein processing in endoplasmic reticulum	8.99E-11	32
		hsa03050	Proteasome	3.69E-09	15
		hsa00970	Aminoacyl-tRNA biosynthesis	1.86E-08	17
		hsa05014	Amyotrophic lateral sclerosis	1.94E-08	45
		hsa05012	Parkinson disease	3.31E-08	35
		hsa05020	Prion disease	3.52E-08	37
		hsa01200	Carbon metabolism	9.11E-08	22
		hsa00020	Citrate cycle (TCA cycle)	1.18E-07	11
		hsa05016	Huntington disease	6.73E-07	37
		hsa05010	Alzheimer disease	3.92E-06	40
		hsa05022	Pathways of neurodegeneration - multiple diseases	1.52E-05	46
		hsa03015	mRNA surveillance pathway	0.000102	15
		hsa05017	Spinocerebellar ataxia	0.000107	19
		hsa01212	Fatty acid metabolism	0.000115	11
		hsa03030	DNA replication	0.000361	8

		hsa01230	Biosynthesis of amino acids	0.000362	12
		hsa04217	Necroptosis	0.000433	19
		hsa03018	RNA degradation	0.002051	11
		hsa00620	Pyruvate metabolism	0.003135	7
		hsa05169	Epstein-Barr virus infection	0.003216	20
		hsa00030	Pentose phosphate pathway	0.003505	6
	sMVs	hsa03010	Ribosome	7.89E-47	65
		hsa03040	Spliceosome	1.79E-14	33
		hsa03013	RNA transport	4.26E-11	32
		hsa04141	Protein processing in endoplasmic reticulum	5.18E-10	29
		hsa03050	Proteasome	7.44E-10	15
		hsa05012	Parkinson disease	2.14E-08	33
		hsa05016	Huntington disease	3.47E-08	37
		hsa05020	Prion disease	6.23E-08	34
		hsa05014	Amyotrophic lateral sclerosis	3.81E-07	39
		hsa05022	Pathways of neurodegeneration - multiple diseases	8.72E-06	43
		hsa05010	Alzheimer disease	9.71E-06	36
		hsa05017	Spinocerebellar ataxia	2.22E-05	19
		hsa04217	Necroptosis	9.75E-05	19
		hsa03015	mRNA surveillance pathway	0.000109	14
		hsa03030	DNA replication	0.000163	8
		hsa01212	Fatty acid metabolism	0.000207	10
		hsa00020	Citrate cycle (TCA cycle)	0.000305	7
		hsa05169	Epstein-Barr virus infection	0.000311	21
		hsa03060	Protein export	0.000435	6
		hsa00970	Aminoacyl-tRNA biosynthesis	0.000703	10
		hsa00620	Pyruvate metabolism	0.001622	7
		hsa01200	Carbon metabolism	0.002551	13
		hsa00071	Fatty acid degradation	0.003327	7

Appendix Table 3.9 Differential protein enrichment analysis of SW480-EVs and SW620-EVs (SW480-EV enriched proteins are with positive log2FC values and SW620-EVs enriched proteins are with negative log2 fold change values)

Protein Access	Description	Genename	Log2 fold change	pvalue
P47895	Aldehyde dehydrogenase family 1 member A3	ALDH1A3	26.69	4.58 E-27
P29034	Protein S100-A2	S100A2	26.03	1.61 E-31
K7EP40	Claudin	CLDN7	25.49	2.99 E-25
Q5MY95	Ectonucleoside triphosphate diphosphohydrolase 8	ENTPD8	24.24	5.46 E-27
Q9BZQ8	Protein Niban	FAM129A	24.04	2.00 E-27
P09471	Guanine nucleotide-binding protein G	GNAO1	23.69	1.03 E-25
P25445	Tumor necrosis factor receptor superfamily member 6	FAS	22.90	2.23 E-24
P21589	5'-nucleotidase	NT5E	22.86	8.98 E-05
Q8N4X5	Actin filament-associated protein 1-like 2	AFAP1L2	21.64	1.79 E-09
P10909	Clusterin	CLU	21.56	1.84 E-09
P09496	Clathrin light chain A	CLTA	20.85	3.27 E-09
Q9H4G4	Golgi-associated plant pathogenesis-related protein 1	GLIPR2	20.78	1.92 E-09
P00533	Epidermal growth factor receptor	EGFR	20.63	6.08 E-04
Q9UNE0	Tumor necrosis factor receptor superfamily member EDAR	EDAR	20.61	1.97 E-09
P21980	Protein-glutamine gamma-glutamyltransferase 2	TGM2	19.23	2.52 E-03
P14222	Perforin-1	PRF1	18.84	7.29 E-07
Q8J025	Protein APCDD1	APCDD1	18.44	7.39 E-07
Q9BX66	Sorbin and SH3 domain-containing protein 1	SORBS1	18.34	1.16 E-04
Q9BUF5	Tubulin beta-6 chain	TUBB6	16.14	1.90 E-03
H3BLV0	Complement decay-accelerating factor	CD55	16.11	2.26 E-05
P05121	Plasminogen activator inhibitor 1	SERPINE1	16.03	2.02 E-04
P08138	Tumor necrosis factor receptor superfamily member 16	NGFR	15.90	2.50 E-05
P22223	Cadherin-3	CDH3	15.80	2.50 E-05
Q5JPT2	SH3 domain-containing kinase-binding protein 1	SH3KBP1	15.36	2.17 E-03
C9JEV6	N-acetyl-D-glucosamine kinase	NAGK	14.98	3.16 E-05
P17096	High mobility group protein HMG-I/HMG-Y	HMGA1	14.75	2.02 E-03
P05362	Intercellular adhesion molecule 1	ICAM1	14.47	3.46 E-04
J3KN16	KIAA0368	KIAA0368	14.19	1.58 E-03
P35998	26S proteasome regulatory subunit 7	PSMC2	14.05	3.96 E-04
O00629	Importin subunit alpha-3	KPNA4	13.79	2.13 E-04

E9PEP6	Neuropilin	NRP1	13.70	2.99 E-04
Q9Y281	Cofilin-2	CFL2	13.67	1.22 E-02
Q9Y295	Developmentally-regulated GTP-binding protein 1	DRG1	13.35	1.31 E-02
R4GMR5	26S proteasome non-ATPase regulatory subunit 8	PSMD8	13.32	3.30 E-04
F8W6X9	Protein sidekick-1	SDK1	13.32	3.03 E-04
P20618	Proteasome subunit beta type-1	PSMB1	13.26	4.99 E-04
Q96L35	EPH receptor B4, isoform CRA_b	EPHB4	13.15	3.34 E-04
Q04721	Neurogenic locus notch homolog protein 2	NOTCH2	13.13	3.06 E-04
P36543	V-type proton ATPase subunit E 1	ATP6V1E 1	13.08	8.42 E-03
Q969P0	Immunoglobulin superfamily member 8	IGSF8	13.05	4.27 E-02
P31350	Ribonucleoside-diphosphate reductase subunit M2	RRM2	12.95	1.59 E-02
A6XND1	Insulin-like growth factor binding protein 3 isoform b	IGFBP3	12.88	3.24 E-04
Q9NQG5	Regulation of nuclear pre-mRNA domain-containing protein 1B	RPRD1B	12.81	3.12 E-04
I3NI44	TOM1-like protein 1	TOM1L1	12.80	1.53 E-02
Q13576	Ras GTPase-activating-like protein IQGAP2	IQGAP2	12.74	7.71 E-04
Q9BXJ9	N-alpha-acetyltransferase 15, NatA auxiliary subunit	NAA15	12.72	1.63 E-02
E1P5H9	Solute carrier organic anion transporter family member	SLCO4A1	12.65	3.14 E-04
Q5HYB6	Epididymis luminal protein 189	DKFZp68 6J1372	12.56	1.84 E-02
P02649	Apolipoprotein E	APOE	12.54	3.16 E-04
Q9UBI1	COMM domain-containing protein 3	COMMD3	12.54	1.08 E-02
Q9NRW1	Ras-related protein Rab-6B	RAB6B	12.38	3.19 E-04
F5GWX2	Heme-binding protein 1	HEBP1	12.34	1.93 E-02
P50453	Serpin B9	SERPINB 9	12.28	3.20 E-04
B1APY4	Receptor tyrosine kinase-like orphan receptor 2, isoform CRA_b	ROR2	12.25	3.21 E-04
P61201	COP9 signalosome complex subunit 2	COPS2	12.21	1.24 E-02
P30711	Glutathione S-transferase theta-1	GSTT1	11.75	3.34 E-04
C9J0K6	Sorcin	SRI	11.75	8.06 E-04
P11387	DNA topoisomerase 1	TOP1	11.58	4.35 E-02
J3KNF8	Cytochrome b5 type B	CYB5B	11.54	3.22 E-03
Q8NFJ5	Retinoic acid-induced protein 3	GPRC5A	11.52	4.82 E-02
P04114	Apolipoprotein B-100	APOB	11.24	2.46 E-02
E7ETK0	40S ribosomal protein S24	RPS24	11.17	3.19 E-03
F8W9W2	Protein-serine/threonine kinase	GRK6	11.08	4.72 E-02
P27816	Microtubule-associated protein 4	MAP4	10.92	2.17 E-03
E9PPJ5	Midkine	MDK	10.88	2.19 E-03
E9PGT1	Translin	TSN	10.83	1.84 E-03
Q13347	Eukaryotic translation initiation factor 3 subunit I	EIF3I	10.79	3.01 E-03
O43242	26S proteasome non-ATPase regulatory subunit 3	PSMD3	10.74	5.12 E-03
B1AMW1	CD58 antigen,	CD58	10.66	2.18 E-03

Q12904	Aminoacyl tRNA synthase complex-interacting multifunctional protein 1	AIMP1	10.62	3.49 E-03
P20645	Cation-dependent mannose-6-phosphate receptor	M6PR	10.60	1.79 E-03
Q15785	Mitochondrial import receptor subunit TOM34	TOMM34	10.59	4.52 E-02
G3V4W0	Heterogeneous nuclear ribonucleoproteins C1/C2	HNRNPC	10.48	6.42 E-03
F8W9Z6	Syntaxin-16	STX16	10.35	2.19 E-03
Q9NY12	H/ACA ribonucleoprotein complex subunit 1	GAR1	10.22	2.20 E-03
A0A1B0GW44	Cathepsin D	CTSD	10.19	3.66 E-02
Q969L2	Protein MAL2	MAL2	10.18	2.20 E-03
Q13641	Trophoblast glycoprotein	TPBG	10.03	3.77 E-03
Q9BQI0	Allograft inflammatory factor 1-like	AIF1L	10.01	2.21 E-03
Q12860	Contactin-1	CNTN1	10.00	2.22 E-03
Q8IZW8	Tensin-4	TNS4	9.93	4.93 E-02
Q00839	Heterogeneous nuclear ribonucleoprotein U	HNRNPU	9.93	1.56 E-02
Q92905	COP9 signalosome complex subunit 5	COPS5	9.92	3.14 E-03
O00505	Importin subunit alpha-4	KPNA3	9.86	2.23 E-03
Q14566	DNA replication licensing factor MCM6	MCM6	9.82	2.27 E-03
P07305	Histone H1.0	H1F0	9.49	1.06 E-02
P15529	Membrane cofactor protein	CD46	9.48	7.04 E-03
P23470	Receptor-type tyrosine-protein phosphatase gamma	PTPRG	9.40	2.27 E-03
P02545	Prelamin-A/C [Cleaved into: Lamin-A/C	LMNA	9.03	6.05 E-03
O00468	Agrin [Cleaved into: Agrin N-terminal 110 kDa subunit; Agrin C-terminal 110 kDa subunit; Agrin C-terminal 90 kDa fragment	AGRN	8.79	1.08 E-02
P43121	Cell surface glycoprotein MUC18	MCAM	8.63	7.39 E-03
P46778	60S ribosomal protein L21	RPL21	8.50	1.75 E-02
Q13155	Aminoacyl tRNA synthase complex-interacting multifunctional protein 2	AIMP2	8.27	1.25 E-02
P46779	60S ribosomal protein L28	RPL28	8.27	1.63 E-02
Q9Y3U8	60S ribosomal protein L36	RPL36	8.11	1.61 E-02
Q9P0L0	Vesicle-associated membrane protein-associated protein A	VAPA	8.09	1.31 E-02
J3QR09	Ribosomal protein L19	RPL19	8.04	1.57 E-02
K7EM56	40S ribosomal protein S15	RPS15	7.99	1.55 E-02
O95347	Structural maintenance of chromosomes protein 2	SMC2	7.97	1.10 E-02
O76021	Ribosomal L1 domain-containing protein 1	RSL1D1	7.94	1.10 E-02
O14579	Coatomer subunit epsilon	COPE	7.90	1.12 E-02
Q9P0M6	Core histone macro-H2A.2	H2AFY2	7.83	1.11 E-02
P27694	Replication protein A 70 kDa DNA-binding subunit	RPA1	7.82	1.11 E-02
C9J4Z3	60S ribosomal protein L37a	RPL37A	7.80	1.73 E-02
P50416	Carnitine O-palmitoyltransferase 1, liver isoform	CPT1A	7.80	1.11 E-02
Q9BXP5	Serrate RNA effector molecule homolog	SRRT	7.71	1.11 E-02
C9JLU1	DNA-directed RNA polymerases I, II, and III subunit RPABC3	POLR2H	7.67	1.12 E-02
J3QR46	Uncharacterized protein KIAA0040	KIAA0040	7.66	1.12 E-02

O60749	Sorting nexin-2	SNX2	7.66	1.12 E-02
E5RK69	Annexin	ANXA6	7.66	1.12 E-02
E9PD53	Structural maintenance of chromosomes protein	SMC4	7.65	1.12 E-02
P13646	Keratin, type I cytoskeletal 13	KRT13	7.64	1.12 E-02
Q14498	RNA-binding protein 39	RBM39	7.61	1.12 E-02
Q2TAP0	Golgin subfamily A member 7B	GOLGA7 B	7.58	1.12 E-02
Q53GS9	U4/U6.U5 tri-snRNP-associated protein 2	USP39	7.56	1.12 E-02
G3XAH6	Poly	PAPOLA	7.55	1.12 E-02
P15927	Replication protein A 32 kDa subunit	RPA2	7.44	1.13 E-02
P23677	Inositol-trisphosphate 3-kinase A	ITPKA	7.41	1.13 E-02
Q13393	Phospholipase D1	PLD1	7.40	1.13 E-02
Q8WV41	Sorting nexin-33	SNX33	7.40	1.13 E-02
P31431	Syndecan-4	SDC4	7.39	1.13 E-02
Q8TEQ6	Gem-associated protein 5	GEMIN5	7.39	1.13 E-02
O75475	PC4 and SFRS1-interacting protein	PSIP1	7.38	1.13 E-02
P13807	Glycogen [starch] synthase, muscle	GYS1	7.38	1.13 E-02
P35754	Glutaredoxin-1	GLRX	7.33	1.13 E-02
Q9Y547	Intraflagellar transport protein 25 homolog	HSPB11	7.33	1.13 E-02
F8VUA7	Oxysterol-binding protein	OSBPL8	7.28	1.13 E-02
E9PKJ0	Protein wntless homolog	WLS	7.24	1.13 E-02
Q9H0A0	RNA cytidine acetyltransferase	NAT10	7.22	1.14 E-02
D6RGI5	Progressive ankylosis protein homolog	ANKH	7.17	1.14 E-02
F5GWX5	Chromodomain-helicase-DNA-binding protein 4	CHD4	6.90	1.15 E-02
P42766	60S ribosomal protein L35	RPL35	6.78	2.49 E-02
Q5T0I0	Gelsolin	GSN	6.76	1.22 E-02
P53794	Sodium/myo-inositol cotransporter	SLC5A3	6.64	3.97 E-03
C9JYN0	Synaptophysin-like protein 1	SYPL1	6.54	1.93 E-02
Q15293	Reticulocalbin-1	RCN1	6.45	4.10 E-02
A0A087W XM6	60S ribosomal protein L17	RPL17	6.29	3.88 E-02
P49720	Proteasome subunit beta type-3	PSMB3	6.26	2.07 E-02
J3QRI7	60S ribosomal protein L26	RPL26	6.09	3.31 E-02
O00487	26S proteasome non-ATPase regulatory subunit 14	PSMD14	5.97	3.22 E-02
O60487	Myelin protein zero-like protein 2	MPZL2	5.60	1.96 E-02
P47897	Glutamine--tRNA ligase	QARS	5.59	4.75 E-02
E9PS44	Interferon-induced transmembrane protein 3	IFITM3	5.57	4.54 E-02
P28074	Proteasome subunit beta type-5	PSMB5	5.40	4.82 E-02
Q9Y237	Peptidyl-prolyl cis-trans isomerase NIMA-interacting 4	PIN4	5.39	4.56 E-02
Q08J23	tRNA	NSUN2	5.36	4.56 E-02
Q9BV40	Vesicle-associated membrane protein 8	VAMP8	5.33	4.56 E-02

O94903	Pyridoxal phosphate homeostasis protein	PROSC	5.31	4.65 E-02
Q86UU1	Pleckstrin homology-like domain family B member 1	PHLDB1	5.27	4.57 E-02
Q99805	Transmembrane 9 superfamily member 2	TM9SF2	5.26	4.58 E-02
Q92769	Histone deacetylase 2	HDAC2	5.26	4.57 E-02
P35659	Protein DEK	DEK	5.24	4.58 E-02
K7ERP4	Glutathione peroxidase	GPX4	5.22	4.58 E-02
Q14156	Protein EFR3 homolog A	EFR3A	5.22	4.58 E-02
P61960	Ubiquitin-fold modifier 1	UFM1	5.21	4.58 E-02
B5MD23	Tetraspanin	TSPAN9	5.18	4.58 E-02
A0A087WTF6	Neural cell adhesion molecule 1	NCAM1	5.18	4.58 E-02
F8WDV0	Importin-11	IPO11	5.15	4.59 E-02
C9JZ87	Transmembrane protein 106B	TMEM106B	5.13	4.59 E-02
P42224	Signal transducer and activator of transcription 1-alpha/beta	STAT1	5.13	4.59 E-02
O94925	Glutaminase kidney isoform, mitochondrial	GLS	5.13	4.60 E-02
B1AHD1	NHP2-like protein 1	SNU13	5.11	4.59 E-02
E7ES10	Calpastatin	CAST	5.09	4.59 E-02
P00966	Argininosuccinate synthase	ASS1	5.09	1.70 E-02
O94826	Mitochondrial import receptor subunit TOM70	TOMM70	5.08	4.60 E-02
C9JW55	N-alpha-acetyltransferase 10	NAA10	5.06	4.60 E-02
Q96PK6	RNA-binding protein 14	RBM14	5.05	4.60 E-02
E7ETB3	Aspartyl aminopeptidase	DNPEP	5.00	4.60 E-02
Q9Y3B4	Splicing factor 3B subunit 6	SF3B6	4.99	4.61 E-02
F5GYQ1	V-type proton ATPase subunit	ATP6V0D1	4.97	4.61 E-02
Q9Y2X3	Nucleolar protein 58	NOP58	4.94	4.62 E-02
P11172	Uridine 5'-monophosphate synthase	UMPS	4.93	4.62 E-02
A0A087WYN0	Transporter	SLC6A6	4.90	4.62 E-02
M0R0V2	Deoxyhypusine synthase	DHPS	4.82	4.63 E-02
A0A0D9SFS3	2-oxoglutarate dehydrogenase, mitochondrial	OGDH	4.82	4.63 E-02
A0A024R412	Neuropilin	NRP2	4.67	4.65 E-02
O60341	Lysine-specific histone demethylase 1A	KDM1A	4.61	4.66 E-02
Q8WYA6	Beta-catenin-like protein 1	CTNNBL1	4.47	4.68 E-02
P04083	Annexin A1	ANXA1	4.40	3.22 E-02
P08133	Annexin A6	ANXA6	3.83	4.52 E-05
P16070	CD44 antigen	CD44	3.78	1.61 E-04
P43007	Neutral amino acid transporter A	SLC1A4	2.62	1.46 E-05
P11166	Solute carrier family 2, facilitated glucose transporter member 1	SLC2A1	2.32	2.84 E-02
P62424	60S ribosomal protein L7a	RPL7A	2.29	4.75 E-02
Q9NZW5	MAGUK p55 subfamily member 6	MPP6	2.29	4.97 E-06
Q9H4G0	Band 4.1-like protein 1	EPB41L1	2.25	1.10 E-04

P29373	Cellular retinoic acid-binding protein 2	CRABP2	2.23	2.63 E-05
B5MCA4	Epithelial cell adhesion molecule	EPCAM	2.17	4.16 E-02
Q15758	Neutral amino acid transporter B	SLC1A5	2.12	1.84 E-06
Q14019	Coactosin-like protein	COTL1	2.04	3.41 E-04
P09455	Retinol-binding protein 1	RBP1	2.01	4.17 E-09
P08243	Asparagine synthetase [glutamine-hydrolyzing]	ASNS	1.98	1.70 E-06
J3KQI6	Stimulated by retinoic acid gene 6 protein homolog	STRA6	1.93	1.69 E-05
P50479	PDZ and LIM domain protein 4	PDLIM4	1.88	2.05 E-05
P43490	Nicotinamide phosphoribosyltransferase	NAMPT	1.83	1.36 E-05
A0A0A0MSY4	Dedicator of cytokinesis protein 9	DOCK9	1.70	3.15 E-02
P02786	Transferrin receptor protein 1	TFRC	1.64	6.09 E-04
Q9NQW6	Anillin	ANLN	1.53	1.94 E-02
A0A0A0MS51	Gelsolin	GSN	1.52	4.36 E-09
Q9Y5Y6	Suppressor of tumorigenicity 14 protein	ST14	1.46	2.27 E-02
P23381	Tryptophan--tRNA ligase, cytoplasmic	WARS	1.46	1.05 E-05
P41091	Eukaryotic translation initiation factor 2 subunit 3	EIF2S3	1.44	2.44 E-02
J3QRU1	Tyrosine-protein kinase	YES1	1.42	6.90 E-03
H0Y7A7	Calmodulin-2	CALM2	1.35	1.88 E-05
F5H6E2	Unconventional myosin-Ic	MYO1C	1.32	4.40 E-02
O43175	D-3-phosphoglycerate dehydrogenase	PHGDH	1.31	4.61 E-04
Q9Y5X1	Sorting nexin-9	SNX9	1.30	4.86 E-07
P18084	Integrin beta-5	ITGB5	1.29	1.82 E-02
Q15008	26S proteasome non-ATPase regulatory subunit 6	PSMD6	1.29	1.65 E-02
P31949	Protein S100-A11	S100A11	1.27	5.30 E-08
P14324	Farnesyl pyrophosphate synthase	FDPS	1.24	4.10 E-05
P08648	Integrin alpha-5	ITGA5	1.23	2.08 E-02
Q92597	Protein NDRG1	NDRG1	1.20	2.33 E-05
Q13200	26S proteasome non-ATPase regulatory subunit 2	PSMD2	1.19	2.36 E-02
P61769	Beta-2-microglobulin [Cleaved into: Beta-2-microglobulin form pI 5.3]	B2M	1.19	2.19 E-02
P06756	Integrin alpha-V	ITGAV	1.19	3.07 E-02
Q92598	Heat shock protein 105 kDa	HSPH1	1.19	1.16 E-04
Q9HCM4	Band 4.1-like protein 5	EPB41L5	1.18	1.66 E-03
O15551	Claudin-3	CLDN3	1.16	1.07 E-02
A0A087WYS1	UTP--glucose-1-phosphate uridylyltransferase	UGP2	1.16	9.56 E-03
P01889	HLA class I histocompatibility antigen, B-7 alpha chain	HLA-B	1.13	4.76 E-03
P31947	14-3-3 protein sigma	SFN	1.13	1.57 E-04
P62760	Visinin-like protein 1	VSNL1	1.12	5.58 E-05
Q01650	Large neutral amino acids transporter small subunit 1	SLC7A5	1.12	8.86 E-05
Q09666	Neuroblast differentiation-associated protein AHNAK	AHNAK	1.11	4.62 E-04

P53004	Biliverdin reductase A	BLVRA	1.11	1.40 E-06
P16152	Carbonyl reductase [NADPH] 1	CBR1	1.09	8.01 E-09
P05023	Sodium/potassium-transporting ATPase subunit alpha-1	ATP1A1	1.06	5.91 E-04
Q9NR45	Sialic acid synthase	NANS	1.02	2.19 E-03
A0A087X 0K9	Tight junction protein ZO-1	TJP1	1.01	3.43 E-03
P30085	UMP-CMP kinase	CMPK1	-1.01	2.35 E-05
A0A0A0M RM8	Unconventional myosin-VI	MYO6	-1.02	3.05 E-04
O75695	Protein XRP2	RP2	-1.05	1.00 E-02
B4DUC8	S-methyl-5'-thioadenosine phosphorylase	MTAP	-1.09	3.23 E-05
Q92747	Actin-related protein 2/3 complex subunit 1A	ARPC1A	-1.10	7.04 E-03
Q14651	Plastin-1	PLS1	-1.10	1.09 E-02
P52788	Spermine synthase	SMS	-1.10	7.87 E-05
Q96N67	Dedicator of cytokinesis protein 7	DOCK7	-1.11	1.13 E-02
P23921	Ribonucleoside-diphosphate reductase large subunit	RRM1	-1.12	2.22 E-02
G5E933	Myotubularin-related protein 5	SBF1	-1.16	4.22 E-03
Q96PY5	Formin-like protein 2	FMNL2	-1.19	2.77 E-02
Q12965	Unconventional myosin-Ie	MYO1E	-1.26	3.73 E-04
O43865	S-adenosylhomocysteine hydrolase-like protein 1	AHCYL1	-1.31	6.58 E-04
P49588	Alanine--tRNA ligase, cytoplasmic	AARS	-1.31	3.02 E-04
P68032	Actin, alpha cardiac muscle 1	ACTC1	-1.41	1.16 E-03
P53992	Protein transport protein Sec24C	SEC24C	-1.41	9.65 E-05
P13797	Plastin-3	PLS3	-1.50	1.53 E-02
P11171	Protein 4.1	EPB41	-1.53	5.25 E-04
Q5JUW8	Discs, large homolog 3	DLG3	-1.58	5.72 E-03
P42345	Serine/threonine-protein kinase mTOR	MTOR	-1.64	6.62 E-04
O75955	Flotillin-1	FLOT1	-1.67	2.77 E-07
P10644	cAMP-dependent protein kinase type I-alpha regulatory subunit	PRKAR1A	-1.71	2.41 E-06
E9PFN4	Anion exchange protein	SLC4A7	-1.94	3.66 E-06
Q68CZ2	Tensin-3	TNS3	-2.03	5.61 E-04
P35580	Myosin-10	MYH10	-2.21	1.10 E-05
P61970	Nuclear transport factor 2	NUTF2	-2.38	8.49 E-06
P06239	Tyrosine-protein kinase Lck	LCK	-2.88	6.41 E-03
Q12929	Epidermal growth factor receptor kinase substrate 8	EPS8	-2.93	9.37 E-03
Q92614	Unconventional myosin-XVIIIa	MYO18A	-3.07	2.77 E-07
E9PMS6	LIM domain only protein 7	LMO7	-3.43	9.04 E-03
Q8IY22	C-Maf-inducing protein	CMIP	-3.73	7.24 E-05
O14907	Tax1-binding protein 3	TAX1BP3	-4.24	4.39 E-02
E9PE17	28S ribosomal protein S17, mitochondrial	MRPS17	-4.43	4.82 E-02
P36551	Oxygen-dependent coproporphyrinogen-III oxidase, mitochondrial	CPOX	-4.51	4.80 E-02

P22413	Ectonucleotide pyrophosphatase/phosphodiesterase family member 1	ENPP1	-4.54	4.55 E-02
H0YML5	Phosphoenolpyruvate carboxykinase [GTP], mitochondrial	PCK2	-4.55	4.79 E-02
H3BPG5	RNA binding protein S1, serine-rich domain, isoform CRA_c	RNPS1	-4.61	4.78 E-02
H3BNW8	Transmembrane channel-like protein	TMC7	-4.62	4.67 E-02
P36894	Bone morphogenetic protein receptor type-1A	BMPR1A	-4.65	4.54 E-02
P53007	Tricarboxylate transport protein, mitochondrial	SLC25A1	-4.65	4.77 E-02
E7EQ72	Transmembrane emp24 domain-containing protein 2	TMED2	-4.67	4.77 E-02
Q9Y2H1	Serine/threonine-protein kinase 38-like	STK38L	-4.68	4.54 E-02
Q13363	C-terminal-binding protein 1	CTBP1	-4.68	4.77 E-02
Q14563	Semaphorin-3A	SEMA3A	-4.68	4.77 E-02
O75396	Vesicle-trafficking protein SEC22b	SEC22B	-4.69	4.77 E-02
Q9Y2W1	Thyroid hormone receptor-associated protein 3	THRAP3	-4.70	4.66 E-02
A8MZF9	Developmentally-regulated GTP-binding protein 2	DRG2	-4.72	4.76 E-02
P10606	Cytochrome c oxidase subunit 5B, mitochondrial	COX5B	-4.73	4.76 E-02
H7C3Z2	Feline leukemia virus subgroup C receptor-related protein 1	FLVCR1	-4.74	4.53 E-02
Q9H269	Vacuolar protein sorting-associated protein 16 homolog	VPS16	-4.76	4.75 E-02
B4DHE8	RNA-binding protein Musashi homolog 2	MSI2	-4.76	4.75 E-02
P51116	Fragile X mental retardation syndrome-related protein 2	FXR2	-4.77	4.66 E-02
P27144	Adenylate kinase 4, mitochondrial	AK4	-4.78	4.75 E-02
V9GYM8	Rho guanine nucleotide exchange factor 2	ARHGEF 2	-4.79	4.63 E-02
Q13242	Serine/arginine-rich splicing factor 9	SRSF9	-4.80	4.75 E-02
Q16850	Lanosterol 14-alpha demethylase	CYP51A1	-4.80	4.65 E-02
O75947	ATP synthase subunit d, mitochondrial	ATP5H	-4.82	4.74 E-02
Q9Y5S9	RNA-binding protein 8A	RBM8A	-4.83	4.74 E-02
K7EQ02	DAZ-associated protein 1	DAZAP1	-4.83	4.74 E-02
P84090	Enhancer of rudimentary homolog	ERH	-4.83	4.74 E-02
Q9UN86	Ras GTPase-activating protein-binding protein 2	G3BP2	-4.84	4.74 E-02
O95793	Double-stranded RNA-binding protein Staufen homolog 1	STAU1	-4.85	4.74 E-02
F8VRH0	Poly	PCBP2	-4.85	4.63 E-02
Q7KYR7	Butyrophilin subfamily 2 member A1	BTN2A1	-4.86	4.52 E-02
H3BRG4	Cytochrome b-c1 complex subunit 2, mitochondrial	UQCRC2	-4.87	4.73 E-02
Q9Y4H2	Insulin receptor substrate 2	IRS2	-4.87	4.52 E-02
P31749	RAC-alpha serine/threonine-protein kinase	AKT1	-4.88	4.52 E-02
K7ELV2	Nucleoporin SEH1	SEH1L	-4.88	4.52 E-02
A0A087X 142	Septin-8	44082	-4.89	4.52 E-02
Q9UKY7	Protein CDV3 homolog	CDV3	-4.89	4.52 E-02
O96008	Mitochondrial import receptor subunit TOM40 homolog	TOMM40	-4.89	4.73 E-02
M0R300	Unconventional myosin-IXb	MYO9B	-4.90	4.52 E-02
O43237	Cytoplasmic dynein 1 light intermediate chain 2	DYNC1LI 2	-4.90	4.73 E-02

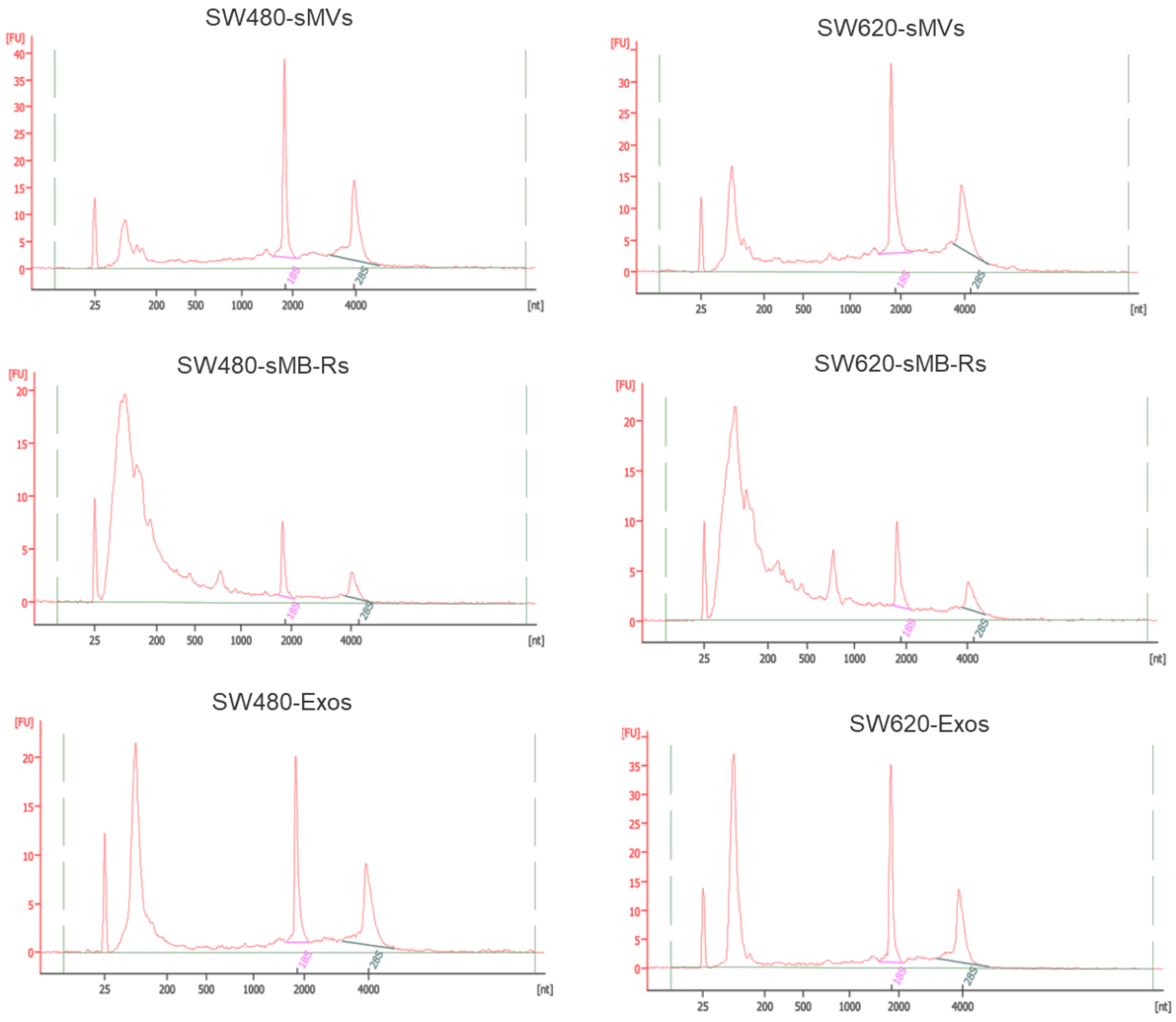
P36542	ATP synthase subunit gamma, mitochondrial	ATP5C1	-4.92	4.72 E-02
P31930	Cytochrome b-c1 complex subunit 1, mitochondrial	UQCRC1	-4.92	4.73 E-02
H0Y547	Phospholipid-transporting ATPase	ATP11A	-4.97	4.51 E-02
Q9Y6C9	Mitochondrial carrier homolog 2	MTCH2	-4.97	4.72 E-02
Q9NS86	LanC-like protein 2	LANCL2	-5.01	4.51 E-02
Q8TE67	Epidermal growth factor receptor kinase substrate 8-like protein 3	EPS8L3	-5.01	2.30 E-02
Q16762	Thiosulfate sulfurtransferase	TST	-5.03	4.62 E-02
F5H7V9	Tenascin	TNC	-5.07	4.62 E-02
J3QL05	Serine/arginine-rich-splicing factor 2	SRSF2	-5.10	4.70 E-02
P13073	Cytochrome c oxidase subunit 4 isoform 1, mitochondrial	COX4I1	-5.11	4.61 E-02
Q96HP0	Dedicator of cytokinesis protein 6	DOCK6	-5.18	1.61 E-02
Q8TEB7	E3 ubiquitin-protein ligase RNF128	RNF128	-5.21	4.49 E-02
P22059	Oxysterol-binding protein 1	OSBP	-5.37	4.87 E-02
O60547	GDP-mannose 4,6 dehydratase	GMDS	-5.47	9.05 E-03
P51812	Ribosomal protein S6 kinase alpha-3	RPS6KA3	-5.49	4.24 E-02
Q9Y6M5	Zinc transporter 1	SLC30A1	-5.52	4.71 E-02
Q8N3E9	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta-3	PLCD3	-5.61	4.84 E-02
P49755	Transmembrane emp24 domain-containing protein 10	TMED10	-5.65	4.15 E-02
Q00535	Cyclin-dependent-like kinase 5	CDK5	-5.68	3.17 E-02
P28838	Cytosol aminopeptidase	LAP3	-5.70	3.63 E-02
P16949	Stathmin	STMN1	-5.80	3.97 E-02
B1AMS2	Septin 6, isoform CRA_b	44080	-5.80	3.36 E-02
P18433	Receptor-type tyrosine-protein phosphatase alpha	PTPRA	-5.86	4.57 E-02
P48147	Prolyl endopeptidase	PREP	-5.89	2.95 E-02
Q9Y5S2	Serine/threonine-protein kinase MRCK beta	CDC42BPB	-5.95	3.30 E-02
P09417	Dihydropteridine reductase	QDPR	-5.95	3.13 E-02
E9PC69	Non-specific serine/threonine protein kinase	MARK2	-6.05	3.45 E-02
Q9UHN6	Cell surface hyaluronidase	TMEM2	-6.10	3.07 E-02
O95832	Claudin-1	CLDN1	-6.21	3.50 E-02
O15020	Spectrin beta chain, non-erythrocytic 2	SPTBN2	-6.35	1.76 E-02
Q9H3M7	Thioredoxin-interacting protein	TXNIP	-6.63	2.00 E-02
E9PDI4	Ladinin-1	LAD1	-6.66	1.37 E-02
Q8N392	Rho GTPase-activating protein 18	ARHGAP18	-6.74	1.90 E-02
Q9NUU7	ATP-dependent RNA helicase DDX19A	DDX19A	-6.78	1.29 E-02
Q8NBJ5	Procollagen galactosyltransferase 1	COLGALT1	-6.89	1.05 E-02
A0A0A0MSA9	Poliovirus receptor	PVR	-6.94	2.90 E-02
Q9Y5L0	Transportin-3	TNPO3	-7.10	2.09 E-02
P35573	Glycogen debranching enzyme	AGL	-7.12	1.04 E-02
P37173	TGF-beta receptor type-2	TGFBR2	-7.23	1.09 E-02

P30084	Enoyl-CoA hydratase, mitochondrial	ECHS1	-7.36	1.04 E-02
Q14764	Major vault protein	MVP	-7.36	1.04 E-02
Q13724	Mannosyl-oligosaccharide glucosidase	MOGS	-7.37	1.04 E-02
Q8WXX5	DnaJ homolog subfamily C member 9	DNAJC9	-7.39	1.04 E-02
B7ZKQ9	SCARB1 protein	SCARB1	-7.42	2.66 E-02
O15431	High affinity copper uptake protein 1	SLC31A1	-7.43	1.06 E-02
Q9NRX4	14 kDa phosphohistidine phosphatase	PHPT1	-7.43	1.59 E-02
Q15283	Ras GTPase-activating protein 2	RASA2	-7.44	1.67 E-02
Q9Y653	Adhesion G-protein coupled receptor G1	ADGRG1	-7.54	2.26 E-02
A0A087X 1K9	Acyl-protein thioesterase 1	LYPLA1	-7.61	1.44 E-02
Q9UMY4	Sorting nexin-12	SNX12	-7.62	1.43 E-02
M0R2P6	SH3KBP1 binding protein 1, isoform CRA_c	SHKBP1	-7.63	1.57 E-02
Q8NEM2	SHC SH2 domain-binding protein 1	SHCBP1	-7.64	1.03 E-02
Q96II5	ARAF protein	ARAF	-7.67	1.57 E-02
Q8WUX1	Sodium-coupled neutral amino acid transporter 5	SLC38A5	-7.69	2.08 E-02
Q13043	Serine/threonine-protein kinase 4	STK4	-7.73	1.08 E-02
Q9Y2L9	Leucine-rich repeat and calponin homology domain-containing protein 1	LRCH1	-7.76	1.24 E-02
P62273	40S ribosomal protein S29	RPS29	-7.77	1.03 E-02
Q4KMP7	TBC1 domain family member 10B	TBC1D10 B	-7.78	1.53 E-02
Q92973	Transportin-1	TNPO1	-7.81	1.45 E-02
O43157	Plexin-B1	PLXNB1	-7.87	1.40 E-02
P36406	E3 ubiquitin-protein ligase TRIM23	TRIM23	-7.95	1.43 E-02
P25098	Beta-adrenergic receptor kinase 1	GRK2	-7.98	8.27 E-03
Q5QPM7	Proteasome inhibitor PI31 subunit	PSMF1	-7.99	1.08 E-02
Q9Y6N7	Roundabout homolog 1	ROBO1	-8.03	1.46 E-02
Q15056	Eukaryotic translation initiation factor 4H	EIF4H	-8.09	1.34 E-02
Q8WUW1	Protein BRICK1	BRK1	-8.10	1.61 E-02
P60983	Glia maturation factor beta	GMFB	-8.16	1.49 E-02
Q92804	TATA-binding protein-associated factor 2N	TAF15	-8.17	1.02 E-02
Q15417	Calponin-3	CNN3	-8.23	5.05 E-03
O95466	Formin-like protein 1	FMNL1	-8.34	7.95 E-03
Q5TG12	Receptor-type tyrosine-protein phosphatase kappa	PTPRK	-8.37	1.30 E-02
Q5T2T1	MAGUK p55 subfamily member 7	MPP7	-8.44	8.91 E-03
Q7L9B9	Endonuclease/exonuclease/phosphatase family domain-containing protein 1	EEPD1	-8.46	6.55 E-03
Q6ZS17	Rho family-interacting cell polarization regulator 1	RIPOR1	-8.58	1.60 E-02
P61457	Pterin-4-alpha-carbinolamine dehydratase	PCBD1	-8.58	7.33 E-03
Q15274	Nicotinate-nucleotide pyrophosphorylase [carboxylating]	QPRT	-8.68	4.83 E-03
Q86YS7	C2 domain-containing protein 5	C2CD5	-8.72	5.31 E-03
Q14195	Dihydropyrimidinase-related protein 3	DPYSL3	-8.80	7.93 E-03

Q9NZQ3	NCK-interacting protein with SH3 domain	NCKIPSD	-9.03	2.17 E-03
P10586	Receptor-type tyrosine-protein phosphatase F	PTPRF	-9.03	1.11 E-02
Q92835	Phosphatidylinositol 3,4,5-trisphosphate 5-phosphatase 1	INPP5D	-9.06	3.63 E-03
Q9Y4D7	Plexin-D1	PLXND1	-9.07	8.20 E-03
P23458	Tyrosine-protein kinase JAK1	JAK1	-9.11	8.83 E-03
P08581	Hepatocyte growth factor receptor	MET	-9.35	2.11 E-02
P01034	Cystatin-C	CST3	-9.37	1.97 E-03
A0A0J9YY01	Unconventional myosin-XVB	MYO15B	-9.39	3.83 E-03
Q9H329	Band 4.1-like protein 4B	EPB41L4B	-9.49	2.14 E-03
Q8IZ21	Phosphatase and actin regulator 4	PHACTR4	-9.50	2.14 E-03
P00505	Aspartate aminotransferase, mitochondrial	GOT2	-9.57	8.08 E-03
P42126	Enoyl-CoA delta isomerase 1, mitochondrial	ECI1	-9.64	1.94 E-03
P55263	Adenosine kinase	ADK	-9.85	4.62 E-03
O15344	E3 ubiquitin-protein ligase Midline-1	MID1	-9.87	2.11 E-03
Q5T6H7	Xaa-Pro aminopeptidase 1	XPNPEP1	-9.90	1.95 E-03
Q9BVC4	Target of rapamycin complex subunit LST8	MLST8	-9.91	2.10 E-03
Q9NX57	Ras-related protein Rab-20	RAB20	-9.96	2.10 E-03
Q9BPW8	Protein NipSnap homolog 1	NIPSNAP1	-10.01	1.98 E-03
Q9H0X4	Protein FAM234A	FAM234A	-10.04	6.09 E-03
H7C3C4	Anion exchange protein	SLC4A7	-10.05	2.02 E-03
Q15437	Protein transport protein Sec23B	SEC23B	-10.10	4.95 E-02
Q8WUJ3	Cell migration-inducing and hyaluronan-binding protein	CEMIP	-10.10	2.01 E-03
Q5VT25	Serine/threonine-protein kinase MRCK alpha	CDC42BPA	-10.37	2.06 E-03
Q9NV96	Cell cycle control protein 50A	TMEM30A	-10.37	3.13 E-03
B1AK53	Espin	ESPN	-10.39	2.06 E-03
Q9NRF8	CTP synthase 2	CTPS2	-10.42	4.84 E-02
Q9BVG4	Protein PBDC1	PBDC1	-10.43	4.97 E-02
P08473	Neprilysin	MME	-10.45	2.80 E-03
P57737	Coronin-7	CORO7	-10.54	2.19 E-03
Q9BRF8	Serine/threonine-protein phosphatase CPPED1	CPPED1	-10.62	1.78 E-03
Q5EBL8	PDZ domain-containing protein 11	PDZD11	-10.74	2.55 E-03
Q8NCR9	Clarin-3	CLRN3	-10.76	2.10 E-03
J3KNV4	Integrin alpha-7	ITGA7	-10.81	4.74 E-03
Q5TCU3	Tropomyosin beta chain	TPM2	-10.92	1.71 E-03
P54886	Delta-1-pyrroline-5-carboxylate synthase	ALDH18A1	-10.99	2.69 E-04
Q08AM6	Protein VAC14 homolog	VAC14	-10.99	1.30 E-03
O43490	Prominin-1	PROM1	-11.03	3.71 E-02
P17612	cAMP-dependent protein kinase catalytic subunit alpha	PRKACA	-11.08	1.21 E-03
Q8N8S7	Protein enabled homolog	ENAH	-11.33	3.17 E-04

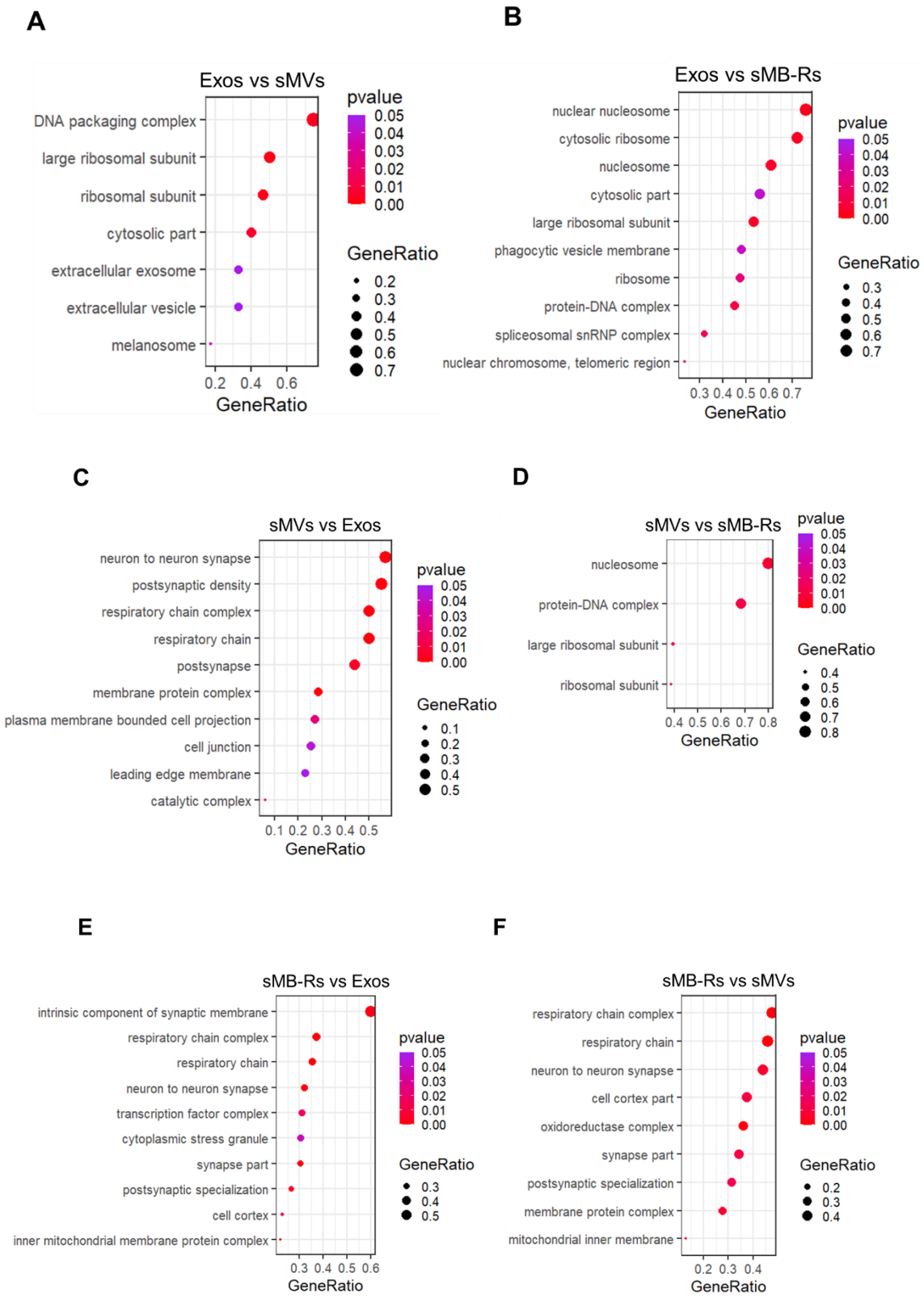
Q9ULE6	Paladin	PALD1	-11.42	1.14 E-03
Q7Z406	Myosin-14	MYH14	-11.47	4.72 E-02
Q86V21	Acetoacetyl-CoA synthetase	AACS	-11.71	2.07 E-02
Q04656	Copper-transporting ATPase 1	ATP7A	-11.75	3.16 E-04
Q9Y2L1	Exosome complex exonuclease RRP44	DIS3	-12.15	2.59 E-04
E7EX17	Eukaryotic translation initiation factor 4B	EIF4B	-12.22	1.31 E-02
E9PGT3	Ribosomal protein S6 kinase	RPS6KA1	-12.22	1.12 E-02
A5YKK6	CCR4-NOT transcription complex subunit 1	CNOT1	-12.24	1.42 E-02
P78504	Protein jagged-1	JAG1	-12.27	1.70 E-02
Q00534	Cyclin-dependent kinase 6	CDK6	-12.28	1.74 E-02
Q9H400	Lck-interacting transmembrane adapter 1	LIME1	-12.31	1.73 E-02
Q5VZK9	F-actin-uncapping protein LRRC16A	CARMIL1	-12.32	1.36 E-02
S4R3Z2	Aldo-keto reductase family 1 member C3	AKR1C3	-12.48	3.74 E-04
E7EWP2	Triple functional domain protein	TRIO	-12.49	1.16 E-02
Q93052	Lipoma-preferred partner	LPP	-12.50	1.65 E-02
Q14145	Kelch-like ECH-associated protein 1	KEAP1	-12.50	3.15 E-04
Q06124	Tyrosine-protein phosphatase non-receptor type 11	PTPN11	-12.56	1.28 E-02
Q96P48	Arf-GAP with Rho-GAP domain, ANK repeat and PH domain-containing protein 1	ARAP1	-12.57	3.02 E-04
Q13153	Serine/threonine-protein kinase PAK 1	PAK1	-12.61	8.15 E-03
Q92900	Regulator of nonsense transcripts 1	UPF1	-12.67	1.40 E-03
P29218	Inositol monophosphatase 1	IMPA1	-12.68	1.49 E-02
P21860	Receptor tyrosine-protein kinase erbB-3	ERBB3	-12.76	1.50 E-02
P15121	Aldose reductase	AKR1B1	-12.77	2.53 E-04
O14979	Heterogeneous nuclear ribonucleoprotein D-like	HNRNPDL	-12.77	2.55 E-04
P30520	Adenylosuccinate synthetase isozyme 2	ADSS	-12.78	1.30 E-02
Q9BQ69	O-acetyl-ADP-ribose deacetylase MACROD1	MACROD1	-12.92	1.57 E-02
O95372	Acyl-protein thioesterase 2	LYPLA2	-13.06	1.37 E-02
F8WA39	Myotubularin-related protein 1	MTMR1	-13.08	3.02 E-04
P51178	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta-1	PLCD1	-13.09	4.81 E-04
Q9UNZ2	NSFL1 cofactor p47	NSFL1C	-13.14	1.31 E-02
P50150	Guanine nucleotide-binding protein G	GNG4	-13.42	4.48 E-02
E7EVJ5	Cytoplasmic FMR1-interacting protein 2	CYFIP2	-13.63	5.24 E-04
Q6ZN28	Metastasis-associated in colon cancer protein 1	MACC1	-13.88	1.66 E-04
Q765P7	MTSS1-like protein	MTSS1L	-14.05	3.45 E-02
C9JQD4	Peptidyl-prolyl cis-trans isomerase	PPIH	-14.42	1.21 E-03
E9PGC0	Ras GTPase-activating protein 1	RASA1	-14.50	2.36 E-05
Q5VW36	Focadhesin	FOCAD	-14.61	3.14 E-04
P62310	U6 snRNA-associated Sm-like protein LSm3	LSM3	-14.69	1.79 E-03
A0A075B6Q2	Formin-binding protein 1-like	FNBP1L	-14.72	3.04 E-05

Q9NRN7	L-aminoadipate-semialdehyde dehydrogenase-phosphopantetheinyl transferase	AASDHPPT	-14.84	3.25 E-05
E7ET40	Urokinase-type plasminogen activator	PLAU	-14.89	2.40 E-03
Q13951	Core-binding factor subunit beta	CBFB	-14.97	2.16 E-03
H0Y6E7	RNA-binding motif protein, X chromosome	RBMX	-15.03	2.00 E-05
Q96DX4	RING finger and SPRY domain-containing protein 1	RSPRY1	-15.08	4.19 E-03
Q96FZ7	Charged multivesicular body protein 6	CHMP6	-15.08	2.03 E-03
Q96FF7	Uncharacterized protein MISP3	MISP3	-15.11	2.79 E-05
Q96BS2	Calcineurin B homologous protein 3	TESC	-15.16	2.02 E-05
A0A0U1RQP1	Dynamin-1	DNM1	-15.78	2.54 E-05
P16150	Leukosialin	SPN	-15.90	1.83 E-03
H7BYY1	Tropomyosin 1	TPM1	-15.94	3.06 E-03
Q6R327	Rapamycin-insensitive companion of mTOR	RICTOR	-15.96	3.03 E-05
Q9H3R2	Mucin-13	MUC13	-16.09	3.53 E-03
D6R938	Calcium/calmodulin-dependent protein kinase	CAMK2D	-16.13	1.97 E-05
Q9HCY8	Protein S100-A14	S100A14	-16.23	3.57 E-03
O95425	Supervillin	SVIL	-17.40	1.33 E-06
F8VYY9	5'-AMP-activated protein kinase subunit gamma-1	PRKAG1	-17.40	7.79 E-05
F8VWW8	Acid sphingomyelinase-like phosphodiesterase 3b	SMPDL3B	-17.59	8.59 E-07
Q5JVZ5	Engulfment and cell motility protein 2	ELMO2	-17.60	1.12 E-04
Q5VV41	Rho guanine nucleotide exchange factor 16	ARHGEF16	-17.62	1.16 E-04
P30047	GTP cyclohydrolase 1 feedback regulatory protein	GCHFR	-17.64	9.16 E-07
X6RM59	5'-nucleotidase	NT5C3A	-17.71	1.23 E-06
Q9NZU5	LIM and cysteine-rich domains protein 1	LMCD1	-17.71	1.08 E-06
G3XAC1	Solute carrier family 26 member 6	SLC26A6	-17.71	1.25 E-06
Q9Y5U5	Tumor necrosis factor receptor superfamily member 18	TNFRSF18	-17.76	9.33 E-05
Q9GZY6	Linker for activation of T-cells family member 2	LAT2	-17.77	1.34 E-04
P15328	Folate receptor alpha	FOLR1	-18.12	1.46 E-06
G8JLA2	Myosin light polypeptide 6	MYL6	-18.52	9.50 E-07
Q96FX8	p53 apoptosis effector related to PMP-22	PERP	-18.62	7.59 E-07
Q68EM7	Rho GTPase-activating protein 17	ARHGAP17	-19.22	1.36 E-04
Q9NVS9	Pyridoxine-5'-phosphate oxidase	PNPO	-20.18	1.77 E-08
P01730	T-cell surface glycoprotein CD4	CD4	-20.24	1.52 E-08
P11279	Lysosome-associated membrane glycoprotein 1	LAMP1	-20.42	1.19 E-08
Q96FC7	Phytanoyl-CoA hydroxylase-interacting protein-like	PHYHIPL	-20.47	4.07 E-07
P17252	Protein kinase C alpha type	PRKCA	-20.52	4.06 E-09



Appendix Figure 4.1 RNA quality data of Exos, sMVs and sMB-Rs derived from SW480 and SW620 cells. (obtained by RNA Bioanalyzer)

Appendix Figure 4.2 Gene Ontology (GO) analysis of Exos, sMVs and sMB-Rs derived from SW480 and SW620 cells. (A) Enriched GO terms of highly-enriched transcripts in Exos compared with sMVs (cutoff: pvalue < 0.05) and **(B)** compared with sMB-Rs (cutoff: pvalue < 0.05). **(C)** Enriched GO terms of highly-enriched transcripts in sMVs compared with Exos (cutoff: pvalue < 0.05) and **(D)** compared with sMB-Rs (cutoff: pvalue < 0.05). **(E)** Enriched GO terms of highly-enriched transcripts in sMB-Rs compared with Exos (cutoff: pvalue < 0.05) and **(F)** compared with sMVs (cutoff: pvalue < 0.05).


A
Exos vs sMVs
DNA packaging complex
large ribosomal subunit
ribosomal subunit
cytosolic part
extracellular exosome
extracellular vesicle
melanosome
0.2 0.4 0.6
GeneRatio
pvalue
0.05
0.04
0.03
0.02
0.01
0.00
GeneRatio
0.2
0.3
0.4
0.5
0.6
0.7
B
Exos vs sMB-Rs
nuclear nucleosome
cytosolic ribosome
nucleosome
cytosolic part
large ribosomal subunit
phagocytic vesicle membrane
ribosome
protein-DNA complex
spliceosomal snRNP complex
nuclear chromosome, telomeric region
0.3 0.4 0.5 0.6 0.7
GeneRatio
pvalue
0.05
0.04
0.03
0.02
0.01
0.00
GeneRatio
0.3
0.4
0.5
0.6
0.7
C
sMVs vs Exos
neuron to neuron synapse
postsynaptic density
respiratory chain complex
respiratory chain
postsynapse
membrane protein complex
plasma membrane bounded cell projection
cell junction
leading edge membrane
catalytic complex
0.1 0.2 0.3 0.4 0.5
GeneRatio
pvalue
0.05
0.04
0.03
0.02
0.01
0.00
GeneRatio
0.1
0.2
0.3
0.4
0.5
D
sMVs vs sMB-Rs
nucleosome
protein-DNA complex
large ribosomal subunit
ribosomal subunit
0.4 0.5 0.6 0.7 0.8
GeneRatio
pvalue
0.05
0.04
0.03
0.02
0.01
0.00
GeneRatio
0.4
0.5
0.6
0.7
0.8
E
sMB-Rs vs Exos
intrinsic component of synaptic membrane
respiratory chain complex
respiratory chain
neuron to neuron synapse
transcription factor complex
cytoplasmic stress granule
synapse part
postsynaptic specialization
cell cortex
inner mitochondrial membrane protein complex
0.3 0.4 0.5 0.6
GeneRatio
pvalue
0.05
0.04
0.03
0.02
0.01
0.00
GeneRatio
0.3
0.4
0.5
F
sMB-Rs vs sMVs
respiratory chain complex
respiratory chain
neuron to neuron synapse
cell cortex part
oxidoreductase complex
synapse part
postsynaptic specialization
membrane protein complex
mitochondrial inner membrane
0.2 0.3 0.4
GeneRatio
pvalue
0.05
0.04
0.03
0.02
0.01
0.00
GeneRatio
0.2
0.3
0.4

Appendix Table 4.1 Formula for FPKM calculation

	Description
FPKM equation	C/((N/1000)*(L/1,000,000))
C	Reads with fragments
N	Total number of sequence reads
L	Gene/transcript length

Appendix Table 4.2 A list of top 50 transcript identifications in Exos, sMVs and sMB-Rs derived from SW480 and SW620 cells

Transcript ID	Transcript name	Transcript biotype	Description	SW480-Exo	SW620-Exo	SW480-sMVs	SW620-sMVs	SW480-sMB-Rs	SW620-sMB-Rs
ENST00000618786	RN7SL1-201	misc_RNA	RNA component of signal recognition particle 7SL1	1,620,941	2,584,103	1,532,467	2,231,568	1,426,656	1,612,142
ENST00000490232	RN7SL2-201	misc_RNA	RNA component of signal recognition particle 7SL2	1,314,236	2,059,865	1,123,595	1,757,799	1,133,148	1,383,263
ENST00000584058	RN7SL4P-201	misc_RNA	RNA, 7SL, cytoplasmic 4, pseudogene	22,963	26,239	17,802	21,079	17,568	19,664
ENST00000233143	TMSB10-201	protein_coding	thymosin beta 10	20,137	18,261	16,609	14,111	11,549	8,712
ENST00000610674	RN7SL3-201	misc_RNA	RNA component of signal recognition particle 7SL3	17,163	19,439	13,478	14,336	11,301	11,238
ENST00000581458	RN7SL5P-201	misc_RNA	RNA, 7SL, cytoplasmic 5, pseudogene	12,650	15,228	6,722	8,796	7,646	6,699
ENST00000331825	FTL-201	protein_coding	ferritin light chain	6,544	2,016	4,925	1,795	4,638	1,298
ENST00000473748	RPS28P7-201	processed_pseudogene	ribosomal protein S28 pseudogene 7	3,276	2,596	2,186	2,409	2,425	2,232
ENST00000614247	HIST1H4D-201	protein_coding	histone cluster 1 H4 family member d	2,423	3,561	1,697	2,298	1,069	881
ENST00000377745	HIST1H4B-201	protein_coding	histone cluster 1 H4 family member b	1,628	4,009	885	1,888	544	574
ENST00000343677	HIST1H1C-201	protein_coding	histone cluster 1 H1 family member c	1,463	2,017	1,232	1,536	1,326	1,107
ENST00000377803	HIST1H4C-201	protein_coding	histone cluster 1 H4 family member c	1,155	2,131	849	1,420	637	727
ENST00000331380	HIST2H2AC-201	protein_coding	histone cluster 2 H2A family member c	1,375	1,647	1,041	1,234	856	764
ENST00000333151	HIST1H2AJ-201	protein_coding	histone cluster 1 H2A family member j	1,466	1,564	1,051	1,175	861	795

ENST00000436459	EEF1A1P5-201	processed_pseudogene	eukaryotic translation elongation factor 1 alpha 1 pseudogene 5	1,740	1,198	1,371	1,006	899	679
ENST00000309311	EEF2-201	protein_coding	eukaryotic translation elongation factor 2	984	1,205	963	1,125	1,104	1,495
ENST00000459748	AC116533.1-201	processed_pseudogene	ribosomal protein L36a (RPL36A) pseudogene	1,382	1,276	1,056	1,395	844	770
ENST00000611927	HIST1H4K-201	protein_coding	histone cluster 1 H4 family member k	1,339	1,620	945	1,184	620	495
ENST00000618305	HIST1H4L-201	protein_coding	histone cluster 1 H4 family member l	882	2,605	540	1,419	312	414
ENST00000621411	HIST1H3B-201	protein_coding	histone cluster 1 H3 family member b	1,031	1,325	850	1,118	946	866
ENST00000356950	HIST1H2BK-201	protein_coding	histone cluster 1 H2B family member k	1,178	1,318	894	972	1,004	737
ENST00000377459	HIST1H2AH-201	protein_coding	histone cluster 1 H2A family member h	1,300	1,458	933	1,018	752	587
ENST00000331442	HIST1H1B-201	protein_coding	histone cluster 1 H1 family member b	1,108	1,122	929	929	1,052	786
ENST00000387405	MT-TC-201	Mt_tRNA	mitochondrially encoded tRNA-Cys (UGU/C)	55	61	428	466	2,689	1,885
ENST00000303910	HIST1H2AE-201	protein_coding	histone cluster 1 H2A family member e	1,069	1,367	754	908	606	484
ENST00000612966	HIST1H3C-201	protein_coding	histone cluster 1 H3 family member c	708	1,471	520	1,075	569	713
ENST00000358739	HIST1H2AI-201	protein_coding	histone cluster 1 H2A family member i	968	922	788	780	831	740
ENST00000304218	HIST1H1E-201	protein_coding	histone cluster 1 H1 family member e	615	853	529	689	926	964
ENST00000616182	HIST1H2BO-201	protein_coding	histone cluster 1 H2B family member o	1,034	959	766	722	636	455
ENST00000618052	HIST1H3F-201	protein_coding	histone cluster 1 H3 family member f	926	1,119	688	754	551	410
ENST00000314133	COX8A-201	protein_coding	cytochrome c oxidase subunit 8A	811	768	715	709	723	625
ENST00000298289	RPL36AL-201	protein_coding	ribosomal protein L36a like	816	648	650	582	502	398
ENST00000314355	CKS2-201	protein_coding	CDC28 protein kinase regulatory subunit 2	800	432	655	492	611	477
ENST00000371437	NDUFA1-201	protein_coding	NADH:ubiquinone oxidoreductase subunit A1	668	638	622	580	492	436

ENST00000495531	RPL21P16-201	processed_pseudogene	ribosomal protein L21 pseudogene 16	1,203	303	784	275	483	155
ENST00000369163	HIST1H3H-201	protein_coding	histone cluster 1 H3 family member h	661	620	509	481	423	306
ENST00000615164	HIST1H4E-201	protein_coding	histone cluster 1 H4 family member e	546	934	382	557	263	238
ENST00000395422	CHCHD2-201	protein_coding	coiled-coil-helix-coiled-coil-helix domain containing 2	576	575	477	476	382	318
ENST00000372134	GHITM-201	protein_coding	growth hormone inducible transmembrane protein	522	593	461	475	338	324
ENST00000377733	HIST1H2BI-201	protein_coding	histone cluster 1 H2B family member i	378	809	291	551	265	304
ENST00000613854	HIST1H3A-201	protein_coding	histone cluster 1 H3 family member a	358	839	266	545	225	280
ENST00000374980	EIF2S2-201	protein_coding	eukaryotic translation initiation factor 2 subunit beta	602	357	499	304	388	240
ENST00000331491	HIST2H3D-201	protein_coding	histone cluster 2 H3 family member d	393	657	294	443	234	243
ENST00000370355	SCD-201	protein_coding	stearoyl-CoA desaturase	490	300	534	304	359	238
ENST00000495392	AC005912.1-201	processed_pseudogene	ribosomal protein S27 (RPS27) pseudogene	463	386	356	372	329	232
ENST00000377401	HIST1H2BL-201	protein_coding	histone cluster 1 H2B family member l	482	409	368	320	308	215
ENST00000233468	SF3B6-201	protein_coding	splicing factor 3b subunit 6	562	257	440	222	317	160
ENST00000541790	HIST1H2BG-201	protein_coding	histone cluster 1 H2B family member g	425	484	293	330	215	179
ENST00000615868	HIST1H2AB-201	protein_coding	histone cluster 1 H2A family member b	346	455	259	367	255	236
ENST00000359193	HIST1H2AG-201	protein_coding	histone cluster 1 H2A family member g	368	491	273	354	213	210

Appendix Table 4.3 Uniquely identified transcripts in both SW480 and SW620-EVs. (FPKM >1.5 in 3 biological replicates, top 5 transcripts are labelled in red)

EV classes	Transcript ID	Transcript name	Transcript biotype	Description
Exos	ENST00000408612	SNORD99-201	snoRNA	small nucleolar RNA, C/D box 99
	ENST00000555662	HSP90AA1-205	retained_intron	heat shock protein 90 alpha family class A member 1
	ENST00000619751	MIR7111-201	miRNA	microRNA 7111
	ENST00000514610	HMGCS1-207	retained_intron	3-hydroxy-3-methylglutaryl-CoA synthase 1
	ENST00000515446	HNRNPH1-226	retained_intron	heterogeneous nuclear ribonucleoprotein H1
	ENST00000504971	FBXO38-208	lncRNA	F-box protein 38
	ENST00000607315	AC026979.2-201	lncRNA	novel transcript
	ENST00000505828	C5orf66-201	lncRNA	chromosome 5 open reading frame 66
	ENST00000403609	TANK-204	protein_coding	TRAF family member associated NFKB activator
	ENST00000314980	LGALS7B-201	protein_coding	galectin 7B
	ENST00000528049	AC135983.3-201	unprocessed_pseudogene	M-phase phosphoprotein 10 (U3 small nucleolar ribonucleoprotein) (MPHOSPH10) pseudogene
	ENST00000297369	AC092681.1-201	unprocessed_pseudogene	pseudogene similar to part of ARP3 actin-related protein 3 homolog B (yeast) ACTR3B
	ENST00000577781	AC112907.3-201	lncRNA	novel transcript, antisense to eukaryotic translation initiation factor 4A, isoform 2 EIF4A2
	ENST00000576718	POLR2A-207	retained_intron	RNA polymerase II subunit A
	ENST00000496387	UQCRH-205	nonsense_mediated_decay	ubiquinol-cytochrome c reductase hinge protein
	ENST00000546566	AC068888.1-203	lncRNA	novel transcript, antisense to TENC1 & EIF4B
	ENST00000423230	MROH1-202	protein_coding	maestro heat like repeat family member 1
	ENST00000481752	NCK1-207	protein_coding	NCK adaptor protein 1
	ENST00000033079	FAM13B-201	protein_coding	family with sequence similarity 13 member B
	ENST00000467263	SMC4-207	retained_intron	structural maintenance of chromosomes 4
	ENST00000546204	TCP1-218	retained_intron	t-complex 1
	ENST00000498658	FXR1-222	retained_intron	FMR1 autosomal homolog 1

	ENST00000383898	RNU6-1-201	snRNA	RNA, U6 small nuclear 1
	ENST00000367284	ELF3-203	protein_coding	E74 like ETS transcription factor 3
	ENST00000616571	AC004801.6-201	lncRNA	novel transcript
	ENST00000462897	GCC2-206	protein_coding	GRIP and coiled-coil domain containing 2
	ENST00000500853	AC100861.1-202	lncRNA	novel transcript, antisense to TNFRSF10A
	ENST00000594885	EEF2-202	retained_intron	eukaryotic translation elongation factor 2
	ENST00000256151	CCDC59-201	protein_coding	coiled-coil domain containing 59
	ENST00000374259	FOXO4-202	protein_coding	forkhead box O4
	ENST00000464011	TFRC-208	lncRNA	transferrin receptor
	ENST00000402314	ATG3-202	protein_coding	autophagy related 3
	ENST00000495810	HNRNPA2B1-207	retained_intron	heterogeneous nuclear ribonucleoprotein A2/B1
	ENST00000457510	AC002467.1-201	lncRNA	novel transcript, antisense to CBLL1
sMVs	ENST00000507813	AC018752.1-201	lncRNA	novel transcript
	ENST00000476081	RPS20P33-201	processed_pseudogene	ribosomal protein S20 pseudogene 33
	ENST00000364473	RF00019.242-201	misc_RNA	To be confirmed
	ENST00000511805	RGS12-219	lncRNA	regulator of G protein signaling 12
	ENST00000528288	C11orf54-207	protein_coding	chromosome 11 open reading frame 54
	ENST00000600646	FKRP-220	lncRNA	fukutin related protein
	ENST00000462305	OR2A1-AS1-201	lncRNA	OR2A1 antisense RNA 1
	ENST00000485465	CHMP3-207	lncRNA	charged multivesicular body protein 3
	ENST00000367816	ATP1B1-202	protein_coding	ATPase Na+/K+ transporting subunit beta 1
sMB-Rs	ENST00000387429	MT-TG-201	Mt_tRNA	mitochondrially encoded tRNA-Gly (GGN)
	ENST00000637026	FASN-211	protein_coding	fatty acid synthase
	ENST00000387441	MT-TH-201	Mt_tRNA	mitochondrially encoded tRNA-His (CAU/C)
	ENST00000636337	SPTAN1-231	protein_coding	spectrin alpha, non-erythrocytic 1
	ENST00000539264	FBRSL1-203	lncRNA	fibrosin like 1
	ENST00000579420	MIR4479-201	miRNA	microRNA 4479
	ENST00000387439	MT-TR-201	Mt_tRNA	mitochondrially encoded tRNA-Arg (CGN)
	ENST00000516664	RF00272.7-201	snoRNA	To be confirmed
	ENST00000579909	MIR4312-201	miRNA	microRNA 4312
	ENST00000387377	MT-TM-201	Mt_tRNA	mitochondrially encoded tRNA-Met (AUA/G)
	ENST00000522148	DMTN-218	protein_coding	dematin actin binding protein
	ENST00000393190	NME2-202	protein_coding	NME/NM23 nucleoside diphosphate kinase 2

	ENST00000415602	TCEA2-205	protein_coding	transcription elongation factor A2
	ENST00000607491	AC233992.3-201	lncRNA	novel transcript, antisense to SLC39A4
	ENST00000590419	DAZAP1-209	lncRNA	DAZ associated protein 1
	ENST00000428691	KCMF1-202	protein_coding	potassium channel modulatory factor 1
	ENST00000526588	AP006621.3-201	lncRNA	novel transcript
	ENST00000641366	AC083906.5-201	unprocessed_pseudogene	RNA binding protein, fox-1 homolog (C. elegans) 1 (RBFOX1) pseudogene
	ENST00000602980	DPY19L2P3-203	lncRNA	DPY19L2 pseudogene 3
	ENST00000361681	MT-ND6-201	protein_coding	mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 6
	ENST00000464438	PPDPF-203	lncRNA	pancreatic progenitor cell differentiation and proliferation factor
	ENST00000385019	MIR199A1-201	miRNA	microRNA 199a-1
	ENST00000002501	DBNDD1-201	protein_coding	dysbindin domain containing 1
	ENST00000580578	LLGL2-216	protein_coding	LLGL scribble cell polarity complex component 2
	ENST00000581153	AC068594.1-203	lncRNA	novel transcript
	ENST00000540260	FKBP4-204	lncRNA	FKBP prolyl isomerase 4
	ENST00000589157	MFSD12-210	retained_intron	major facilitator superfamily domain containing 12
	ENST00000330877	ADSSL1-201	protein_coding	adenylosuccinate synthase like 1
	ENST00000468682	HRAS-207	lncRNA	HRas proto-oncogene, GTPase
	ENST00000577370	MIR5189-201	miRNA	microRNA 5189
	ENST00000589397	AC002398.1-201	lncRNA	novel transcript
	ENST00000440934	ZNF142-204	protein_coding	zinc finger protein 142
	ENST00000506465	PRR7-AS1-202	lncRNA	PRR7 antisense RNA 1
	ENST00000560763	EIF5-215	protein_coding	eukaryotic translation initiation factor 5
	ENST00000603893	PPP1R26-AS1-203	lncRNA	PPP1R26 antisense RNA 1
	ENST00000444482	AL353804.1-201	lncRNA	novel transcript
	ENST00000541078	ARHGDIA-203	protein_coding	Rho GDP dissociation inhibitor alpha
	ENST00000439870	AC079807.1-202	lncRNA	novel transcript
	ENST00000450469	LRP8-205	lncRNA	LDL receptor related protein 8
	ENST00000306329	MTCL1-201	protein_coding	microtubule crosslinking factor 1
	ENST00000414130	HMGA1P8-201	processed_pseudogene	high mobility group AT-hook 1 pseudogene 8
	ENST00000377390	SF1-204	protein_coding	splicing factor 1
	ENST00000567216	CAPN15-205	lncRNA	calpain 15

	ENST00000397333	DDX51-202	protein_coding	DEAD-box helicase 51
	ENST00000514507	CEP72-204	lncRNA	centrosomal protein 72
	ENST00000600056	AC016586.2-201	lncRNA	novel transcript, sense intronic to ZBTB7A
	ENST00000637419	GSE1-216	protein_coding	Gse1 coiled-coil protein
	ENST00000446475	COL18A1-AS2-201	lncRNA	COL18A1 antisense RNA 2
	ENST00000591510	PRKCSH-213	lncRNA	protein kinase C substrate 80K-H
	ENST00000538780	FAM222A-202	protein_coding	family with sequence similarity 222 member A
	ENST00000572681	CIC-204	protein_coding	capicua transcriptional repressor
	ENST00000491719	IGF2BP3-213	lncRNA	insulin like growth factor 2 mRNA binding protein 3
	ENST00000623846	AP006621.5-201	TEC	TEC
	ENST00000435627	AC007383.2-202	lncRNA	novel transcript
	ENST00000607286	SLC9A3-AS1-207	lncRNA	SLC9A3 antisense RNA 1
	ENST00000613359	RNA5-8SN3-201	rRNA	RNA, 5.8S ribosomal N3
	ENST00000373137	AL591845.1-201	lncRNA	novel transcript, antisense to EVA1B
	ENST00000593393	AC012313.1-201	lncRNA	novel transcript
	ENST00000643721	TRIM8-209	protein_coding	tripartite motif containing 8
	ENST00000561970	BCAR1-210	protein_coding	BCAR1 scaffold protein, Cas family member
	ENST00000508416	SKI-204	lncRNA	SKI proto-oncogene
	ENST00000340356	SOX18-201	protein_coding	SRY-box 18
	ENST00000418929	PRR12-201	protein_coding	proline rich 12
	ENST00000582295	C1QTNF1-210	lncRNA	C1q and TNF related 1
	ENST00000574260	AC004771.3-201	lncRNA	novel transcript, antisense to INCA1
	ENST00000380243	CCDC85C-201	protein_coding	coiled-coil domain containing 85C
	ENST00000651080	GAS5-231	lncRNA	growth arrest specific 5
	ENST00000518895	ERICH1-204	lncRNA	glutamate rich 1
	ENST00000441358	BCAN-204	protein_coding	brevican
	ENST00000635733	FASN-208	lncRNA	fatty acid synthase
	ENST00000319449	RUFY1-201	protein_coding	RUN and FYVE domain containing 1
	ENST00000215770	DDTL-201	protein_coding	D-dopachrome tautomerase like
	ENST00000431118	CYP20A1-204	nonsense_mediated_decay	cytochrome P450 family 20 subfamily A member 1
	ENST00000294008	SLX4-201	protein_coding	SLX4 structure-specific endonuclease subunit
	ENST00000331559	MEIS3-201	protein_coding	Meis homeobox 3

	ENST00000594828	LSM4-204	protein_coding	LSM4 homolog, U6 small nuclear RNA and mRNA degradation associated
	ENST00000377022	CASZ1-202	protein_coding	castor zinc finger 1
	ENST00000564272	SHCBP1-203	lncRNA	SHC binding and spindle associated 1
	ENST00000381150	SDC1-202	protein_coding	syndecan 1
	ENST00000360105	NFIX-202	protein_coding	nuclear factor I X
	ENST00000612193	IKBKGP1-201	unprocessed_pseudogene	inhibitor of nuclear factor kappa B kinase subunit gamma pseudogene 1
	ENST00000525223	EEF1D-210	protein_coding	eukaryotic translation elongation factor 1 delta
	ENST00000450909	AC120114.1-201	lncRNA	novel transcript, antisense to KCTD13
	ENST00000598112	AC024075.2-201	lncRNA	novel transcript, sense intronic C19orf42
	ENST00000597721	PSPN-202	protein_coding	persephin
	ENST00000610122	AC092171.3-201	lncRNA	uncharacterized LOC100129484
	ENST00000404338	ARHGAP35-201	protein_coding	Rho GTPase activating protein 35
	ENST00000406213	MIF-AS1-201	lncRNA	MIF antisense RNA 1
	ENST00000467148	HELZ2-204	protein_coding	helicase with zinc finger 2
	ENST00000611267	ZNF701-207	protein_coding	zinc finger protein 701
	ENST00000624273	AL513497.1-201	TEC	TEC
	ENST00000444760	ELP6-207	protein_coding	elongator acetyltransferase complex subunit 6
	ENST00000340305	TMEM201-201	protein_coding	transmembrane protein 201
	ENST00000379095	NGRN-202	protein_coding	neugrin, neurite outgrowth associated
	ENST00000438796	LRCH3-206	protein_coding	leucine rich repeats and calponin homology domain containing 3
	ENST00000368918	GABPB2-202	protein_coding	GA binding protein transcription factor subunit beta 2
	ENST00000607025	AC141002.1-201	lncRNA	novel transcript
	ENST00000447221	SNHG7-204	lncRNA	small nucleolar RNA host gene 7
	ENST00000648844	AHCTF1-207	protein_coding	AT-hook containing transcription factor 1
	ENST00000341976	ZNF70-201	protein_coding	zinc finger protein 70
	ENST00000590508	ZSWIM4-202	protein_coding	zinc finger SWIM-type containing 4
	ENST00000534164	AC138932.1-201	transcribed_unprocessed_pseudogene	polycystic kidney disease 1 (autosomal dominant) (PKD1) pseudogene
	ENST00000490103	THOC5-216	protein_coding	THO complex 5
	ENST00000401528	CLDN15-202	protein_coding	claudin 15
	ENST00000399413	AC129492.2-201	lncRNA	novel transcript, antisense to ALOX12B
	ENST00000223210	ZNF862-201	protein_coding	zinc finger protein 862

	ENST00000371586	CC2D1B-202	protein_coding	coiled-coil and C2 domain containing 1B
	ENST00000516768	SCARNA20-201	scaRNA	small Cajal body-specific RNA 20
	ENST00000615670	PRAG1-201	protein_coding	PEAK1 related, kinase-activating pseudokinase 1
	ENST00000262613	SLC9A3R1-201	protein_coding	SLC9A3 regulator 1
	ENST00000609436	JMJD1C-AS1-201	lncRNA	JMJD1C antisense RNA 1
	ENST00000465971	AC018638.4-201	processed_pseudogene	novel pseudogene
	ENST00000256474	VHL-201	protein_coding	von Hippel-Lindau tumor suppressor
	ENST00000440097	LIMD1-202	protein_coding	LIM domains containing 1
	ENST00000371608	BCAS4-202	protein_coding	breast carcinoma amplified sequence 4
	ENST00000304449	NKRF-201	protein_coding	NFKB repressing factor
	ENST00000526795	RAB30-DT-202	lncRNA	RAB30 divergent transcript
	ENST00000555524	AL583722.4-201	lncRNA	novel transcript, antisense to INF2
	ENST00000254325	RFX1-201	protein_coding	regulatory factor X1
	ENST00000456374	STAG3L1-202	transcribed_unprocessed_pseudogene	stromal antigen 3-like 1 (pseudogene)
	ENST00000499988	RAD51-AS1-201	retained_intron	RAD51 antisense RNA 1
	ENST00000515707	ABCC3-217	protein_coding	ATP binding cassette subfamily C member 3
	ENST00000573058	SLC38A10-210	retained_intron	solute carrier family 38 member 10
	ENST00000400376	TTYH3-202	protein_coding	tweety family member 3
	ENST00000455245	DUX4L50-201	unprocessed_pseudogene	double homeobox 4 like 50 (pseudogene)
	ENST00000587595	ERC1-220	lncRNA	ELKS/RAB6-interacting/CAST family member 1
	ENST00000602597	AC135178.5-201	lncRNA	novel transcript
	ENST00000335999	NCKAP5L-201	protein_coding	NCK associated protein 5 like
	ENST00000609755	AC102953.2-201	lncRNA	novel transcript
	ENST00000378061	ZNF425-201	protein_coding	zinc finger protein 425
	ENST00000588040	AC093227.1-201	lncRNA	novel transcript
	ENST00000561275	OIP5-AS1-210	lncRNA	OIP5 antisense RNA 1
	ENST00000553851	PRAF2-202	protein_coding	PRA1 domain family member 2
	ENST00000465356	PCNT-203	retained_intron	pericentrin

Appendix Table 4.4 Transcript biotype classification

Category	Sub-category
lncRNA	lncRNA
mRNA	protein-coding
Pseudogene	pseudogene
	processed pseudogene
	IG_V_pseudogene
	TR_J_pseudogene
	rRNA_pseudogene
	transcribed_processed_pseudogene
	transcribed_unitary_pseudogene
	transcribed_unprocessed_pseudogene
	translated_processed_pseudogene
	translated_unprocessed_pseudogene
	unitary_pseudogene
	unprocessed_pseudogene
Other	IG_V_gene
	miRNA
	misc_RNA
	Mt_rRNA
	Mt_tRNA
	non_stop_decay
	nonsense_mediated_decay
	retained_intron
	ribozyme
	rRNA
	scaRNA
	scRNA
	snoRNA
	snRNA
	TEC
	TR_C_gene
	TR_D_gene
	TR_J_gene
	TR_V_gene

Appendix Table 4.5 Differential transcript expression analysis of Exos and sMVs. (negative log2foldchange indicates enriched transcripts in Exos, positive log2foldchange indicates enriched transcripts in sMVs)

Transcript ID	Transcript name	Transcript biotypes	log2Fold change	pvalue	Description
ENST00000465768	SAMM50-202	lncRNA	-17.21	4.78E-09	SAMM50 sorting and assembly machinery component
ENST00000548583	C12orf45-203	protein_coding	-3.49	7.93E-03	chromosome 12 open reading frame 45
ENST00000523339	NPM1-210	retained_intron	-3.26	1.64E-02	nucleophosmin 1
ENST00000617193	MIR6747-201	miRNA	-3.24	3.16E-02	microRNA 6747
ENST00000531529	NUF2-211	retained_intron	-3.05	2.11E-02	NUF2 component of NDC80 kinetochore complex
ENST00000474740	NDUFV2-205	lncRNA	-3.03	8.78E-03	NADH:ubiquinone oxidoreductase core subunit V2
ENST00000615842	RF00003.27-201	snRNA	-2.98	0.00E+00	0
ENST00000593229	DUS3L-215	retained_intron	-2.95	2.79E-03	dihydrouridine synthase 3 like
ENST00000450304	NALT1-202	lncRNA	-2.87	3.86E-02	NOTCH1 associated lncRNA in T cell acute lymphoblastic leukemia 1
ENST00000483541	NSMCE4A-210	retained_intron	-2.81	2.55E-02	NSE4 homolog A, SMC5-SMC6 complex component
ENST00000464079	PRRC2A-205	retained_intron	-2.78	7.59E-04	proline rich coiled-coil 2A
ENST00000516241	RF00322.6-201	snoRNA	-2.73	2.67E-02	0
ENST00000384187	RF00019.449-201	misc_RNA	-2.64	3.93E-02	0
ENST00000483853	FUS-204	retained_intron	-2.61	8.88E-04	FUS RNA binding protein
ENST00000510426	KAT7-212	retained_intron	-2.57	2.56E-03	lysine acetyltransferase 7
ENST00000408612	SNORD99-201	snoRNA	-2.54	5.35E-03	small nucleolar RNA, C/D box 99
ENST00000471995	RBM5-215	retained_intron	-2.51	6.56E-04	RNA binding motif protein 5
ENST00000555662	HSP90AA1-205	retained_intron	-2.37	3.34E-09	heat shock protein 90 alpha family class A member 1
ENST00000488849	ZNF92-205	retained_intron	-2.35	3.05E-03	zinc finger protein 92
ENST00000521864	OXR1-213	retained_intron	-2.20	1.56E-02	oxidation resistance 1
ENST00000581776	MSI2-214	nonsense_mediated_decay	-2.14	4.12E-02	musashi RNA binding protein 2
ENST00000586193	TPM4-203	retained_intron	-2.14	7.92E-03	tropomyosin 4
ENST00000365448	RF00019.354-201	misc_RNA	-2.14	2.11E-02	0
ENST00000569687	SF3B3-213	protein_coding	-2.12	1.92E-02	splicing factor 3b subunit 3
ENST00000521710	NPM1-209	retained_intron	-2.10	1.76E-03	nucleophosmin 1
ENST00000364915	SNORD117-201	snoRNA	-2.09	4.13E-02	small nucleolar RNA, C/D box 117
ENST00000490189	NKTR-216	retained_intron	-2.09	2.42E-02	natural killer cell triggering receptor
ENST00000431293	AL731563.2-201	lncRNA	-2.09	2.89E-02	novel transcript, antisense to P4HA1
ENST00000616097	FP325318.1-201	lncRNA	-2.04	6.87E-03	novel transcript
ENST00000422509	AL158156.1-201	processed_pseudogene	-2.04	1.59E-02	ribosomal protein L21 (RPL21) pseudogene
ENST00000431653	RPL7-205	protein_coding	-2.00	2.14E-03	ribosomal protein L7
ENST00000405331	TRIB2-203	protein_coding	-1.87	3.74E-02	tribbles pseudokinase 2
ENST00000516365	RNU7-18P-201	snRNA	-1.87	1.33E-02	RNA, U7 small nuclear 18 pseudogene
ENST00000493798	BAD-205	protein_coding	-1.87	3.28E-02	BCL2 associated agonist of cell death

ENST00000557089	HSP90AA1-207	retained_intron	-1.86	2.90E-04	heat shock protein 90 alpha family class A member 1
ENST00000458455	RPL11-203	protein_coding	-1.83	1.55E-03	ribosomal protein L11
ENST00000478196	RPS27A-208	retained_intron	-1.82	3.66E-03	ribosomal protein S27a
ENST00000481846	SLMAP-218	lncRNA	-1.80	4.78E-02	sarcolemma associated protein
ENST00000437952	UFM1-203	protein_coding	-1.73	5.40E-03	ubiquitin fold modifier 1
ENST00000520595	UBE2V2-203	nonsense_mediated_decay	-1.71	1.92E-03	ubiquitin conjugating enzyme E2 V2
ENST00000619751	MIR7111-201	miRNA	-1.70	3.28E-02	microRNA 7111
ENST00000411366	RNY3P8-201	misc_RNA	-1.69	4.07E-02	RNY3 pseudogene 8
ENST00000518420	AC027801.3-201	lncRNA	-1.69	4.08E-02	novel transcript
ENST00000415255	PABPC4-AS1-201	lncRNA	-1.65	2.51E-02	PABPC4 antisense RNA 1
ENST00000595874	SHKBP1-210	retained_intron	-1.61	2.51E-02	SH3KBP1 binding protein 1
ENST00000442341	RPL35A-204	protein_coding	-1.61	1.18E-02	ribosomal protein L35a
ENST00000609428	AL021707.8-201	lncRNA	-1.61	1.85E-02	novel transcript, sense intronic to SUN2
ENST00000592934	AC005256.1-202	lncRNA	-1.59	4.90E-02	novel transcript
ENST00000549183	TUBA1C-204	protein_coding	-1.57	1.69E-03	tubulin alpha 1c
ENST00000614578	CENPF-206	protein_coding	-1.56	2.96E-02	centromere protein F
ENST00000567969	ARL6IP1-207	retained_intron	-1.53	2.03E-03	ADP ribosylation factor like GTPase 6 interacting protein 1
ENST00000471565	KRT19-205	retained_intron	-1.53	4.98E-05	keratin 19
ENST00000463639	BRD2-207	retained_intron	-1.50	1.54E-04	bromodomain containing 2
ENST00000556043	PRMT5-221	protein_coding	-1.50	1.27E-02	protein arginine methyltransferase 5
ENST00000541671	SLC2A3-208	retained_intron	-1.48	8.53E-03	solute carrier family 2 member 3
ENST00000418803	AL021707.3-201	lncRNA	-1.48	1.83E-03	novel transcript, antisense to SUN2
ENST00000364931	RNU5E-4P-201	snRNA	-1.48	1.22E-02	RNA, U5E small nuclear 4, pseudogene
ENST00000559647	ANXA2-223	retained_intron	-1.48	1.75E-04	annexin A2
ENST00000552785	AC079313.1-201	lncRNA	-1.47	3.12E-02	novel transcript, antisense to ZNF385A, GPR84, ITGA5 and GTSF1
ENST00000471991	SFPQ-207	lncRNA	-1.46	3.93E-02	splicing factor proline and glutamine rich
ENST00000579810	ZNF207-211	retained_intron	-1.45	1.89E-03	zinc finger protein 207
ENST00000562683	RBBP6-206	protein_coding	-1.44	2.94E-02	RB binding protein 6, ubiquitin ligase
ENST00000460860	ECT2-213	retained_intron	-1.43	7.93E-03	epithelial cell transforming 2
ENST00000553730	PPP2R5C-207	retained_intron	-1.42	1.41E-03	protein phosphatase 2 regulatory subunit B'gamma
ENST00000568525	SPINT1-AS1-205	lncRNA	-1.40	1.88E-02	SPINT1 antisense RNA 1
ENST00000492262	ZNF638-223	retained_intron	-1.40	3.19E-02	zinc finger protein 638
ENST00000582898	KPNA2-204	lncRNA	-1.39	3.62E-03	karyopherin subunit alpha 2
ENST00000549592	EIF4B-206	retained_intron	-1.39	2.28E-03	eukaryotic translation initiation factor 4B
ENST00000468711	COX7A2L-206	protein_coding	-1.38	4.46E-02	cytochrome c oxidase subunit 7A2 like
ENST00000363823	RF00012.9-201	snoRNA	-1.38	1.62E-02	0
ENST00000411377	RNU6-722P-201	snRNA	-1.38	4.70E-02	RNA, U6 small nuclear 722, pseudogene
ENST00000473843	MACF1-215	lncRNA	-1.36	2.71E-02	microtubule actin crosslinking factor 1
ENST00000468885	GNL3-206	retained_intron	-1.36	3.37E-02	G protein nucleolar 3
ENST00000564276	PKM-211	retained_intron	-1.36	1.38E-02	pyruvate kinase M1/2
ENST00000529173	RPS3-212	retained_intron	-1.36	2.71E-04	ribosomal protein S3

ENST00000514610	HMGCS1-207	retained_intron	-1.35	2.17E-04	3-hydroxy-3-methylglutaryl-CoA synthase 1
ENST00000483432	ANAPC2-203	lncRNA	-1.35	5.88E-10	anaphase promoting complex subunit 2
ENST00000549253	TGIF1-212	protein_coding	-1.34	8.26E-03	TGFB induced factor homeobox 1
ENST00000472516	DIAPH1-207	retained_intron	-1.34	5.59E-03	diaphanous related formin 1
ENST00000498025	IARS-212	lncRNA	-1.33	1.85E-02	isoleucyl-tRNA synthetase
ENST00000480307	SUN2-217	retained_intron	-1.33	2.76E-02	Sad1 and UNC84 domain containing 2
ENST00000474894	FANCG-206	retained_intron	-1.32	3.76E-02	FA complementation group G
ENST00000477635	SRSF7-211	retained_intron	-1.32	1.39E-02	serine and arginine rich splicing factor 7
ENST00000600533	PIK3R2-206	retained_intron	-1.32	3.74E-02	phosphoinositide-3-kinase regulatory subunit 2
ENST00000484250	IVD-205	retained_intron	-1.31	3.24E-02	isovaleryl-CoA dehydrogenase
ENST00000515612	KLHL5-207	protein_coding	-1.31	4.49E-02	kelch like family member 5
ENST00000496205	CLK1-214	retained_intron	-1.29	2.74E-02	CDC like kinase 1
ENST00000507608	PPWD1-205	retained_intron	-1.29	4.26E-02	peptidylprolyl isomerase domain and WD repeat containing 1
ENST00000618702	SEC22B2-201	unprocessed_pseudogene	-1.29	2.50E-02	SEC22 homolog B2
ENST00000291527	TFF1-201	protein_coding	-1.26	6.04E-03	trefoil factor 1
ENST00000534645	RCE1-208	lncRNA	-1.25	3.79E-02	Ras converting CAAX endopeptidase 1
ENST00000598087	TMSB15B-207	protein_coding	-1.25	1.45E-02	thymosin beta 15B
ENST00000468863	YTHDF2-202	lncRNA	-1.25	1.40E-02	YTH N6-methyladenosine RNA binding protein 2
ENST00000518935	DST-220	protein_coding	-1.25	1.44E-05	dystonin
ENST00000356884	BICD2-201	protein_coding	-1.25	4.36E-30	BICD cargo adaptor 2
ENST00000482167	HSPD1-212	retained_intron	-1.25	1.94E-02	heat shock protein family D (Hsp60) member 1
ENST00000524933	FAR1-202	retained_intron	-1.24	2.26E-02	fatty acyl-CoA reductase 1
ENST00000463397	RN7SL674P-201	misc_RNA	-1.24	0.00E+00	RNA, 7SL, cytoplasmic 674, pseudogene
ENST00000466248	WWC3-202	lncRNA	-1.24	3.05E-02	WWC family member 3
ENST00000425699	NANOS1-202	protein_coding	-1.21	3.93E-06	nanos C2HC-type zinc finger 1
ENST00000582591	AP000919.3-201	lncRNA	2.55	1.75E-03	novel transcript, antisense to LPIN2
ENST00000428733	AC004975.1-201	processed_pseudogene	2.55	7.82E-03	ribosomal protein L36 (RPL36A) pseudogene
ENST00000442115	AL137077.1-201	processed_pseudogene	2.56	6.31E-03	ribosomal protein L17 (RPL17) pseudogene
ENST00000589464	AC060780.3-201	unprocessed_pseudogene	2.56	2.62E-02	breast cancer 1, early onset (BRCA1) pseudogene
ENST00000569006	BCAR1-221	protein_coding	2.58	2.77E-02	BCAR1 scaffold protein, Cas family member
ENST00000566222	AC105036.1-201	processed_pseudogene	2.59	1.92E-02	small nuclear ribonucleoprotein polypeptide F (SNRPF) pseudogene
ENST00000585921	#N/A	#N/A	2.59	9.01E-03	#N/A
ENST00000438601	CR391992.1-201	processed_pseudogene	2.59	1.19E-02	DnaJ (Hsp40) homolog, subfamily C member 8 (DNAJC8) pseudogene
ENST00000548666	AC008147.3-201	transcribed_processed_pseudogene	2.61	3.26E-02	40S ribosomal protein S26 (RPS26) pseudogene
ENST00000577097	BAIAP2-230	retained_intron	2.62	6.52E-03	BAI1 associated protein 2
ENST00000613401	AL121845.4-201	lncRNA	2.63	8.07E-03	novel transcript, antisense to ZGPAT
ENST00000504625	AHRR-201	protein_coding	2.63	2.19E-03	aryl-hydrocarbon receptor repressor
ENST00000525891	AP2A2-206	lncRNA	2.64	1.18E-02	adaptor related protein complex 2 subunit alpha 2
ENST00000455694	PRSS48-202	protein_coding	2.66	8.47E-04	serine protease 48
ENST00000624910	AC002044.2-201	TEC	2.67	2.09E-03	novel transcript

ENST00000440224	AC003989.2-201	processed_pseudogene	2.69	2.49E-02	pseudogene similar to part of adenosine deaminase, tRNA-specific 2, TAD2 homolog (S. cerevisiae) (ADAT2)
ENST00000526496	CKAP5-205	protein_coding	2.70	1.99E-03	cytoskeleton associated protein 5
ENST00000384471	RNU6-45P-201	snRNA	2.70	1.95E-02	RNA, U6 small nuclear 45, pseudogene
ENST00000516818	RNU6-122P-201	snRNA	2.71	3.69E-02	RNA, U6 small nuclear 122, pseudogene
ENST00000495121	CABIN1-212	lncRNA	2.71	2.72E-03	calcineurin binding protein 1
ENST00000560731	MARK3-222	lncRNA	2.72	8.83E-03	microtubule affinity regulating kinase 3
ENST00000578255	MIR3177-201	miRNA	2.73	9.47E-03	microRNA 3177
ENST00000497848	RF00019.654-201	misc_RNA	2.73	6.19E-03	0
ENST00000362954	RNU6-199P-201	snRNA	2.73	6.86E-03	RNA, U6 small nuclear 199, pseudogene
ENST00000483938	LPP-215	lncRNA	2.76	3.15E-02	LIM domain containing preferred translocation partner in lipoma
ENST00000543422	MACROD1-207	lncRNA	2.76	1.69E-02	mono-ADP ribosylhydrolase 1
ENST00000445956	BMP7-AS1-201	lncRNA	2.76	5.83E-03	BMP7 antisense RNA 1
ENST00000469001	RPL31P58-201	processed_pseudogene	2.77	1.67E-02	ribosomal protein L31 pseudogene 58
ENST00000565511	RAB40C-209	nonsense_mediated_decay	2.77	3.97E-03	RAB40C, member RAS oncogene family
ENST00000361567	MT-ND5-201	protein_coding	2.77	5.70E-33	mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 5
ENST00000484182	AC106707.2-201	processed_pseudogene	2.78	1.71E-02	ribosomal protein S12 (RPS12) pseudogene
ENST00000433041	BRD3-204	protein_coding	2.79	1.39E-02	bromodomain containing 3
ENST00000312310	COX6B1P3-201	processed_pseudogene	2.79	1.68E-02	cytochrome c oxidase subunit 6B1 pseudogene 3
ENST00000413175	POU5F1P4-201	processed_pseudogene	2.79	6.68E-04	POU class 5 homeobox 1 pseudogene 4
ENST00000620377	AC073534.2-201	lncRNA	2.80	5.20E-03	novel transcript
ENST00000520405	KCTD9-206	retained_intron	2.80	4.05E-02	potassium channel tetramerization domain containing 9
ENST00000614396	AL110115.1-201	lncRNA	2.80	2.18E-02	novel transcript, sense intronic to HM13
ENST00000387461	MT-TP-201	Mt_tRNA	2.81	1.21E-18	mitochondrially encoded tRNA-Pro (CCN)
ENST00000560126	ABHD17C-204	lncRNA	2.81	1.91E-02	abhydrolase domain containing 17C
ENST00000375350	NSD1-203	nonsense_mediated_decay	2.81	1.34E-02	nuclear receptor binding SET domain protein 1
ENST00000422306	AL162430.2-201	lncRNA	2.82	7.81E-03	novel transcript
ENST00000361624	MT-CO1-201	protein_coding	2.83	4.79E-41	mitochondrially encoded cytochrome c oxidase I
ENST00000630128	AL355480.3-201	pseudogene	2.84	8.99E-03	0
ENST00000618902	SHMT1-211	protein_coding	2.86	2.77E-02	serine hydroxymethyltransferase 1
ENST00000450016	LINC01952-202	lncRNA	2.88	1.32E-02	long intergenic non-protein coding RNA 1952
ENST00000581568	LPIN2-202	lncRNA	2.88	9.05E-03	lipin 2
ENST00000583739	MIR5690-201	miRNA	2.88	4.16E-02	microRNA 5690
ENST00000529229	AP003390.2-201	lncRNA	2.88	4.35E-03	novel transcript
ENST00000570544	AC036164.1-201	processed_pseudogene	2.89	1.48E-03	chromosome 19 open reading frame 70 (C19orf70) pseudogene
ENST00000625131	AL353662.3-201	pseudogene	2.89	3.64E-03	0
ENST00000578534	MIR4694-201	miRNA	2.90	2.05E-02	microRNA 4694
ENST00000442564	AC007000.2-201	processed_pseudogene	2.93	2.18E-02	pseudogene similar to part of SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 1 SMARCA1
ENST00000583905	MIR4755-201	miRNA	2.95	1.27E-02	microRNA 4755
ENST00000624885	AC073592.9-201	TEC	2.95	8.59E-03	novel transcript
ENST00000362079	MT-CO3-201	protein_coding	2.97	4.99E-42	mitochondrially encoded cytochrome c oxidase III

ENST00000387405	MT-TC-201	Mt_tRNA	2.97	2.27E-27	mitochondrially encoded tRNA-Cys (UGU/C)
ENST00000361739	MT-CO2-201	protein_coding	2.99	4.40E-37	mitochondrially encoded cytochrome c oxidase II
ENST00000617368	LINC01632-202	lncRNA	2.99	1.74E-02	long intergenic non-protein coding RNA 1632
ENST00000384816	MIR25-201	miRNA	2.99	5.18E-03	microRNA 25
ENST00000436122	TIMM9P1-201	processed_pseudogene	3.00	3.80E-03	TIMM9 pseudogene 1
ENST00000530976	AP001767.1-201	processed_pseudogene	3.00	1.23E-02	ribosomal protein L10 (RPL10) pseudogene
ENST00000361335	MT-ND4L-201	protein_coding	3.00	3.79E-15	mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 4L
ENST00000384721	RNU6-606P-201	snRNA	3.01	3.22E-02	RNA, U6 small nuclear 606, pseudogene
ENST00000472214	RPL12P21-201	processed_pseudogene	3.03	1.00E-02	ribosomal protein L12 pseudogene 21
ENST00000612810	RF00017.142-201	misc_RNA	3.03	7.35E-03	0
ENST00000497468	RN7SL558P-201	misc_RNA	3.04	9.75E-03	RNA, 7SL, cytoplasmic 558, pseudogene
ENST00000438019	AL121601.2-201	processed_pseudogene	3.05	2.32E-04	par-6 partitioning defective 6 homolog beta (C. elegans) (PARD6B) pseudogene
ENST00000603159	AC097639.2-201	processed_pseudogene	3.07	8.47E-03	small integral membrane protein 11, pseudogene
ENST00000637725	RF00026.110-201	snRNA	3.09	8.78E-03	0
ENST00000637592	EP300-212	retained_intron	3.10	1.77E-02	E1A binding protein p300
ENST00000587052	CEP295NL-203	retained_intron	3.11	2.53E-03	CEP295 N-terminal like
ENST00000492975	FO393419.2-201	lncRNA	3.11	3.01E-03	novel transcript
ENST00000603908	FO393422.1-201	processed_pseudogene	3.13	7.14E-04	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1, methenyltetrahydrofolate cyclohydrolase, formyltetrahydrofolate synthetase (MTHFD1) pseudogene
ENST00000625310	AC006001.3-204	lncRNA	3.19	1.64E-03	RAB guanine nucleotide exchange factor (GEF) 1 (RABGEF1) pseudogene
ENST00000563659	WWP2-206	protein_coding	3.19	7.26E-04	WW domain containing E3 ubiquitin protein ligase 2
ENST00000505029	AC113410.1-201	processed_pseudogene	3.20	2.65E-02	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 4, 9kDa (NDUFA4) pseudogene
ENST00000361390	MT-ND1-201	protein_coding	3.20	1.01E-13	mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 1
ENST00000385248	MIR613-201	miRNA	3.20	3.42E-03	microRNA 613
ENST00000361381	MT-ND4-201	protein_coding	3.20	8.11E-27	mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 4
ENST00000387392	MT-TA-201	Mt_tRNA	3.21	1.04E-09	mitochondrially encoded tRNA-Ala (GCN)
ENST00000361789	MT-CYB-201	protein_coding	3.23	1.32E-38	mitochondrially encoded cytochrome b
ENST00000603716	ELOCP33-201	processed_pseudogene	3.24	5.46E-05	elongin C pseudogene 33
ENST00000588954	ZNF440-206	protein_coding	3.26	1.54E-03	zinc finger protein 440
ENST00000578566	MIR3183-201	miRNA	3.30	3.15E-03	microRNA 3183
ENST00000387409	MT-TY-201	Mt_tRNA	3.30	6.84E-04	mitochondrially encoded tRNA-Tyr (UAU/C)
ENST00000522494	AC009686.1-201	lncRNA	3.32	6.72E-03	novel transcript
ENST00000428429	AL162727.2-201	lncRNA	3.34	1.11E-03	novel transcript
ENST00000579892	MIR5192-201	miRNA	3.35	3.38E-03	microRNA 5192
ENST00000427475	AL353807.1-201	processed_pseudogene	3.39	4.25E-03	claudin 6 (CLDN6) pseudogene
ENST00000387400	MT-TN-201	Mt_tRNA	3.40	8.59E-29	mitochondrially encoded tRNA-Asn (AAU/C)
ENST00000387441	MT-TH-201	Mt_tRNA	3.43	1.38E-03	mitochondrially encoded tRNA-His (CAU/C)

ENST00000361453	MT-ND2-201	protein_coding	3.43	1.30E-18	mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 2
ENST00000470590	RN7SL594P-201	misc_RNA	3.43	6.08E-03	RNA, 7SL, cytoplasmic 594, pseudogene
ENST00000364083	RF00019.202-201	misc_RNA	3.45	6.52E-04	0
ENST00000616367	RF00017.155-201	misc_RNA	3.63	6.91E-04	0
ENST00000387449	MT-TS2-201	Mt_tRNA	4.21	6.40E-11	mitochondrially encoded tRNA-Ser (AGU/C) 2
ENST00000387460	MT-TT-201	Mt_tRNA	4.25	4.74E-06	mitochondrially encoded tRNA-Thr (ACN)
ENST00000414273	MTCO1P12-201	unprocessed_pseudogene	4.30	3.79E-06	MT-CO1 pseudogene 12
ENST00000578274	MIR4711-201	miRNA	15.47	3.04E-07	microRNA 4711
ENST00000615130	RF01518.8-201	misc_RNA	19.55	0.00E+00	0

Appendix Table 4.6 Differential transcript expression analysis of Exos and sMB-Rs (negative log2foldchange indicates enriched transcripts in Exos, positive log2foldchange indicates enriched transcripts in sMB-Rs)

Transcript ID	Transcript name	Transcript biotypes	log2Fold change	pvalue	Description
ENST00000466217	ATG9A-218	retained_intron	-4.89	4.89E-06	autophagy related 9A
ENST00000559597	RCOR1-203	retained_intron	-4.42	3.31E-05	REST corepressor 1
ENST00000483853	FUS-204	retained_intron	-4.39	1.73E-05	FUS RNA binding protein
ENST00000481561	RN7SL273P-201	misc_RNA	-4.15	1.65E-03	RNA, 7SL, cytoplasmic 273, pseudogene
ENST00000555662	HSP90AA1-205	retained_intron	-4.10	1.25E-12	heat shock protein 90 alpha family class A member 1
ENST00000611125	MYO19-206	retained_intron	-4.04	5.45E-05	myosin XIX
ENST00000616097	FP325318.1-201	lncRNA	-3.96	1.93E-05	novel transcript
ENST00000615842	RF00003.27-201	snRNA	-3.95	NA	0
ENST00000470259	ITGA6-209	retained_intron	-3.65	1.46E-04	integrin subunit alpha 6
ENST00000617193	MIR6747-201	miRNA	-3.59	1.70E-02	microRNA 6747
ENST00000464079	PRRC2A-205	retained_intron	-3.55	9.68E-05	proline rich coiled-coil 2A
ENST00000532146	CELF1-216	retained_intron	-3.51	1.73E-04	CUGBP Elav-like family member 1
ENST00000460806	BHLHE40-202	retained_intron	-3.49	8.46E-04	basic helix-loop-helix family member e40
ENST00000516365	RNU7-18P-201	snRNA	-3.48	4.61E-04	RNA, U7 small nuclear 18 pseudogene
ENST00000497697	SMTN-220	retained_intron	-3.48	5.20E-04	smoothelin
ENST00000586193	TPM4-203	retained_intron	-3.44	5.75E-04	tropomyosin 4
ENST00000431088	AC016712.1-201	processed_pseudogene	-3.44	1.11E-02	ribosomal protein L31 (RPL31) pseudogene
ENST00000485090	RABGAP1-211	retained_intron	-3.39	6.65E-03	RAB GTPase activating protein 1
ENST00000496205	CLK1-214	retained_intron	-3.34	4.40E-05	CDC like kinase 1
ENST00000497835	KIF1B-210	retained_intron	-3.31	3.06E-02	kinesin family member 1B
ENST00000365484	RNY3-201	misc_RNA	-3.26	6.19E-06	RNA, Ro60-associated Y3
ENST00000486622	RNF144B-202	lncRNA	-3.21	7.14E-03	ring finger protein 144B
ENST00000561350	DUT-211	lncRNA	-3.11	1.84E-02	deoxyuridine triphosphatase
ENST00000619751	MIR7111-201	miRNA	-3.08	2.19E-03	microRNA 7111
ENST00000549183	TUBA1C-204	protein_coding	-3.08	4.89E-07	tubulin alpha 1c
ENST00000473924	DDX56-209	retained_intron	-3.06	2.50E-03	DEAD-box helicase 56
ENST00000458455	RPL11-203	protein_coding	-3.03	2.33E-05	ribosomal protein L11
ENST00000462613	SF3B1-206	lncRNA	-3.03	4.29E-02	splicing factor 3b subunit 1
ENST00000460860	ECT2-213	retained_intron	-3.02	3.67E-05	epithelial cell transforming 2
ENST00000521710	NPM1-209	retained_intron	-2.96	2.52E-04	nucleophosmin 1
ENST00000391130	RF00072.1-201	snoRNA	-2.96	8.88E-03	0
ENST00000454222	MTERF1-206	lncRNA	-2.94	1.65E-03	mitochondrial transcription termination factor 1
ENST00000531529	NUF2-211	retained_intron	-2.93	2.62E-02	NUF2 component of NDC80 kinetochore complex
ENST00000606420	SNORA51-201	snoRNA	-2.92	2.43E-02	small nucleolar RNA, H/ACA box 51
ENST00000363624	RF00019.154-201	misc_RNA	-2.89	2.16E-05	0
ENST00000483541	NSMCE4A-210	retained_intron	-2.89	2.44E-02	NSE4 homolog A, SMC5-SMC6 complex component
ENST00000485158	NUP188-206	retained_intron	-2.88	1.60E-02	nucleoporin 188
ENST00000548583	C12orf45-203	protein_coding	-2.88	2.82E-02	chromosome 12 open reading frame 45
ENST00000493798	BAD-205	protein_coding	-2.86	2.80E-03	BCL2 associated agonist of cell death

ENST00000478196	RPS27A-208	retained_intron	-2.85	8.71E-05	ribosomal protein S27a
ENST00000530086	ARRB1-207	retained_intron	-2.84	1.98E-02	arrestin beta 1
ENST00000468163	CEP85-204	lncRNA	-2.80	9.17E-03	centrosomal protein 85
ENST00000418764	AC004543.1-201	lncRNA	-2.78	6.75E-03	novel transcript
ENST00000460660	BZW1-210	retained_intron	-2.77	2.01E-02	basic leucine zipper and W2 domains 1
ENST00000494682	SFT2D1-206	lncRNA	-2.77	1.71E-02	SFT2 domain containing 1
ENST00000494277	DKK1-204	lncRNA	-2.74	6.49E-05	dickkopf WNT signaling pathway inhibitor 1
ENST00000584696	ZNF207-217	retained_intron	-2.72	1.06E-05	zinc finger protein 207
ENST00000459374	RF00019.638-201	misc_RNA	-2.71	4.35E-02	0
ENST00000469631	EZH2-204	retained_intron	-2.70	2.88E-03	enhancer of zeste 2 polycomb repressive complex 2 subunit
ENST00000428586	RPL21P40-201	processed_pseudogene	-2.70	1.15E-02	ribosomal protein L21 pseudogene 40
ENST00000534490	RPL38-207	protein_coding	-2.68	7.65E-03	ribosomal protein L38
ENST00000574158	POLR2A-203	retained_intron	-2.64	5.25E-03	RNA polymerase II subunit A
ENST00000482040	RN7SL809P-201	misc_RNA	-2.63	9.78E-06	RNA, 7SL, cytoplasmic 809, pseudogene
ENST00000614578	CENPF-206	protein_coding	-2.63	2.03E-03	centromere protein F
ENST00000584340	PSMD11-210	nonsense_mediated_decay	-2.62	1.22E-03	proteasome 26S subunit, non-ATPase 11
ENST00000555922	EGLN3-AS1-201	lncRNA	-2.62	5.86E-05	EGLN3 antisense RNA 1
ENST00000472516	DIAPH1-207	retained_intron	-2.61	1.58E-05	diaphanous related formin 1
ENST00000456545	PDLIM2-215	protein_coding	-2.59	1.53E-03	PDZ and LIM domain 2
ENST00000524244	PTTG1-207	retained_intron	-2.59	3.26E-02	PTTG1 regulator of sister chromatid separation, securin
ENST00000518935	DST-220	protein_coding	-2.58	1.25E-14	dystonin
ENST00000565723	RPL4-209	lncRNA	-2.58	1.34E-04	ribosomal protein L4
ENST00000565701	MEAK7-205	retained_intron	-2.58	3.65E-03	MTOR associated protein, eak-7 homolog
ENST00000471991	SFPQ-207	lncRNA	-2.58	2.48E-03	splicing factor proline and glutamine rich
ENST00000564276	PKM-211	retained_intron	-2.57	7.98E-05	pyruvate kinase M1/2
ENST00000583453	MIR4470-201	miRNA	-2.57	4.06E-02	microRNA 4470
ENST00000417726	ZNF143-207	protein_coding	-2.57	2.69E-02	zinc finger protein 143
ENST00000446556	COX7CP1-201	processed_pseudogene	-2.56	1.24E-02	cytochrome c oxidase subunit 7C pseudogene 1
ENST00000468885	GNL3-206	retained_intron	-2.55	9.88E-04	G protein nucleolar 3
ENST00000473022	MYH9-207	retained_intron	-2.55	3.58E-03	myosin heavy chain 9
ENST00000514610	HMGCS1-207	retained_intron	-2.54	6.07E-09	3-hydroxy-3-methylglutaryl-CoA synthase 1
ENST00000523071	AZIN1-210	retained_intron	-2.53	3.86E-05	antizyme inhibitor 1
ENST00000431653	RPL7-205	protein_coding	-2.51	5.65E-04	ribosomal protein L7
ENST00000480806	PRRC2C-209	lncRNA	-2.51	4.63E-02	proline rich coiled-coil 2C
ENST00000431026	AC009238.2-201	processed_pseudogene	-2.49	1.30E-02	pseudogene similar to part of Cdon homolog (mouse) (CDON)
ENST00000483047	SEPT10-212	retained_intron	-2.48	1.71E-02	septin 10
ENST00000364228	RNY1-201	misc_RNA	-2.46	1.30E-06	RNA, Ro60-associated Y1
ENST00000495111	AC015911.1-201	processed_pseudogene	-2.45	3.47E-02	ribosomal protein L39 (RPL39) pseudogene
ENST00000427794	APCDD1L-DT-204	lncRNA	-2.44	4.83E-02	APCDD1L divergent transcript
ENST00000638074	TBL1XR1-236	lncRNA	-2.42	1.53E-02	transducin beta like 1 X-linked receptor 1
ENST00000408612	SNORD99-201	snoRNA	-2.39	8.33E-03	small nucleolar RNA, C/D box 99
ENST00000474740	NDUFV2-205	lncRNA	-2.39	3.32E-02	NADH:ubiquinone oxidoreductase core subunit V2
ENST00000498025	IARS-212	lncRNA	-2.38	3.59E-04	isoleucyl-tRNA synthetase
ENST00000490763	SLC2A3-206	retained_intron	-2.38	1.51E-03	solute carrier family 2 member 3

ENST00000567969	ARL6IP1-207	retained_intron	-2.37	3.87E-05	ADP ribosylation factor like GTPase 6 interacting protein 1
ENST00000391305	SNORA3B-201	snoRNA	-2.35	1.62E-02	small nucleolar RNA, H/ACA box 3B
ENST00000443196	UPK1A-AS1-201	lncRNA	-2.34	8.40E-04	UPK1A antisense RNA 1
ENST00000529085	ZNF195-221	retained_intron	-2.34	4.80E-04	zinc finger protein 195
ENST00000383876	SNORA22B-201	snoRNA	-2.34	4.70E-03	small nucleolar RNA, H/ACA box 22B
ENST00000596494	FKBP8-205	retained_intron	-2.34	5.18E-05	FKBP prolyl isomerase 8
ENST00000364507	RNY4P23-201	misc_RNA	-2.33	3.03E-02	RNY4 pseudogene 23
ENST00000479143	CCAR1-203	nonsense_mediated_decay	-2.31	1.38E-03	cell division cycle and apoptosis regulator 1
ENST00000514928	MRPL27-207	retained_intron	-2.28	8.10E-05	mitochondrial ribosomal protein L27
ENST00000405331	TRIB2-203	protein_coding	-2.25	1.56E-02	tribbles pseudokinase 2
ENST00000524243	MTSS1-217	lncRNA	-2.25	2.94E-02	MTSS I-BAR domain containing 1
ENST00000553730	PPP2R5C-207	retained_intron	-2.25	7.51E-06	protein phosphatase 2 regulatory subunit B'gamma
ENST00000600533	PIK3R2-206	retained_intron	-2.23	2.67E-03	phosphoinositide-3-kinase regulatory subunit 2
ENST00000491646	CLASP1-214	retained_intron	-2.22	3.13E-03	cytoplasmic linker associated protein 1
ENST00000559647	ANXA2-223	retained_intron	-2.22	4.21E-07	annexin A2
ENST00000472832	PTEN-203	protein_coding	-2.22	1.60E-04	phosphatase and tensin homolog
ENST00000440417	AC073052.1-201	processed_pseudogene	-2.19	1.47E-03	ankyrin repeat domain 49 (ANKRD49) pseudogene
ENST00000620845	AC133540.1-201	lncRNA	3.98	3.52E-06	novel transcript
ENST00000395152	U73169.1-201	lncRNA	3.99	2.61E-06	novel transcript, sense intronic to GNAI2
ENST00000312310	COX6B1P3-201	processed_pseudogene	4.00	3.78E-04	cytochrome c oxidase subunit 6B1 pseudogene 3
ENST00000547455	CERS5-210	nonsense_mediated_decay	4.00	1.66E-05	ceramide synthase 5
ENST00000445956	BMP7-AS1-201	lncRNA	4.00	3.37E-05	BMP7 antisense RNA 1
ENST00000610177	LINC02604-201	lncRNA	4.00	3.29E-17	long intergenic non-protein coding RNA 2604
ENST00000613359	RNA5-8SN3-201	rRNA	4.00	6.01E-06	RNA, 5.8S ribosomal N3
ENST00000582772	MSI2-216	lncRNA	4.02	1.22E-04	musashi RNA binding protein 2
ENST00000560126	ABHD17C-204	lncRNA	4.04	3.93E-04	abhydrolase domain containing 17C
ENST00000544948	AP006333.2-201	lncRNA	4.04	3.44E-11	novel transcript, antisense to MACROD1
ENST00000384471	RNU6-45P-201	snRNA	4.06	1.92E-04	RNA, U6 small nuclear 45, pseudogene
ENST00000365219	RNU6-1306P-201	snRNA	4.06	1.71E-02	RNA, U6 small nuclear 1306, pseudogene
ENST00000587655	SBNO2-205	protein_coding	4.07	2.38E-05	strawberry notch homolog 2
ENST00000509695	WDR1-214	lncRNA	4.07	4.43E-05	WD repeat domain 1
ENST00000504625	AHRR-201	protein_coding	4.10	6.97E-07	aryl-hydrocarbon receptor repressor
ENST00000478223	SKI-202	lncRNA	4.10	5.49E-06	SKI proto-oncogene
ENST00000516599	RF00212.3-201	snoRNA	4.11	1.11E-03	0
ENST00000496478	RN7SL285P-201	misc_RNA	4.11	1.24E-04	RNA, 7SL, cytoplasmic 285, pseudogene
ENST00000563298	PPCDC-204	lncRNA	4.11	9.63E-04	phosphopantothenoylcysteine decarboxylase
ENST00000589397	AC002398.1-201	lncRNA	4.12	6.50E-11	novel transcript
ENST00000442115	AL137077.1-201	processed_pseudogene	4.13	4.15E-06	ribosomal protein L17 (RPL17) pseudogene
ENST00000591472	SEPT9-238	protein_coding	4.13	9.92E-08	septin 9
ENST00000563593	AC120498.1-201	lncRNA	4.15	6.57E-05	novel transcript, antisense to CACNA1H
ENST00000393264	Z97192.2-201	lncRNA	4.15	3.00E-12	novel transcript
ENST00000359574	LFNG-203	protein_coding	4.15	6.49E-10	LFNG O-fucosylpeptide 3-beta-N-acetylglucosaminyltransferase
ENST00000576309	CLUH-213	protein_coding	4.17	5.41E-06	clustered mitochondria homolog
ENST00000584169	LINC00334-202	lncRNA	4.17	5.91E-05	long intergenic non-protein coding RNA 334

ENST00000385248	MIR613-201	miRNA	4.18	7.31E-05	microRNA 613
ENST00000574688	BAIAP2-217	lncRNA	4.18	2.80E-05	BAI1 associated protein 2
ENST00000494918	GOSR1-209	retained_intron	4.18	8.81E-06	golgi SNAP receptor complex member 1
ENST00000581792	MIR3648-2-201	miRNA	4.19	1.38E-05	microRNA 3648-2
ENST00000495121	CABIN1-212	lncRNA	4.19	1.67E-06	calcineurin binding protein 1
ENST00000418612	LINC01219-201	lncRNA	4.19	8.53E-07	long intergenic non-protein coding RNA 1219
ENST00000428790	ASLP1-201	unprocessed_pseudogene	4.21	1.08E-08	argininosuccinate lyase pseudogene 1
ENST00000387392	MT-TA-201	Mt_tRNA	4.21	4.47E-16	mitochondrially encoded tRNA-Ala (GCN)
ENST00000635859	DNM1-218	lncRNA	4.21	2.08E-11	dynamin 1
ENST00000563249	AL138756.1-203	lncRNA	4.23	6.96E-05	novel transcript, overlapping SUSD1
ENST00000375350	NSD1-203	nonsense_mediated_decay	4.23	9.81E-05	nuclear receptor binding SET domain protein 1
ENST00000525891	AP2A2-206	lncRNA	4.25	1.64E-05	adaptor related protein complex 2 subunit alpha 2
ENST00000582040	CCDC57-214	lncRNA	4.26	6.14E-06	coiled-coil domain containing 57
ENST00000433041	BRD3-204	protein_coding	4.27	7.74E-05	bromodomain containing 3
ENST00000592069	MAU2-214	lncRNA	4.28	9.97E-06	MAU2 sister chromatid cohesion factor
ENST00000580500	LINC00511-209	lncRNA	4.31	2.00E-05	long intergenic non-protein coding RNA 511
ENST00000497663	TNRC18-210	retained_intron	4.32	9.05E-05	trinucleotide repeat containing 18
ENST00000543674	MARK2-215	protein_coding	4.32	9.97E-06	microtubule affinity regulating kinase 2
ENST00000636396	MIR941-1-201	miRNA	4.34	2.03E-04	microRNA 941-1
ENST00000625118	AC073592.3-201	TEC	4.35	2.86E-09	novel transcript
ENST00000361851	MT-ATP8-201	protein_coding	4.35	1.04E-05	mitochondrially encoded ATP synthase membrane subunit 8
ENST00000569006	BCAR1-221	protein_coding	4.36	7.69E-05	BCAR1 scaffold protein, Cas family member
ENST00000485207	FAM207A-205	lncRNA	4.39	2.56E-05	family with sequence similarity 207 member A
ENST00000534229	NAV2-214	lncRNA	4.39	6.75E-07	neuron navigator 2
ENST00000577097	BAIAP2-230	retained_intron	4.42	1.25E-06	BAI1 associated protein 2
ENST00000387461	MT-TP-201	Mt_tRNA	4.44	1.23E-46	mitochondrially encoded tRNA-Pro (CCN)
ENST00000568884	AC120498.3-201	lncRNA	4.44	1.77E-06	novel transcript, antisense to CACNA1H
ENST00000608521	MIR663AHG-229	lncRNA	4.45	3.31E-07	MIR663A host gene
ENST00000492975	FO393419.2-201	lncRNA	4.45	9.36E-06	novel transcript
ENST00000563659	WWP2-206	protein_coding	4.46	1.08E-06	WW domain containing E3 ubiquitin protein ligase 2
ENST00000620782	MIR6075-201	miRNA	4.51	4.00E-05	microRNA 6075
ENST00000614492	FP671120.1-201	miRNA	4.52	9.89E-19	0
ENST00000492234	AGAP3-223	lncRNA	4.52	4.91E-08	ArfGAP with GTPase domain, ankyrin repeat and PH domain 3
ENST00000568140	AC009054.1-201	lncRNA	4.54	2.48E-07	novel transcript, sense intronic to CFDP1
ENST00000612810	RF00017.142-201	misc_RNA	4.54	2.24E-05	0
ENST00000582210	MIR3176-201	miRNA	4.57	5.75E-10	microRNA 3176
ENST00000588954	ZNF440-206	protein_coding	4.58	3.87E-06	zinc finger protein 440
ENST00000428429	AL162727.2-201	lncRNA	4.58	3.70E-06	novel transcript
ENST00000538550	AP000721.2-202	lncRNA	4.60	5.27E-05	novel transcript, antisense to MACROD1
ENST00000361899	MT-ATP6-201	protein_coding	4.62	3.66E-62	mitochondrially encoded ATP synthase membrane subunit 6
ENST00000529229	AP003390.2-201	lncRNA	4.63	1.79E-06	novel transcript
ENST00000616367	RF00017.155-201	misc_RNA	4.67	6.65E-06	0

ENST00000361227	MT-ND3-201	protein_coding	4.71	7.35E-31	mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 3
ENST00000615609	RF00017.66-201	misc_RNA	4.75	3.32E-05	0
ENST00000570544	AC036164.1-201	processed_pseudogene	4.78	3.02E-08	chromosome 19 open reading frame 70 (C19orf70) pseudogene
ENST00000543422	MACROD1-207	lncRNA	4.79	7.66E-06	mono-ADP ribosylhydrolase 1
ENST00000528520	AP006259.1-201	lncRNA	4.84	2.45E-07	novel transcript
ENST00000625131	AL353662.3-201	pseudogene	4.88	2.69E-07	0
ENST00000578566	MIR3183-201	miRNA	4.92	3.58E-06	microRNA 3183
ENST00000624885	AC073592.9-201	TEC	4.98	1.96E-06	novel transcript
ENST00000577370	MIR5189-201	miRNA	5.01	2.40E-06	microRNA 5189
ENST00000508416	SKI-204	lncRNA	5.06	2.73E-09	SKI proto-oncogene
ENST00000457540	MTND2P28-201	unprocessed_pseudogene	5.06	4.51E-12	MT-ND2 pseudogene 28
ENST00000362079	MT-CO3-201	protein_coding	5.08	1.57E-120	mitochondrially encoded cytochrome c oxidase III
ENST00000361624	MT-CO1-201	protein_coding	5.10	1.54E-129	mitochondrially encoded cytochrome c oxidase I
ENST00000625310	AC006001.3-204	lncRNA	5.10	1.17E-07	RAB guanine nucleotide exchange factor (GEF) 1 (RABGEF1) pseudogene
ENST00000361335	MT-ND4L-201	protein_coding	5.27	1.06E-45	mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 4L
ENST00000387439	MT-TR-201	Mt_tRNA	5.28	1.38E-07	mitochondrially encoded tRNA-Arg (CGN)
ENST00000387400	MT-TN-201	Mt_tRNA	5.32	1.31E-69	mitochondrially encoded tRNA-Asn (AAU/C)
ENST00000387405	MT-TC-201	Mt_tRNA	5.36	1.15E-85	mitochondrially encoded tRNA-Cys (UGU/C)
ENST00000361567	MT-ND5-201	protein_coding	5.36	1.93E-121	mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 5
ENST00000361739	MT-CO2-201	protein_coding	5.46	9.21E-121	mitochondrially encoded cytochrome c oxidase II
ENST00000361453	MT-ND2-201	protein_coding	5.49	4.19E-47	mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 2
ENST00000361789	MT-CYB-201	protein_coding	5.52	1.69E-111	mitochondrially encoded cytochrome b
ENST00000361381	MT-ND4-201	protein_coding	5.64	6.11E-81	mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 4
ENST00000387460	MT-TT-201	Mt_tRNA	5.76	2.23E-10	mitochondrially encoded tRNA-Thr (ACN)
ENST00000361390	MT-ND1-201	protein_coding	5.76	1.37E-42	mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 1
ENST00000387441	MT-TH-201	Mt_tRNA	5.87	5.77E-09	mitochondrially encoded tRNA-His (CAU/C)
ENST00000414273	MTCO1P12-201	unprocessed_pseudogene	5.97	5.06E-11	MT-CO1 pseudogene 12
ENST00000587052	CEP295NL-203	retained_intron	6.03	1.54E-10	CEP295 N-terminal like
ENST00000387409	MT-TY-201	Mt_tRNA	6.05	8.66E-11	mitochondrially encoded tRNA-Tyr (UAU/C)
ENST00000387449	MT-TS2-201	Mt_tRNA	7.40	3.47E-31	mitochondrially encoded tRNA-Ser (AGU/C) 2
ENST00000578274	MIR4711-201	miRNA	15.34	4.06E-07	microRNA 4711

Appendix Table 4.7 Gene Ontology analysis (GO) of Exos and sMVs

(negative enrichment score indicates enriched GO terms in Exos, positive enrichment score indicates enriched GO terms in sMVs)

ID	Description	Normalised enrichment score	pvalue	rank
GO:0000788	nuclear nucleosome	-2.70	0.00275	514
GO:0000786	nucleosome	-2.93	0.00398	566
GO:0044815	DNA packaging complex	-2.92	0.00394	566
GO:0015934	large ribosomal subunit	-1.95	0.00275	483
GO:0032993	protein-DNA complex	-2.42	0.00450	566
GO:0044391	ribosomal subunit	-1.97	0.00316	242
GO:0044445	cytosolic part	-1.77	0.00676	171
GO:0042470	melanosome	-1.58	0.03812	85
GO:0048770	pigment granule	-1.58	0.03812	85
GO:0005840	ribosome	-1.73	0.00694	242
GO:0000790	nuclear chromatin	-1.73	0.00758	514
GO:0000785	chromatin	-1.89	0.00559	566
GO:1990904	ribonucleoprotein complex	-1.57	0.00645	242
GO:0044427	chromosomal part	-1.50	0.00694	566
GO:0070062	extracellular exosome	-1.30	0.04762	448
GO:0043230	extracellular organelle	-1.30	0.04839	448
GO:1903561	extracellular vesicle	-1.30	0.04839	448
GO:0005615	extracellular space	-1.30	0.02083	448
GO:0044421	extracellular region part	-1.24	0.02381	448
GO:0005576	extracellular region	-1.24	0.02500	448
GO:0005886	plasma membrane	1.27	0.04262	657
GO:0071944	cell periphery	1.32	0.02178	657
GO:0031090	organelle membrane	1.37	0.01360	452
GO:0042995	cell projection	1.39	0.03571	352
GO:0005783	endoplasmic reticulum	1.43	0.01980	757
GO:0030054	cell junction	1.39	0.04452	352
GO:1902494	catalytic complex	1.45	0.02123	45
GO:0120025	plasma membrane bounded cell projection	1.45	0.02354	390
GO:0044432	endoplasmic reticulum part	1.55	0.01022	924
GO:0005739	mitochondrion	1.61	0.00332	88
GO:0043005	neuron projection	1.58	0.01296	417
GO:0044459	plasma membrane part	1.71	0.00111	657
GO:0098590	plasma membrane region	1.65	0.00596	484
GO:0009986	cell surface	1.50	0.03748	657
GO:0045202	synapse	1.79	0.00352	441
GO:0097458	neuron part	1.87	0.00114	417
GO:0098978	glutamatergic synapse	1.55	0.02846	405
GO:0031967	organelle envelope	1.93	0.00114	88
GO:0031975	envelope	1.93	0.00114	88
GO:0031224	intrinsic component of membrane	2.17	0.00104	762

GO:0098794	postsynapse	1.79	0.00518	405
GO:0016021	integral component of membrane	2.21	0.00104	762
GO:0044456	synapse part	1.94	0.00122	413
GO:0044429	mitochondrial part	2.05	0.00117	88
GO:0045211	postsynaptic membrane	1.52	0.04734	405
GO:0031256	leading edge membrane	1.53	0.04819	167
GO:0031300	intrinsic component of organelle membrane	1.62	0.02346	618
GO:0098793	presynapse	1.83	0.00278	520
GO:0099572	postsynaptic specialization	1.89	0.00278	364
GO:0043235	receptor complex	1.64	0.02439	625
GO:0031226	intrinsic component of plasma membrane	2.18	0.00127	653
GO:0005887	integral component of plasma membrane	2.19	0.00127	653
GO:0098796	membrane protein complex	2.31	0.00121	308
GO:0097060	synaptic membrane	1.88	0.00571	405
GO:0005788	endoplasmic reticulum lumen	1.95	0.00280	657
GO:0031301	integral component of organelle membrane	1.80	0.01183	500
GO:0005740	mitochondrial envelope	2.37	0.00121	88
GO:0031966	mitochondrial membrane	2.36	0.00122	88
GO:0014069	postsynaptic density	2.05	0.00142	364
GO:0032279	asymmetric synapse	2.05	0.00142	364
GO:0098984	neuron to neuron synapse	2.18	0.00139	364
GO:0019866	organelle inner membrane	2.52	0.00127	88
GO:0005743	mitochondrial inner membrane	2.58	0.00128	88
GO:0098798	mitochondrial protein complex	2.69	0.00135	12
GO:0044455	mitochondrial membrane part	2.75	0.00142	88
GO:0098800	inner mitochondrial membrane protein complex	2.81	0.00148	12
GO:0070469	respiratory chain	2.98	0.00152	24
GO:0098803	respiratory chain complex	2.89	0.00157	12

Appendix Table 4.8 Gene Ontology analysis (GO) of Exos and sMB-Rs

(negative enrichment score indicates enriched GO terms in Exos, positive enrichment score indicates enriched GO terms in sMB-Rs)

ID	Description	Normalised enrichment score	pvalue	rank
GO:0097458	neuron part	1.54	0.00101	1022
GO:0044456	synapse part	1.70	0.00103	932
GO:0098796	membrane protein complex	1.68	0.00103	1023
GO:0031966	mitochondrial membrane	1.59	0.00104	22
GO:0019866	organelle inner membrane	1.88	0.00105	22
GO:0005743	mitochondrial inner membrane	1.99	0.00106	22
GO:0005887	integral component of plasma membrane	1.64	0.00106	1439
GO:0098984	neuron to neuron synapse	1.80	0.00109	913
GO:0098798	mitochondrial protein complex	2.21	0.00112	11
GO:0044455	mitochondrial membrane part	2.44	0.00118	11
GO:0098800	inner mitochondrial membrane protein complex	2.64	0.00124	11
GO:1990204	oxidoreductase complex	2.61	0.00130	10
GO:0070469	respiratory chain	2.91	0.00131	22
GO:0005746	mitochondrial respiratory chain	2.71	0.00134	10
GO:0098803	respiratory chain complex	2.88	0.00134	10
GO:0045202	synapse	1.54	0.00202	1166
GO:0005740	mitochondrial envelope	1.59	0.00207	22
GO:0098794	postsynapse	1.58	0.00211	932
GO:0099240	intrinsic component of synaptic membrane	1.89	0.00275	866
GO:0000788	nuclear nucleosome	-3.04	0.00362	921
GO:0099572	postsynaptic specialization	1.68	0.00441	781
GO:0032279	asymmetric synapse	1.67	0.00443	1145
GO:0022626	cytosolic ribosome	-1.86	0.00503	1164
GO:0044459	plasma membrane part	1.45	0.00503	1426
GO:0015934	large ribosomal subunit	-1.75	0.00526	656
GO:0031226	intrinsic component of plasma membrane	1.55	0.00530	1439
GO:0000786	nucleosome	-3.13	0.00654	976
GO:0014069	postsynaptic density	1.67	0.00672	1145
GO:0044815	DNA packaging complex	-3.08	0.00709	976
GO:0044448	cell cortex part	1.79	0.00778	1032
GO:0044391	ribosomal subunit	-1.77	0.00781	1190
GO:0022625	cytosolic large ribosomal subunit	-1.87	0.00781	1258
GO:0098978	glutamatergic synapse	1.62	0.00868	1166
GO:0098590	plasma membrane region	1.48	0.00921	1204
GO:0032993	protein-DNA complex	-2.38	0.01176	994
GO:0042641	actomyosin	1.67	0.01228	1149
GO:0005938	cell cortex	1.59	0.01253	426
GO:0097525	spliceosomal snRNP complex	-1.65	0.01569	130
GO:0005667	transcription factor complex	1.53	0.01774	971
GO:0043005	neuron projection	1.41	0.01945	1461
GO:0000784	nuclear chromosome, telomeric region	-1.48	0.02010	521
GO:0120025	plasma membrane bounded cell projection	1.32	0.02104	1209

GO:0098793	presynapse	1.50	0.02215	913
GO:0044463	cell projection part	1.35	0.02222	1021
GO:0120038	plasma membrane bounded cell projection part	1.35	0.02222	1021
GO:0005840	ribosome	-1.33	0.02247	883
GO:0043235	receptor complex	1.58	0.02594	1145
GO:0031256	leading edge membrane	1.57	0.02618	1204
GO:0097060	synaptic membrane	1.48	0.03167	1158
GO:0042995	cell projection	1.28	0.03213	1209
GO:0042734	presynaptic membrane	1.60	0.03279	1022
GO:0030532	small nuclear ribonucleoprotein complex	-1.50	0.03361	130
GO:0030054	cell junction	1.32	0.03636	1215
GO:0017053	transcriptional repressor complex	1.53	0.03747	1038
GO:0010494	cytoplasmic stress granule	1.55	0.03878	536
GO:0030670	phagocytic vesicle membrane	-1.53	0.03929	1080
GO:0009986	cell surface	1.44	0.03943	1166
GO:0044445	cytosolic part	-1.30	0.04237	1164
GO:0005911	cell-cell junction	1.43	0.04239	1337
GO:0098797	plasma membrane protein complex	1.48	0.04327	1623
GO:0000781	chromosome, telomeric region	-1.33	0.04706	521

Appendix Table 4.9 Differential transcript expression analysis of sMVs and sMB-Rs (negative log2foldchange indicates enriched transcripts in sMVs, positive log2foldchange indicates enriched transcripts in sMB-Rs)

Transcript ID	Transcript name	Transcript biotypes	log2Fold change	pvalue	Description
ENST00000466217	ATG9A-218	retained_intron	-3.92	3.42E-04	autophagy related 9A [Source:HGNC Symbol;Acc:HGNC:22408]
ENST00000486622	RNF144B-202	lncRNA	-3.36	4.41E-03	ring finger protein 144B [Source:HGNC Symbol;Acc:HGNC:21578]
ENST00000559597	RCOR1-203	retained_intron	-3.28	3.12E-03	REST corepressor 1 [Source:HGNC Symbol;Acc:HGNC:17441]
ENST00000470259	ITGA6-209	retained_intron	-3.12	1.38E-03	integrin subunit alpha 6 [Source:HGNC Symbol;Acc:HGNC:6142]
ENST00000417726	ZNF143-207	protein_coding	-3.09	6.07E-03	zinc finger protein 143 [Source:HGNC Symbol;Acc:HGNC:12928]
ENST00000428586	RPL21P40-201	processed_pseudogene	-3.01	4.31E-03	ribosomal protein L21 pseudogene 40 [Source:HGNC Symbol;Acc:HGNC:36189]
ENST00000384028	SNORA30-201	snoRNA	-2.88	2.78E-02	small nucleolar RNA, H/ACA box 30 [Source:HGNC Symbol;Acc:HGNC:32620]
ENST00000532146	CELF1-216	retained_intron	-2.86	2.64E-03	CUGBP Elav-like family member 1 [Source:HGNC Symbol;Acc:HGNC:2549]
ENST00000611125	MYO19-206	retained_intron	-2.85	5.85E-03	myosin XIX [Source:HGNC Symbol;Acc:HGNC:26234]
ENST00000460660	BZW1-210	retained_intron	-2.84	1.64E-02	basic leucine zipper and W2 domains 1 [Source:HGNC Symbol;Acc:HGNC:18380]
ENST00000494907	EXTL2-206	lncRNA	-2.82	2.86E-03	exostosin like glycosyltransferase 2 [Source:HGNC Symbol;Acc:HGNC:3516]
ENST00000524243	MTSS1-217	lncRNA	-2.75	6.52E-03	MTSS I-BAR domain containing 1 [Source:HGNC Symbol;Acc:HGNC:20443]
ENST00000561350	DUT-211	lncRNA	-2.74	4.07E-02	deoxyuridine triphosphatase [Source:HGNC Symbol;Acc:HGNC:3078]
ENST00000637681	TBL1XR1-234	lncRNA	-2.67	3.16E-02	transducin beta like 1 X-linked receptor 1 [Source:HGNC Symbol;Acc:HGNC:29529]
ENST00000530086	ARRB1-207	retained_intron	-2.61	3.26E-02	arrestin beta 1 [Source:HGNC Symbol;Acc:HGNC:711]
ENST00000491326	SYTL4-205	lncRNA	-2.59	1.57E-02	synaptotagmin like 4 [Source:HGNC Symbol;Acc:HGNC:15588]
ENST00000473924	DDX56-209	retained_intron	-2.55	1.31E-02	DEAD-box helicase 56 [Source:HGNC Symbol;Acc:HGNC:18193]
ENST00000516176	RNU6-1161P-201	snRNA	-2.53	1.02E-02	RNA, U6 small nuclear 1161, pseudogene [Source:HGNC Symbol;Acc:HGNC:48124]
ENST00000460806	BHLHE40-202	retained_intron	-2.50	2.04E-02	basic helix-loop-helix family member e40 [Source:HGNC Symbol;Acc:HGNC:1046]
ENST00000446556	COX7CP1-201	processed_pseudogene	-2.48	1.54E-02	cytochrome c oxidase subunit 7C pseudogene 1 [Source:HGNC Symbol;Acc:HGNC:2293]
ENST00000391130	RF00072.1-201	snoRNA	-2.43	3.46E-02	0
ENST00000497697	SMTN-220	retained_intron	-2.39	2.09E-02	smoothelin [Source:HGNC Symbol;Acc:HGNC:11126]
ENST00000564344	RSL24D1-204	lncRNA	-2.36	3.72E-02	ribosomal L24 domain containing 1 [Source:HGNC Symbol;Acc:HGNC:18479]
ENST00000363624	RF00019.154-201	misc_RNA	-2.35	6.75E-04	0
ENST00000553453	SIPA1L1-204	lncRNA	-2.34	1.40E-02	signal induced proliferation associated 1 like 1 [Source:HGNC Symbol;Acc:HGNC:20284]
ENST00000596494	FKBP8-205	retained_intron	-2.30	6.53E-05	FKBP prolyl isomerase 8 [Source:HGNC Symbol;Acc:HGNC:3724]

ENST00000557187	TTC7B-213	lncRNA	-2.24	1.33E-03	tetratricopeptide repeat domain 7B [Source:HGNC Symbol;Acc:HGNC:19858]
ENST00000571562	NPLOC4-205	lncRNA	-2.21	3.21E-02	NPL4 homolog, ubiquitin recognition factor [Source:HGNC Symbol;Acc:HGNC:18261]
ENST00000609153	AC097534.2-202	lncRNA	-2.16	5.88E-04	novel transcript, antisense to SAP30
ENST00000494277	DKK1-204	lncRNA	-2.15	2.11E-03	dickkopf WNT signaling pathway inhibitor 1 [Source:HGNC Symbol;Acc:HGNC:2891]
ENST00000365484	RNY3-201	misc_RNA	-2.14	4.11E-03	RNA, Ro60-associated Y3 [Source:HGNC Symbol;Acc:HGNC:10243]
ENST00000542474	EMP1-210	protein_coding	-2.13	3.47E-02	epithelial membrane protein 1 [Source:HGNC Symbol;Acc:HGNC:3333]
ENST00000540747	YBX3-209	lncRNA	-2.10	7.30E-03	Y-box binding protein 3 [Source:HGNC Symbol;Acc:HGNC:2428]
ENST00000483047	SEPT10-212	retained_intron	-2.10	4.64E-02	septin 10 [Source:HGNC Symbol;Acc:HGNC:14349]
ENST00000418764	AC004543.1-201	lncRNA	-2.10	4.75E-02	novel transcript
ENST00000496205	CLK1-214	retained_intron	-2.05	1.60E-02	CDC like kinase 1 [Source:HGNC Symbol;Acc:HGNC:2068]
ENST00000431026	AC009238.2-201	processed_pseudogene	-2.04	4.45E-02	pseudogene similar to part of Cdon homolog (mouse) (CDON)
ENST00000495160	HESX1-203	protein_coding	-2.04	9.53E-04	HESX homeobox 1 [Source:HGNC Symbol;Acc:HGNC:4877]
ENST00000517404	RMDN1-203	retained_intron	-2.01	1.66E-02	regulator of microtubule dynamics 1 [Source:HGNC Symbol;Acc:HGNC:24285]
ENST00000584696	ZNF207-217	retained_intron	-2.01	1.49E-03	zinc finger protein 207 [Source:HGNC Symbol;Acc:HGNC:12998]
ENST00000412563	AC007091.1-201	lncRNA	-2.00	6.51E-03	novel transcript
ENST00000618460	AL691403.2-201	lncRNA	-1.94	1.94E-02	novel transcript
ENST00000544780	CCDC91-219	lncRNA	-1.93	1.49E-03	coiled-coil domain containing 91 [Source:HGNC Symbol;Acc:HGNC:24855]
ENST00000555724	POLE2-209	lncRNA	-1.91	4.62E-02	DNA polymerase epsilon 2, accessory subunit [Source:HGNC Symbol;Acc:HGNC:9178]
ENST00000456545	PDLIM2-215	protein_coding	-1.91	2.24E-02	PDZ and LIM domain 2 [Source:HGNC Symbol;Acc:HGNC:13992]
ENST00000410489	RF00019.599-201	misc_RNA	-1.88	8.49E-03	0
ENST00000479651	CAPG-213	lncRNA	-1.85	4.37E-02	capping actin protein, gelsolin like [Source:HGNC Symbol;Acc:HGNC:1474]
ENST00000442746	SEPT10-209	protein_coding	-1.85	2.65E-04	septin 10 [Source:HGNC Symbol;Acc:HGNC:14349]
ENST00000520625	ASAP1-208	lncRNA	-1.79	5.36E-03	ArfGAP with SH3 domain, ankyrin repeat and PH domain 1 [Source:HGNC Symbol;Acc:HGNC:2720]
ENST00000449056	MPHOSPH8-203	protein_coding	-1.77	1.14E-02	M-phase phosphoprotein 8 [Source:HGNC Symbol;Acc:HGNC:29810]
ENST00000628181	CHD2-214	lncRNA	-1.75	3.46E-02	chromodomain helicase DNA binding protein 2 [Source:HGNC Symbol;Acc:HGNC:1917]
ENST00000446547	RAB5A-205	nonsense_mediated_decay	-1.75	4.13E-04	RAB5A, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9783]
ENST00000555662	HSP90AA1-205	retained_intron	-1.73	5.54E-03	heat shock protein 90 alpha family class A member 1 [Source:HGNC Symbol;Acc:HGNC:5253]
ENST00000383861	RNU1-28P-201	snRNA	-1.68	3.30E-04	RNA, U1 small nuclear 28, pseudogene [Source:HGNC Symbol;Acc:HGNC:37498]
ENST00000384158	SNORA75-201	snoRNA	-1.65	1.17E-02	small nucleolar RNA, H/ACA box 75 [Source:HGNC Symbol;Acc:HGNC:32661]
ENST00000635481	AC110994.2-201	processed_pseudogene	-1.65	2.05E-02	basic transcription factor 3 (BTF3) pseudogene
ENST00000529085	ZNF195-221	retained_intron	-1.64	1.75E-02	zinc finger protein 195 [Source:HGNC Symbol;Acc:HGNC:12986]

ENST00000610252	AC009237.15-201	lncRNA	-1.63	4.53E-03	novel transcript
ENST00000644716	GRB10-223	lncRNA	-1.63	1.39E-02	growth factor receptor bound protein 10 [Source:HGNC Symbol;Acc:HGNC:4564]
ENST00000488706	HKDC1-203	lncRNA	-1.62	2.52E-02	hexokinase domain containing 1 [Source:HGNC Symbol;Acc:HGNC:23302]
ENST00000560546	ANXA2-235	retained_intron	-1.62	3.59E-04	annexin A2 [Source:HGNC Symbol;Acc:HGNC:537]
ENST00000545388	IST1-222	retained_intron	-1.61	1.64E-02	IST1 factor associated with ESCRT-III [Source:HGNC Symbol;Acc:HGNC:28977]
ENST00000584906	RPL26-211	nonsense_mediated_decay	-1.61	3.88E-02	ribosomal protein L26 [Source:HGNC Symbol;Acc:HGNC:10327]
ENST00000555922	EGLN3-AS1-201	lncRNA	-1.60	1.70E-02	EGLN3 antisense RNA 1 [Source:HGNC Symbol;Acc:HGNC:49077]
ENST00000565723	RPL4-209	lncRNA	-1.60	2.33E-02	ribosomal protein L4 [Source:HGNC Symbol;Acc:HGNC:10353]
ENST00000443196	UPK1A-AS1-201	lncRNA	-1.60	2.72E-02	UPK1A antisense RNA 1 [Source:HGNC Symbol;Acc:HGNC:40603]
ENST00000460860	ECT2-213	retained_intron	-1.59	4.03E-02	epithelial cell transforming 2 [Source:HGNC Symbol;Acc:HGNC:3155]
ENST00000478348	S100A10-203	lncRNA	-1.59	6.00E-03	S100 calcium binding protein A10 [Source:HGNC Symbol;Acc:HGNC:10487]
ENST00000477189	MYH9-209	retained_intron	-1.58	3.55E-02	myosin heavy chain 9 [Source:HGNC Symbol;Acc:HGNC:7579]
ENST00000504511	STIM2-210	retained_intron	-1.58	2.60E-02	stromal interaction molecule 2 [Source:HGNC Symbol;Acc:HGNC:19205]
ENST00000490763	SLC2A3-206	retained_intron	-1.56	4.46E-02	solute carrier family 2 member 3 [Source:HGNC Symbol;Acc:HGNC:11007]
ENST00000363485	RF00139.1-201	snoRNA	-1.55	7.15E-03	0
ENST00000531421	RRAS2-206	protein_coding	-1.54	4.10E-05	RAS related 2 [Source:HGNC Symbol;Acc:HGNC:17271]
ENST00000576840	DVL2-213	retained_intron	-1.54	3.20E-05	dishevelled segment polarity protein 2 [Source:HGNC Symbol;Acc:HGNC:3086]
ENST00000584500	DDX5-229	lncRNA	-1.54	1.21E-03	DEAD-box helicase 5 [Source:HGNC Symbol;Acc:HGNC:2746]
ENST00000466800	AC133134.1-201	processed_pseudogene	-1.54	4.16E-02	ribosomal protein S20 (RPS20) pseudogene
ENST00000414896	AC005104.1-201	lncRNA	-1.54	2.74E-02	novel transcript
ENST00000609131	CDC37L1-DT-202	lncRNA	-1.53	9.41E-03	CDC37L1 divergent transcript [Source:HGNC Symbol;Acc:HGNC:49735]
ENST00000412485	LINC02154-201	lncRNA	-1.52	3.83E-04	long intergenic non-protein coding RNA 2154 [Source:HGNC Symbol;Acc:HGNC:53015]
ENST00000482040	RN7SL809P-201	misc_RNA	-1.52	1.48E-02	RNA, 7SL, cytoplasmic 809, pseudogene [Source:HGNC Symbol;Acc:HGNC:46825]
ENST00000510645	DHX15-206	lncRNA	-1.52	1.77E-02	DEAH-box helicase 15 [Source:HGNC Symbol;Acc:HGNC:2738]
ENST00000549183	TUBA1C-204	protein_coding	-1.51	1.87E-02	tubulin alpha 1c [Source:HGNC Symbol;Acc:HGNC:20768]
ENST00000578224	THOC1-204	retained_intron	-1.51	3.01E-02	THO complex 1 [Source:HGNC Symbol;Acc:HGNC:19070]
ENST00000515280	FAM135A-215	protein_coding	-1.49	8.84E-03	family with sequence similarity 135 member A [Source:HGNC Symbol;Acc:HGNC:21084]
ENST00000567649	NPW-203	protein_coding	-1.49	1.62E-02	neuropeptide W [Source:HGNC Symbol;Acc:HGNC:30509]
ENST00000532624	IMMP1L-210	lncRNA	-1.48	2.83E-03	inner mitochondrial membrane peptidase subunit 1 [Source:HGNC Symbol;Acc:HGNC:26317]
ENST00000402380	ATXN10-203	protein_coding	-1.46	5.18E-06	ataxin 10 [Source:HGNC Symbol;Acc:HGNC:10549]
ENST00000477802	DHX9-204	lncRNA	-1.45	9.78E-03	DExH-box helicase 9 [Source:HGNC Symbol;Acc:HGNC:2750]

ENST00000534470	EIF4G2-230	retained_intron	-1.44	1.59E-02	eukaryotic translation initiation factor 4 gamma 2 [Source:HGNC Symbol;Acc:HGNC:3297]
ENST00000415536	AC003092.1-201	lncRNA	-1.44	3.62E-02	novel transcript
ENST00000502839	LRBA-202	protein_coding	-1.43	2.63E-04	LPS responsive beige-like anchor protein [Source:HGNC Symbol;Acc:HGNC:1742]
ENST00000496969	SF1-219	lncRNA	-1.43	4.66E-02	splicing factor 1 [Source:HGNC Symbol;Acc:HGNC:12950]
ENST00000626934	HERC2P7-202	unprocessed_pseudogene	-1.42	1.50E-03	hect domain and RLD 2 pseudogene 7 [Source:HGNC Symbol;Acc:HGNC:4875]
ENST00000547479	NAP1L1-207	protein_coding	-1.41	3.72E-02	nucleosome assembly protein 1 like 1 [Source:HGNC Symbol;Acc:HGNC:7637]
ENST00000394536	PCDH1-203	protein_coding	-1.40	8.40E-06	protocadherin 1 [Source:HGNC Symbol;Acc:HGNC:8655]
ENST00000549261	LINC02373-201	lncRNA	-1.39	7.84E-04	long intergenic non-protein coding RNA 2373 [Source:HGNC Symbol;Acc:HGNC:53295]
ENST00000491395	APP-216	lncRNA	-1.38	3.38E-02	amyloid beta precursor protein [Source:HGNC Symbol;Acc:HGNC:620]
ENST00000568539	SF3B3-212	retained_intron	-1.38	4.21E-02	splicing factor 3b subunit 3 [Source:HGNC Symbol;Acc:HGNC:10770]
ENST00000490188	BCS1L-219	retained_intron	-1.37	3.52E-02	BCS1 homolog, ubiquinol-cytochrome c reductase complex chaperone [Source:HGNC Symbol;Acc:HGNC:1020]
ENST00000552030	CORO1C-218	lncRNA	-1.36	3.04E-03	coronin 1C [Source:HGNC Symbol;Acc:HGNC:2254]
ENST00000511193	ZDHHC11-208	protein_coding	2.70	2.82E-02	zinc finger DHHC-type containing 11 [Source:HGNC Symbol;Acc:HGNC:19158]
ENST00000646046	AL449043.2-201	transcribed_unprocessed_pseudogene	2.70	9.67E-03	phosphoglucomutase 5 (PGM5) pseudogene
ENST00000594445	SUGP2-205	protein_coding	2.71	1.50E-04	SURP and G-patch domain containing 2 [Source:HGNC Symbol;Acc:HGNC:18641]
ENST00000516241	RF00322.6-201	snoRNA	2.72	2.78E-02	0
ENST00000364251	RNY1P12-201	misc_RNA	2.72	1.28E-02	RNY1 pseudogene 12 [Source:HGNC Symbol;Acc:HGNC:50875]
ENST00000492234	AGAP3-223	lncRNA	2.72	2.50E-05	ArfGAP with GTPase domain, ankyrin repeat and PH domain 3 [Source:HGNC Symbol;Acc:HGNC:16923]
ENST00000220003	CSK-201	protein_coding	2.73	2.30E-16	C-terminal Src kinase [Source:HGNC Symbol;Acc:HGNC:2444]
ENST00000418851	AC008280.1-201	processed_pseudogene	2.73	8.08E-03	ribosomal protein L21 (RPL21) pseudogene
ENST00000320521	SLC16A8-201	protein_coding	2.74	7.54E-06	solute carrier family 16 member 8 [Source:HGNC Symbol;Acc:HGNC:16270]
ENST00000484625	AC093663.2-201	processed_pseudogene	2.74	4.63E-02	ribosomal protein L36a (RPL36A) pseudogene
ENST00000625118	AC073592.3-201	TEC	2.74	1.40E-06	novel transcript
ENST00000644787	SOCS1-202	protein_coding	2.75	3.64E-08	suppressor of cytokine signaling 1 [Source:HGNC Symbol;Acc:HGNC:19383]
ENST00000445833	AL133260.2-201	lncRNA	2.75	9.15E-08	novel transcript
ENST00000387409	MT-TY-201	Mt_tRNA	2.75	1.07E-07	mitochondrially encoded tRNA-Tyr (UAU/C) [Source:HGNC Symbol;Acc:HGNC:7502]
ENST00000560906	AC013553.1-201	processed_pseudogene	2.77	1.42E-02	ribosomal protein L17 (RPL17) pseudogene
ENST00000596631	AC245884.9-201	lncRNA	2.78	2.28E-02	novel transcript, sense intronic to TTYH1
ENST00000488586	MZT2A-205	lncRNA	2.79	3.30E-03	mitotic spindle organizing protein 2A [Source:HGNC Symbol;Acc:HGNC:33187]
ENST00000586713	CEP295NL-202	protein_coding	2.80	2.36E-03	CEP295 N-terminal like [Source:HGNC Symbol;Acc:HGNC:44659]
ENST00000620071	RF00017.205-201	misc_RNA	2.80	1.21E-02	0

ENST00000485207	FAM207A-205	lncRNA	2.80	1.65E-03	family with sequence similarity 207 member A [Source:HGNC Symbol;Acc:HGNC:15811]
ENST00000586521	SEPT9-215	protein_coding	2.81	3.71E-06	septin 9 [Source:HGNC Symbol;Acc:HGNC:7323]
ENST00000443837	RNF216-IT1-201	lncRNA	2.81	1.53E-04	RNF216 intronic transcript 1 [Source:HGNC Symbol;Acc:HGNC:41463]
ENST00000544948	AP006333.2-201	lncRNA	2.81	1.39E-09	novel transcript, antisense to MACROD1
ENST00000450968	AL161672.1-201	processed_pseudogene	2.82	1.87E-04	novel pseudogene
ENST00000584169	LINC00334-202	lncRNA	2.83	1.07E-03	long intergenic non-protein coding RNA 334 [Source:HGNC Symbol;Acc:HGNC:16425]
ENST00000453213	LINC00963-206	lncRNA	2.83	2.05E-02	long intergenic non-protein coding RNA 963 [Source:HGNC Symbol;Acc:HGNC:48716]
ENST00000560011	AC020891.3-201	lncRNA	2.83	6.81E-03	novel transcript, intronic to CYP19A1
ENST00000387456	MT-TL2-201	Mt_tRNA	2.83	4.97E-02	mitochondrially encoded tRNA-Leu (CUN) 2 [Source:HGNC Symbol;Acc:HGNC:7491]
ENST00000635619	LINC01346-203	lncRNA	2.83	2.18E-03	long intergenic non-protein coding RNA 1346 [Source:HGNC Symbol;Acc:HGNC:50563]
ENST00000429246	ARPC1B-205	protein_coding	2.85	3.09E-05	actin related protein 2/3 complex subunit 1B [Source:HGNC Symbol;Acc:HGNC:704]
ENST00000546514	CERS5-207	nonsense_mediated_decay	2.85	1.70E-04	ceramide synthase 5 [Source:HGNC Symbol;Acc:HGNC:23749]
ENST00000495282	NR2C2-212	retained_intron	2.85	8.91E-03	nuclear receptor subfamily 2 group C member 2 [Source:HGNC Symbol;Acc:HGNC:7972]
ENST00000414464	AC018638.1-201	processed_pseudogene	2.85	1.67E-32	novel pseudogene
ENST00000384195	RNU6-80P-201	snRNA	2.86	2.66E-02	RNA, U6 small nuclear 80, pseudogene [Source:HGNC Symbol;Acc:HGNC:42570]
ENST00000623341	AC010531.8-201	TEC	2.87	2.27E-03	novel transcript
ENST00000637467	MECP2-221	lncRNA	2.89	3.52E-02	methyl-CpG binding protein 2 [Source:HGNC Symbol;Acc:HGNC:6990]
ENST00000613285	AC005391.1-201	lncRNA	2.89	9.50E-05	novel transcript, sense intronic to ARID3A
ENST00000362168	MIR210-201	miRNA	2.89	6.23E-03	microRNA 210 [Source:HGNC Symbol;Acc:HGNC:31587]
ENST00000615959	MIR3648-1-201	miRNA	2.90	5.86E-11	microRNA 3648-1 [Source:HGNC Symbol;Acc:HGNC:38941]
ENST00000483893	CERCAM-210	lncRNA	2.92	3.74E-03	cerebral endothelial cell adhesion molecule [Source:HGNC Symbol;Acc:HGNC:23723]
ENST00000587052	CEP295NL-203	retained_intron	2.92	2.88E-06	CEP295 N-terminal like [Source:HGNC Symbol;Acc:HGNC:44659]
ENST00000474396	GPR146-204	lncRNA	2.92	7.69E-04	G protein-coupled receptor 146 [Source:HGNC Symbol;Acc:HGNC:21718]
ENST00000494134	RN7SL653P-201	misc_RNA	2.93	3.95E-02	RNA, 7SL, cytoplasmic 653, pseudogene [Source:HGNC Symbol;Acc:HGNC:46669]
ENST00000400376	TTYH3-202	protein_coding	2.93	1.63E-09	tweety family member 3 [Source:HGNC Symbol;Acc:HGNC:22222]
ENST00000581792	MIR3648-2-201	miRNA	2.93	2.92E-04	microRNA 3648-2 [Source:HGNC Symbol;Acc:HGNC:50843]
ENST00000568399	RSRP1-218	lncRNA	2.94	6.39E-03	arginine and serine rich protein 1 [Source:HGNC Symbol;Acc:HGNC:25234]
ENST00000455552	PTMAP1-205	processed_pseudogene	2.94	1.87E-03	prothymosin alpha pseudogene 1 [Source:HGNC Symbol;Acc:HGNC:9624]
ENST00000635859	DNM1-218	lncRNA	2.94	2.35E-09	dynamin 1 [Source:HGNC Symbol;Acc:HGNC:2972]
ENST00000589397	AC002398.1-201	lncRNA	2.95	2.65E-09	novel transcript
ENST00000494918	GOSR1-209	retained_intron	2.96	1.45E-04	golgi SNAP receptor complex member 1 [Source:HGNC Symbol;Acc:HGNC:4430]

ENST00000582772	MSI2-216	lncRNA	2.97	1.49E-03	musashi RNA binding protein 2 [Source:HGNC Symbol;Acc:HGNC:18585]
ENST00000562466	AC092720.1-201	lncRNA	2.97	1.13E-06	novel transcript
ENST00000480095	RN7SL698P-201	misc_RNA	2.98	1.82E-03	RNA, 7SL, cytoplasmic 698, pseudogene [Source:HGNC Symbol;Acc:HGNC:46714]
ENST00000563593	AC120498.1-201	lncRNA	3.00	8.67E-04	novel transcript, antisense to CACNA1H
ENST00000423240	HSPB1P1-201	processed_pseudogene	3.02	1.07E-02	heat shock protein family B (small) member 1 pseudogene 1 [Source:HGNC Symbol;Acc:HGNC:5251]
ENST00000609620	SNORD91A-202	snoRNA	3.02	3.37E-03	small nucleolar RNA, C/D box 91A [Source:HGNC Symbol;Acc:HGNC:32752]
ENST00000416167	IGF2-205	protein_coding	3.04	1.37E-20	insulin like growth factor 2 [Source:HGNC Symbol;Acc:HGNC:5466]
ENST00000420909	HIP1-203	protein_coding	3.07	7.58E-03	huntingtin interacting protein 1 [Source:HGNC Symbol;Acc:HGNC:4913]
ENST00000583440	MIR3174-201	miRNA	3.07	1.78E-02	microRNA 3174 [Source:HGNC Symbol;Acc:HGNC:38264]
ENST00000390220	MIR760-201	miRNA	3.15	4.93E-02	microRNA 760 [Source:HGNC Symbol;Acc:HGNC:33666]
ENST00000591472	SEPT9-238	protein_coding	3.15	5.62E-07	septin 9 [Source:HGNC Symbol;Acc:HGNC:7323]
ENST00000582841	AC127024.3-201	lncRNA	3.15	2.29E-04	novel transcript
ENST00000387449	MT-TS2-201	Mt_tRNA	3.19	2.78E-12	mitochondrially encoded tRNA-Ser (AGU/C) 2 [Source:HGNC Symbol;Acc:HGNC:7498]
ENST00000574139	ABR-225	protein_coding	3.20	7.74E-08	ABR activator of RhoGEF and GTPase [Source:HGNC Symbol;Acc:HGNC:81]
ENST00000625100	AC007787.2-201	TEC	3.24	6.56E-03	TEC
ENST00000421194	AL590627.1-203	transcribed_processed_pseudogene	3.25	3.71E-04	forkhead box H1 (FOXH1) pseudogene
ENST00000505338	SLC2A13-204	retained_intron	3.25	4.93E-02	solute carrier family 2 member 13 [Source:HGNC Symbol;Acc:HGNC:15956]
ENST00000569468	AC105036.2-201	unprocessed_pseudogene	3.27	6.01E-03	microtubule associated serine/threonine kinase 2 (MAST2) pseudogene
ENST00000498244	SLC22A18-212	retained_intron	3.29	1.02E-02	solute carrier family 22 member 18 [Source:HGNC Symbol;Acc:HGNC:10964]
ENST00000492020	CAMK2G-214	lncRNA	3.33	5.36E-03	calcium/calmodulin dependent protein kinase II gamma [Source:HGNC Symbol;Acc:HGNC:1463]
ENST00000615609	RF00017.66-201	misc_RNA	3.39	2.65E-03	0
ENST00000458354	PNCK-216	protein_coding	3.40	4.51E-05	pregnancy up-regulated nonubiquitous CaM kinase [Source:HGNC Symbol;Acc:HGNC:13415]
ENST00000397787	COL18A1-AS1-201	lncRNA	3.41	7.47E-17	COL18A1 antisense RNA 1 [Source:HGNC Symbol;Acc:HGNC:23132]
ENST00000363424	RF00340.1-201	snoRNA	3.44	1.58E-03	0
ENST00000447103	FSCN1-204	protein_coding	3.45	7.62E-10	fascin actin-bundling protein 1 [Source:HGNC Symbol;Acc:HGNC:11148]
ENST00000579667	MIR3619-201	miRNA	3.46	1.72E-03	microRNA 3619 [Source:HGNC Symbol;Acc:HGNC:38998]
ENST00000625598	FP236383.2-201	lncRNA	3.48	2.89E-03	novel transcript, similar to YY1 associated myogenesis RNA 1 YAM1
ENST00000465886	MICAL3-209	lncRNA	3.50	1.30E-03	microtubule associated monooxygenase, calponin and LIM domain containing 3 [Source:HGNC Symbol;Acc:HGNC:24694]
ENST00000415656	AC091729.2-201	lncRNA	3.54	8.55E-04	novel transcript
ENST00000384899	MIR641-201	miRNA	3.54	3.36E-03	microRNA 641 [Source:HGNC Symbol;Acc:HGNC:32897]
ENST00000619471	RNA5-8SN1-201	rRNA	3.57	1.33E-02	RNA, 5.8S ribosomal N1 [Source:HGNC Symbol;Acc:HGNC:53517]

ENST00000526416	PLEC-209	protein_coding	3.58	6.42E-04	plectin [Source:HGNC Symbol;Acc:HGNC:9069]
ENST00000606947	HDAC2-AS2-218	lncRNA	3.62	7.16E-04	HDAC2 and HS3ST5 antisense RNA 2 [Source:HGNC Symbol;Acc:HGNC:43590]
ENST00000637662	EHMT1-245	lncRNA	3.64	2.93E-03	euchromatic histone lysine methyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:24650]
ENST00000587649	UBALD1-203	protein_coding	3.65	2.07E-05	UBA like domain containing 1 [Source:HGNC Symbol;Acc:HGNC:29576]
ENST00000477214	NOTUM-203	protein_coding	3.67	2.36E-13	notum, palmitoleoyl-protein carboxylesterase [Source:HGNC Symbol;Acc:HGNC:27106]
ENST00000621667	MIR941-5-201	miRNA	3.71	5.63E-04	microRNA 941-5 [Source:HGNC Symbol;Acc:HGNC:50845]
ENST00000492187	LHPP-208	lncRNA	3.74	1.44E-02	phospholysine phosphohistidine inorganic pyrophosphate phosphatase [Source:HGNC Symbol;Acc:HGNC:30042]
ENST00000576364	BAIAP2-226	lncRNA	3.77	3.62E-03	BAI1 associated protein 2 [Source:HGNC Symbol;Acc:HGNC:947]
ENST00000527021	AP006621.4-201	lncRNA	3.78	8.68E-07	novel transcript, antisense to EPS8L2
ENST00000627981	FP236383.3-201	lncRNA	3.80	7.56E-07	novel transcript, similar to YY1 associated myogenesis RNA 1 YAM1
ENST00000591009	FBXL12-211	protein_coding	3.84	2.56E-04	F-box and leucine rich repeat protein 12 [Source:HGNC Symbol;Acc:HGNC:13611]
ENST00000533346	NUP98-220	protein_coding	4.05	2.35E-03	nucleoporin 98 [Source:HGNC Symbol;Acc:HGNC:8068]
ENST00000528401	ZC3H3-202	protein_coding	4.07	9.87E-05	zinc finger CCCH-type containing 3 [Source:HGNC Symbol;Acc:HGNC:28972]
ENST00000529566	TSPAN4-218	retained_intron	4.12	1.20E-04	tetraspanin 4 [Source:HGNC Symbol;Acc:HGNC:11859]
ENST00000614492	FP671120.1-201	miRNA	4.15	4.23E-17	0
ENST00000387439	MT-TR-201	Mt_tRNA	4.41	1.06E-05	mitochondrially encoded tRNA-Arg (CGN) [Source:HGNC Symbol;Acc:HGNC:7496]
ENST00000608521	MIR663AHG-229	lncRNA	4.57	1.29E-07	MIR663A host gene [Source:HGNC Symbol;Acc:HGNC:27662]
ENST00000615130	RF01518.8-201	misc_RNA	10.45	NA	0
ENST00000465768	SAMM50-202	lncRNA	16.95	9.25E-09	SAMM50 sorting and assembly machinery component [Source:HGNC Symbol;Acc:HGNC:24276]

Appendix Table 4.10 Gene Ontology analysis (GO) of sMVs and sMB-Rs

(positive enrichment score indicates enriched GO terms in sMB-Rs, negative enrichment score indicates enriched GO terms in sMVs)

ID	Description	Normalised enrichment score	pvalue	rank
GO:0098798	mitochondrial protein complex	1.97	0.00110	45
GO:0044455	mitochondrial membrane part	2.25	0.00113	45
GO:0098800	inner mitochondrial membrane protein complex	2.46	0.00120	45
GO:0070469	respiratory chain	2.68	0.00126	108
GO:1990204	oxidoreductase complex	2.53	0.00126	54
GO:0098803	respiratory chain complex	2.66	0.00128	45
GO:0005746	mitochondrial respiratory chain	2.48	0.00130	45
GO:0005743	mitochondrial inner membrane	1.61	0.00628	177
GO:0098984	neuron to neuron synapse	1.62	0.00653	1148
GO:0098796	membrane protein complex	1.48	0.00705	744
GO:0000786	nucleosome	-2.58	0.00746	1436
GO:0044815	DNA packaging complex	-2.52	0.00769	1436
GO:0044391	ribosomal subunit	-1.79	0.00917	742
GO:0044448	cell cortex part	1.72	0.01090	727
GO:0044456	synapse part	1.45	0.01113	955
GO:0032993	protein-DNA complex	-2.05	0.01205	1437
GO:0014069	postsynaptic density	1.56	0.01425	711
GO:0099572	postsynaptic specialization	1.60	0.01522	711
GO:0015934	large ribosomal subunit	-1.67	0.01734	673
GO:0032279	asymmetric synapse	1.56	0.01745	711
GO:0019866	organelle inner membrane	1.48	0.01765	177

Appendix Table 4.11 RNA-binding proteins that might bind to lncRNA and pseudogene transcripts in Exos, sMVs and sMB-Rs derived from SW480 and SW620 cells

Protein Accession (UniProt)	Protein description	Gene name
Q86V81	THO complex subunit 4	ALYREF*
E9PID8	Cleavage stimulation factor subunit 2	CSTF2*
A0A0D9SFB3	ATP-dependent RNA helicase DDX3X	DDX3X*
Q8TDD1	ATP-dependent RNA helicase DDX54	DDX54*
Q08211	ATP-dependent RNA helicase A	DHX9*
Q15717	ELAV-like protein 1	ELAVL1*
P35637	RNA-binding protein FUS	FUS*
F8W6I7	Heterogeneous nuclear ribonucleoprotein A1	HNRNPA1*
P22626	Heterogeneous nuclear ribonucleoproteins A2/B1	HNRNPA2B1*
G3V576	Heterogeneous nuclear ribonucleoproteins C1/C2	HNRNPC*
P61978	Heterogeneous nuclear ribonucleoprotein K	HNRNPK*
P52272	Heterogeneous nuclear ribonucleoprotein M	HNRNPM*
Q9NZI8	Insulin-like growth factor 2 mRNA-binding protein 1	IGF2BP1*
F8W930	Insulin-like growth factor 2 mRNA-binding protein 2	IGF2BP2*
O00425	Insulin-like growth factor 2 mRNA-binding protein 3	IGF2BP3*
J3KTL2	Serine/arginine-rich-splicing factor 1	SRSF1*
Q92804	TATA-binding protein-associated factor 2N	TAF15*
Q13148	TAR DNA-binding protein 43	TARDBP*
K7ENG2	Splicing factor U2AF 65 kDa subunit	U2AF2*
Q8WYQ5	Microprocessor complex subunit DGCR8	DGCR8
Q06787	Fragile X mental retardation protein 1 homolog	FMR1
Q6ZN17	Protein lin-28 homolog B	LIN28B
Q86U44	Methyltransferase-like protein 3	METTL3
Q9HCE1	Moloney leukemia virus 10 protein	MOV10
O43251	RNA binding protein fox-1 homolog 2	RBFOX2
Q96MU7	YTH domain-containing protein 1	YTHDC1
Q15007	Pre-mRNA-splicing regulator WTAP	WTAP

*proteins co-identified in proteome of Exos, sMVs and sMB-Rs

Appendix Table 4.12 Gene ontology analysis of RBPs (ENCORI database) that might bind to lncRNA and pseudogenic transcripts

term_id	term_name	adjusted_p_value	intersections
GO:0005654	nucleoplasm	2.05E-11	ALYREF,CSTF2T,DDX3X,DDX54,DGCR8,DHX9,ELAVL1,FMR1,FUS,HNRNPA1,HNRNPA2B1,HNRNPC,HNRNPK,HNRNPM,IGF2BP1,LIN28B,METTL3,RBFOX2,SRSF1,TAF15,TARDBP,U2AF2,WTAP,YTHDC1
GO:1990904	ribonucleoprotein complex	2.06E-11	DDX3X,DHX9,ELAVL1,FMR1,FUS,HNRNPA1,HNRNPA2B1,HNRNPC,HNRNPK,HNRNPM,IGF2BP1,SRSF1,TARDBP,U2AF2
GO:0010494	cytoplasmic stress granule	8.29E-10	DDX3X,ELAVL1,FMR1,HNRNPK,IGF2BP1,MOV10,TARDBP
GO:0031981	nuclear lumen	1.24E-09	ALYREF,CSTF2T,DDX3X,DDX54,DGCR8,DHX9,ELAVL1,FMR1,FUS,HNRNPA1,HNRNPA2B1,HNRNPC,HNRNPK,HNRNPM,IGF2BP1,LIN28B,METTL3,RBFOX2,SRSF1,TAF15,TARDBP,U2AF2,WTAP,YTHDC1
GO:0005634	nucleus	3.13E-09	ALYREF,CSTF2T,DDX3X,DDX54,DGCR8,DHX9,ELAVL1,FMR1,FUS,HNRNPA1,HNRNPA2B1,HNRNPC,HNRNPK,HNRNPM,IGF2BP1,IGF2BP2,IGF2BP3,LIN28B,METTL3,MOV10,RBFOX2,SRSF1,TAF15,TARDBP,U2AF2,WTAP,YTHDC1
GO:0036464	cytoplasmic ribonucleoprotein granule	1.27E-07	DDX3X,DHX9,ELAVL1,FMR1,HNRNPK,IGF2BP1,MOV10,TARDBP
GO:0031974	membrane-enclosed lumen	1.55E-07	ALYREF,CSTF2T,DDX3X,DDX54,DGCR8,DHX9,ELAVL1,FMR1,FUS,HNRNPA1,HNRNPA2B1,HNRNPC,HNRNPK,HNRNPM,IGF2BP1,LIN28B,METTL3,RBFOX2,SRSF1,TAF15,TARDBP,U2AF2,WTAP,YTHDC1
GO:0070013	intracellular organelle lumen	1.55E-07	ALYREF,CSTF2T,DDX3X,DDX54,DGCR8,DHX9,ELAVL1,FMR1,FUS,HNRNPA1,HNRNPA2B1,HNRNPC,HNRNPK,HNRNPM,IGF2BP1,LIN28B,METTL3,RBFOX2,SRSF1,TAF15,TARDBP,U2AF2,WTAP,YTHDC1
GO:0043233	organelle lumen	1.55E-07	ALYREF,CSTF2T,DDX3X,DDX54,DGCR8,DHX9,ELAVL1,FMR1,FUS,HNRNPA1,HNRNPA2B1,HNRNPC,HNRNPK,HNRNPM,IGF2BP1,LIN28B,METTL3,RBFOX2,SRSF1,TAF15,TARDBP,U2AF2,WTAP,YTHDC1
GO:0035770	ribonucleoprotein granule	1.80E-07	DDX3X,DHX9,ELAVL1,FMR1,HNRNPK,IGF2BP1,MOV10,TARDBP
GO:0071013	catalytic step 2 spliceosome	3.81E-07	HNRNPA1,HNRNPA2B1,HNRNPC,HNRNPK,HNRNPM,SRSF1
GO:0016604	nuclear body	9.90E-07	ALYREF,DHX9,FMR1,HNRNPA2B1,HNRNPM,METTL3,SRSF1,TARDBP,U2AF2,WTAP,YTHDC1
GO:0005681	spliceosomal complex	1.50E-06	HNRNPA1,HNRNPA2B1,HNRNPC,HNRNPK,HNRNPM,SRSF1,U2AF2
GO:0043231	intracellular membrane-bounded organelle	1.24E-04	ALYREF,CSTF2T,DDX3X,DDX54,DGCR8,DHX9,ELAVL1,FMR1,FUS,HNRNPA1,HNRNPA2B1,HNRNPC,HNRNPK,HNRNPM,IGF2BP1,IGF2BP2,IGF2BP3,LIN28B,METTL3,MOV10,RBFOX2,SRSF1,TAF15,TARDBP,U2AF2,WTAP,YTHDC1

GO:0016607	nuclear speck	2.48E-04	ALYREF,METTL3,SRSF1,TARDBP,U2AF2,WTAP,YTHDC1
GO:0019034	viral replication complex	2.52E-03	FMR1,MOV10
GO:1902494	catalytic complex	2.59E-03	DGCR8,DHX9,HNRNPA1,HNRNPA2B1,HNRNPC,HNRNPK,HNRNPM,METTL3,SRSF1,WTAP
GO:0044094	host cell nuclear part	4.19E-03	FMR1,MOV10
GO:0005726	perichromatin fibrils	4.19E-03	DHX9,TARDBP
GO:0043227	membrane-bounded organelle	4.77E-03	ALYREF,CSTF2T,DDX3X,DDX54,DGCR8,DHX9,ELAVL1,FMR1,FUS,HNRNPA1,HNRNPA2B1,HNRNPC,HNRNPK,HNRNPM,IGF2BP1,IGF2BP2,IGF2BP3,LIN28B,METTL3,MOV10,RBFOX2,SRSF1,TAF15,TARDBP,U2AF2,WTAP,YTHDC1
GO:0070937	CRD-mediated mRNA stability complex	6.28E-03	DHX9,IGF2BP1
GO:0043229	intracellular organelle	7.01E-03	ALYREF,CSTF2T,DDX3X,DDX54,DGCR8,DHX9,ELAVL1,FMR1,FUS,HNRNPA1,HNRNPA2B1,HNRNPC,HNRNPK,HNRNPM,IGF2BP1,IGF2BP2,IGF2BP3,LIN28B,METTL3,MOV10,RBFOX2,SRSF1,TAF15,TARDBP,U2AF2,WTAP,YTHDC1
GO:0042025	host cell nucleus	8.78E-03	FMR1,MOV10
GO:0036396	RNA N6-methyladenosine methyltransferase complex	1.17E-02	METTL3,WTAP
GO:0045293	mRNA editing complex	1.17E-02	METTL3,WTAP
GO:0032991	protein-containing complex	1.25E-02	CSTF2T,DDX3X,DGCR8,DHX9,ELAVL1,FMR1,FUS,HNRNPA1,HNRNPA2B1,HNRNPC,HNRNPK,HNRNPM,IGF2BP1,METTL3,SRSF1,TARDBP,U2AF2,WTAP
GO:0033648	host intracellular membrane-bounded organelle	1.50E-02	FMR1,MOV10
GO:0033647	host intracellular organelle	1.50E-02	FMR1,MOV10
GO:0005844	polysome	3.11E-02	DHX9,FMR1,FUS
GO:0043226	organelle	4.85E-02	ALYREF,CSTF2T,DDX3X,DDX54,DGCR8,DHX9,ELAVL1,FMR1,FUS,HNRNPA1,HNRNPA2B1,HNRNPC,HNRNPK,HNRNPM,IGF2BP1,IGF2BP2,IGF2BP3,LIN28B,METTL3,MOV10,RBFOX2,SRSF1,TAF15,TARDBP,U2AF2,WTAP,YTHDC1

Appendix Table 4.13 Fusion genes from cell lysate, Exos, sMVs and sMB-Rs derived from SW480 and SW620 cells

Fusiongene_name	SW480_Cell	SW620_Cell	SW480-Exos	SW620-Exos	SW480-sMVs	SW620-sMVs	SW480-sMB-Rs	SW620-sMB-Rs
PLAGL1-MSH2	0.00	0.00	235.67	1555.33	602.33	2803.00	8661.33	17079.67
CDK6-VWC2	0.00	0.00	1016.00	334.67	599.33	159.67	10832.00	1856.00
METRNL-MSH2	0.00	0.00	121.00	486.00	218.67	904.00	3494.33	5448.33
HNRPLL-MSH2	0.00	0.00	48.33	309.33	151.67	736.33	1937.33	4384.67
XYLT1-MSH2	0.00	0.00	40.33	225.67	128.33	452.33	2096.33	3228.00
DNAJC1-MSH2	6.00	0.00	32.33	334.00	42.67	185.67	1242.67	4269.00
CDK6-NAT8L	0.00	0.00	0.00	209.67	0.00	0.00	5686.00	0.00
WDR62-ARL4C	0.00	0.00	57.00	129.33	170.67	354.00	1239.33	2242.33
MZT2B-CCDC160	0.00	0.00	26.00	91.33	88.00	208.00	1392.33	1727.00
WDR62-SSX2IP	5.00	0.00	45.67	102.33	130.00	292.00	1153.67	1665.67
PIP4K2B-MMP15	0.00	0.00	40.00	135.00	170.00	66.67	1254.33	1540.67
CACNA1B-MSH2	0.00	0.00	12.00	44.67	69.33	237.00	764.00	1341.00
WDR62-HABP4	0.00	0.00	20.67	58.67	68.00	146.33	787.33	1117.00
PIP4K2B-WDR62	0.00	0.00	51.33	78.00	65.67	152.33	847.33	782.00
MZT2B-RNF214	0.00	0.00	0.00	10.67	14.67	27.33	857.00	1048.33
FHL3-CCDC160	0.00	0.00	29.00	57.33	88.33	166.00	708.33	832.00
MZT2B-SLC8A2	0.00	0.00	25.00	53.00	56.33	127.33	557.33	717.67
GNE-SMPD4	0.00	0.00	19.67	61.00	34.00	65.00	562.67	705.67
GNE-ARL4C	0.00	0.00	0.00	22.67	0.00	41.33	565.33	806.33
RAB12-VWC2	0.00	0.00	9.67	99.67	32.67	35.00	383.33	850.33
RAB12-NAT8L	0.00	0.00	11.33	137.00	42.67	37.67	262.33	896.00
C16orf52-NOS1AP	0.00	0.00	5.33	152.33	37.67	61.33	158.33	912.33
KIAA1217-MSH2	0.00	0.00	4.67	59.67	22.67	99.00	336.00	593.00
PIP4K2B-GNE	0.00	0.00	7.67	49.00	36.00	55.67	322.33	639.67
C16orf52-LOXL4	0.00	0.00	0.00	102.67	18.67	58.67	142.67	780.33
ATP2C1-ANKDD1B	59.00	21.67	9.33	31.33	26.00	83.33	341.67	493.00
LHFPL2-MSH2	0.00	0.00	9.33	42.33	23.67	85.67	347.00	479.67
FHL3-SLC8A2	0.00	0.00	17.67	32.67	42.33	99.00	285.67	383.67
ATXN7-SLC8A2	0.00	0.00	7.33	30.33	39.00	100.67	300.00	362.00
RRBP1-CAMK1D	8.00	0.00	92.67	37.67	149.33	34.00	197.67	307.67
RRBP1-NANS	8.00	0.00	92.67	37.33	149.33	34.00	198.00	306.67
PIP4K2B-ADRBK2	0.00	0.00	7.00	23.67	37.00	61.67	320.67	361.33
SFXN3-CAMK1D	20.33	0.00	30.33	83.00	0.00	59.00	466.00	110.33
SFXN3-NANS	20.00	0.00	30.33	83.00	0.00	59.00	466.00	110.00
ADD3-MAFK	0.00	0.00	144.00	9.67	8.00	0.00	464.33	100.33
RAB12-FMNL1	0.00	0.00	0.00	61.00	10.67	14.00	139.33	490.33
FHL3-RNF214	0.00	0.00	11.67	25.67	13.00	26.00	467.00	129.00
ATXN7-CCDC160	0.00	0.00	3.67	16.00	22.67	51.00	279.00	294.33
GNE-HABP4	0.00	0.00	14.00	29.33	22.00	42.67	247.33	290.67
GNE-SSX2IP	0.00	0.00	6.67	13.00	0.00	18.00	198.33	391.00
EHMT1-MSH2	0.00	0.00	0.00	21.00	9.00	65.67	185.33	335.33
GRINA-C16orf52	0.00	0.00	12.67	13.00	13.00	23.67	405.00	142.00
B3GNT4-MSH2	0.00	0.00	0.00	84.00	182.00	5.00	79.67	256.00

DNAJC1-LHFPL2	7.67	0.00	28.00	14.00	30.67	5.00	253.67	262.00
RARA-ARL4C	0.00	0.00	30.00	18.00	35.00	79.67	216.00	185.33
MZT2B-RNF126	0.00	0.00	0.00	11.00	13.00	21.67	290.67	210.00
RNF135-SEPHS1	0.00	0.00	3.67	28.33	13.00	33.67	190.33	267.67
TGFBR3L-PGP	0.00	0.00	0.00	31.33	0.00	30.33	98.67	354.00
FHL3-RNF126	0.00	0.00	9.33	13.33	19.33	27.67	265.67	167.33
PIP4K2B-TRIM39	0.00	0.00	0.00	17.33	0.00	22.67	88.00	373.00
RAB12-TGFBR3L	0.00	0.00	4.33	20.67	10.00	22.00	242.00	158.33
COX20-ZC3H4	0.00	0.00	0.00	23.33	14.00	26.00	157.67	221.33
RAB12-PLXNC1	0.00	0.00	5.67	57.33	17.33	17.00	109.00	233.67
B3GNT4-LHFPL2	0.00	0.00	49.67	34.00	115.00	44.33	79.33	110.33
PLEKHG5-ARL4C	0.00	0.00	0.00	7.67	8.33	26.67	176.67	189.67
COX20-ELOVL1	0.00	0.00	0.00	21.33	0.00	25.00	177.00	184.33
OPN3-CHML	51.00	93.67	34.67	51.67	35.67	48.00	35.33	40.67
PTMS-ATXN7	0.00	0.00	14.67	4.33	0.00	11.33	288.33	66.00
RARA-METRNL	0.00	0.00	8.00	11.33	23.33	20.00	166.33	122.67
MX1-ZC3H4	0.00	0.00	0.00	25.33	0.00	15.67	101.00	201.67
DLG3-MAFK	0.00	0.00	64.33	0.00	13.00	0.00	218.67	45.33
SPPL3-C16orf52	0.00	0.00	7.67	9.67	0.00	5.00	236.00	82.33
B3GNT4-KIAA1217	0.00	0.00	29.67	28.00	53.00	15.00	6.00	205.67
LOC284801-AP1M1	38.00	22.33	3.67	11.67	6.67	44.33	79.00	128.67
C1QTNF4-HABP4	0.00	0.00	0.00	4.33	4.67	18.00	155.67	146.00
TGFBR3L-TCF15	0.00	0.00	0.00	31.33	0.00	20.00	66.67	207.67
RARA-SSX2IP	0.00	0.00	16.00	5.67	12.67	28.67	150.67	110.33
NFKBIL1-MAP2K2	0.00	0.00	4.33	29.33	4.33	10.00	57.00	218.67
TCF15-SEMA3C	0.00	0.00	0.00	33.33	0.00	20.33	50.33	215.33
PLEKHG5-SSX2IP	0.00	0.00	0.00	8.33	10.33	14.00	148.33	137.00
B3GNT4-PLAGL1	8.33	0.00	12.33	26.33	65.00	7.33	59.67	135.00
HRK-MAP2K2	0.00	0.00	9.33	8.67	4.67	8.67	192.33	88.00
HRK-IRX2	0.00	0.00	0.00	12.67	0.00	7.00	119.00	156.33
JA668105,ANKRD30BL-RCAN2	32.00	0.00	0.00	5.67	0.00	20.33	45.67	189.67
ETNK2-IRX2	0.00	0.00	0.00	8.67	0.00	10.00	169.33	99.00
ANKRD30BL-RCAN2	24.67	48.67	6.33	12.67	19.67	31.67	53.00	87.00
MX1-ELOVL1	0.00	0.00	0.00	21.33	0.00	21.67	84.33	153.33
C1QTNF4-ARL4C	0.00	0.00	7.00	5.33	6.00	0.00	232.67	24.00
PRPSAP2-TNPO2	0.00	0.00	0.00	0.00	14.33	34.33	110.67	112.33
PLAGL1-PPFIA3	0.00	0.00	5.00	10.33	9.00	19.00	160.67	67.00
AKAP13-PDE8A	30.00	23.33	28.67	58.33	31.67	42.00	31.00	22.33
RBM42-C16orf52	0.00	0.00	3.67	0.00	0.00	0.00	204.67	49.67
COX20-PTMS	0.00	0.00	0.00	12.33	10.67	6.67	116.00	110.67
LOC100130000-LINC00623	0.00	21.00	28.33	53.33	31.33	44.33	28.67	42.67
C16orf52-PPP1R36	0.00	0.00	0.00	26.33	3.67	8.67	31.00	171.00
HNRNPU-CDK6	0.00	0.00	0.00	11.33	0.00	17.67	48.33	160.67
B3GNT4-XYLT1	0.00	0.00	19.67	16.67	18.33	7.33	127.00	44.33
PLEKHG5-SMPD4	0.00	0.00	0.00	7.67	8.00	10.00	171.67	34.00
PDS5A-NOS1AP	0.00	0.00	0.00	16.67	0.00	7.33	24.00	181.33

HNRNPU-B3GALT6	0.00	0.00	0.00	3.67	0.00	8.33	52.33	158.67
PPP1R13L-IRX2	0.00	0.00	0.00	0.00	3.67	0.00	163.00	52.33
PLEKHG5-HABP4	0.00	0.00	0.00	0.00	0.00	4.00	118.00	92.00
MX1-PTMS	0.00	0.00	0.00	17.33	0.00	4.67	78.00	111.33
WDR62-MMP15	0.00	0.00	0.00	8.00	11.67	20.33	71.00	95.67
SFXN3-FHL3	5.00	0.00	0.00	12.67	29.33	9.33	71.33	76.00
BROX-AK025140	46.67	41.33	19.67	20.67	24.33	23.33	8.33	8.00
ATXN7-C1orf115	0.00	0.00	0.00	43.33	6.67	4.00	15.33	120.00
MX1-SIM2	0.00	0.00	0.00	29.33	0.00	13.33	12.67	132.33
TRIO-YY1	0.00	0.00	0.00	0.00	0.00	5.67	0.00	177.67
TRIO-LOC441204	9.33	0.00	5.67	0.00	0.00	3.67	124.00	40.00

Appendix Table 4.14 Differential transcript expression analysis of SW480-EVs and SW620-EVs (negative log2foldchange indicates enriched transcripts in SW480-EVs, positive log2foldchange indicates enriched transcripts in SW620-EVs)

Transcript ID	Transcript name	Transcript biotypes	log2Fold change	pvalue	Description
ENST00000531559	LINC02547-201	lncRNA	4.33	1.65E-50	metastasis associated lung adenocarcinoma transcript 1
ENST00000563414	ADGRG1-222	protein_coding	4.23	4.20E-06	microRNA 6859-1
ENST00000602632	UBE2E3-208	protein_coding	4.18	2.24E-06	MET transcriptional regulator MACC1
ENST00000552826	LINC02588-204	lncRNA	4.10	7.54E-08	MT-ND2 pseudogene 28
ENST00000490100	POU2F1-210	retained_intron	4.08	2.65E-05	MT-CO1 pseudogene 12
ENST00000635931	ALG13-238	lncRNA	4.03	8.71E-06	MT-ATP6 pseudogene 1
ENST00000260227	MMP7-201	protein_coding	3.98	1.23E-11	KCNQ1 opposite strand/antisense transcript 1
ENST00000431455	AC017048.2-201	lncRNA	3.94	1.38E-06	NOC2 like nucleolar associated transcriptional repressor
ENST00000318426	LINC01559-201	lncRNA	3.93	7.77E-93	SRY-box 4
ENST00000648428	LINC00511-238	lncRNA	3.71	9.20E-07	SRY-box 2
ENST00000588803	AC009271.1-202	lncRNA	3.68	7.69E-05	agrin
ENST00000363519	RNU6-509P-201	snRNA	3.67	1.15E-04	splicing factor proline and glutamine rich
ENST00000523716	CALB1-211	protein_coding	3.64	4.99E-06	FUS RNA binding protein
ENST00000511417	LINC02144-201	lncRNA	3.64	7.29E-06	SMAD family member 6
ENST00000483237	KHDC4-209	lncRNA	3.62	1.19E-04	matrix metallopeptidase 7
ENST00000515142	AC112206.4-201	processed_pseudogene	3.61	1.51E-08	cyclin D1
ENST00000636856	GRIN2B-207	lncRNA	3.60	6.88E-04	CD8b2 molecule
ENST00000461918	APLP2-205	retained_intron	3.60	2.18E-06	microRNA 429
ENST00000362165	MIR103A1-201	miRNA	3.60	1.10E-03	novel transcript
ENST00000556200	PRC1-AS1-202	lncRNA	3.58	9.22E-06	Wnt family member 9A
ENST00000606105	AC091946.2-201	lncRNA	3.57	3.14E-04	notch receptor 2
ENST00000442496	MFNG-207	protein_coding	3.55	5.92E-10	ubiquitin conjugating enzyme E2 J2
ENST00000623500	AC022140.2-201	TEC	3.55	6.22E-04	ArfGAP with coiled-coil, ankyrin repeat and PH domains 3
ENST00000365488	RNU6-652P-201	snRNA	3.55	7.61E-04	ArfGAP with coiled-coil, ankyrin repeat and PH domains 3
ENST00000363105	RN7SKP243-201	misc_RNA	3.53	3.94E-04	ArfGAP with coiled-coil, ankyrin repeat and PH domains 3
ENST00000572036	AC127496.5-202	lncRNA	3.53	1.63E-08	integrator complex subunit 11
ENST00000480950	FTCD-208	lncRNA	3.53	1.52E-03	novel transcript
ENST00000484398	CAB39-205	lncRNA	3.51	1.56E-04	TIMP metallopeptidase inhibitor 3
ENST00000516973	RNU6-341P-201	snRNA	3.50	1.78E-04	APC down-regulated 1
ENST00000463821	RN7SL168P-201	misc_RNA	3.49	2.14E-06	dishevelled segment polarity protein 1
ENST00000566062	SUB1P4-201	processed_pseudogene	3.48	4.46E-04	apoptosis antagonizing transcription factor
ENST00000471679	TBC1D5-223	lncRNA	3.47	3.50E-03	epidermal growth factor receptor
ENST00000604995	AL353807.4-201	processed_pseudogene	3.45	1.53E-05	novel transcript
ENST00000510214	OR51K1P-201	unprocessed_pseudogene	3.42	1.85E-16	APC down-regulated 1

ENST00000516217	RNU6-1262P-201	snRNA	3.41	5.05E-04	ATPase family AAA domain containing 3B
ENST00000427881	#N/A	#N/A	3.41	7.19E-07	NEDD4 binding protein 3
ENST00000437313	TSPAN32-205	nonsense_mediated_decay	3.41	4.96E-12	ADAM metallopeptidase domain 19
ENST00000418637	Z98742.3-201	processed_pseudogene	3.40	1.67E-04	NOTCH regulated ankyrin repeat protein
ENST00000558749	AC092868.2-201	lncRNA	3.37	2.17E-04	ATPase family AAA domain containing 3A
ENST00000495070	ARHGEF2-216	protein_coding	3.37	2.24E-07	SSU72 homolog, RNA polymerase II CTD phosphatase
ENST00000566559	RIPOR1-214	protein_coding	3.37	1.80E-12	Wnt family member 6
ENST00000461336	KCNIP3-204	lncRNA	3.36	1.86E-05	fibroblast growth factor 20
ENST00000450055	PRELID1P3-201	processed_pseudogene	3.36	5.25E-06	polypeptide N-acetylgalactosaminyltransferase 5
ENST00000636167	DOCK7-224	lncRNA	3.35	6.37E-04	sphingolipid transporter 2
ENST00000529858	AP2A2-218	lncRNA	3.34	4.27E-06	tubulin beta class I pseudogene 6
ENST00000512888	SPON2-218	retained_intron	3.34	6.27E-29	inositol 1,4,5-trisphosphate receptor type 2
ENST00000502313	HMGB1P21-201	processed_pseudogene	3.34	3.46E-07	hexokinase domain containing 1
ENST00000539026	LINC01559-205	lncRNA	3.32	3.66E-28	interleukin 1 receptor associated kinase 2
ENST00000453112	VOPP1-210	lncRNA	3.29	7.44E-05	G protein subunit beta 1
ENST00000513453	SIL1-214	protein_coding	3.29	1.82E-03	glucuronidase, beta (GUSB) pseudogene
ENST00000507608	PPWD1-205	retained_intron	3.29	1.22E-07	chromosome 9 open reading frame 152
ENST00000461944	NGEF-208	retained_intron	3.28	1.24E-05	protein kinase C zeta
ENST00000516665	RNU6-539P-201	snRNA	3.28	2.48E-04	protein kinase C zeta
ENST00000480060	CCHCR1-213	lncRNA	3.27	4.37E-05	PRKCZ antisense RNA 1
ENST00000529298	AP002340.1-201	lncRNA	3.27	2.38E-03	PRKCZ antisense RNA 1
ENST00000597717	RF00017.57-201	misc_RNA	3.26	5.96E-03	S100 calcium binding protein P
ENST00000434255	AL023693.1-201	lncRNA	3.26	2.59E-03	long intergenic non-protein coding RNA 1559
ENST00000518498	TFF3-204	protein_coding	3.26	5.33E-45	long intergenic non-protein coding RNA 1572
ENST00000635619	LINC01346-203	lncRNA	3.26	6.32E-05	arginine and glutamate rich 1
ENST00000494355	TRAF5-207	lncRNA	3.25	5.00E-04	novel transcript, sense intronic to MORN1
ENST00000491186	TCTEX1D2-205	lncRNA	3.23	1.16E-04	retention in endoplasmic reticulum sorting receptor 1
ENST00000623576	AC011471.3-201	TEC	3.23	1.38E-03	peroxisomal biogenesis factor 10
ENST00000495378	AP002812.1-201	processed_pseudogene	3.23	5.86E-04	fyn related Src family tyrosine kinase
ENST00000468602	MSH5-214	lncRNA	3.23	1.85E-06	T cell receptor beta constant 2
ENST00000409905	AC106876.1-201	lncRNA	3.22	4.33E-28	microRNA 4251
ENST00000364447	RNU4-59P-201	snRNA	3.22	1.08E-02	poly(ADP-ribose) glycohydrolase pseudogene 1
ENST00000365074	RF00139.3-201	snoRNA	3.21	1.49E-06	transport and golgi organization 6 homolog
ENST00000514406	CALB1-207	protein_coding	3.21	2.08E-08	Rho guanine nucleotide exchange factor 16
ENST00000624521	AC090506.2-201	TEC	3.21	1.83E-03	novel transcript
ENST00000565904	AC009063.3-201	lncRNA	3.20	3.75E-04	microRNA 551a
ENST00000408707	MIR1276-201	miRNA	3.20	5.72E-06	WD repeat containing, antisense to TP73
ENST00000365219	RNU6-1306P-201	snRNA	3.19	2.10E-02	general transcription factor IIi pseudogene 7
ENST00000578836	SNORA70D-201	snoRNA	3.19	3.04E-03	golgi glycoprotein 1
ENST00000544277	ERC1-216	lncRNA	3.18	1.61E-03	centrosomal protein 104
ENST00000426104	AC126124.1-201	processed_pseudogene	3.18	5.33E-05	SHC adaptor protein 3

ENST00000446871	CYP4F25P-201	unprocessed_pseudogene	3.18	9.89E-07	pyruvate dehyrogenase phosphatase catalytic subunit 2
ENST00000521373	LINC01942-201	lncRNA	3.17	4.18E-08	Putative beta-glucuronidase-like protein FLJ75287 pseudogene
ENST00000412485	LINC02154-201	lncRNA	3.17	2.55E-18	chondroitin sulfate N-acetylgalactosaminyltransferase 2
ENST00000442753	LINC02621-201	lncRNA	3.16	2.15E-08	TERF2 interacting protein
ENST00000488093	MEST-216	retained_intron	3.16	2.22E-04	microRNA 4689
ENST00000609126	ARHGEF2-219	lncRNA	3.16	1.34E-19	general transcription factor IIi pseudogene 7
ENST00000597045	AC020908.3-201	lncRNA	3.15	1.38E-03	tensin 3
ENST00000565442	CHD9-215	protein_coding	3.14	2.24E-10	teratocarcinoma-derived growth factor 1
ENST00000391625	Z97192.1-201	lncRNA	3.14	1.70E-17	chromosome 1 open reading frame 115
ENST00000557908	AC021739.3-201	lncRNA	3.14	1.47E-09	splicing factor 3b subunit 3
ENST00000558443	AC091231.1-201	lncRNA	3.13	1.84E-07	opioid receptor delta 1
ENST00000569407	AC109597.2-201	lncRNA	3.13	4.38E-05	cingulin like 1
ENST00000418471	AL009181.1-201	lncRNA	3.13	5.02E-05	hes family bHLH transcription factor 2
ENST00000590309	AC004223.2-201	lncRNA	3.13	1.15E-04	tet methylcytosine dioxygenase 1
ENST00000385283	MIR645-201	miRNA	3.13	6.00E-03	nucleolar protein 9
ENST00000487254	DNAH1-205	retained_intron	3.12	1.26E-03	RNA, U6 small nuclear 731, pseudogene
ENST00000620068	MIR7854-201	miRNA	3.12	1.43E-04	leucine rich adaptor protein 1 like
ENST00000259056	GALNT5-201	protein_coding	3.11	1.41E-148	glucuronidase, beta (GUSB) pseudogene
ENST00000584153	MIR4677-201	miRNA	3.11	5.13E-03	THAP domain containing 3
ENST00000573448	RPH3AL-210	protein_coding	3.11	7.46E-03	forkhead box A3
ENST00000565003	CMTM3-209	protein_coding	3.11	6.72E-05	long intergenic non-protein coding RNA 1239
ENST00000607372	AL136985.2-201	lncRNA	3.11	9.37E-03	CAMTA1 divergent transcript
ENST00000565768	MT1L-201	lncRNA	3.11	2.24E-27	zinc finger protein 775
ENST00000637875	GRIN2B-209	lncRNA	3.10	9.66E-04	coiled-coil domain containing 144C (CCDC144C) pseudogene
ENST00000532841	LINC01559-204	lncRNA	3.09	4.49E-11	jagged canonical Notch ligand 1
ENST00000433903	KCNG1-203	protein_coding	-5.87	2.78E-15	novel transcript
ENST00000469484	WNT5A-AS1-201	lncRNA	-5.88	7.66E-16	H2B histone family member S
ENST00000378075	LRRC10B-201	protein_coding	-5.89	2.76E-19	RNA, U1 small nuclear 106, pseudogene
ENST00000552864	LIN7A-204	protein_coding	-5.89	3.98E-60	pyridoxal kinase
ENST00000267294	ZIC5-201	protein_coding	-5.89	2.55E-45	pyridoxal kinase
ENST00000528183	AC010768.4-201	lncRNA	-5.91	2.99E-15	pyridoxal kinase
ENST00000537269	MIR200CHG-201	lncRNA	-5.94	1.38E-28	pyridoxal kinase
ENST00000425678	AC011747.1-201	lncRNA	-5.95	9.66E-56	0
ENST00000450938	AL161630.1-201	lncRNA	-5.96	5.05E-17	1-acylglycerol-3-phosphate O-acyltransferase 3
ENST00000458420	CLSTN2-201	protein_coding	-5.96	4.90E-91	platelet derived growth factor subunit A
ENST00000263734	EPAS1-201	protein_coding	-5.97	8.98E-162	high mobility group box 1 pseudogene 1
ENST00000546835	AC125603.2-201	lncRNA	-5.97	2.27E-20	1-acylglycerol-3-phosphate O-acyltransferase 3
ENST00000417157	SLC13A3-205	protein_coding	-6.01	4.05E-16	pregnancy up-regulated nonubiquitous CaM kinase
ENST00000436942	TCERG1L-AS1-201	lncRNA	-6.01	1.49E-16	angiotensin II receptor associated protein
ENST00000297436	DEFA6-201	protein_coding	-6.04	2.48E-16	stromal interaction molecule 2
ENST00000606796	AL096865.1-201	lncRNA	-6.04	7.89E-16	RNA, U5E small nuclear 4, pseudogene

ENST00000334232	EDARADD-201	protein_coding	-6.07	3.59E-16	purine nucleoside phosphorylase
ENST00000582531	AP005209.1-201	lncRNA	-6.11	7.60E-18	ladybird homeobox 2
ENST00000396721	SDR16C5-202	protein_coding	-6.14	6.85E-20	cyclin E2
ENST00000482448	NSMF-209	protein_coding	-6.15	4.80E-18	novel transcript, antisense to RDH13 and GP6
ENST00000286201	FZD7-201	protein_coding	-6.17	1.08E-57	TSPEAR antisense RNA 2
ENST00000304698	FAM171B-201	protein_coding	-6.17	3.83E-136	long intergenic non-protein coding RNA 2522
ENST00000438447	PDZD2-202	protein_coding	-6.20	5.21E-95	ubiquitin conjugating enzyme E2 G2
ENST00000529694	KCNJ5-202	protein_coding	-6.23	1.34E-109	PTTG1 interacting protein
ENST00000376335	ZIC2-201	protein_coding	-6.28	5.49E-46	inhibitor of growth family member 5
ENST00000556405	LINC02302-201	lncRNA	-6.32	6.72E-19	PTTG1 interacting protein
ENST00000319420	SHISA2-201	protein_coding	-6.32	2.40E-29	family with sequence similarity 207 member A
ENST00000377474	KCTD12-201	protein_coding	-6.33	4.87E-122	novel transcript
ENST00000615130	RF01518.8-201	misc_RNA	-6.37	NA	proteasome 26S subunit, non-ATPase 5
ENST00000267119	KRT71-201	protein_coding	-6.40	4.61E-24	kelch like family member 24
ENST00000372469	PRRX2-201	protein_coding	-6.44	3.41E-19	long intergenic non-protein coding RNA 334
ENST00000651999	AC016074.2-201	lncRNA	-6.45	1.59E-271	TMEM105 long non-coding RNA
ENST00000610809	AC245041.2-204	lncRNA	-6.46	8.33E-20	microRNA 4740
ENST00000502827	ADAMTS19-AS1-201	lncRNA	-6.48	2.34E-18	collagen type XVIII alpha 1 chain
ENST00000241125	GJA3-201	protein_coding	-6.48	4.47E-36	RNA, U6 small nuclear 7
ENST00000308497	STOX2-201	protein_coding	-6.49	1.08E-94	ribosomal protein S12 pseudogene 31
ENST00000257359	ADAMTS8-201	protein_coding	-6.51	1.59E-60	novel transcript
ENST00000529547	MRVI1-208	protein_coding	-6.52	1.89E-19	microRNA 6815
ENST00000511217	SH3RF2-206	protein_coding	-6.52	2.75E-64	ribosomal protein L35 pseudogene 1
ENST00000438795	AC092013.1-201	processed_pseudogene	-6.54	1.45E-19	alpha 1,4-galactosyltransferase (P blood group)
ENST00000368738	S100A9-201	protein_coding	-6.55	2.19E-19	dehydrogenase/reductase 12
ENST00000610908	AC008556.1-201	lncRNA	-6.60	7.84E-21	S100 calcium binding protein A2
ENST00000583088	KCNJ12-202	protein_coding	-6.60	2.67E-44	small nucleolar RNA, H/ACA box 11
ENST00000624982	AP002800.1-201	TEC	-6.71	2.00E-23	lanosterol synthase
ENST00000290354	CBR3-201	protein_coding	-6.73	2.83E-91	novel transcript
ENST00000457591	ANKRD20A10P-201	processed_pseudogene	-6.77	3.79E-21	MCM3AP antisense RNA 1
ENST00000378825	CCDC3-201	protein_coding	-6.79	1.34E-21	MCM3AP antisense RNA 1
ENST00000366621	KCNK1-202	protein_coding	-6.82	2.34E-21	minichromosome maintenance complex component 3 associated protein
ENST00000619297	LHX1-203	lncRNA	-6.87	6.79E-22	novel transcript
ENST00000414126	AC073316.2-201	lncRNA	-6.87	3.83E-22	novel transcript, antisense to MCM3APAS
ENST00000375992	NUDT11-201	protein_coding	-6.88	2.69E-22	CD44 molecule (Indian blood group)
ENST00000265634	NPTX2-201	protein_coding	-6.88	8.28E-80	novel transcript, antisense to MSTO1
ENST00000652087	AL591074.2-201	lncRNA	-6.88	1.41E-21	ENO1 antisense RNA 1
ENST00000367511	FAM129A-201	protein_coding	-6.92	1.48E-267	RNA, U6 small nuclear 396, pseudogene
ENST00000282146	KCNK13-201	protein_coding	-6.93	7.00E-23	protein arginine methyltransferase 2
ENST00000304710	CST5-201	protein_coding	-6.94	3.42E-22	DSTN pseudogene 1
ENST00000471328	FAM20C-202	lncRNA	-6.94	1.76E-23	novel transcript
ENST00000483722	TREML2-201	protein_coding	-6.97	1.98E-30	TMC3 antisense RNA 1

ENST00000437877	MAD1L1-208	protein_coding	-6.98	1.61E-39	mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 1
ENST00000555868	TMEM30B-202	protein_coding	-7.00	1.08E-24	mitochondrially encoded tRNA-Ile (AUU/C)
ENST00000306087	SOX14-201	protein_coding	-7.01	2.37E-23	mitochondrially encoded tRNA-Gln (CAA/G)
ENST00000428534	GLI3-202	retained_intron	-7.01	3.51E-22	mitochondrially encoded tRNA-Met (AUA/G)
ENST00000369663	TBX18-202	protein_coding	-7.05	2.09E-68	mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 2
ENST00000180166	FGF20-201	protein_coding	-7.05	5.14E-44	mitochondrially encoded tRNA-Trp (UGA/G)
ENST00000579474	LINC02563-201	lncRNA	-7.07	3.76E-23	death associated protein 3 pseudogene 2
ENST00000262426	FOXF1-201	protein_coding	-7.14	1.66E-24	mitochondrially encoded tRNA-Asn (AAU/C)
ENST00000519049	NKAIN3-201	lncRNA	-7.16	5.35E-24	mitochondrially encoded tRNA-Cys (UGU/C)
ENST00000238994	PPP1R3C-201	protein_coding	-7.27	1.73E-25	mitochondrially encoded cytochrome c oxidase I
ENST00000612499	IFI27-205	retained_intron	-7.29	7.47E-32	mitochondrially encoded tRNA-Ser (UCN) 1
ENST00000556569	AL355916.1-202	lncRNA	-7.29	7.05E-25	mitochondrially encoded tRNA-Asp (GAU/C)
ENST00000229465	AL355916.1-201	lncRNA	-7.38	2.88E-26	mitochondrially encoded cytochrome c oxidase II
ENST00000371697	ANKRD1-201	protein_coding	-7.42	2.81E-104	mitochondrially encoded ATP synthase membrane subunit 8
ENST00000332822	NOG-201	protein_coding	-7.42	1.41E-27	APCDD1L divergent transcript
ENST00000245903	CD70-201	protein_coding	-7.50	1.61E-26	mitochondrially encoded tRNA-Gly (GGN)
ENST00000614239	LHX1-201	protein_coding	-7.52	1.09E-32	ribosomal protein L31 (RPL31) pseudogene
ENST00000456783	ZNF492-201	protein_coding	-7.53	1.32E-27	mitochondrially encoded tRNA-Arg (CGN)
ENST00000631321	AL136962.1-201	lncRNA	-7.56	4.36E-27	proline rich coiled-coil 2A
ENST00000511685	TENM3-204	protein_coding	-7.56	4.77E-42	mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 4
ENST00000274487	ADAMTS19-201	protein_coding	-7.57	9.59E-59	mitochondrially encoded tRNA-His (CAU/C)
ENST00000368744	PRR9-201	protein_coding	-7.59	1.13E-33	ribosomal protein L7a (RPL7A) pseudogene
ENST00000504349	AC034245.1-201	lncRNA	-7.67	6.42E-28	mitochondrially encoded tRNA-Leu (CUN) 2
ENST00000304749	CST1-201	protein_coding	-7.72	6.89E-29	mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 5
ENST00000248564	GNG11-201	protein_coding	-7.80	1.40E-29	mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 6
ENST00000275493	EGFR-201	protein_coding	-7.87	1.29E-140	mitochondrially encoded tRNA-Glu (GAA/G)
ENST00000371225	TACSTD2-201	protein_coding	-7.94	1.10E-15	mitochondrially encoded cytochrome b
ENST00000621428	LHX1-DT-204	lncRNA	-8.00	2.05E-30	mitochondrially encoded tRNA-Thr (ACN)
ENST00000261275	FAM189A1-201	protein_coding	-8.03	1.28E-32	novel transcript
ENST00000315691	DIPK2A-201	protein_coding	-8.05	8.07E-31	novel transcript
ENST00000317012	ZBED2-201	protein_coding	-8.11	3.09E-31	leucine rich repeat containing 37B
ENST00000395925	GLI3-201	protein_coding	-8.18	8.00E-77	0
ENST00000252242	KRT5-201	protein_coding	-8.25	9.86E-198	0
ENST00000567574	LINC01992-201	lncRNA	-8.30	7.76E-33	0
ENST00000261254	CCND2-201	protein_coding	-8.40	0.00E+00	novel transcript
ENST00000614891	MAL2-204	protein_coding	-8.43	3.65E-87	Homo sapiens protein DGCR6 (LOC102724770), mRNA.
ENST00000344099	ZNF14-201	protein_coding	-8.79	3.46E-37	RNA, 5.8S ribosomal N5
ENST00000320892	RNF144A-201	protein_coding	-8.81	3.18E-37	novel transcript

ENST00000651482	LINC00861-205	lncRNA	-8.89	8.63E-38	39S ribosomal protein L23, mitochondrial-like
ENST00000359175	SPTSSB-201	protein_coding	-8.89	1.30E-41	autophagy related 9B
ENST00000623646	NKAIN3-IT1-202	lncRNA	-9.03	4.24E-48	novel transcript
ENST00000370357	PASD1-201	protein_coding	-9.07	1.27E-39	RAD51 paralog B

Appendix Table 4.15 Reactome pathway analysis of SW480-EVs and SW620-EVs (negative enrichment score indicates enriched pathways in SW480-EVs, positive enrichment score indicates enriched pathways in SW620-EVs)

ID	Description	Normalised enrichmen score	pvalue	rank
R-HSA-373080	Class B/2 (Secretin family receptors)	-2.13	0.001587	817
R-HSA-1296071	Potassium Channels	-2.06	0.001681	254
R-HSA-112316	Neuronal System	-1.89	0.002541	363
R-HSA-112314	Neurotransmitter receptors and postsynaptic signal transmission	-1.74	0.002786	363
R-HSA-6811558	PI5P, PP2A and IER3 Regulate PI3K/AKT Signaling	-1.89	0.002874	644
R-HSA-72695	Formation of the ternary complex, and subsequently, the 43S complex	1.90	0.003058	2153
R-HSA-72649	Translation initiation complex formation	2.05	0.003135	2188
R-HSA-72662	Activation of the mRNA upon binding of the cap-binding complex and eIFs, and subsequent binding to 43S	2.05	0.003135	2188
R-HSA-2219530	Constitutive Signaling by Aberrant PI3K in Cancer	-1.96	0.003175	644
R-HSA-72702	Ribosomal scanning and start codon recognition	2.08	0.003226	2188
R-HSA-72689	Formation of a pool of free 40S subunits	1.82	0.003472	2209
R-HSA-156842	Eukaryotic Translation Elongation	1.78	0.003509	2209
R-HSA-156827	L13a-mediated translational silencing of Ceruloplasmin expression	1.85	0.003636	2209
R-HSA-72613	Eukaryotic Translation Initiation	1.86	0.00369	2209
R-HSA-72706	GTP hydrolysis and joining of the 60S ribosomal subunit	1.86	0.00369	2209
R-HSA-72737	Cap-dependent Translation Initiation	1.86	0.00369	2209
R-HSA-2219528	PI3K/AKT Signaling in Cancer	-1.82	0.004367	644
R-HSA-72163	mRNA Splicing - Major Pathway	2.07	0.004651	1686
R-HSA-72172	mRNA Splicing	2.03	0.004902	1686
R-HSA-72203	Processing of Capped Intron-Containing Pre-mRNA	1.85	0.005464	1705
R-HSA-5654727	Negative regulation of FGFR2 signaling	-1.83	0.005505	129
R-HSA-5654732	Negative regulation of FGFR3 signaling	-1.83	0.005505	129
R-HSA-199418	Negative regulation of the PI3K/AKT network	-1.80	0.005839	644
R-HSA-6809371	Formation of the cornified envelope	-1.80	0.00624	190
R-HSA-72187	mRNA 3'-end processing	1.75	0.006452	1473
R-HSA-72764	Eukaryotic Translation Termination	1.75	0.006557	2380
R-HSA-156902	Peptide chain elongation	1.76	0.006645	2631
R-HSA-192823	Viral mRNA Translation	1.68	0.006645	2380
R-HSA-2408557	Selenocysteine synthesis	1.70	0.006897	2631
R-HSA-9634638	Estrogen-dependent nuclear events downstream of ESR-membrane signaling	-1.80	0.006908	542
R-HSA-9010553	Regulation of expression of SLITs and ROBOs	1.67	0.007491	2380

R-HSA-112315	Transmission across Chemical Synapses	-1.64	0.008054	226
R-HSA-372790	Signaling by GPCR	-1.56	0.009412	901
R-HSA-8848021	Signaling by PTK6	-1.79	0.0096	494
R-HSA-9006927	Signaling by Non-Receptor Tyrosine Kinases	-1.79	0.0096	494
R-HSA-500792	GPCR ligand binding	-1.66	0.009695	817
R-HSA-975956	Nonsense Mediated Decay (NMD) independent of the Exon Junction Complex (EJC)	1.69	0.009836	2631
R-HSA-5654741	Signaling by FGFR3	-1.71	0.009917	129
R-HSA-2408522	Selenoamino acid metabolism	1.57	0.010345	2209
R-HSA-6805567	Keratinization	-1.69	0.010401	190
R-HSA-8953854	Metabolism of RNA	1.74	0.012658	1733
R-HSA-3906995	Diseases associated with O-glycosylation of proteins	-1.68	0.013223	228
R-HSA-927802	Nonsense-Mediated Decay (NMD)	1.61	0.013746	2631
R-HSA-975957	Nonsense Mediated Decay (NMD) enhanced by the Exon Junction Complex (EJC)	1.61	0.013746	2631
R-HSA-176409	APC/C:Cdc20 mediated degradation of mitotic proteins	1.64	0.014881	1274
R-HSA-176814	Activation of APC/C and APC/C:Cdc20 mediated degradation of mitotic proteins	1.64	0.014881	1274
R-HSA-73856	RNA Polymerase II Transcription Termination	1.77	0.015385	1680
R-HSA-77595	Processing of Intronless Pre-mRNAs	1.67	0.017284	661
R-HSA-6807070	PTEN Regulation	1.47	0.018868	1653
R-HSA-983712	Ion channel transport	1.61	0.018987	888
R-HSA-381753	Olfactory Signaling Pathway	1.66	0.019753	64
R-HSA-397014	Muscle contraction	-1.68	0.020378	425
R-HSA-1643713	Signaling by EGFR in Cancer	-1.64	0.020725	1627
R-HSA-8986944	Transcriptional Regulation by MECP2	1.65	0.021333	2154
R-HSA-5654733	Negative regulation of FGFR4 signaling	-1.64	0.023451	129
R-HSA-75067	Processing of Capped Intronless Pre-mRNA	1.64	0.024	1680
R-HSA-168643	Nucleotide-binding domain, leucine rich repeat containing receptor (NLR) signaling pathways	1.67	0.02457	1176
R-HSA-2404192	Signaling by Type 1 Insulin-like Growth Factor 1 Receptor (IGF1R)	-1.62	0.028571	193
R-HSA-2428924	IGF1R signaling cascade	-1.62	0.028571	193
R-HSA-2428928	IRS-related events triggered by IGF1R	-1.62	0.028571	193
R-HSA-72766	Translation	1.39	0.029412	2209
R-HSA-376176	Signaling by ROBO receptors	1.49	0.032	2262
R-HSA-432040	Vasopressin regulates renal water homeostasis via Aquaporins	-1.62	0.032815	218
R-HSA-445717	Aquaporin-mediated transport	-1.62	0.032815	218
R-HSA-9009391	Non-genomic estrogen signaling	-1.57	0.033046	542
R-HSA-6802957	Oncogenic MAPK signaling	1.54	0.033846	1033
R-HSA-2672351	Stimuli-sensing channels	1.57	0.034667	888
R-HSA-3899300	SUMOylation of transcription cofactors	1.56	0.036011	1286
R-HSA-72306	tRNA processing	1.50	0.039007	2087

R-HSA-179419	APC:Cdc20 mediated degradation of cell cycle proteins prior to satisfation of the cell cycle checkpoint	1.50	0.039514	1274
R-HSA-5576891	Cardiac conduction	-1.55	0.04	425
R-HSA-168273	Influenza Viral RNA Transcription and Replication	1.38	0.04	2594
R-HSA-163125	Post-translational modification: synthesis of GPI-anchored proteins	-1.57	0.040336	1700
R-HSA-422356	Regulation of insulin secretion	-1.59	0.042122	1031
R-HSA-5654736	Signaling by FGFR1	-1.57	0.043062	129
R-HSA-174143	APC/C-mediated degradation of cell cycle proteins	1.50	0.045752	1274
R-HSA-453276	Regulation of mitotic cell cycle	1.50	0.045752	1274
R-HSA-176408	Regulation of APC/C activators between G1/S and early anaphase	1.49	0.046584	1274
R-HSA-1839126	FGFR2 mutant receptor activation	-1.57	0.046632	1091
R-HSA-168255	Influenza Life Cycle	1.36	0.048246	2594

Appendix Article 1

communications biology

ARTICLE

https://doi.org/10.1038/s42003-021-01882-z OPEN

Secreted midbody remnants are a class of extracellular vesicles molecularly distinct from exosomes and microparticles

Alin Rai [1,2]✉, David W. Greening [1,2], Rong Xu [1], Maoshan Chen [1], Wittaya Suwakulsiri[1] & Richard J. Simpson [1]✉

During the final stages of cell division, newly-formed daughter cells remain connected by a thin intercellular bridge containing the midbody (MB), a microtubule-rich organelle responsible for cytokinetic abscission. Following cell division the MB is asymmetrically inherited by one daughter cell where it persists as a midbody remnant (MB-R). Accumulating evidence shows MB-Rs are secreted (sMB-Rs) into the extracellular medium and engulfed by neighbouring non-sister cells. While much is known about intracellular MB-Rs, sMB-Rs are poorly understood. Here, we report the large-scale purification and biochemical characterisation of sMB-Rs released from colon cancer cells, including profiling of their proteome using mass spectrometry. We show sMB-Rs are an abundant class of membrane-encapsulated extracellular vesicle (200-600 nm) enriched in core cytokinetic proteins and molecularly distinct from exosomes and microparticles. Functional dissection of sMB-Rs demonstrated that they are engulfed by, and accumulate in, quiescent fibroblasts where they promote cellular transformation and an invasive phenotype.

[1] Department of Biochemistry and Genetics, La Trobe Institute for Molecular Science (LIMS), La Trobe University, Melbourne, VIC 3086, Australia. [2] Baker Heart and Diabetes Institute, Melbourne, VIC 3004, Australia. ✉email: alin.rai@baker.edu.au; richard.simpson@latrobe.edu.au

At the terminal stage of cell division, the prospective daughter cells remain connected by a thin intercellular bridge containing the midbody (MB), a transient organelle responsible for mediating final abscission during cytokinesis[1–3]. The MB is then inherited by one of the newly-formed daughter cells where they perform non-mitotic functions[3–8]. While MB-Rs are degraded by autophagy[3] they can also accumulate intracellularly and influence cell fate[8]. An alternative fate of MB-Rs involves extracellular secretion as sMB-Rs[9–11]. Over the past 30 years the prevailing view has been that sMB-Rs are degraded extracellularly[11]. However, recent evidence shows that sMB-Rs can also be engulfed by non-sister cells[8,12]. Because MB-Rs can influence cell signalling and cell fate in daughter cells[8–10,13,14], it has been speculated that they might perturb cell signalling at distal sites[15], however, this question remains largely unexplored.

Previously, we observed that MKLP1 (also known as KIF23 or ZEN-4)[16], a component of the centraspindlin complex, co-purified with shed microvesicles (sMVs, also referred to as microparticles and ectosomes[17]) isolated from the cell culture medium of the human colorectal cancer (CRC) cell line LIM1863[18]. Because centraspindlin is a core component of MBs[16] and MB-Rs[8], we reasoned that MB-Rs might be released from cell lines into culture medium (CM) in sufficient quantities to permit their biochemical and functional characterisation. The focus of this paper is directed at the large-scale preparation of sMB-Rs that would allow us to undertake their biophysical and functional characterisation, to ask whether in vitro release of MKLP1 from LIM1863 CRC cells is a general phenomenon or a cancer-cell specific process, and whether MKLP1 is an indicative marker of sMB-Rs.

Results and discussion

Cancer cell-derived midbody remnants are shed into the extracellular space. To determine whether sMB-Rs are secreted into the extracellular space, we first investigated whether MKLP1 could be used as a reliable marker for MB/MB-R detection. For this purpose we used CRC cells (SW620) grown in 2D culture. Fluorescent microscopy revealed that the MKLP1 antibody readily stained punctate structures nested in the middle of the intercellular bridge between dividing cells (revealed using beta-tubulin antibody) typical of MBs[1–3] (Supplementary Fig. 1). Next we examined whether MKLP1 stained for MB-Rs. During the final stages of cytokinesis the inner leaflet phosphatidylserine (PS) of the MB membrane flips to the outer leaflet[19,20] resulting in PS enrichment in the outer leaflet of the MB-R membrane. Importantly, PS on outer leaflet of sMB-Rs is required for their engulfment by cells[20]. In our study, fluorescent microscopy revealed that MBs localised between dividing SW620 cells did not stain with annexin V (Fig. 1a, *top panel*), whereas annexin V- and MKLP1-staining structures similar in size to MB-Rs associate with non-dividing cells. This observation is consistent with the presence of a MB-R inherited by one of the daughter cells post cytokinesis (Fig. 1a, *middle panel*). Moreover, co-staining of annexin V- and MKLP1 of MB-Rs was detected in the extracellular space (Fig. 1a, *bottom panel*) consistent with these particles being sMB-Rs. These anti-MKLP1-staining MB-R/sMB-R punctate structures were distinct from cellular debris as evidenced by lack of genomic DNA staining (Hoechst stain); their size range (~500 nm particle diameter) is consistent with previous report for sMB-Rs[9,10,19]. Collectively, these data show that MKLP1 can be used as a stereotypic marker for CRC cell MB/sMB-Rs. To provide additional proof that MB-Rs are secreted from cell lines into the extracellular space we generated fluorescently-labelled MB-Rs. For this purpose, we constructed SW620 cells stably-expressing plasma membrane-targeting GAP43 (1-20 a.a.)[21] fused to GFP (SW620-GAP-GFP cells) (Supplementary Fig. 2). In Fig. 1b it can be seen that fluorescently-labelled/ MKLP1-positive MB-Rs are shed extracellularly.

MB-Rs have been reported to accumulate in small subpopulations of cancer cells and associate with stemness[8]. It can be seen in Supplementary Fig. 3a, b that while only a small population of SW620 and SW480 cells (0.9–1.6%) grown in culture accumulate MB-Rs, a large pool of non-cell associated MB-Rs are observed, a finding consistent with shedding of MB-Rs into the extracellular medium. This observation was corroborated using an antibody to the centraspindlin component RACGAP1 (Supplementary Fig. 3a). To rule out the possibility that MB-R shedding is an artefact of 2-D cell culturing, we prepared 3-D cultures of SW620 and SW480 cells in Matrigel^{TM} matrix (Fig. 1c) and tested for MB-R shedding. These data show that both SW620 and SW480 cells grown as spheroids also shed MB-Rs into their extracellular space (Fig. 1c, *upper and middle panels*).

Next, we set out to determine whether cancer cells shed MB-Rs in vivo. For this purpose, we established SW620-GAP-GFP cells as subcutaneous xenografts in mice. Immunohistochemical analysis of ectopic tumours from these mice using MKLP1 antibodies revealed GFP-tagged MB-Rs in the extracellular space (Supplementary Fig. 4). To determine whether sMB-Rs could be detected in human colon cancer tissues, we analysed MKLP1-or RACGAP1-antibody based immunohistochemical images of human colon cancer tissues publicly available from the Human Protein Atlas (http://www.proteinatlas.org/) (Fig. 1d, Supplementary Figs. 5 and 6). Strikingly, large pools of MKLP1- or RACGAP1-staining punctate structures were detectable in the extracellular space (lumen) of CRC tissues in greater abundance compared to non-disease colon tissue (Supplementary Fig. 7). Further, we also show that in stark contrast to non-polarised SW480- / SW620-spheroids that shed their sMB-Rs non-directionally (Fig. 1c), highly-polarised spheroids SW1222 cells (Fig. 1c, *lower panel*) or organoids cultured from mouse-derived intestinal crypts (Supplementary Fig. 8) shed their MB-Rs into the central lumen (stained by filamentous actin). These observations are consistent with the emerging role of MB-Rs in cell polarity[6,7] and boost the notion that cancer cells actively shed MB-Rs into the extracellular space.

sMB-Rs can be isolated from the culture medium of SW620 cells in high yield. As a first step towards purifying sMB-Rs from CRC SW620 cell line culture medium it was important to establish which of the two major EV classes sMB-Rs belong to - exosomes or shed microvesicles/microparticles[17]. sMB-Rs from CRC SW620 cell line culture medium it was important to establish which of the two major EV classes sMB-Rs belong to - exosomes or shed microvesicles/microparticles[17]. EV classes differ in size range (exosomes typically 30–200 nm and sMVs, 50–1300 nm) and mode of biosynthesis (exosomes being of endosomal origin and sMVs forming by plasma membrane budding)[17]. Using a differential centrifugation strategy[22], we first separated sMVs (which pellet at 10,000 × *g*) from exosomes which pellet at 100,000 × *g* (Fig. 2a) and showed sMB-Rs co-pellet with crude sMVs, but not exosomes as evidenced by western blot analysis using RACGAP1 antibody (Supplementary Fig. 9). Similar results were obtained using SW480 and LIM1863 cell lines (Supplementary Fig. 9). We next used an orthogonal step, isopycnic (iodixanol-density) centrifugation[23,24], to further fractionate sMVs based on buoyant density (Fig. 2b, c). Two well-separated sMV fractions with distinct buoyant densities were found – low-density sMVs (sMV-LD fractions 7&8, 1.13–1.14 g ml^{-1}) and high-density sMVs (sMV-HD fractions 9&10, 1.22–1.30 g ml^{-1}) (Fig. 2a–c, Supplementary Fig. 10). Centraspindlin markers MKLP1 and RACGAP1 identified the sMV-HD fraction as highly enriched in sMB-Rs; this

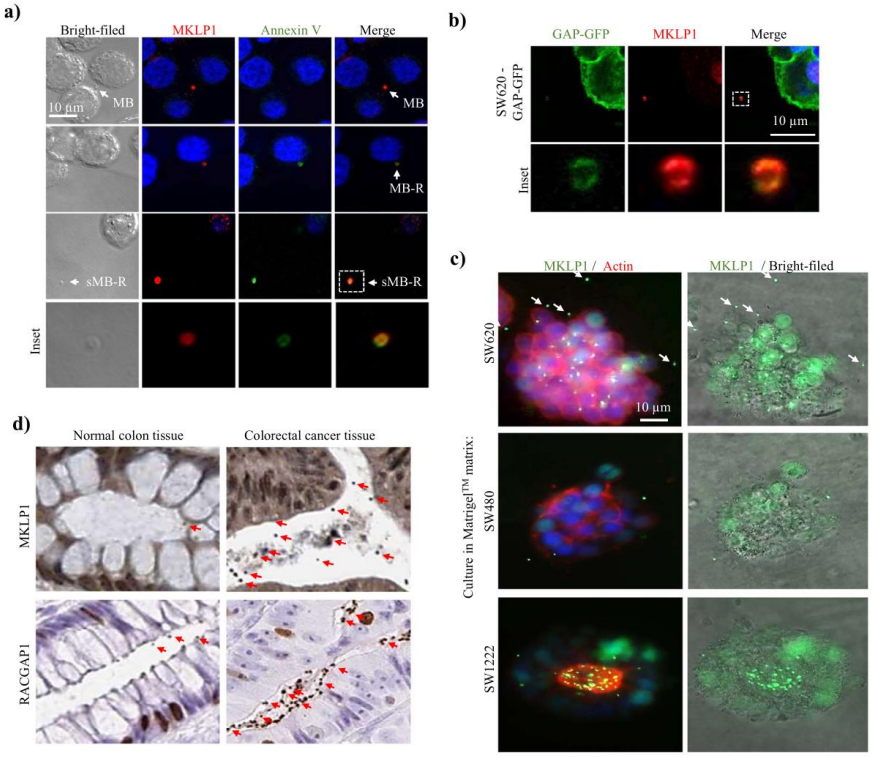


Fig. 1 Secretion of cancer cells shed midbody remnants into extracellular space. a Immunofluorescence microscopy analysis of SW620 cells using anti-MKLP1 antibodies. SW620 cells cultured on glass coverslips were stained with annexin-V. Midbodies (MB) between prospective daughter cells do not stain with annexin-V (*top panel*). In contrast, midbody remnants (MB-R) associated with one of the cells stained with annexin V (*middle panel*). MB-Rs are also detected extracellularly as shed midbody remnants (sMB-Rs) (*bottom panel*). Inset panels represent enlarged images of sMB-Rs. Scale bar, 10 µm. **b** Immunofluorescence microscopy analysis of SW620-GAP-GFP cells using anti-MKLP1 antibodies. Nuclei (blue) were stained with Hoechst stain. Inset: higher magnification of GFP-tagged sMB-Rs in the extracellular space. Scale bar, 10 µm. **c** Immunofluorescence microscopy analysis of SW620, SW480 and SW1222 cells cultured in Matrigel™ matrix using anti-MKLP1 antibodies (green) and Alexa Fluor 594 Phalloidin (red) to stain actin. Nuclei (blue) were stained with Hoechst stain. White arrow heads point show sMB-Rs. Scale bar, 10 µm. **d** Immunohistochemistry analysis of normal human colon tissue and colon cancer tissue (adenocarcinoma) using anti-MKLP1 and anti-RACGAP1 antibodies. Red arrows indicate anti-MKLP1 or anti-RACGAP1 staining extracellular sMB-Rs. Images obtained from Human Protein Atlas (http://www.proteinatlas.org/) with permission.

fraction was subjected to further biochemical and functional characterisation.

Both sMV-HD and -LD fractions displayed low or non-detectable levels of ALIX, TSG101, CD81 and CD82 and CD63 using Western Blotting (Fig. 2d, Supplementary Figs. 11 and 12), however, both ALIX and TSG101 were detected in sMB-R proteome data set (Supplementary Data 1). The yield of sMB-Rs (i.e., sMV-HD fraction) from 450 ml of SW620 cells grown in continuous culture (10 days, 3 Bioreactor Cell Line™ flasks) was 1080 µg protein (~67.986×10^6 sMB-R particles). Cryo-electron microscopy revealed sMB-Rs (sMV-HD) with particle diameters in the range 200–600 nm partially overlapping with sMV-LD particles, but significantly larger than exosomes (30–200 nm) (Fig. 2e, f). This finding is comparable to particle size determinations obtained using nanoparticle-tracking analysis (Supplementary Fig. 13) and sMB-Rs (~300 nm) based on conventional electron microscopy[9]. The presence of MKLP1-positive sMB-Rs in sMV-HD fractions, but not sMV-LD or exosome fractions, was further validated by aldehyde sulfate latex bead capture/fluorescence microscopy (Fig. 2g). In RACGAP1 immunoprecipitation analysis of the sMV-HD fraction, MS-based proteomics identified MKLP1, an integral component of the MB centraspindlin complex of RACGAP1 and MKLP1/KIF23[16] and other midbody components known to interact with RACGAP1, including PLK1, RHOA, CIT, KIF14 and KIF1A (Supplementary Data 2, Supplementary Fig. 14). We further showed that MB-R shedding by cancer cells is a widespread phenomenon, being observed for primary (e.g., SW480, LIM1863) as well as metastatic CRC cell lines (COLO 205 and T84) and the breast cancer cell line MDA MB 231 (Supplementary Figs. 15 and 16). Moreover, these data indicate that sMB-Rs (sMVs-HD) represent a significant portion of total secreted EVs (exosomes and sMVs-LD fractions) depending on cancer cell type (Fig. 2h).

Protein profiling of shed midbody remnants. Next, we performed a comparative proteome analysis of SW620 cell-derived sMB-Rs (sMV-HD), sMV-LD and exosomes using a label-free quantitative mass spectrometry (MS) approach[18]. A total of 2300

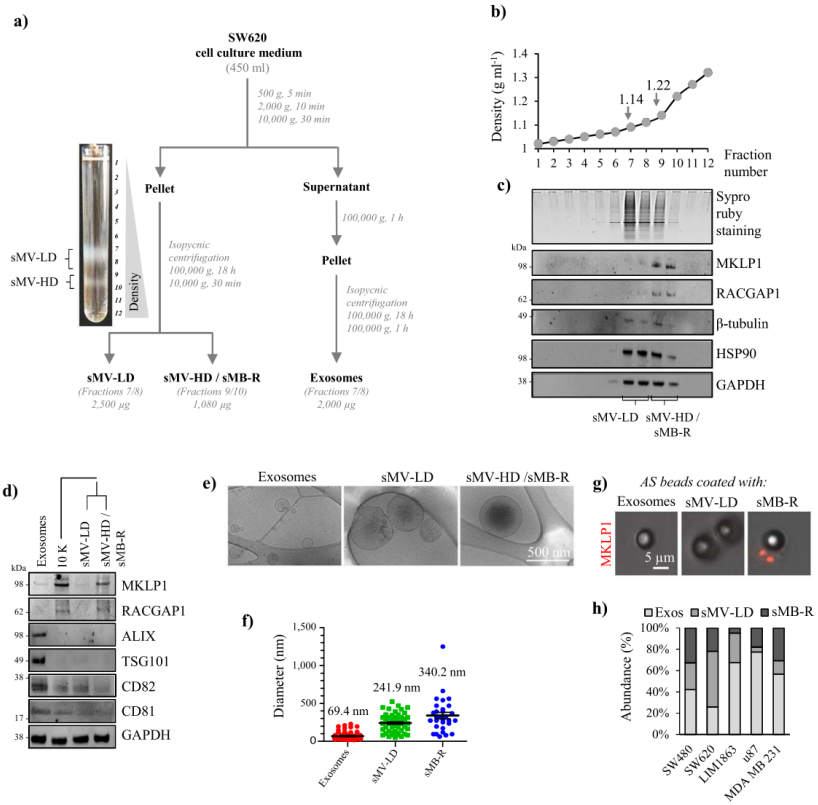


Fig. 2 Isolation and characterisation of shed midbody remnants (sMB-Rs). a Experimental workflow for purification of sMB-Rs from SW620 cell culture medium (CM). CM was subjected to differential centrifugation to obtain crude sMVs (10,000 × *g* pellet) and exosomes (100,000 × *g* pellet) that were further fractionated using isopycnic density gradient centrifugation. Photographic image shows that crude sMVs (10,000 × *g* pellet) floated in two major fractions: low-density fractions 7/8 (sMV-LD) and high-density fractions 9/10 (sMV-HD/sMB-Rs). **b** The buoyant densities of twelve 1-mL fractions collected for each preparation were determined by absorbance at 244 nm using a molar extinction coefficient of 320 L $g^{-1}cm^{-1}$. **c** SDS-PAGE of 12 OptiPrep™ fractions. Protein quantitation was determined by SYPRO Ruby staining and western blot analysis performed using indicated antibodies. **d** Western blot analysis of SW620 cell-derived Exos, 10,000 × *g* EVs (crude sMVs), sMV-LD (fractions 7-8) and sMB-R (sMV-HD) (fractions 9-10) using indicated antibodies ($n = 2$ biological replicates). **e** Cryo-electron microscopic analysis of SW620 cell-derived Exos, sMV-LD and sMB-Rs. **f** Histogram represents the measurements of diameter of Exos, sMV-LD and sMB-R based on cryo-EM images. Data presented as mean ± s.e.m (standard error of mean). **g** Fluorescence microscopic analysis of Exos, sMVs-LD and sMB-Rs derived from SW620-GAP-GFP cells loaded onto aldehyde sulfate (AS) latex beads and immunostained with anti-MKLP1 antibodies (in red). **h** Bar plot showing protein yield (based on SYPRO Ruby protein quantitation) of Exos, sMVs-LD and sMB-R (sMVs-HD) secreted from five different cancer cell lines.

proteins were identified in sMB-Rs, 2153 in sMVs-LD, and 1929 in exosomes (Supplementary Data 1) and 382, 144 and 236 proteins, respectively are uniquely identified (Venn diagram, Fig. 3a) indicating that these three vesicle types are molecularly distinct from one another.

The relative abundance of proteins in each EV subtype, based on normalised spectral counts, is shown in the heatmap (Fig. 3b). Notably, proteins associated with cytokinesis such as microtubule-bundling proteins[25], the centraspindlin complex[16] and chromosomal passenger complex[26] are selectively enriched in sMB-Rs (the sMV-HDs fraction), but not in sMV-LD and exosome fractions. These cytokinesis-signature proteins found exclusively in sMB-Rs boost our argument that sMB-Rs represent a new category of EV, hitherto undescribed in the EV literature.

Next, using the STRING database (version 10.5) we identified 982 high-abundance SW620 cellular proteins in sMB-Rs, compared to exosomes and sMV-LD (Supplementary Data 3 (highlighted in red in heatmap, Fig. 3b)). Using this list, we constructed a protein-protein interaction network for sMB-Rs proteins (Fig. 3c). GO analysis identified protein clusters implicated in biological processes such as RNA regulation (e.g., "ribosome", "aminoacyl-tRNA synthase", "eIF3 complex" and "spliceosome"), Protein degradation (e.g., "proteasome complex"), and Vesicle transport (e.g., "COPI-coated vesicle membrane") (Fig. 3c). Strikingly, these clusters include proteins important in biological processes such as translation[27,28], protein degradation[29–32], and vesicle transport[1,2,33] - processes reported to be tightly-regulated in MBs and critical for faithful cytokinesis[1,2,27–36].

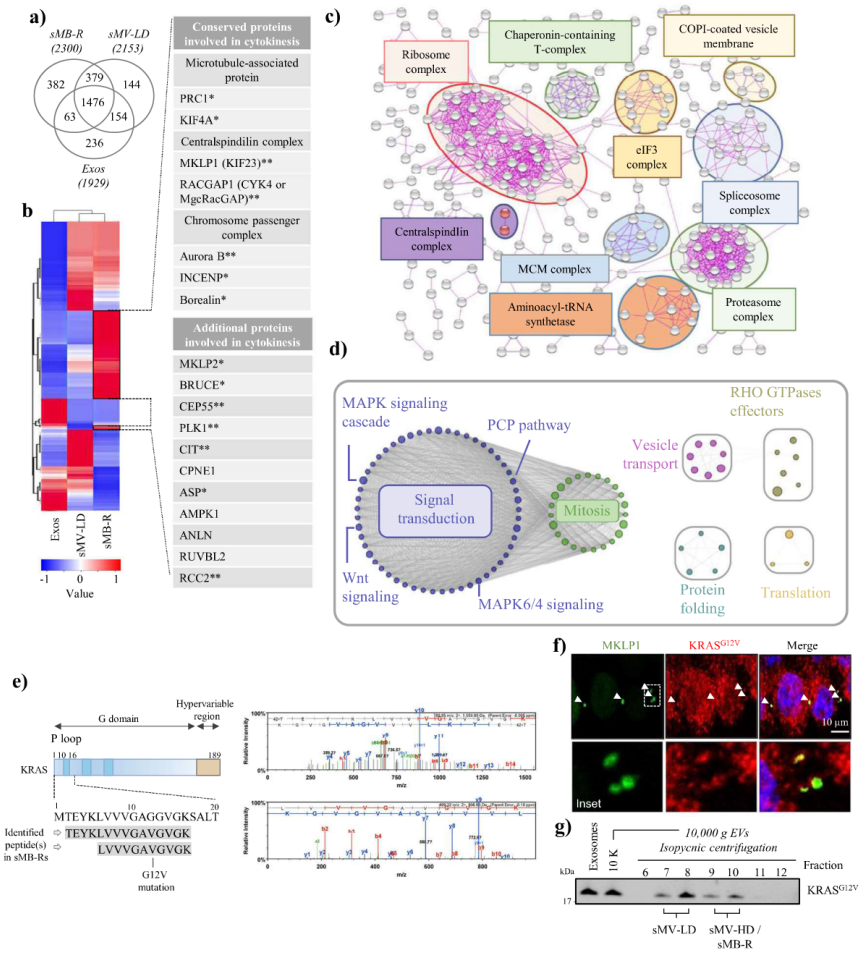


Fig. 3 Proteome analysis of shed midbody remnants. a Venn diagram of proteins identified in SW620 cell-derived sMB-R (sMV-HD), sMV-LD and Exos. **b** Heatmap illustration of proteins identified in sMB-R (sMV-HD), sMV-LD and Exos. Proteins present in higher abundance in sMB-R (red) as compared to sMV-LD and Exos include conserved cytokinetic proteins as well as additional cytokinetic proteins. *Proteins uniquely identified in sMB-Rs. **Proteins enriched (fold change >2) in sMB-R compared to sMV-LD and Exos. **c** STRING-based protein-protein interaction network analysis of 928 enriched proteins in sMB-Rs (sMV-HD) compared to sMV-LD and Exos. The interactions were "evidence"-based, with "experiments" as active interaction source and interaction threshold set at 0.900 (highest confidence). Disconnected nodes in the network are hidden. Proteins identified under biological processes or molecular processes (Gene Ontology) are indicated. Centralspindlin complex components (RACGAP1 and KIF23/MKLP1) are also indicated. **d** EnrichmentMap of Reactome pathways enriched in 456 proteins commonly identified in SW620 cell-derived sMB-R proteome (2300 proteins) with the proteome of MB-Rs shed by Hela cells reported recently by Peterman et al. 2019[20]. **e** Mass spectrometry-based identification of KRAS peptides (UniProtKB ID RASK_HUMAN) in sMB-Rs. Two peptides (TEYKLVVVGAGGVGK and LVVVGAGGVGK) spanning Gly-12/ Val-12 substitution in KRAS protein. Peptide spectral profiles are displayed on the right. **f** Immunofluorescence microscopy of SW620 cells using anti-MKLP1 and anti-KRAS^{G12V} antibodies. Nuclei (blue) were stained with Hoechst stain. White arrows indicate the position of MB and MB-Rs. Inset represents higher magnification. Scale bar, 10 µm. **g** Western blot analysis of exosomes, crude 10,000 x *g* sMVs, and isopycnic (iodixanol-density) gradient centrifugation fractions of sMV-LD and -HD/ sMB-Rs using anti-KRAS^{G12V} antibody.

Furthermore, 32/982 of high-abundance sMB-Rs proteins are listed in the MiCroKITS-v4.0 database of proteins 'experimentally-verified to temporally and spatially localise to midbody, centromere, kinetochore, telomere or spindle structures during cell division'[37] (http://microkit.biocuckoo.org, Supplementary Data 4).

To further address the functionality of sMB-R proteins, we conducted a gene-annotation enrichment and pathway analysis (DAVID[38]) (version 6.8) using the Gene Ontology (GO) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) databases (Supplementary Data 5). This analysis revealed 207 proteins in

sMB-Rs involved in regulation of "signal transduction"; amongst these 'MAPK signalling", "Ras signalling pathway" and "Pathways in cancer" are preeminent (Supplementary Fig. 17).

We next compared SW620 cell derived sMB-R proteome (2300 proteins) with the proteome of MB-Rs shed by Hela cells reported recently by Peterman et al.[20]. A total of 456 proteins were commonly identified (Supplementary Fig. 18, listed in Supplementary Data 6). We next performed Reactome pathways analysis on 456 common proteins (Supplementary Data 7). EnrichmentMap of the top 100 (q-value < 0.0001) pathways (Fig. 3d) revealed significant enrichment of mitotic processes (39 proteins including centralspindlin complex proteins MKLP1 and RACGAP1), signal transduction pathways (including 38 MAPK signalling pathway proteins such as MEK1/2/3, IQGAP1, PAK1/2, RAC1, SEPTINN7, SPTAN1, TLN1, XPO1, YWHAB), and RHO GTPase effectors (IQGAP1, IQGAP3, ITGB1, PAK1/2, RAC1/2, RHOA/G, XPO1, YWHAB/E/H/Q/Z)). A complete list of proteins involved in these pathways is given in Supplementary Data 8. We validated the expression of one such protein ($KRAS^{G12V}$ oncoprotein (Fig. 3e)) in SW620 cell-derived sMB-Rs by mass spectrometric analysis as well as by fluorescence microscopy and western blotting (Fig. 3f, g).

Shed midbody remnants are taken up by fibroblasts. To address whether sMB-Rs from cancer cells can influence non-cancer cells, like exosomes and microparticles, we treated NIH3T3 fibroblasts with purified sMB-Rs for 2 h and then used MLKP1 and RACGAP1 antibodies to evaluate vesicle uptake (Fig. 4a). Compared to untreated fibroblasts, sMB-R-treated fibroblasts displayed an approximate 4-fold increase in uptake and accumulation of MKLP1/RACGAP1-positive sMB-Rs (in green, up to eight sMB-R green puncta per recipient cell) (Fig. 4a, b). Uptake of sMB-Rs was evident within 1 h (Supplementary Fig. 19). Confocal microscopy revealed that sMB-Rs were internalised by NIH3T3 fibroblasts (Fig. 4c) and can deliver their protein cargo (Fig. 4d).

We envision that several cargo proteins in sMB-Rs collectively signal in recipient cells. These include MEK1/2/3, IQGAP1, PAK1/2, RAC1, SEPTINN7, SPTAN1, TLN1, XPO1 and YWHAB, which are effectors of various signalling pathways. Additionally, in our proteomic data set we also detect FGF3 and its receptor FGFR4. Because sMB-Rs released into the extracellular space during symmetric abscission contain the membrane envelope that originates from the plasma membrane of the parent cell, it is conceivable that MB-Rs might deliver functional FGF3/FGFR4 signalling complex to the recipient cells and activate its

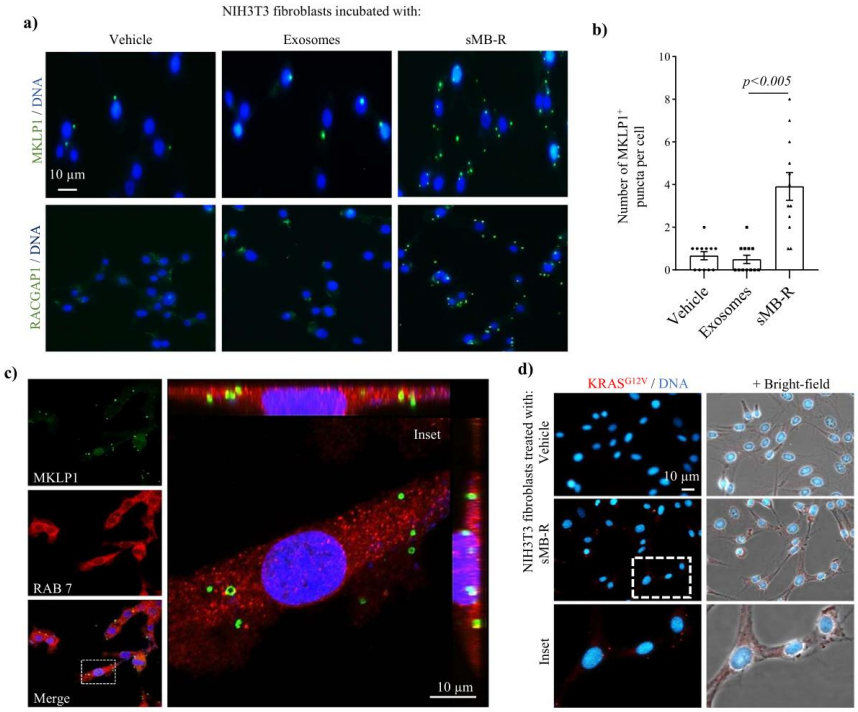


Fig. 4 Internalisation of cancer cell-derived shed midbody remnants by fibroblasts. a Uptake of sMB-Rs by fibroblasts. Fluorescence microscopy analysis of NIH3T3 fibroblasts incubated with/without SW620 cell-derived sMB-Rs or Exos (50 µg ml^{-1}) for 2 h using anti-MKLP1 and anit-RACGAP1 antibodies. **b** Uptake and accumulation of sMB-Rs in NIH3T3 fibroblasts was quantified by counting $MKLP1^+$ puncta per cell; data represented as mean ± s.e.m. Nuclei (blue) were stained with Hoechst. Scale bar, 10 µm. **c** Internalisation of sMB-Rs by fibroblasts. Confocal microscopy of NIH3T3 fibroblasts incubated with sMB-Rs using anti-MKLP1 (in green) and anti-RAB7 (in red) antibodies. Confocal microscopy analysis along Z-axis (inset) reveal internalisation of sMB-Rs following uptake. Scale bar, 10 µm. **d** Intercellular transfer of sMB-R $KRAS^{G12V}$ into NIH3T3 cells. Fluorescence microscopy of NIH3T3 fibroblasts incubated with SW620 cell-derived sMB-Rs (5 µg) for 2 h using anti-$KRAS^{G12V}$ antibodies. Nuclei were stained with Hoechst stain (blue). Right panel represents fluorescence signals from left panel overlaid onto bright-field images. Inset represents enlarged image. Scale bar, 10 µm.

downstream signalling pathways[39]. It is well documented that soluble secreted signalling molecules bound to their cognate receptors can be loaded onto the EV-surface (for example, TFGβ-1[40]). Non-specific binding of particles on EV surfaces has also been observed in physiological conditions such as binding of lipoprotein particles to blood EVs[41]. Whether specific or non-specific binding of factors on vesicular surfaces is of physiological significance remains an open question. Thus, we anticipate that several players collectively signal upon sMB-R uptake in recipient cells. In addition, although sMB-Rs are internalised by fibroblasts, downstream signalling could be mediated by engagement of surface receptors[20].

SW620 cell-derived sMB-Rs promote anchorage independent and invasive phenotype in fibroblasts. Because cancer EVs play an important role in cellular transformation, such as acquisition of invasive phenotype and anchorage independent cell growth capacity[42], we reasoned that SW620-derived sMB-R uptake by NIH3T3 fibroblasts might promote cell invasion and anchorage-independent growth capacity. To test whether sMB-Rs promote cell invasion, NIH3T3 fibroblasts were incubated with sMB-Rs (30 μg ml^{-1}) for 2 h, overlaid onto Matrigel^{TM} matrix coated inserts on a Transwell invasion assay plate, and the number of invading cells quantified (Fig. 5a). In contrast to control (untreated) fibroblasts that failed to invade, sMB-R-treated fibroblasts displayed significantly higher invasive capacity (>14-fold increase, $p < 0.005$, Fig. 5a); this invasive capacity was attenuated by pre-treatment of NIH3T3 fibroblasts with MEK inhibitor selumetinib (AZD6244). Next, to test whether sMB-Rs promote anchorage-independent growth, NIH3T3 fibroblasts were incubated with sMB-Rs (30 μg ml^{-1}) for 2 h and grown in 0.6% soft-agar suspension in a soft agar colony formation assay (Fig. 5b). Compared to control fibroblasts treated with vehicle alone that failed to form colonies, sMB-R-treated fibroblasts formed significantly greater numbers of colonies (>14-fold increase, $p < 0.005$) on soft agar, which was attenuated by pre-treatment of NIH3T3 fibroblasts with selumetinib. This fibroblast

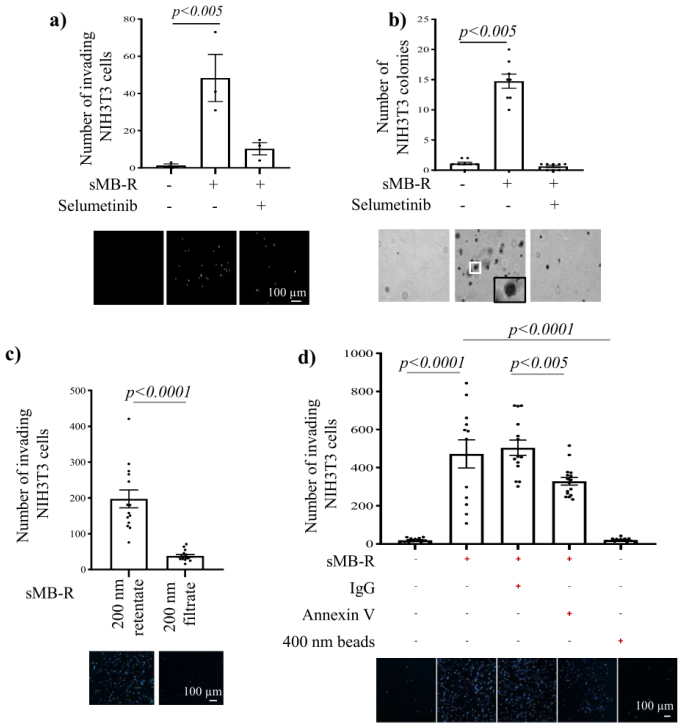


Fig. 5 Cancer cell-derived shed midbody remnants induce invasive/transformed phenotype in NIH3T3 fibroblasts. a sMB-Rs confer invasive capability in NIH3T3 fibroblasts. Transwell-Matrigel^{TM} invasion assay of NIH3T3 cells treated with SW620-derived sMB-Rs NIH3T3 fibroblasts were incubated with SW620 cell-derived sMB-Rs (30 μg ml^{-1}) for 2 h and overlaid onto Matrigel^{TM} matrix coated inserts on a Transwell invasion assay and the number of invading cells quantified. Data represented as mean ± s.e.m. **b** sMB-Rs confer anchorage-independent growth capability in NIH3T3 fibroblasts. Soft agar colony formation assay of NIH3T3 treated with SW620-derived sMB-Rs. NIH3T3 fibroblasts were incubated with SW620 cell-derived sMB-Rs (200 μg ml^{-1}) for 2 h and then grown in 0.6% soft-agar suspension in a soft agar colony formation assay. Data represented as mean ± s.e.m. **c** Transwell-Matrigel^{TM} invasion assay of NIH3T3 cells treated sMB-Rs obtained using size-based purification (0.22 μm pore size filter). The filtrate (flowthrough that contains small EVs (30–150 nm) and soluble protein contaminants) and the retentate (containing larger sMB-Rs) were used to assess fibroblast invasion. Data represented as mean ± s.e.m. Lower panel: fluorescence microscopy images of nuclei (stained with Hoechst) of invasive fibroblasts. **d** Transwell-Matrigel^{TM} invasion assay of fibroblasts treated with by sMB-R fraction depleted of sMB-Rs using biotin-Annexin V or 400 nm Polystyrene latex bead. Data represented as mean ± s.e.m. Lower panel: fluorescence microscopy images of nuclei (stained with Hoechst) of invasive fibroblasts.

cell transforming capacity of sMB-Rs is comparable to that of exosomes and sMVs-LD (Supplementary Fig. 20). Furthermore, we subjected purified sMB-Rs to size-based filtration (220 nm) to remove any remaining exosomes (Fig. 5c). The pro-invasive signalling capacity on fibroblasts by purified sMB-Rs was solely observed in the retentate that contains larger sMB-Rs *versus* the flowthrough, which contains residual exosomes and soluble secretome. We also show using 400 nm polystyrene latex beads that the acquired invasive capacity is not due to mere exposure to particles (Fig. 5d). Additionally, depletion of sMB-Rs using Annexin V resulted in ~30% reduction in the invasive phenotype of NIH3T3 fibroblasts (Fig. 5d). Although Annexin V has been shown to promote invasion in cancer cells[43,44], Peterman et al.[20] showed that PS on the outer leaflet of sMB-Rs is required for their engulfment by cells, thus any residual Annexin V carrying over to the invasion assay, albeit in very low amounts, would potentially reduce sMB-R-uptake by fibroblasts and thereby reduce their invasive capacity.

The mechanism by which sMB-Rs exert long-term effect is currently not understood. Similar to our findings, one-time treatment of MDA-MB-231 cells with sMB-Rs from Hela cells resulted in enhanced soft agar colony forming capacity (as evident from larger colonies over 14 days)[20]. Further, Peterman et al. showed that MB-Rs taken up by recipient cells were still present after 48 h post-feeding, potentially allowing for continuous signalling through at least two different pathways: αVβ3-FAK-Src and EGF-EGFR. Similar observations has also been made in the EV field where a single stimulation of U373 cells with EGFRvIII-containing EV caused a 2-fold increase in anchorage-independent soft-agar colony formation of U373 cells (over a period of 3 weeks), whereas exposure to the equivalent amount of microvesicles devoid of EGFRvIII resulted in no significant increase in colony forming capacity[45].

These findings suggest that upon uptake of sMB-Rs, initial signalling in recipient cells is sufficient to support or enhance a phenotype, in this case anchorage independent growth, for up to 10–14 days. Alternatively, MB-Rs may persist longer than 48 h to continually signal. In this regard, several studies have shown that MB-Rs persist in cells for extended periods. For example, accumulation of asymmetrically inherited MB remnants (up to 20 MB-Rs per cell) was shown to persist in stem cells and cancer cells[8]. Exogenously supplemented sMB-Rs have also been shown to persist for up to 48 h in cells following uptake[20]. While, phagocytosed particles are rapidly degraded within 2–5 h by fusing with lysosomes, internalised sMB-Rs can persist within actin-coated endosomes and thereby evade lysosomal degradation[8,20]. In our study, we show that sMB-Rs can persist in recipient cells for up to 48 h (Supplementary Fig. 21). However, it is conceivable that in vivo, continuous exposure to sMB-Rs might be required to support reprogrammed phenotype in recipient cells, as is the case of exosomes within the TME[45]. Although EVs have been previously shown to transfer mutated proteins that function in recipient cells, whether sMB-Rs deliver functional mutant KRAS warrants future investigation. On the other hand, sMB-Rs might be able to mediate epigenetic reprogramming of cells in a manner similar to exosomes[46]. In conclusion, characterising the biochemical properties of sMB-Rs is an indispensable first step towards understanding their underlying functionality. In this work we show that colon cancer-derived sMB-Rs can be isolated in high yield from the culture medium of SW620 cells grown in continuous culture tanks. Using a combination of differential centrifugation (10,000 × *g*) and OptiPrep™ (iodixanol) density gradient centrifugation ~1 mg of highly-purified sMB-Rs (based on protein concentration) exhibiting a range of particle diameter (200–600 nm) and buoyant density (in the range 1.22–1.30 g ml^{-1}) were obtained. GeLC-MS/MS analysis shows that sMB-Rs contain "cytokinesis signature proteins" (microtubule-bundling proteins, the centraspindlin complex (MKLP1/KIF23and RACGAP1) and chromosomal passenger complex proteins) not seen in exosomes and sMVs/microparticles. Functional studies show that sMB-Rs, like exosomes and low-density sMVs (buoyant density 1.08–1.14 g ml^{-1}) can be taken up and accumulate in quiescent fibroblasts where they promote cellular transformation and a pro-invasive phenotype[47,48]. Collectively, our findings show for the first time that sMB-Rs represent a third major class of EV molecularly distinct from exosomes and shed microvesicles/microparticles (Fig. 6). Our findings provide significant insights into sMB-R biology, suggesting this class of circulating EV may not merely constitute remnants of cytokinesis, but might also possess an unexpected role in cancer biology.

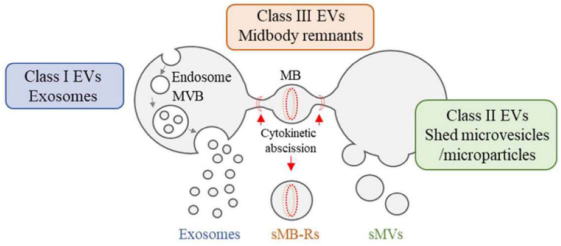
Methods

Cell lines and cell culture. SW620 (CCL-227, ATCC), SW480 (CCL-228, ATCC), LIM1863 cells[49] (Ludwig Institute for Cancer Research, Melbourne), COLO 205 (CCL-222, ATCC), T84 (CCL-248, ATCC), SW1463 (CCL-234, ATCC), SW1222 (12022910, CellBank Australia), LIM2405 (12062003, Sigma Aldrich) and LIM2408 (The Ludwig Institute for Cancer Research, Melbourne) cells were cultured in RPMI-1640 (Life Technologies). NIH3T3 fibroblasts (CRL-1658, ATCC), MDA MB 231 (HTB-26, ATCC), U87 (HTB-14, ATCC), HCT15 (CCL-225, ATCC), HCT116 (CCL-247, ATCC), HT29 (HTB-38, ATCC) and HCA7 (06061902, CellBank Australia) were cultured in DMEM (Life Technologies). Media were supplemented with 10% (v/v) Foetal Bovine Serum (FBS) (Life Technologies) and 1% (v/v) Penicillin Streptomycin (Pen/Strep, Life Technologies) and maintained at 37 °C with 10% CO_2. Cells were routinely passaged using trypsin-EDTA (Gibco).

Isolation of exosomes, sMV-LD and sMV-HD/sMB-Rs. Cells (SW620, SW480, MDA MB 231 and U87) were cultured in CELLine AD-1000 Bioreactor classic flasks (Integra Biosciences)[50]. SW620 or SW480 cells (3×10^7) in 15 ml of RPMI media (supplemented with EV-depleted 10% FBS and 1% Pen/Strep), or MDA MB 231 or U87 cells (3×10^7) in 15 ml of DMEM medium (supplemented with EV-depleted 10% FBS, 1% Pen/Strep) were added to the lower cell-cultivation chamber. The upper nutrient supply chamber contained 500 ml RPMI or DMEM (5% FBS, 1% Pen/Strep) that was replaced every 4 days. Cells in the lower cell cultivation chamber were allowed to adhere for 72 h at 37 °C with 10% CO_2. Thereafter, the lower chamber was gently washed with serum-free RMPI or DMEM medium. For SW620 or SW480 cells, 15 ml of RPMI media (supplemented with 0.5% (v/v) insulin transferrin selenium (Invitrogen) and 1% Pen/Strep) was added. For MDA MB 231 or U87 cells, 15 ml of DMEM medium (supplemented with EV-depleted 10% FBS, 1% Pen/Strep) were added to the lower cell-cultivation chamber. Thereafter, culture medium (CM) in the cell cultivation chamber was replaced each day. CM was sequentially centrifuged at 500 × *g* for 5 min (4 °C) and 2000 × *g* for 10 min (4 °C). The resultant supernatant was centrifuged at 10,000 × *g* (30 min) to pellet crude sMVs. The supernatant was then centrifuged at 100,000 × *g* (1 h) to pellet crude exosomes. Crude sMVs and crude exosomes were resuspended in 500 µl PBS and subjected to isopycnic (iodixanol-density) ultracentrifugation[24,51]. Briefly, a discontinuous gradient of OptiPrep™ (iodixanol solution) was prepared by layering 40% (3 ml), 20% (3 ml), 10% (3 ml) and 5% (2.5 ml) of iodixanol solution in a 14 × 89 mm polyallomer tube (Beckman Coulter). Dilutions of iodixanol solution were made in 0.25 M sucrose/10 mM Tris (pH 7.5) solution. Crude sMVs and crude exosomes (in 500 µl PBS) were overlaid and subjected to centrifugation at 100,000 × *g* for 18 h (4 °C). Next, twelve 1-ml fractions were collected, diluted in PBS (2 ml) and centrifuged at 100,000 × *g* for 1 h. The supernatant was discarded and pellets were further washed in PBS (500 µl) with final resuspension made in 100 µl of PBS and stored at −80 °C until further use. Human colon carcinoma LIM1863 cells were cultured in T175 flasks in RPMI medium supplemented with 0.5% (v/v) insulin transferrin selenium and 1% Pen/Strep and CM harvested as previously described[18] and subjected to EV isolation strategy as described above.

Size-based purification of sMB-Rs was performed using Ultrafree-CL Centrifugal Filter (Millipore, 0.22 µm pore size, hydrophilic PVDF). Briefly, purified sMB-Rs (150 µg in 600 µl PBS) was filtered using Ultrafree-CL Centrifugal Filter at 1000 × *g* (5 min, 4 °C). The filtrate (flowthrough that contains small EVs (30–200 nm) and soluble protein contaminants) and the retentate were then centrifuged at 10,000 × *g* (30 min, 4 °C). The pellets were resuspended in 150 µl PBS used for Transwell-Matrigel™ invasion assay (see below) at 1:1 (Fig. 5c).

Protein quantification and western blotting. Protein samples were quantified by 1D SDS-PAGE/SYPRO Ruby protein staining-based densitometry[18]. Western blotting was performed on protein samples (10–20 µg) as previously described[18]. Rabbit antibodies raised against GAPDH (Cell Signalling), β-tubulin (Cell Signalling), $KRAS^{G12V}$ mutant specific (Cell Signalling), RAB7 (Abcam) and GFP (Abcam) were used. Mouse antibodies against MKLP1 (Santa Cruz), RACGAP1 (Santa Cruz), ALIX (BD Biosciences), TSG101 (BD Biosciences), CD63



	Exosomes	sMB-Rs	sMVs
Biogenesis mechanism	Endosomal pathway	Cytokinesis (cytokinetic abscission)	Plasma membrane blebbing
Stereotypic protein marker	CD63, CD81, ALIX, TSG101	MKLP1 RACGAP1	???
Particle diameter range	30-200 nm	200-600 nm	100-500 nm
Buoyant density	1.08-1.14 g/ml	1.22-1.30 g/ml	1.08-1.14 g/ml

Fig. 6 Schematic illustration of three distinct classes of extracellular vesicles – exosomes, shed microvesicles/microparticles and shed midbody remnants (sMB-Rs). Exosomes (class I EVs) are of endosomal origin (formed by invagination of multivesicular bodies (MVB)), shed microvesicles/microparticles (class II extracellular vesicles) are formed via direct outward blebbing of plasma membrane and midbody remnants (MB-Rs) (class III extracellular vesicles) are generated by cytokinetic abscission of the interconnecting bridge between dividing daughter cells (post completion of cytokinesis). Current understanding of biophysical properties and stereotypic marker proteins for exosomes, sMVs (microparticles) and sMB-Rs are listed in the table. Centraspindlin complex proteins MKLP1 and RACGAP1 enable distinction of sMB-Rs from exosomes and sMVs.

(Santa Cruz), CD81 (Santa Cruz), RAB2A (Thermo Fisher), FLOT1 (BD Biosciences), α-actinin (Abcam), CD9 (Santa Cruz) and HSP90 (BD Biosciences) were used. Secondary antibodies used were IRDye 800 goat anti-mouse IgG or IRDye 700 goat anti-rabbit IgG (1:15000, LI-COR Biosciences).

Cryo-EM and NTA. Cryo-electron microscopy (Tecnai G2 F30) on EVs (2 μg) was performed as described[18]. Vesicle particle size was determined using a NanoSight NS300, Nanoparticle tracking analysis (NTA) (Malvern) system fitted with a NS300 flow-cell top plate with a 405 nm laser. Samples (1 μg μl^{-1}) were diluted in 500 μl PBS (1:10,000) and injected using 1 ml syringes (BD Biosciences) (detection threshold = 10, flowrate = 50, temperature = 25 °C). Each analysis consisted of 60 s video captures. Data was analysed using NTA software 3.0 (Malvern).

Immunofluorescence assay. Immunofluorescence was performed, as previously described[47]. Briefly, cells were cultured on Nunc® Lab-Tek® Chamber Slide™ (Sigma-Aldrich) system to 60–80% confluency. Cells were washed, fixed (4% formaldehyde for 10 min), permeabilized (0.2% (v/v) Triton X-100 in TTBS, 5 min) and blocked (3% (w/v) bovine serum albumin (BSA, Sigma) in TTBS (0.2% (v/v) Triton X-100) (blocking solution) for 30 min at room temperature. Cells were then incubated with primary antibodies (1:100) (mouse anti-MKLP1 (Santa Cruz Biotechnology), mouse anti-RACGAP1 (Santa Cruz), rabbit anti-$KRAS^{G12V}$ mutant specific (Cell Signalling), rabbit anti-RAB7 (Abcam) and rabbit anti-β-tubulin (Cell Signalling) in blocking solution for 1 h at room temperature. Cells were washed and incubated with secondary antibodies (1:200) (Alexa Fluor 488-conjugated goat anti-mouse IgG or Alexa Fluor 568-conjugated goat anti-rabbit IgG (Invitrogen) in blocking solution for 20 min at room temperature (in the dark). Cells were washed 3× in TTBS. Where indicated, nuclei were stained with Hoechst stain (10 μg ml^{-1}) for 1 min and actin labelled with Alexa Fluor 555 Phalloidin (Cell Signalling). Cells were imaged using a Zeiss AxioObserver Z1 microscope (Zeiss) or Zeiss Confocal LSM 780 PicoQuant FLIM (Zeiss) and images were analysed using Zen 2011 (Blue edition, Zeiss). For annexin V staining, live cells were labelled with Annexin V, Alexa Fluor™ 488 conjugate (ThermoFisher) according to the manufacturer's instructions. Briefly, cells were washed in cold PBS, and incubated with annexin binding buffer (5 μl of the annexin V conjugate in 100 μl of 10 mM HEPES, 140 mM NaCl, and 2.5 mM $CaCl_2$, pH 7.4) for 15 min at room temperature. Cells were washed with PBS, followed by fixation and immunofluorescence assay as above.

For 3-D culture, 500 cells were mixed with 50 μl Growth Factor-Reduced Matrigel™ matrix (Corning) and overlaid onto a Nunc® Lab-Tek® Chamber Slide™ system. The matrix was allowed to polymerise at 37 °C for 1 h and gently overlaid with growth medium. After 4–8 days, 3-D cultures were fixed with 2% aqueous formaldehyde and subjected to immunofluorescence assay.

Animal experiments were performed in accordance with La Trobe University Ethics committee guidelines. SW620 GAP GFP cells (1×10^6 cells/site) were subcutaneously injected into both inguinal regions of two NOD/SCID male mice to establish a total of four tumour xenografts. After 4 weeks, mice were killed, and tumours were excised, fixed in 4% aqueous formaldehyde, incubated in 20% sucrose solution for 48 h, embedded in optimum cutting temperature solution (Tissue-Tek®) and frozen (using isopentane). Sections (20 μm) were then subjected to immunofluorescence assay using mouse anti-MKLP1 antibody (1:100).

Isolation and culturing of intestinal crypts (as organoids) from small intestine or colon of C57BL/6 mice was performed using Gentle Cell Dissociation Reagent (STEMCELL™ Technologies) and IntestiCult™ Organoid Growth Medium (Mouse) (STEMCELL™ Technologies) as per manufacturer's instructions. Organoids were cultured in Growth Factor-Reduced Matrigel™ matrix for 7–10 days, fixed with 2% aqueous formaldehyde and subjected to immunofluorescence assay.

To quantify sMB-R particle number, 10 μg of sMB-Rs were subjected to immunofluorescence labelling as described above, using either mouse anti-MKLP1 (1:100, Santa Cruz Biotechnology) or mouse IgG isotype matched antibody (Abcam), and probed with Alexa Fluor 488-conjugated goat anti-mouse IgG (1:200). Labelled sMB-Rs were embedded in Matrigel™, and MKLP1 + particles imaged using Zeiss AxioObserver Z1 microscope and numbers quantified using Image J software v1.49e.

RACGAP1 immunoprecipitation assay. Dynabeads™ Protein G (Life Technologies) (10 μl) were conjugated with 1 μg RACGAP1 antibody (Santa Cruz) or mouse IgG isotype matched antibody (BD Biosciences) for 15 min at room temperature under continuous rotation. Antibody-bead conjugates were collected using a magnet. Next, sMB-Rs (200 μg) were solubilized in 0.5% TX-100-PBS (supplemented with Complete™ EDTA-free Protease Inhibitor Cocktail (Roche) and PhosSTOP™ Phosphatase inhibitor (Roche)) on ice for 30 min. Samples were centrifuged at 5000 × *g* for 1 min. The resultant supernatant was then incubated with antibody-conjugated Dynabeads™ Protein G for 2 h at 4 °C. Beads were washed 3× with 0.2% TX-100-PBS and proteins eluted in SDS sample buffer and analysed by GeLC-MS/MS.

EV loading on aldehyde/sulphate latex beads. EVs (30 μg) were incubated with 1 μl aldehyde/sulphate latex beads (Invitrogen) (total volume, 500 μl PBS) for 15 min (room temperature) with continuous rotation. The reaction was stopped using 100 mM glycine and 2% BSA in PBS for 30 min (room temperature) with continuous rotation. Beads were centrifuged at 5000 × *g* (2 min), washed again with 300 μl PBS, permeabilized (0.2% TX-100 in PBS for 5 min), blocked (0.2% TX-100, 10% BSA in PBS for 10 min), washed and incubated with anti-MKLP1 antibody (1:50 dilution in 0.2% TX-100/2% BSA in PBS for 1 h) at room temperature with continuous rotation. Beads were washed 3× and incubated with Alexa Fluor 488-conjugated goat anti-mouse IgG (Invitrogen) at 1:100 dilution in 0.2% TX-100/2%

BSA in PBS for 20 min under continuous rotation. Beads were washed 3×, resuspended in PBS and imaged using Zeiss AxioObserver Z1 microscope (Zeiss).

GeLC-MS/MS and data analysis. Proteomic experiments were performed in two independent biological replicates (with technical duplicates) using GeLC-MS/MS for each sample (sMB-Rs, exosomes, sMV-LD) as described previously[18]. Raw data was processed using Proteome Discoverer (v2.1, Thermo Fischer Scientific) and searched with Mascot (Matrix Science, London, UK; v2.5), Sequest (Thermo Fisher Scientific, San Jose, CA, v1.4.0.288), and X! Tandem (v2010.12.01.1) against the UniProt Human database comprising 71,785 entries. Data was searched with a parent tolerance of 10 ppm, fragment tolerance of 0.5 Da and minimum peptide length 7, with FDR 1% at the peptide and protein levels. Peptide spectral matches were validated using Percolator based on q-values at a 1% false discovery rate (FDR)[52]. Scaffold Q+/Q + S (Proteome Software Inc., Portland, OR, v4.8.7) was employed to validate MS/MS-based peptide and protein identifications from database searching[53]. Initial peptide identifications were accepted if they could be established at greater than 95% probability as specified by the Peptide Prophet algorithm. Protein identifications were accepted, if they reached greater than 99% probability and contained at least two identified unique peptides. Protein probabilities were assigned using Protein Prophet[54]. The relative abundance of a protein within a sample was determined using normalised spectral count (Nsc)[24].

GO enrichment analysis of proteins was conducted using DAVID Bioinformatics Resources 6.8 (https://david.ncifcrf.gov/). KEGG pathway analysis was conducted as previously described[50]. Protein-protein interaction networks were generated using STRING (http://string-db.org)[55]. Heatmaps were generated using R-package software.

sMB-R uptake and KRAS^{G12V} transfer assay. NIH3T3 fibroblasts were grown on Nunc® Lab-Tek® Chamber Slide™ system to 70% confluency. The medium was supplemented with sMB-R (5 µg), exosomes (5 µg) or PBS vehicle and cells further cultured at 37 °C for 2 h to allow uptake. The ratio of sMB-R particles to cells are 50:1 (Fig. 4d). Cells were then subjected to immunofluorescence microscopy analysis using anti-MKLP1 or anti-KRAS^{G12V} mutant-specific antibodies.

Soft-agar colony formation assay. NIH3T3 fibroblasts (20,000 cells) in 100 µl DMEM (1% Pen/Strep) were stimulated with 20 µg of SW620-sMB-R or 20 µl PBS vehicle for 2 h at 37 °C. The ratio of sMB-R particles to cells are 250:1 (Fig. 5b). Where indicated, experiments were performed in the presence of 10 nM selumetinib or DMSO vehicle. Fibroblasts were then mixed with 300 µl 0.3% agarose (in DMEM with 10% FBS, 1% Pen/Strep) that was pre-warmed to 40 °C in a water bath. The mixture was overlaid onto wells of a 24-well plate pre-coated with 300 µl 0.6% agarose (in DMEM with 10% FBS, 1% Pen/Strep). The mixture was allowed to solidify at 37 °C for 15 min. The wells were then gently overlaid with 500 µl DMEM (5% FBS, 1% Pen/Strep) supplemented with 10 nM selumetinib (AZD6244) or DMSO vehicle and maintained at 37 °C for 10 days. Culture medium was replaced every 2 days. Colonies were imaged using Zeiss AxioObserver Z1 microscope (Zeiss) under bright-field.

Transwell-Matrigel^{TM} invasion assay. Transwell-Matrigel^{TM} invasion assay was performed as previously described[47]. Briefly, Transwell inserts (8 µm pore size, Corning) were coated with 100 µl of 1 mg ml^{-1} growth factor reduced Matrigel^{TM} and allowed to polymerise for 4 h at 37 °C. NIH3T3 fibroblasts (50,000 cells) in DMEM (1% Pen/Strep) were incubated with either sMB-R (30 µg ml^{-1}) or PBS alone for 2 h at 37 °C. The ratio of sMB-R particles to cells are 75:1 (Fig. 5a, c, d). Where indicated, 400 nm Polystyrene latex beads (Thermo Fisher Scientific) were incubated with cells at 75:1. Cells were then carefully overlaid onto Matrigel^{TM}-coated inserts. The inserts were placed into wells of 24-well plate companion plate (Corning) that contained DMEM (5% FCS, 1% Pen/Strep) supplemented with either sMB-R (30 µg ml^{-1}) or PBS alone. Invasion chambers were incubated overnight (~16 h) at 37 °C to facilitate invasion. Experiments were performed in the presence of 10 nM selumetinib or DMSO vehicle, as indicated. Inserts were washed, cells fixed (4% (v/v) formaldehyde, 5 min), and nuclei stained with Hoechst stain (10 µg ml^{-1}) for 20 min. Non-invading cells were removed from the upper side of the inserts using cotton swab. Nuclei of fibroblasts that invaded to the lower side of the insert were imaged using Zeiss AxioObserver Z1 microscope. Centre of the membrane was imaged for each inset. Images were quantified using Image J software v1.49e.

Depletion of sMB-Rs was performed using Biotin-Annexin V (BD Biosciences). Briefly, we incubated purified sMB-Rs (150 µg) with Biotin-Annexin V (5 µl, BD Biosciences) or biotin-IgG (5 µg per 5 µl, Sigma Aldrich) in 500 µl 10 mM Hepes (pH 7.4), 140 mM NaCl, and 2.5 mM $CaCl_2$ solution for 30 min at room temperature using gentle rotation and centrifuged at 10,000 × g (30 min, 4 °C). The pellet was resuspended in 150 µl PBS with Dynabeads® M-280 Streptavidin (100 µg, Thermo Fisher Scientific), and incubated at room temperature for 30 min using gentle rotation. The magnetic beads were separated with a magnet for 3 min. The unbound fractions were then used for Transwell-Matrigel^{TM} invasion assay at 1:1.

Generation of SW620-GAP-GFP cells. pE-Growth-associated protein (GAP) (1-20 a.a., MLCCMRRTKQVEKNDEDQKI)-GFP plasmid was transfected using Lipofectamine^{TM} 2000 (Invitrogen) into SW620 cells that were seeded to 70% confluency in 6-well plate. Briefly, 10 µg of plasmid was mixed with 10 µl of Lipofectamine^{TM} 2000 in 500 µl RPMI medium at room temperature for 20 min. Cells were washed and overlaid with 1.5 mL RPMI (10% FBS). Plasmid-Lipofectamine^{TM} mixture was then overlaid onto the cells. Cells were incubated at 37 °C with 10% CO_2. Cells with stable expression of GAP-GFP fusion proteins were selected following multiple rounds of single cell cloning into wells of 96-well plate. Expression of the GAP-GFP fusion protein in the expanded colonies was monitored using a Zeiss AxioObserver Z1 microscope and analysed by BD FACSCanto II HTS (BD Biosciences) using FlowJo software (TreeStar).

Statistics and reproducibility. Quantitative data represented as mean ± standard error of mean (s.e.m.). Statistical analyses were performed using GraphPad Prism software (one-way ANOVA (Turkey test)) with $P < 0.05$ considered as statistically significant. No method of randomisation was used. Investigators were not blinded to allocation during experiments or outcome assessment.

For proteomics analysis, experiments were performed in biological duplicate (with technical duplicates). Data were searched with a parent tolerance of 10 ppm, fragment tolerance of 0.5 Da and minimum peptide length 7, with FDR 1% at the peptide and protein levels. Peptide spectral matches were validated using Percolator based on q-values at a 1% false discovery rate (FDR)[52]. Scaffold (Proteome Software Inc., Portland, OR, v 4.3.4) was employed to validate MS/MS-based peptide and protein identifications from database searching[54]. Initial peptide identifications were accepted if they could be established at greater than 95% probability as specified by Peptide Prophet[53]. Protein identifications were accepted, if they reached greater than 99% probability and contained at least two identified unique peptides. Protein probabilities were assigned by Protein Prophet[54].

Fluorescence microscopy analysis of MB-R shedding was performed ≥3× for SW620 cells and 2× for SW480 cells (Fig. 1a and Supplementary Fig. 3). Fluorescence microscopy analysis of MB-R shedding by SW620, SW480, LIM1215 (Fig. 1c) and mouse intestinal organoids (Supplementary Fig. 8) was performed 3×. Shedding of MB-R in vivo was analysed in 4 SW620-GAP-GFP tumour xenografts (Fig. 1d). Cryo-EM imaging in Fig. 2e was performed ≥3× for Exos, 3× for sMV-LD and 2× for sMB-R with similar results. Numbers of EVs counted from Cryo-EM analysis in Fig. 2f were 60 for exos, 50 for sMV-LD and 30 for sMB-R. NTA analysis of three EV subtypes in Supplementary Fig. 13 was performed ≥3×. Immunofluorescence detection of sMB-R on aldehyde sulphate latex beads (Fig. 2g) was performed 2×. Co-IP experiment was performed 2×. Relative protein abundance of three EV subtype was performed in duplicate from five different cell lines (Fig. 2h).

Western blot-based validation of specific proteins for each EV-subtype was performed 2× for Fig. 2d, Supplementary Figs. 9, 12, and 15. Western blot analysis for Fig. S10 was performed 1×. Isolation of three EV subtypes from CM was performed ≥3× for SW620, SW480 and LIM1863 cells and 2× for MBA MB 231 and U87 cells. Western blot analysis of MB-R shedding into CM by multiple cell types in Fig. S16 was performed 1×. Western blot analysis of Supplementary Fig. 10 was performed 1×.

Immunofluorescence microscopy analysis of KRAS^{G12V} in SW620 MB/ MB-R (Fig. 3f) was performed 3×. Western blot detection of KRAS^{G12V} in SW620 sMB-R (Fig. 3g) was performed 3×. Uptake of sMB-Rs by NIH3T3 in Fig. 4a and Fig. S19 was performed ≥3×. Confocal microscopy of sMB-R uptake and internalisation by NIH3T3 was performed 3× (Fig. 4c). Transfer of KRAS^{G12V} by sMB-R to NIH3T3 fibroblasts (Fig. 4d) was performed 2×. Anchorage independent growth assay (Fig. 5b, Supplementary Fig. 20) and Transwell invasion assays (Fig. 5a, c, d) were performed ≥3×.

For Western blot, antibodies were validated as noted on manufacturer's website. For immuno-fluorescence, antibodies were validated as noted on manufacturer's website. Uncropped/entire Western blot images are provided in Supplementary Fig. 22.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Raw mass spectrometry data is deposited in the PeptideAtlas: #PASS01206 or can be accessed at http://www.peptideatlas.org/PASS/PASS01206

Source data for Figs. 2f, i, 4b, 5a–d, S7 and S20b have been provided as Supplementary Data 9.

MKLP1- or RACGAP1-antibodies based immunohistochemical images of human colon cancer tissues are publicly available from the Human Protein Atlas (http://www.proteinatlas.org/) (Fig. 1d, Supplementary Figs. 5 and 6) and published here with permission. MiCroKITS (version 4.0) manually-curated database contains experimentally-verified midbody, centromere, kinetochore, telomere and spindle proteins[12] (http://microkit.biocuckoo.org).

Any remaining information can be obtained from the corresponding author upon reasonable request.

Received: 29 August 2019; Accepted: 15 February 2021;
Published online: 25 March 2021

References

1. Gromley, A. et al. Centriolin anchoring of exocyst and SNARE complexes at the midbody is required for secretory-vesicle-mediated abscission. *Cell* **123**, 75–87 (2005).
2. Goss, J. W. & Toomre, D. K. Both daughter cells traffic and exocytose membrane at the cleavage furrow during mammalian cytokinesis. *J. Cell Biol.* **181**, 1047–1054 (2008).
3. Pohl, C. & Jentsch, S. Midbody ring disposal by autophagy is a post-abscission event of cytokinesis. *Nat. Cell Biol.* **11**, 65–70 (2009).
4. Lujan, P., Rubio, T., Varsano, G. & Kohn, M. Keep it on the edge: the post-mitotic midbody as a polarity signal unit. *Commun. Integr. Biol.* **10**, e1338990 (2017).
5. Pollarolo, G., Schulz, J. G., Munck, S. & Dotti, C. G. Cytokinesis remnants define first neuronal asymmetry in vivo. *Nat. Neurosci.* **14**, 1525–1533 (2011).
6. Bernabe-Rubio, M. et al. Novel role for the midbody in primary ciliogenesis by polarized epithelial cells. *J. Cell Biol.* **214**, 259–273 (2016).
7. Lujan, P. et al. PRL-3 disrupts epithelial architecture by altering the post-mitotic midbody position. *J. Cell Sci.* **129**, 4130–4142 (2016).
8. Kuo, T. C. et al. Midbody accumulation through evasion of autophagy contributes to cellular reprogramming and tumorigenicity. *Nat. Cell Biol.* **13**, 1214–1223 (2011).
9. Ettinger, A. W. et al. Proliferating versus differentiating stem and cancer cells exhibit distinct midbody-release behaviour. *Nat. Commun.* **2**, 503 (2011).
10. Dubreuil, V., Marzesco, A. M., Corbeil, D., Huttner, W. B. & Wilsch-Brauninger, M. Midbody and primary cilium of neural progenitors release extracellular membrane particles enriched in the stem cell marker prominin-1. *J. Cell Biol.* **176**, 483–495 (2007).
11. Mullins, J. M. & Biesele, J. J. Terminal phase of cytokinesis in D-98s cells. *J. Cell Biol.* **73**, 672–684 (1977).
12. Crowell, E. F., Gaffuri, A. L., Gayraud-Morel, B., Tajbakhsh, S. & Echard, A. Engulfment of the midbody remnant after cytokinesis in mammalian cells. *J. Cell Sci.* **127**, 3840–3851 (2014).
13. Marzesco, A. M. et al. Release of extracellular membrane particles carrying the stem cell marker prominin-1 (CD133) from neural progenitors and other epithelial cells. *J. Cell Sci.* **118**, 2849–2858 (2005).
14. Kuo, J. C., Han, X., Hsiao, C. T., Yates, J. R. 3rd & Waterman, C. M. Analysis of the myosin-II-responsive focal adhesion proteome reveals a role for beta-Pix in negative regulation of focal adhesion maturation. *Nat. cell Biol.* **13**, 383–393 (2011).
15. Dionne, L. K., Wang, X. J. & Prekeris, R. Midbody: from cellular junk to regulator of cell polarity and cell fate. *Curr. Opin. Cell Biol.* **35**, 51–58 (2015).
16. Mishima, M., Kaitna, S. & Glotzer, M. Central spindle assembly and cytokinesis require a kinesin-like protein/RhoGAP complex with microtubule bundling activity. *Dev. Cell* **2**, 41–54 (2002).
17. Xu, R. et al. Extracellular vesicles in cancer - implications for future improvements in cancer care. *Nat. Rev. Clin. Oncol.* **15**, 617–638 (2018).
18. Xu, R., Greening, D. W., Rai, A., Ji, H. & Simpson, R. J. Highly-purified exosomes and shed microvesicles isolated from the human colon cancer cell line LIM1863 by sequential centrifugal ultrafiltration are biochemically and functionally distinct. *Methods* **87**, 11–25 (2015).
19. Arai, Y. et al. Lipidome of midbody released from neural stem and progenitor cells during mammalian cortical neurogenesis. *Front Cell Neurosci.* **9**, 325 (2015).
20. Peterman, E. et al. The post-abscission midbody is an intracellular signaling organelle that regulates cell proliferation. *Nat. Commun.* **10**, 3181 (2019).
21. Zuber, M. X., Strittmatter, S. M. & Fishman, M. C. A membrane-targeting signal in the amino terminus of the neuronal protein GAP-43. *Nature* **341**, 345–348 (1989).
22. Tauro, B. J. et al. Comparison of ultracentrifugation, density gradient separation, and immunoaffinity capture methods for isolating human colon cancer cell line LIM1863-derived exosomes. *Methods* **56**, 293–304 (2012).
23. Graham, J., Ford, T. & Rickwood, D. The preparation of subcellular organelles from mouse liver in self-generated gradients of iodixanol. *Anal. Biochem* **220**, 367–373 (1994).
24. Ji, H. et al. Proteome profiling of exosomes derived from human primary and metastatic colorectal cancer cells reveal differential expression of key metastatic factors and signal transduction components. *Proteomics* **13**, 1672–1686 (2013).
25. Kurasawa, Y., Earnshaw, W. C., Mochizuki, Y., Dohmae, N. & Todokoro, K. Essential roles of KIF4 and its binding partner PRC1 in organized central spindle midzone formation. *EMBO J.* **23**, 3237–3248 (2004).
26. Carmena, M., Wheelock, M., Funabiki, H. & Earnshaw, W. C. The chromosomal passenger complex (CPC): from easy rider to the godfather of mitosis. *Nat. Rev. Mol. Cell Biol.* **13**, 789–803 (2012).
27. Wilker, E. W. et al. 14-3-3sigma controls mitotic translation to facilitate cytokinesis. *Nature* **446**, 329–332 (2007).
28. Hofmann, J. C., Husedzinovic, A. & Gruss, O. J. The function of spliceosome components in open mitosis. *Nucleus* **1**, 447–459 (2010).
29. Lindon, C. & Pines, J. Ordered proteolysis in anaphase inactivates Plk1 to contribute to proper mitotic exit in human cells. *J. Cell Biol.* **164**, 233–241 (2004).
30. Hong, K. U. et al. Functional importance of the anaphase-promoting complex-Cdh1-mediated degradation of TMAP/CKAP2 in regulation of spindle function and cytokinesis. *Mol. Cell Biol.* **27**, 3667–3681 (2007).
31. Seki, A. & Fang, G. CKAP2 is a spindle-associated protein degraded by APC/C-Cdh1 during mitotic exit. *J. Biol. Chem.* **282**, 15103–15113 (2007).
32. Pohl, C. & Jentsch, S. Final stages of cytokinesis and midbody ring formation are controlled by BRUCE. *Cell* **132**, 832–845 (2008).
33. Wilson, G. M. et al. The FIP3-Rab11 protein complex regulates recycling endosome targeting to the cleavage furrow during late cytokinesis. *Mol. Biol. Cell* **16**, 849–860 (2005).
34. Kittler, R. et al. An endoribonuclease-prepared siRNA screen in human cells identifies genes essential for cell division. *Nature* **432**, 1036–1040 (2004).
35. Kittler, R. et al. Genome-scale RNAi profiling of cell division in human tissue culture cells. *Nat. Cell Biol.* **9**, 1401–1412 (2007).
36. Neumann, B. et al. Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature* **464**, 721–727 (2010).
37. Huang, Z. et al. MiCroKiTS 4.0: a database of midbody, centrosome, kinetochore, telomere and spindle. *Nucleic Acids Res* **43**, D328–D334 (2015).
38. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
39. Bai, Y. P. et al. FGF-1/-3/FGFR4 signaling in cancer-associated fibroblasts promotes tumor progression in colon cancer through Erk and MMP-7. *Cancer Sci.* **106**, 1278–1287 (2015).
40. Shelke, G. V. et al. Endosomal signalling via exosome surface TGFbeta-1. *J. Extracell. Vesicles* **8**, 1650458 (2019).
41. Sodar, B. W. et al. Low-density lipoprotein mimics blood plasma-derived exosomes and microvesicles during isolation and detection. *Sci. Rep.* **6**, 24316 (2016).
42. Pylayeva-Gupta, Y., Grabocka, E. & Bar-Sagi, D. RAS oncogenes: weaving a tumorigenic web. *Nat. Rev. Cancer* **11**, 761–774 (2011).
43. Wu, L. et al. Annexin A5 promotes invasion and chemoresistance to temozolomide in glioblastoma multiforme cells. *Tumour Biol.* **35**, 12327–12337 (2014).
44. Wehder, L. et al. Annexin A5 is involved in migration and invasion of oral carcinoma. *Cell Cycle* **8**, 1552–1558 (2009).
45. Al-Nedawi, K. et al. Intercellular transfer of the oncogenic receptor EGFRvIII by microvesicles derived from tumour cells. *Nat. Cell Biol.* **10**, 619–624 (2008).
46. Camussi, G. et al. Exosome/microvesicle-mediated epigenetic reprogramming of cells. *Am. J. Cancer Res* **1**, 98–110 (2011).
47. Rai, A. et al. Exosomes derived from human primary and metastatic colorectal cancer cells contribute to functional heterogeneity of activated fibroblasts by reprogramming their proteome. *Proteomics* **19**, e1800148 (2019).
48. Rai, A., Greening, D. W., Xu, R., Suwakulsiri, W. & Simpson, R. J. Exosomes derived from the human primary colorectal cancer cell line SW480 orchestrate fibroblast-led cancer invasion. *Proteomics* **20**, e2000016 (2020).
49. Whitehead, R. H., Jones, J. K., Gabriel, A. & Lukies, R. E. A new colon carcinoma cell line (LIM1863) that grows as organoids with spontaneous differentiation into crypt-like structures in vitro. *Cancer Res* **47**, 2683–2689 (1987).
50. Ji, H. et al. Deep sequencing of RNA from three different extracellular vesicle (EV) subtypes released from the human LIM1863 colon cancer cell line uncovers distinct miRNA-enrichment signatures. *PLoS One* **9**, e110314 (2014).
51. Ford, T., Graham, J. & Rickwood, D. Iodixanol: a nonionic iso-osmotic centrifugation medium for the formation of self-generated gradients. *Anal. Biochem* **220**, 360–366 (1994).
52. Greening, D. W., Kapp, E. A., Ji, H., Speed, T. P. & Simpson, R. J. Colon tumour secretopeptidome: insights into endogenous proteolytic cleavage events in the colon tumour microenvironment. *Biochim Biophys. Acta* **1834**, 2396–2407 (2013).
53. Keller, A., Nesvizhskii, A. I., Kolker, E. & Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**, 5383–5392 (2002).
54. Nesvizhskii, A. I. & Aebersold, R. Interpretation of shotgun proteomic data: the protein inference problem. *Mol. Cell Proteom.* **4**, 1419–1440 (2005).
55. Szklarczyk, D. et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* **39**, D561–D568 (2011).

Acknowledgements
A.R., D.W.G., M.C., R.X., W.S., and R.J.S. acknowledge funding support from La Trobe University, Melbourne, Australia. The authors thank Dr. Jacqueline Orian (La Trobe University) for assistance with immunohistochemistry.

Author contributions
A.R. and R.J.S. conceptualised the idea, designed the experiments and wrote the manuscript. A.R., D.W.G., R.X., W.S. performed experiments. A.R., D.W.G. and M.C. performed bioinformatics analysis.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-021-01882-z.

Correspondence and requests for materials should be addressed to A.R. or R.J.S.

Appendix Article 2

Extracellular vesicles in cancer — implications for future improvements in cancer care

Rong Xu[1], *Alin Rai*[1], *Maoshan Chen*[1], *Wittaya Suwakulsiri*[1], *David W. Greening*[1] *and Richard J. Simpson*[1]*

Abstract | The sustained growth, invasion, and metastasis of cancer cells depend upon bidirectional cell–cell communication within complex tissue environments. Such communication predominantly involves the secretion of soluble factors by cancer cells and/or stromal cells within the tumour microenvironment (TME), although these cell types have also been shown to export membrane-encapsulated particles containing regulatory molecules that contribute to cell–cell communication. These particles are known as extracellular vesicles (EVs) and include species of exosomes and shed microvesicles. EVs carry molecules such as oncoproteins and oncopeptides, RNA species (for example, microRNAs, mRNAs, and long non-coding RNAs), lipids, and DNA fragments from donor to recipient cells, initiating profound phenotypic changes in the TME. Emerging evidence suggests that EVs have crucial roles in cancer development, including pre-metastatic niche formation and metastasis. Cancer cells are now recognized to secrete more EVs than their nonmalignant counterparts, and these particles can be isolated from bodily fluids. Thus, EVs have strong potential as blood-based or urine-based biomarkers for the diagnosis, prognostication, and surveillance of cancer. In this Review, we discuss the biophysical properties and physiological functions of EVs, particularly their pro-metastatic effects, and highlight the utility of EVs for the development of cancer diagnostics and therapeutics.

Extracellular vesicle (EV). Lipid membrane-encapsulated particle released by cells into the intercellular space and/or circulation that functions in bidirectional cell–cell communication; EVs comprise at least two major subclasses — exosomes and shed microvesicles — with distinct cargo profiles of proteins, RNAs, DNA, and lipids

[1]*Department of Biochemistry and Genetics, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, Victoria, Australia.*

[2]*These authors contributed equally: Rong Xu, Alin Rai.*

**e-mail:* richard.simpson@latrobe.edu.au

https://doi.org/10.1038/s41571-018-0036-9

Bidirectional communication between cells and their microenvironment is crucial for both normal and pathological physiology. Such crosstalk is traditionally known to occur via direct cell–cell contact or the secretion of soluble factors (for example, cytokines, chemokines, and growth factors)[1–3]; however, over the past 10 years, a new paradigm involving extracellular vesicle (EV) trafficking has emerged as a mechanism of cell–cell communication[4,5]. EVs are secreted by many eukaryotic cell types in vitro and have been found in bodily fluids including blood, urine, bile, breast milk, and synovial, lacrimal, seminal, ascites, and bronchoalveolar lavage fluids, as well as in faeces[6]. EVs can be released in response to cell activation (for example, platelet activation), changes in pH, hypoxia, irradiation, injury, exposure to complement proteins, and cellular stress[7–9]. Interestingly, EVs are also secreted by plant cells[10] and pathogens, including bacteria, mycobacteria, archaea, and fungi[11], suggesting an important evolutionarily conserved mechanism of intercellular signalling. Accruing literature reveals at least two major classes of EV: exosomes and shed microvesicles (sMVs). Exosomes are endocytic in origin, being formed as intraluminal vesicles (ILVs) by inward budding of the limiting membrane of late endosomes (multivesicular bodies (MVBs)). MVBs subsequently traffic to and fuse with the plasma membrane, whereupon they release their ILV contents (now referred to as exosomes) into the extracellular environment (FIG. 1a). By contrast, sMVs are formed through the direct budding of the plasma membrane (FIG. 1b). EVs contain diverse cargos (for example, proteins, RNA, DNA, and lipids; FIG. 1c–d) that can be trafficked between cells as a means of intercellular communication at both paracrine and systemic levels (FIG. 1e).

EVs participate in a variety of normal physiological processes, including blood coagulation[12], innate and/or acquired immunity and immunomodulation[13,14], stem cell differentiation[15], tissue regeneration and angiogenesis[16], autophagy[17], embryo implantation[18], placental physiology, semen regulatory function, reproductive biology[19], and pregnancy[20]. Furthermore, EVs have been proposed to be novel mediators of cell–cell communication during the normal development and physiology of the nervous system and regeneration of normal neurons[21,22].

Key points

- Exosomes and shed microvesicles are two classes of small lipid-encapsulated extracellular vesicles (EVs) that transmit molecular messengers (functional proteins and nucleic acids) between cells to alter the phenotype of recipient cells.
- Each class of EVs has a distinct mechanism of biogenesis, and within each class, subtypes (subpopulations) exist that can be distinguished by their distinct protein and RNA signatures.
- The participation of exosomes in signalling between tumour cells and the microenvironment aids the establishment of the pre-metastatic niche and facilitates tumour progression.
- Circulating exosomes containing tumour-specific molecular signatures (oncoproteins, mRNAs, long non-coding RNAs, and DNA fragments) have clinical utility as next-generation biomarkers for liquid biopsy in cancer diagnosis and management.
- Standardized isolation protocols for EV subpopulations are required to enable interlaboratory data comparison and for the advancement of their clinical utility.
- Exosomes have potential as vehicles for the delivery of therapeutic agents and also as anticancer vaccines and could possibly guide changes in clinical practice.

In addition to their function in normal physiology, EVs also participate in pathological processes such as the progression of neurodegenerative diseases[23] and cancer[24–26]. In the context of cancer, EVs are involved in a wide range of processes that underlie cancer progression — the so-called 'hallmarks of cancer' (REF.[27]) — including inflammatory responses, angiogenesis, lymphogenesis, cell migration, cell proliferation, immune suppression, invasion, epithelial-to-mesenchymal transition, and metastasis. Unsurprisingly, because exosome biogenesis shares many similarities with virion assembly, the host exosome pathway is manipulated for viral pathogenesis[28]. EVs have been implicated in many facets of cancer development and progression[29,30] and are, therefore, ideal candidates as biomarkers and/or therapeutic tools for anticancer treatment.

In this Review, we discuss the biophysical properties and biogenesis of the two major EV classes, exosomes and sMVs; describe their cargos, including oncoproteins and oncopeptides, RNA species (for example, microRNAs (miRNAs), mRNAs, and long non-coding RNAs (lncRNAs)), lipids, and DNA fragments; describe their role in the tumour microenvironment (TME) and pre-metastatic niche; and highlight the utility of EVs for the development of cancer diagnostics and therapeutics. Platelet-derived vesicles[31] and other vesicle types, such as apoptotic bodies (50–2,000 nm in diameter)[32] that are generated from cells undergoing programmed cell death, will not be discussed in this Review.

EV classes, biogenesis, and cargos

At least two different classes of EVs have been identified on the basis of their mechanism of biogenesis: exosomes and sMVs. sMVs have also commonly been referred to as microvesicles, ectosomes, and microparticles; for clarity, we will use the term sMVs throughout this Review. These two classes of EVs are often distinguished on the basis of their size — with exosomes averaging between 30 nm and 150 nm in diameter and sMVs typically ranging in diameter from 50 nm to ~1,300 nm (REF.[33–35]) — and size discrimination alone has been

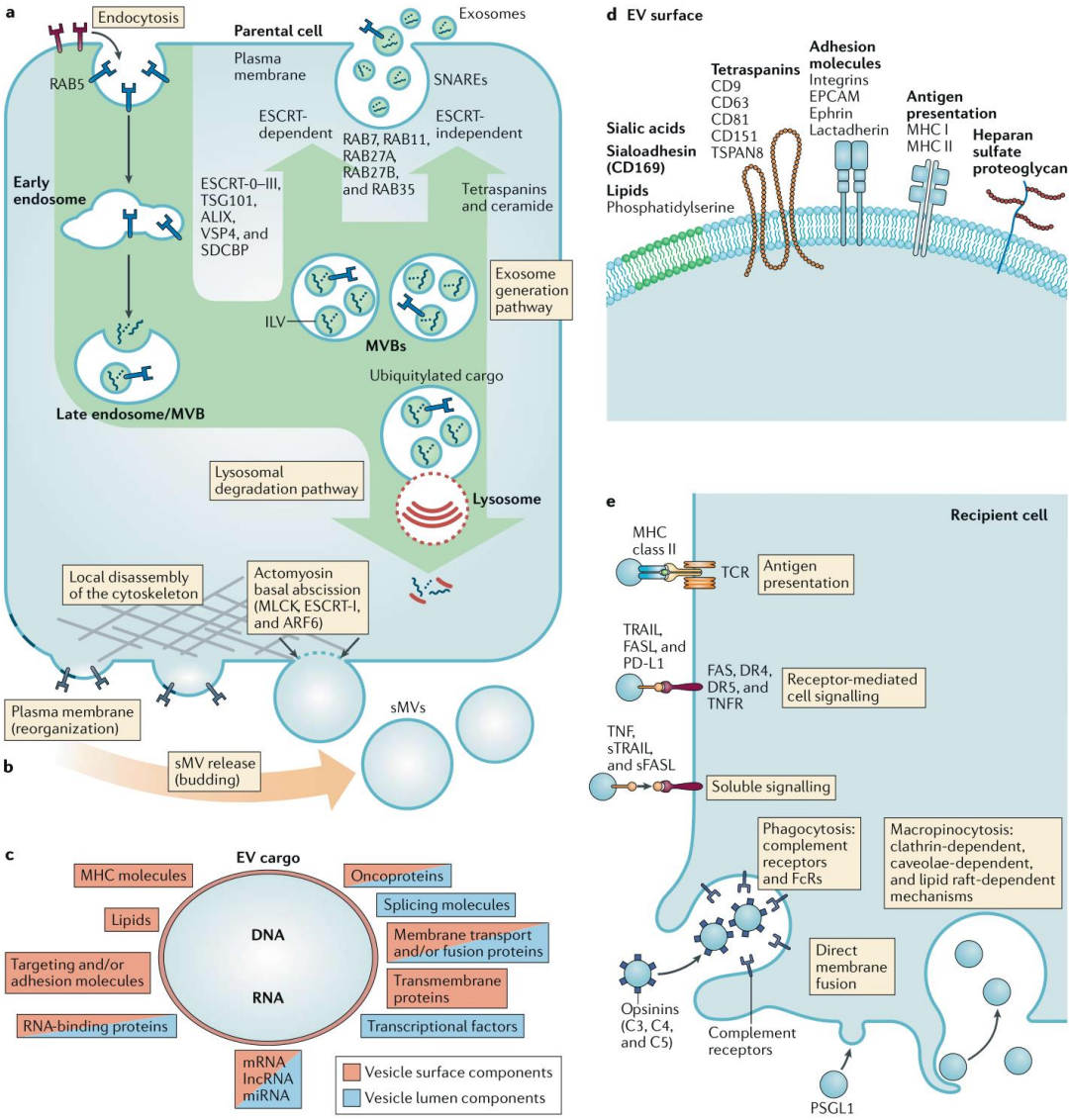
Exosomes
A major class of extracellular vesicle (typically 30–150 nm in diameter) of endocytic origin released by all cell types following fusion of multivesicular bodies with the plasma membrane.

Shed microvesicles
(sMVs). A major class of extracellular vesicle (typically, 50–1,300 nm in diameter) formed by direct budding from the plasma membrane; sMVs are also known as microparticles and ectosomes.

Tumour microenvironment
(TME). The area immediately surrounding a tumour that typically comprises nonmalignant lymphoid and/or myeloid cells, fibroblasts, pericytes, endothelial cells, lymphoid vessels, and extracellular matrix (collectively referred to as the stroma). The interaction between stromal cells and tumour cells has a critical role in cancer growth and metastasis.

Pre-metastatic niche
A microenvironment induced by factors released from the primary tumour in a distant organ that supports metastatic cell seeding, survival, and outgrowth.

Fig. 1 | **EV biogenesis and cargo contents. a** | Exosome biogenesis and release are coordinated through selected intracellular pathways. Endocytosis is an active process by which cells internalize material in the extracellular fluid by invagination and pinching of the plasma membrane to form internal vesicles (early and late endosomes). Intraluminal vesicles (ILVs) are formed through the inward budding of the late endosomal membrane (now referred to as multivesicular bodies (MVBs)). MVBs traffic to and fuse with the plasma membrane to release ILVs into the extracellular space — thereafter, the ILVs are referred to as exosomes. The exosome generation pathway can be regulated by either endosomal sorting complex required for transport (ESCRT) complexes (referred to as the ESCRT-dependent pathway) or via an ESCRT-independent pathway; these two pathways are thought to operate in parallel. Fusion of MVBs with the plasma membrane, and thus exosome release, is regulated by several RAB GTPases (including RAS-related protein RAB7A, RAB11, RAB27A, RAB27B, and RAB35) as well as membrane-fusion soluble *N*-ethylmaleimide-sensitive factor attachment protein receptor (SNARE) complex proteins. Alternatively, the MVBs can fuse with lysosomes, resulting in the degradation of ILVs. **b** | The biogenesis and release of shed microvesicles (sMVs) follows a distinct pathway. Key events include plasma membrane reorganization and the redistribution of phospholipids, the repositioning of phosphatidylserine to the outer leaflet of the plasma membrane, the local disassembly of the cytoskeleton network, and the contraction of the actin–myosin machinery via the activation of MLCK, ESCRT-I, and ADP ribosylation factor 6 (ARF6). **c** | Extracellular vesicles (EVs) can deliver nucleic acids (DNA and/or RNA, including mRNA, microRNA (miRNA), and long non-coding RNA (lncRNA)), proteins, and lipids that can be functional in recipient cells, including various oncoproteins, fusion and/or splice variant genes, transcriptional factors, and RNA-binding proteins. For a publicly accessible, and fully annotated, database of exosomal proteins, RNA, and lipids, refer to Exocarta2012 (REF.[231]) and Vesiclepedia[232] (a compendium of EVs with continuous community annotation). **d** | The specificity of EV-recipient cell targeting probably occurs by interactions between various cognate ligands and receptors presented on the cell surface and on EVs, for example, lipids (such as phosphatidylserine), tetraspanins, adhesion molecules (such as integrins), signalling receptors, and molecules involved in antigen presentation and membrane trafficking. **e** | EVs can be internalized by recipient cells via different mechanisms, including phagocytosis (dynamin-dependent, PI3K-dependent, and actin polymerization-dependent mechanisms), macropinocytosis, endocytosis (clathrin-dependent and caveolae-dependent mechanisms, and lipid raft-dependent endocytosis), and direct membrane fusion. Ligand–receptor interactions on the recipient cell surface can elicit biological responses and can also mediate targeting of vesicles to specific cell types. ALIX, ALG-2-interacting protein X; C3, complement C3; C4, complement component C4; C5, complement component C5; DR4, death receptor 4 (also known as TNFRSF10A); DR5, death receptor 5 (also known as TNFRSF10B); EPCAM, epithelial cell adhesion molecule; FAS, apoptosis-mediating surface antigen FAS; FASL, FAS antigen ligand (also known as FASLG); MHC, major histocompatibility complex; PD-L1, programmed cell death 1 ligand 1; PSGL1, P-selectin glycoprotein ligand 1; SDCBP, syntenin 1; sFASL, soluble FASL; sTRAIL, soluble TRAIL; TCR, T cell receptor; TNFR, TNF receptor; TRAIL, TNF-related apoptosis-inducing ligand; TSG101, tumour susceptibility gene 101 protein; TSPAN8, tetraspanin 8; VPS4, vacuolar protein sorting-associated protein 4.



considered useful in defining vesicle type; however, awareness of the size overlap between these two classes, especially in the smaller particle range (30–150 nm), is increasing[33,34]. Notably, the size of some EV classes is also affected by storage conditions. For example, in contrast to exosomes, which are stable upon repeated freeze–thawing, the size of sMVs is markedly reduced upon storage at −80 °C and repeated freeze–thawing (A.R. and R.J.S., unpublished observations). A list of biophysical attributes of EVs (for example, the range of particle diameters and their buoyant densities) is shown in Supplementary Figure 1.

Accumulating evidence from in vitro studies using cell cultures and samples of bodily fluids indicates that more than one exosome subtype exists[36–38]. Comprehensive protein and RNA profiling studies have revealed substantial differences in cargos between exosome subtypes, although whether marked functional differences exist between the subtypes is unclear and requires further experimentation using highly purified and well-characterized EVs. Other types of EVs, such as platelet-derived microparticles (~130–500 nm in diameter; also referred to as 'platelet dust') that are released upon platelet activation[12,39], are undoubtedly the most abundant vesicle type in blood.

Exosome biogenesis. Exosomes and sMVs differ not only in content but also in their mechanism of formation. The first step in exosome biogenesis involves the inward budding of the plasma membrane to form a membrane-bound vacuole (early endosome), which undergoes several changes as it matures to form a late endosome (FIG. 1a). The limiting membrane of the late endosome then buds inward and pinches off to form membrane-enclosed vesicles known as ILVs within the late endosome, which is referred to as a MVB thereafter[40]. At this juncture, the primary role of ILV-loaded MVBs is to function as intermediates in the lysosomal degradation pathway, whereby MVBs fuse with lysosomes, resulting in the degradation of the ILV contents by nucleases, proteases, lipases, and other hydrolytic enzymes within the lumen of the lysosome[41]. MVBs targeted for the lysosomal degradation pathway display specific surface proteins, such as the tumour suppressor His domain-containing protein tyrosine phosphatase (HD-PTP; also known as PTPN23), the HSP70–HSP90 organizing protein (HOP) complex (a co-chaperone of the HSP70–HSP90 complex), the GTPase RAS-related protein RAB7A, and members of the membrane-fusion soluble *N*-ethylmaleimide-sensitive factor attachment protein receptor (SNARE) complex, including vesicle-associated membrane protein 7 (VAMP7), which mediate vesicle transport through interaction with t-SNAREs homologue 1B (VTI1B), syntaxin 7 (STX7), and syntaxin 8 (STX8)[42,43].

Alternatively, MVBs destined for the formation of exosomes are integrated into the endosomal recycling system, resulting in their trafficking to and fusion with the plasma membrane, releasing their cargo of ILVs (at this point referred to as exosomes) into the extracellular space[44,45]. Four endosomal sorting complexes required for transport (ESCRT-0–III)[46–49] are known to be key drivers of MVB–ILV-based exosome formation, a process that involves reversible protein ubiquitylation in the selection of ILV protein cargos[50–52]. In addition to ESCRT-dependent formation of MVBs and exosomes, ESCRT-independent pathways involving neutral sphingomyelinase (N-SMase)-dependent ceramide formation[53], as well as ADP ribosylation factor 6 (ARF6) and phospholipase D2 (PLD2)[54], have been reported. Selected proteins, including syntenin 1 (SDCBP), are potential regulators of the endosomal targeting of exosome cargo to endosomal membranes and MVBs[55]. The ESCRT-dependent and ESCRT-independent machineries for MVB and exosome biogenesis can vary between tissues (or even cell types) depending on their specific lipid composition[23], but whether the two alternative mechanisms of exosome biogenesis can take place in a single MVB, or whether different populations of MVBs reflective of the two exosome biogenesis mechanisms can coexist within the cell, is unclear; however, this important aspect of EV biogenesis has not been explored in detail and warrants further study.

Shed microvesicle biogenesis. In comparison with exosome biogenesis, much less is known about sMV formation. In contrast to exosomes, sMV release is MVB-independent and does not require exocytosis; sMV egress from the cell occurs via direct budding from the plasma membrane through ARF6 (REF.[56]) and RHOA-dependent rearrangement of the actin cytoskeleton[57]. Similar to viral budding, sMV plasma membrane budding exploits the ESCRT (FIG. 1b). For example, the ESCRT-I subunit tumour susceptibility gene 101 protein (TSG101) has been shown to traffic to the plasma membrane and interact with accessory proteins ALG-2-interacting protein X (ALIX; also known as PDCD6IP) and arrestin domain-containing protein-1 (ARRDC1) during the later stages of sMV release, a process that is also implicated in sMV cargo sorting[58]. ESCRT-III and ALIX are also implicated in Gag-mediated budding of HIV virions from cells[59], as well as in cytokinetic abscission[60,61]. Interestingly, activation of acid sphingomyelinase (A-SMase), a downstream event in activation of the ionotropic ATP receptor P2X purinoceptor 7 (P2X7), triggers the release of sMVs from glial cells and astrocytes, an event that is thought to be a crucial mediator of neuroinflammatory disease[62]. This observation, along with the demonstration that N-SMase modulates the release of exosomes from oligodendrocytes[53], indicates that different members of the SMase family are key molecular effectors of EV formation and that inhibitors of these enzymes might provide new strategies for the treatment of neuroinflammatory disease.

Despite differences in their mechanism of biogenesis and membrane of origin — the MVB limiting membrane in the case of exosomes and the plasma membrane for sMVs — the two classes of EVs seem to function similarly after they are released into the extracellular space[63]. This aspect of EV biology is, however, far from being resolved, and elucidation of any functional differences will depend on further experimentation.

EV classes and cargo profiles. During their biogenesis, EVs are selectively enriched with an array of cellular bioactive cargo molecules (FIG. 1c). Interestingly, in the human colorectal cancer (CRC) SW480 and SW620 cell lines, our preliminary data indicate that many miRNAs undetectable in parent cell lysates using high-throughput next-generation sequencing (<5 transcripts per million reads (TPM)) are uniquely enriched (>1,000 TPM) in exosomes and sMVs secreted by these cells (M.C. and R.J.S., unpublished observations). This observation is important given the current global efforts to identify RNA signatures in cancer biopsy tissues using deep sequencing for the development of diagnostic biomarkers. In general, although much is known about the trafficking of cellular cargo molecules to EVs[64], our understanding of the underlying mechanism of cargo selection remains very much in its infancy. In addition, during EV biogenesis, diverse surface proteins that are characteristic of the parent cell are selectively displayed on secreted EVs (FIG. 1d). These proteins include RNA-binding proteins, ribonucleoproteins, signalling receptors, and integrins, the latter of which have a crucial role in the recognition of the target recipient cell and/or in EV uptake by recipient cells through various processes, including direct fusion, endocytosis (lipid raft-dependent, clathrin-dependent, and caveolae-dependent mechanisms, micropinocytosis, and phagocytosis), and T cell receptor (TCR)–major histocompatibility complex (MHC) and lymphocyte

function-associated antigen 1 (LFA1)–intercellular adhesion molecule 1 (ICAM1) interactions[65] (FIG. 1e).

The need to purify EVs

Over the past decade, awareness of the need to rigorously isolate specific populations of EVs for research purposes has been increasing within the EV research community. Foremost, highly purified EVs will be crucial to better understand the fundamental biochemistry of EVs (for example, to define their bioactive cargo) as a first step towards further elucidation of the mechanisms of EV biogenesis and functionality. Additionally, a pressing need exists to characterize proteins exposed on the EV surface for the purpose of generating monoclonal antibodies (mAbs) that would enable the discrimination of EV classes and/or subtypes (that is, using stereotypical protein markers that are unique to each subtype) and enable the large-scale purification of EVs for clinical applications, such as EV vaccines[66] and for the presentation of tumour-associated antigens to the immune system[67,68]. For a summary of commonly used EV purification methods for stringent biochemical analyses, diagnostic application (for example, the isolation of EVs from bodily fluids), and large-scale production for therapeutic studies, see BOX 1, Supplementary Box 1, and previous reviews[4,69].

Exosomes in the tumour microenvironment

The molecular composition of EVs secreted by diverse cell types is known to vary markedly, although phenotypic changes in recipient cells induced by exosome uptake — exosomes, specifically, given the paucity of functional data for sMVs — have in the past decade been directly attributed to the action of specific exosomal proteins and RNA molecules. Evidence supports a role for specific exosomal cargo molecules in crosstalk between constituent cells of the TME, for example, bidirectional cancer cell–stromal cell and/or cancer cell–cancer cell communication (TABLE 1).

The presence of occult tumours in situ is a widespread phenomenon (for example, in breast[70], thyroid[71], lung[72], and pancreatic cancer[73]) and has been extensively studied over the past 50 years, particularly with respect to why these tumours do not always become malignant. Despite possessing oncogenic mutations, neoplastic cells do not form tumours when injected into a nonmalignant microenvironment (for example, developing embryos)[74], an observation that is thought to be due to the ability of nonmalignant tissues to restrain aberrant growth and progression to malignancy[75,76]. This phenomenon led to a paradigm shift whereby cancer is now considered more than a disease defined by molecular (genetic and epigenetic) events within a cell and is also considered an 'ecological disease' that is modulated by components of the TME[77,78].

Over the past decade, tumour cell-derived exosomes have been shown to influence non-cancer cells to generate a TME that is permissive for tumour growth and metastasis (TABLE 1). For instance, tumour-derived exosomes influence endothelial cells to support neoangiogenesis, a process that fuels tumour growth[79] and induces vascular permeability, which facilitates metastasis[30]. Exosomes also trigger the differentiation of fibroblasts into pro-angiogenic and pro-tumorigenic cancer-associated fibroblasts (CAFs)[80,81]. Moreover, tumour-derived exosomes can initially suppress immune cells to evade immune detection of tumours[82,83] and, subsequently, as the cancer progresses, modify immune cells towards pro-tumorigenic[84] and pro-metastatic[30] phenotypes. Exosomes induce these phenotypic changes by transferring functional oncoproteins to recipient cells, in which they activate downstream signalling pathways, such as the MAPK and PI3K–AKT–mTOR pathways (TABLE 1). Exosomes also transfer miRNAs to

Box 1 | Methods for purifying EVs for biophysical studies and clinical use

The large body of literature describing protocols for extracellular vesicle (EV) purification attests to the technical challenges associated with this task[4,69] and the lack of a universally accepted approach (gold standard method). This problem is further confounded by emerging evidence that at least two major classes of EVs exist (exosomes and shed microvesicles) and that subtypes exist within each class. The majority of rapid, one-step approaches for isolating EVs does not address the possible mixture of vesicle classes and/or subtypes and co-isolated contaminants, such as high-molecular mass protein oligomers and protein–RNA complexes (for example, high-density lipoprotein (HDL), low-density lipoprotein (LDL), and protein Argonaute 2 (AGO2) complexes). Varying methodologies for purifying (enriching) EVs include differential centrifugation; density (sucrose, Percoll, or iodixanol) gradient centrifugation (DGC); high-performance liquid chromatography gel permeation (size-exclusion) chromatography; affinity chromatography using biospecific reagents (for example, monoclonal antibodies (mAbs) targeting glycoprotein A33 (GPA33), epithelial cell adhesion molecule (EPCAM), major histocompatibility complex class II (MHC II) antigens, CD45, CD63, CD81, CD9, CD1b, CD1d, CD14, CD24, or HER2) covalently fused to either magnetic or agarose beads; membrane ultrafiltration devices using low centrifugal force; microfluidic devices; and synthetic polymer-based precipitation reagents[4,200]. The choice of these diverse methods for EV isolation very much depends on the research question.

If the research aim is to purify EVs for stringent biochemical analysis (for example, to define their diverse luminal cargo molecules and surface-exposed proteins) and specific functionality, then rigorous fractionation strategies are critical. Evidence suggests that affinity chromatography using mAb-coated magnetic beads results in a higher exosomal yield than differential centrifugation and DGC methods[201].

In the absence of a suitable mAb, targeted EV capture methods that rely upon synthetic peptides with specific affinity for heat shock proteins[202] and vesicle-surface heparan sulfate proteoglycans[203,204] have been reported. Other successful methods for enriching EV classes include low gravitational (*g*) force sequential centrifugal membrane ultrafiltration[33] and differential centrifugation when used in combination with DGC (top-loaded)[18,37,205,206]. For a discussion on clinically relevant approaches in EV isolation, see Supplementary Box 1.

Table 1 | **Evidence supporting a role for exosomal cargos in cancer cell–cell communication**

Donor cell (exosome source)	Recipient cells	Functional cargo	Approaches for confirming cargo functionality	Key findings
Cancer cell to non-cancer cell transfer				
Mouse melanoma (B16F10) or human melanoma (A375) cells expressing type 2 tumour suppressor PEDF	Mouse bone marrow monocytes	PEDF	Overexpression of PEDF in metastatic melanoma cells; anti-PEDF antibody	PEDF-containing exosomes from cancer cells with poor metastatic capacity recruited natural killer cells and macrophages to the pre-metastatic niche, which in turn killed metastatic cancer cells that disseminated to pre-metastatic niche sites[234]
Mouse melanoma (B16F10) cells	Mouse BMDCs	MET	*Met* shRNA, *Rab27a* shRNA, and crizotinib (MET inhibitor)	Exosomes transferred MET to BMDCs, which resulted in their mobilization into the circulation and subsequent homing to the lung, where they generated a pre-metastatic niche to increase metastasis[30]
Mouse PDAC cells (R6560B, PAN02) and human PDAC cells (BxPC-3)	Human and mouse Kupffer cells	MIF	*MIF* shRNA	Exosomes containing MIF homed to the liver in mice, where they induced Kupffer cells to release TGFβ, which in turn stimulated hepatic stellate cells to deposit FN. FN then arrested circulating BMDCs in the liver to generate a pre-metastatic niche that increased liver metastasis[120]
Human prostate cancer cells (PC3)	Human primary fibroblasts (AG02262)	TGFβ	Anti-TGFβ antibody and TGFβ–SMAD inhibitor (SB431542)	Exosomes with high TGFβ expression induced SMAD3 signalling and α-smooth muscle actin expression in fibroblasts, resulting in differentiation of fibroblasts to myofibroblasts[80]
Human breast cancer cells (MDA-MB-231) and human glioma cells (U87)	Mouse fibroblasts (NIH3T3)	TG and FN	TG-encoding gene siRNA, TG inhibitor (T101) and FN inhibitor (RGD peptide)	Exosomal TG and FN activated mitogenic signalling, increased anchorage-independent growth, and increased survival in fibroblasts[90]
Human gastric cancer cells (SGC7901)	Mouse Kupffer cells and hepatic stellate cells	EGFR	*EGFR* siRNA	Tumour-derived EGFR-containing exosomes activated HGF by suppressing the microRNAs miR-26a and miR-26b in Kupffer cells; upregulated HGF increased liver metastasis of MET-expressing cancer cells[235]
Human glioblastoma cells (U87)	Human HMVECs	DLL4	DLL4 overexpression	Exosomal DLL4 transferred to endothelial cells inhibited Notch target genes (*HES1* and *HEY1*), resulting in loss of Notch receptor expression, which promoted angiogenesis[236]
Human primary glioblastoma cells	HMVECs	*GLuc* mRNA	*GLuc* mRNA overexpression	Transfer of functional exosomal mRNA to HMVECs was confirmed by expression of GLuc protein. Exosome-treated endothelial cells showed increased tubule formation; the identity of the functional mRNA causing angiogenesis has not been determined[29]
Breast cancer cells (MDA-MA-231-D3H2LN, BMD1a, and BMD2a and BMD2b)	Human HUVECs, pericytes, and astrocytes	miR-181c	miR-181c transfection and overexpression	Exosomes containing miR-181c compromised the blood–brain barrier by downregulating PDPK1 in endothelial cells, resulting in abnormal actin localization; this facilitated breast cancer metastasis in the brain[237]
Lung adenocarcinoma cells (CL1-5)	HUVECs	miR-23a	miR-23a inhibitor (transfection)	Exosomes derived from hypoxic lung cancer cells contained miR-23a, which suppressed EGLN2 and EGLN1 in endothelial cells, resulting in accumulation of HIF1α, which increased angiogenesis; exosomal miR-23a inhibited tight junction protein ZO-1 in endothelial cells, resulting in increased vascular permeability and transendothelial migration by cancer cells[79]
Human prostate cancer cells (Du145)	HUVECs	TGFβ	Inhibitor of TGFβ signalling (SB431542) and *RAB27A* knockdown	Exosomal TGFβ1 transformed fibroblasts to myofibroblasts, leading to increased angiogenesis in vitro and accelerated tumour growth in vivo[81]
Human multiple myeloma cells (RPMI8226, KMS-11, and U266)	HUVECs	miR-135b	miR-135b inhibitor (transfection)	Hypoxia-resistant multiple myeloma cancer cells that mimic the in vivo conditions of hypoxic bone marrow were shown to release exosomal miR-135b, which increased endothelial tube formation under hypoxia via the HIF1α inhibitor signalling pathway[238]
Cancer cell to cancer cell transfer				
Sunitinib-resistant RCC cells (7Su3rd)	Sunitinib-sensitive RCC cells (786-O)	lncRNA-ARSR	Locked nucleic acid targeting lncRNA-ARSR	Sunitinib-resistant RCC cells released exosomes containing lncRNA-ARSR, which upon uptake by sunitinib-sensitive RCC cells confer sunitinib resistance; lncRNA-ARSR was shown to bind to miR-34 and/or miR-449, resulting in increased expression of AXL and MET[89]

Table 1 (cont.) | **Evidence supporting a role for exosomal cargos in cancer cell–cell communication**

Donor cell (exosome source)	Recipient cells	Functional cargo	Approaches for confirming cargo functionality	Key findings
Cancer cell to cancer cell transfer (cont.)				
Human breast cells (MDA-MB-231) and B16 melanoma cells	Human breast cancer cells (T47D and MCF-7); mouse MMTV-PyMT metastatic mammary tumour model	Cre recombinase mRNA	Cre mRNA overexpression	Exosomes secreted by malignant tumour cells transferred Cre recombinase mRNA that was shown to be translated into functional protein in less malignant recipient tumour cells; exosome uptake conferred motility and metastatic capacity to recipient cells (the identity of functional proteins was not determined)[239]
Human glioma cells expressing EGFRvIII (U373vIII)	Human glioma cells lacking EGFRvIII (U373)	EGFRvIII	Annexin V (blocks recipient cell–exosome interaction) and CI-1033 (an irreversible pan-ERBB inhibitor)	Exosomes transferred EGFRvIII from EGFRvIII-expressing glioma cells to glioma cells lacking EGFRvIII to alter cancer cell morphology and increase anchorage-independent growth capacity in recipient cells in vitro[93]
Human colon cancer cells (mutant *KRAS* allele only; DKO-1)	Human colon cancer cells (wild-type *KRAS* allele only; DKs-8) and Rat intestinal epithelial cells (RIE1)	KRAS (mutant)	MRM mass spectrometry was used to detect mutant KRAS-G13D peptide (LVVVGAGDVGK) in DKs-8 recipient cells	Transfer of exosomes containing mutant $KRAS^{G13D}$ induced anchorage-independent growth in non-transformed colon cancer cells (wild-type *KRAS* allele) in vitro[96]
Non-cancer cell to cancer cell transfer				
Primary: human breast CAFs; secondary: human breast cancer cells (MDA-MB-231)	Primary: human breast cancer cells (MDA-MB-231); secondary: human breast cancer cells (MDA-MB-231)	CD81	*CD81* shRNA	Induction of breast cancer cell protrusion, motility (through PCP–WNT signalling), and metastasis in an exosomal CD81-dependent manner[101]
Human primary CAFs	Human breast cancer cells (MDA-MB-231)	RN7SL1 (signal recognition particle RNA)	POL3 inhibitor and siRNA knockdown of POL3	Protein-free RN7SL1 in CAF-derived exosomes stimulated DDX58 signalling in breast cancer cells, leading to increased chemoresistance and proliferation[104]
Mouse primary dermal fibroblasts	Human breast cancer cells (MDA-MB-231)	ADAM10	ADAM10 inhibitor (GI254023), ADAM17 and ADAM10 inhibitor (GW280264), ADAM inhibitor (TAPI-1), and metalloproteinase inhibitor (BB94); *ADAM10* knockdown (shRNA), and *Timp*-knockout mice	Exosomes secreted by *Timp*-knockout fibroblasts (which lack expression of metalloproteinase inhibitor 1) had increased ADAM10 levels and increased breast cancer cell motility[102]
Human BM-MSCs	Human breast cancer cells (BM2)	miR-23b	*MARCKS* siRNA and overexpression of *MARCKS* and miR-23b	miR-23b-containing exosomes derived from BM-MSCs induced a dormant phenotype via the suppression of the miR-23b target gene *MARCKS*[140]
Mouse primary astrocytes	Human breast cancer cells (MDA-MB-231) and mouse breast cancer cells (4T1)	miR-19a	Mutant miR-19a, *PTEN* shRNA, and *CCL2* shRNA	Astrocyte exosome-derived miR-19a induced loss of PTEN expression in metastatic breast cancer cells that had disseminated to the brain and potentiated metastasis[103]

ADAM, disintegrin and metalloproteinase domain-containing protein; BMDC, bone marrow-derived cells; BM-MSC, bone marrow mesenchymal stem cell; CAF, cancer-associated fibroblast; DDX58, probably ATP-dependent RNA helicase DDX58; DLL4, Delta-like protein 4; EGFRvIII, EGFR variant III; EGLN, EGL nine homologue; FN, fibronectin; GLuc, Gaussia luciferase; HIF1α, hypoxia-inducible factor 1α; HMVEC, microvascular endothelial cell; HUVEC, umbilical vein endothelial cell; lncRNA-ARSR, long non-coding RNA activated in RCC with sunitinib resistance; MIF, macrophage migration inhibitory factor; MRM, multiple reaction monitoring; PCP, planar cell polarity; PDAC, pancreatic ductal adenocarcinoma; PEDF, pigment epithelium-derived factor; PTEN, phosphatase and tensin homologue; RCC, renal cell carcinoma; RGD, Arg–Gly–Asp; RN7SL1, RNA, 7SL, cytoplasmic 1; shRNA, short hairpin RNA; siRNA, small interfering RNA; SMAD, mothers against decapentaplegic; TG, transglutaminase.

recipient cells, resulting in attenuated expression of target genes[79,85–88]. Furthermore, tumour-derived exosomal lncRNAs have been shown to be functionally important[89]. Taken together, the aforementioned studies — albeit, conducted using mouse models — demonstrate that 'exosome-transformed' non-cancer cells promote primary tumour growth[81,90].

A number of investigators have reported that human cancer cell line-derived EVs can transform nonmalignant, non-tumour cells in experimental systems

in vitro and in vivo. For example, CRC cell line-derived exosomes transform patient-derived stromal cells[91]; prostate cancer cell-derived EVs transform patient-derived adipose stem cells[92]; glioblastoma–astrocytoma cell line-derived exosomes containing mutant EGFR variant III (EGFRvIII) can transform glioblastoma cells lacking EGFRvIII expression[93]; nonmalignant mouse fibroblast cells form tumours in immunocompromised mice when treated with pro-tumorigenic breast cancer cell-derived microvesicles[90]; and nonmalignant human breast cells form tumours in immunocompromised mice upon treatment with breast cancer cell-derived exosomes or exosomes isolated from the sera of patients with breast cancer[94]. Whether such transformations occur in vivo, which has been speculated in the field[95], remains an open question.

Exosomes also transfer oncogenic entities, such as mutated proteins[93,96], fusion gene mRNAs (for example, *EML4–ALK*)[97], and oncogenic lncRNAs[89], from cancer cells to neighbouring cells in the TME. For instance, glioblastoma-derived exosomes transfer EGFRvIII — and its oncogenic signalling activity — to other indolent cancer cells to drive malignancy[93]. Apart from cancer 'hallmark-enabling' capabilities, exosomes can also transfer miRNAs that confer drug resistance[98]. In fact, the cargo content of cancer-derived exosomes can change in response to external cues, such as hypoxia[79], to induce angiogenic responses in endothelial cells, and in response to therapeutic agents (for example, sunitinib)[89] to confer drug resistance in neighbouring cells.

Increasing evidence also indicates that intratumoural heterogeneity (both spatial and temporal)[99] is not restricted to cancer cells. For example, stromal cells have been reported to co-evolve with tumour cells, such that both cell types continuously participate in driving cancer progression[100]. During this process, stromal cell-derived exosomes continually traffic to cancer cells to transfer functional proteins and RNAs that support tumour growth, invasion, and metastasis (TABLE 1). Moreover, stromal cell-derived exosomes can dictate progression to malignancy[101,102] and successful metastatic dissemination[103], and can also confer drug resistance[104,105], thereby influencing therapeutic outcomes in clinical settings.

Notably, most studies to investigate the role of exosomes in generating a permissive TME involved the use of mouse models and/or exosomes derived from human cancer cell lines grown in vitro. Importantly, emerging data suggest that exosomes isolated from patients with cancer are indeed functional. For example, exosomes isolated from malignant ascites[106] or the sera of patients with cancer[82,107] have been shown to be functional in that they modified nonmalignant and/or stromal cells to support tumour progression in vitro[82,106,107] and in vivo[94]. Taken together, these studies indicate that signalling reciprocity between cancer and non-cancer cells is an important aspect of cancer biology[77], with exosome-mediated signalling now attracting increasing attention[108] (TABLE 1).

Exosomes and the pre-metastatic niche

A major breakthrough in our understanding of the generation of predetermined metastatic microenvironments, referred to as pre-metastatic niches[24,109], arose from the pivotal observation that primary tumours can secrete factors ('fertilizer') that migrate to preferred metastatic sites, before the dissemination of cancer cells ('seeds'), to actively remodel these sites (that is, to generate a favourable 'soil') and promote metastasis[110] (Supplementary Box 2). Unlike nonmalignant tissues, which have an innate ability to suppress the outgrowth of tumour cells[76], pre-metastatic niches are shaped by the secretome and acquire traits of the primary tumour, which enable these distant sites to recapitulate the primary TME — these traits include vascular leakiness[111–113], inflammation[114], immune suppression[115], coagulation[116], stromal cell activation[104], and other features that support tumour outgrowth.

Numerous tumour-secreted factors[110,117] have been implicated in pre-metastatic niche generation. For example, primary tumour-derived soluble factors — including VEGFA, PlGF[110], the chemoattractants S100-A8 and S100-A9 (REF. 117), and tumour-secreted granulocyte colony-stimulating factor (G-CSF)[118] — are implicated in lung pre-metastatic niche generation[3,109,119]. Herein, we restrict our remarks to the contribution of primary tumour-derived exosomes in pre-metastatic niche generation[30,120,121]; the involvement of soluble protein and/or peptide factors has been described elsewhere[3,24,109].

Exosomal construction of the pre-metastatic niche. Our understanding of the biology of pre-metastatic niche generation is very much in its infancy, although observations in transgenic and orthotopic mouse models suggest that pre-metastatic niche construction occurs in a step-wise manner, a process that is similar across different cancer types[24]. Pre-metastatic niche construction begins with primary tumour-derived exosomes entering the circulation, via which they encounter the vascular beds of distant secondary organs, which are potential future metastatic sites. The blood vessels that drain tumours probably influence the pattern of secondary organ involvement, and exosomes have been shown to display organ tropism; for instance, exosomes derived from melanomas or breast cancers home primarily to the lung, liver, bone, and brain, whereas CRC-derived exosomes predominantly home to the liver (FIGS. 2,3). Very little is known about the exosomal surface proteins that dictate this organotropic homing behaviour, but emerging evidence implicates integral exosomal membrane proteins, such as integrins[121]. Intriguingly, systemic biodistribution studies in mice[30,120,121] have revealed that specific human cancer cell-derived exosomes disseminate to organs that mirror the parental cancer-specific metastases in the clinic (FIG. 2). Moreover, the vast majority of cancer cell-derived exosomes tested to date distribute to the bone marrow, suggesting that engagement between exosomes and bone marrow-derived cells (BMDCs) is a common feature of metastasis[24,30,120–122].

A crucial initial step in the generation of pre-metastatic niches within target organ tissues involves vascular leakiness, which is induced by a combination of disseminated cancer cell-derived exosomes[30,121] and soluble protein factors[111–113] that influence local

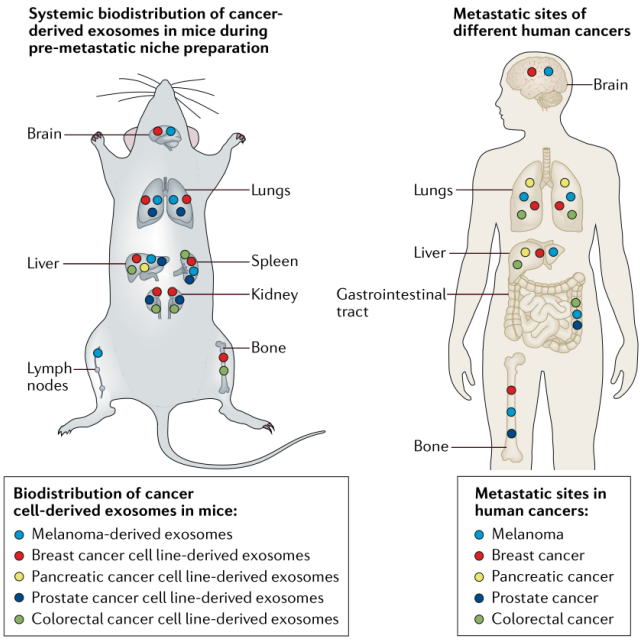


Fig. 2 | **Biodistribution of cancer exosomes in mice and common cancer metastatic sites in humans.** The left panel illustrates the biodistribution in mice of exosomes derived from melanoma[30] and breast[121,122], colorectal (A.R. and R.J.S., unpublished observations), pancreatic[120], and prostate[233] cancer cells. The schematic in the right panel shows the common metastatic sites of the same cancer types in humans on the basis of data from the National Cancer Institute.

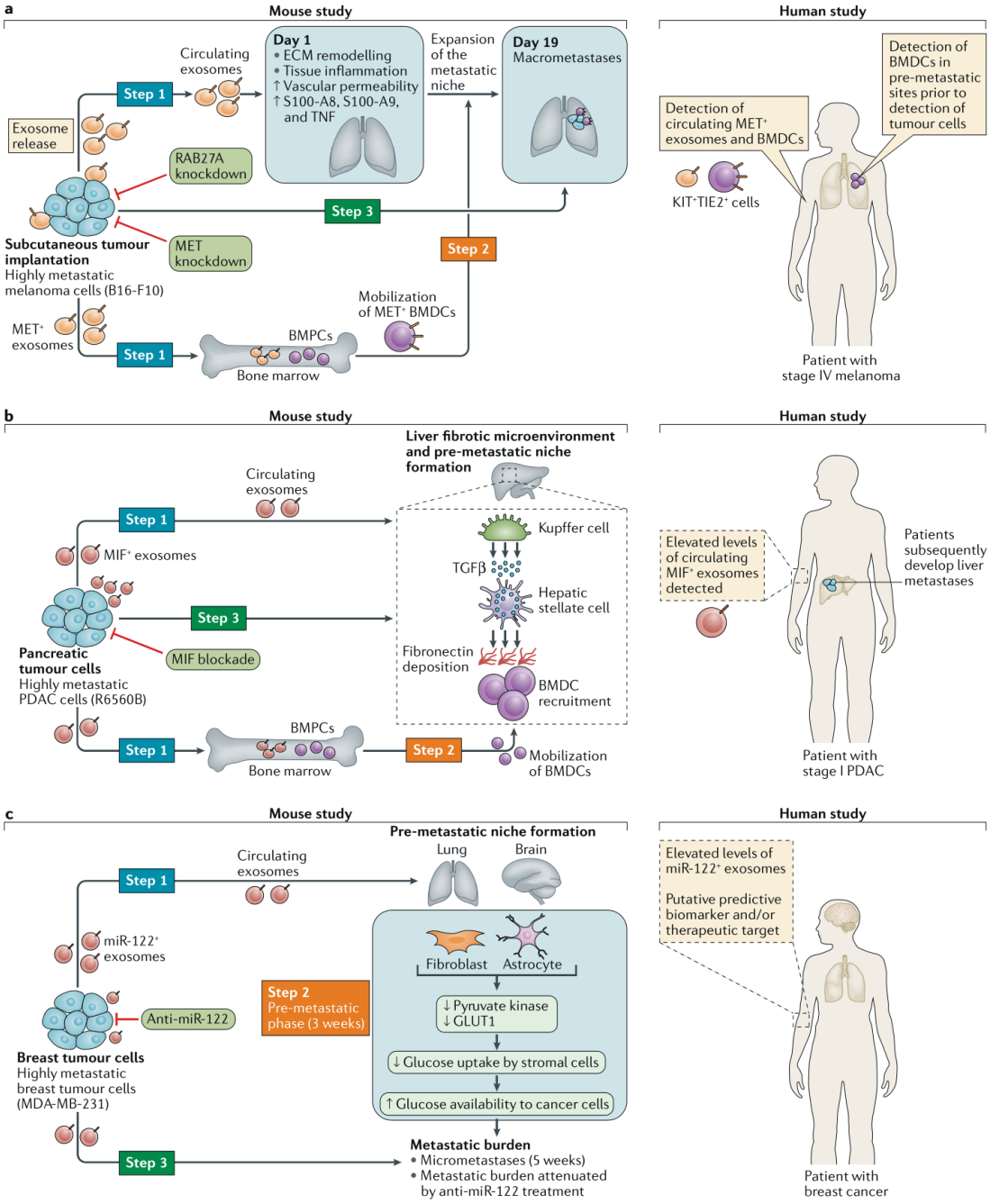
stromal cells. The nature of the recipient cell type varies depending on the cancer type and is specific to the secondary metastatic organ site. For example, pancreatic cancer-derived exosomes are taken up by Kupffer cells in the liver[120,121], whereas breast cancer-derived exosomes are taken up by fibroblasts[121,122] and/or epithelial cells[121] in the lung and by astrocytes[122] and/or endothelial cells[121] in the brain. The uptake of exosomes by stromal cells results in reprogramming of these cells[30,114,120,122] and activation of signalling pathways[114], which, in turn, can alter the local chemokine repertoire of the TME[30], remodel the composition of the extracellular matrix (ECM)[114,120,123], increase nutrient availability[122], and elicit neoangiogenesis[124] and lymphogenesis[123].

During the initial phase of pre-metastatic niche development, a pool of tumour-derived exosomes in the circulation home to bone marrow, where they are taken up by BMDCs, increasing the mobilization of these cells into the peripheral circulation[30,114,120]. The altered microenvironment in secondary organs attracts these circulating BMDCs[30,120], where upon infiltration, they secrete soluble factors that generate a local pro-inflammatory millieu[30,120] and/or exert pro-tumorigenic immunosuppression. Overall, these events culminate in the generation of a receptive pre-metastatic niche that attracts circulating cancer cells, therefore promoting metastasis.

Exosomal cargo in pre-metastatic niche development. The molecular composition of the pre-metastatic niche is directly influenced by components of the exosomal cargo, such as signalling proteins (for example, the receptor tyrosine kinase MET[30] and macrophage migration inhibitory factor (MIF))[120], RNA species[114,122], and potentially DNA[125]. For example, in an experimental model of melanoma metastasis[30], mouse melanoma (B16F10) cells — which are capable of metastasizing to the lung — released MET-containing exosomes (FIG. 3a). The MET receptor was transferred to BMDCs upon uptake of these exosomes, resulting in increased expression of KIT and TIE2, and promoted the subsequent mobilization of MET-containing BMDCs into the circulation[30]. Circulating tumour-derived exosomes also caused vascular leakiness in the lung, dysregulated ECM remodelling, and increased expression of inflammatory genes (for example, those encoding S100-A8 and S100-A9), which resulted in the recruitment of KIT^+TIE2^+ BMDCs, creating a pre-metastatic niche that supports lung metastasis. Genetic ablation of *Met* in the tumour cells from which the exosomes were derived or interfering with MET function using the pharmacological inhibitor crizotinib reduced the levels of circulating KIT^+TIE2^+ BMDCs in mice and attenuated the exosome-dependent effect on metastatic burden. In another experimental mouse model of pancreatic cancer metastasis to the liver[120], MIF-containing exosomes were shown to establish the pre-metastatic niche by inducing TGFβ expression and secretion in recipient Kupffer cells; TGFβ then activated hepatic stellate cells to deposit fibronectin, which subsequently facilitated recruitment of pro-tumorigenic bone marrow macrophages ($F4/80^+$), ultimately increasing liver metastatic burden (FIG. 3b).

Exosomal miRNAs are also important in pre-metastatic niche generation. For example, in an experimental mouse model of breast cancer metastasis to the brain and lungs[122], circulating breast cancer-derived exosomes containing high levels of miR-122 attenuated levels of pyruvate kinase in recipient astrocytes in the brain and in the lung fibroblasts. The downregulation of pyruvate kinase reduced the expression of glucose transporter type 1, erythrocyte/brain (GLUT1; also known as SLC2A1) and suppressed glucose uptake in niche tissues (brain and lung), thereby increasing glucose availability to cancer cells and resulting in increased metastatic outgrowth (FIG. 3c).

Sentinel lymph node pre-metastatic niche development. Most mouse models of metastasis focus on pre-metastatic niche generation via the circulation[24] (for example, through tail vein[30] or retro-orbital[120] injection of tumour cell-derived exosomes); however, metastasis also occurs via local lymphatic drainage, given that the presence of sentinel lymph nodes that are histologically positive for tumour cells is an indicator of a poor prognosis[126]. In mice bearing melanoma tumours, exosomes released by the primary tumour were more abundant in tumour-draining lymph nodes than in blood and other organs, such as the bone, lungs, and liver[127]. Alteration in the pre-metastatic sentinel lymph node microenvironment occurs in many cancer types[128–130] and might be a necessary step for the development of lymph node



metastases[131]. Indeed, melanoma cell-derived exosomes injected into mice via the footpad have been shown to foster a pre-metastatic niche in lymph nodes[123]. In the primed lymph node, exosomes resulted in upregulated expression of genes implicated in cell recruitment, ECM remodelling, and angiogenesis and also influenced the distribution pattern of disseminated tumour cells (DTCs).

Fig. 3 | **Exosomes and pre-metastatic niche formation.** Our current understanding of pre-metastatic niche formation has been guided by the use of orthotopic (part **a**) and transgenic (parts **a**, **b**, and **c**) mouse models. **a** | Exosomes from highly metastatic melanoma (B16-F10 cell line) increase the metastatic behaviour of primary tumours by permanently educating bone marrow progenitor cells (BMPCs) through the receptor tyrosine kinase MET[30]. Melanoma-derived exosomes home to the lung, where they induce vascular leakiness, dysregulate extracellular matrix (ECM) remodelling, and increase expression of inflammatory genes, such as the gene encoding TNF, and the chemoattractants S100-A8 and S100-A9 (step 1). Melanoma-derived exosomes also home to the bone marrow, where they educate BMPCs by transfer of MET (step 1). Exosome-educated $MET^+KIT^+TIE2^+$ bone marrow-derived cells (BMDCs) home to the lungs, where they facilitate generation and expansion of the pre-metastatic niche (step 2). This altered environment in the lung attracts circulating cancer cells to increase the metastatic burden (step 3). Exosome-mediated pre-metastatic niche formation is abrogated by knockdown of MET or RAS-related protein RAB27A (a GTPase important for exosome biogenesis) in tumour cells or by using the pharmacological inhibitor crizotinib, which has activity against MET. MET^+ exosomes and $MET^+KIT^+TIE2^+$ BMDCs have been detected in the circulation of patients with stage IV melanoma; BMDCs have also been detected in pre-metastatic sites in patients with melanoma[129]. **b** | Primary pancreatic ductal adenocarcinoma (PDAC)-derived exosomes (released from the R6560B cell line) induce liver pre-metastatic niche formation in tumour-free mice and consequently increase liver metastatic burden upon intrasplenic administration of cancer cells[120]. Uptake of circulating PDAC-derived exosomes by Kupffer cells causes the release of TGFβ, which activates hepatic stellate cells to deposit fibronectin (step 1). Subsequently, fibronectin deposition promotes the recruitment of BMDCs (step 2) and increases metastatic burden (step 3). Blockade of macrophage migration inhibitory factor (MIF), which is expressed at high levels in PDAC-derived exosomes, abrogates liver pre-metastatic niche formation and metastasis. Notably, MIF was found to be present at markedly higher levels in circulating exosomes derived from patients with stage I PDAC who later developed liver metastasis than in those derived from patients with pancreatic tumours that did not progress[120]. These findings suggest that exosomal MIF primes the liver for metastasis and is a prognostic marker for the development of liver metastasis in PDAC. **c** | Exosomes containing the microRNA miR-122 released from human breast cancer cells (MDA-MB-231 or MCFDCIS cell lines) home to the lungs and brain, where they are transferred to fibroblasts and astrocytes, respectively[122] (step 1). This generates a pre-metastatic niche by suppressing the glycolytic enzyme pyruvate kinase and glucose transporter type 1, erythrocyte/brain (GLUT1), which in turn results in decreased glucose uptake by stromal cells leading to increased glucose availability to cancer cells (step 2). This favourable niche is associated with an increased metastatic burden (step 3). This process can be reversed by anti-miR-122 treatment of the exosome-producing breast cancer cell lines. Notably, high exosomal miR-122 levels in the circulation have been associated with metastasis in patients with breast cancer[122].

Popliteal lymph node
A deep lymph node posterior to the knee embedded in the popliteal fossa that is moderately small in size, close to the popliteal vessels and superficial vessels, and functions as part of the lymphatic system of the lower leg and feet.

Active metastatic niches
Microenvironments in a distant organ that are conducive to metastasis but exist independently of the influence of the primary tumour.

Sleepy niches
Specialized microenvironments in which tumour cells survive in a dormant state, thereby extensively delaying the development of overt metastases.

In a mouse model of pancreatic cancer, the injection of exosomes — from highly metastatic pancreatic cancer cells — into the footpad induced the metastasis of poorly metastatic tumour cells to the popliteal lymph node[132]. This phenomenon is thought to occur via the exosomal transfer of functional miRNAs to stromal cells, resulting in alteration of their adhesion protein profile, secretion of chemokine ligands and proteases, and expression of genes that promote the cell cycle and angiogenesis[133].

Physiological exosomes in cancer

A pre-metastatic site, primed for metastasis, might not always require the primary tumour for its generation. For example, changes in the tissue environment owing to ageing[134] or infections[135] can potentially foster niches in the lung that are permissive to metastasis. Thus, some organs are intrinsically capable of supporting the metastatic outgrowth of DTCs without the influence of the primary tumour. For example, cancer cells directly injected into tumour-free mice successfully established metastatic foci[103]; these sites could be referred to as active metastatic niches, which, in contrast to pre-metastatic niche sites, do not require pre-conditioning by a primary tumour.

Active metastatic niche formation. One key line of evidence supporting the involvement of exosomes derived from non-tumour cells (physiological exosomes) in active metastatic niche formation was outlined in a study from 2015 (REF.[103]). In a mouse model of metastasis, breast cancer cells lost expression of an important tumour suppressor — phosphatase and tensin homologue (PTEN) — after dissemination to the brain, but not to other organs[103]. This loss of PTEN expression in the DTCs was shown to be mediated by the transfer of *PTEN*-targeting miR-19a from exosomes released by resident astrocytes. This adaptive loss of PTEN in cells that metastasize to the brain increased secretion of CC-motif chemokine 2 (CCL2), leading to the recruitment of ionized calcium-binding adaptor molecule 1 (IBA1; also known as AIF1)-expressing myeloid cells, which reciprocally promote the outgrowth of brain metastatic cells by increasing their proliferation and reducing apoptosis. Such niches exist independently of the influence of primary tumours and represent active metastatic niches.

Cancer cell dormancy. Given that the tumorigenic potential of oncogenes is context-dependent[76] and that secondary organs are generally inhospitable to DTCs, the pre-metastatic niche microenvironment attempts to recapitulate hallmark-enabling traits of the primary TME that initially supported the expansion of tumour cells. Cancer cells that have successfully disseminated to distant organs do not, however, always arrive in the pre-metastatic niche or active metastatic niches. As a result, tumour cells that enter tissues with unfavourable environments can enter long-term dormancy[136].

Secreted factors that regulate dormancy seem to be organ-specific[137,138]. Additionally, the ECM composition that DTCs encounter at distinct sites also induces dormancy[139] (such sites are often referred to as sleepy niches[24]). A growing body of evidence highlights the role of physiological exosomes in this process[136,140]. For example, exosomes released from bone marrow mesenchymal stem cells induced dormancy in breast cancer cells that disseminated to the bone by transferring miR-23b, which targets the gene encoding the cell division and motility-related protein myristoylated alanine-rich C-kinase substrate (MARCKS)[140].

Clinical utility of exosomes

The discovery of reliable biomarkers for the early detection of cancer is the 'holy grail' in diagnostic cancer research. Ideally, a useful biomarker should be specific for a given tumour type and universally detectable at pre-metastatic stages using non-invasive techniques. Despite a steadily increasing number of biomarker reports, few FDA-approved biomarkers have reached the clinic to date[141,142]. With the aim of improving biomarker identification, stringent guidelines for sample number,

Box 2 | **EVs — a source of novel and specific biomarkers of cancer?**

Extracellular vesicles (EVs) contain rare mRNA transcripts (of oncogenes, fusion genes, and tumour suppressor genes), DNA fragments (with cancer driver mutation genes), and long non-coding RNAs (lncRNAs) that could be a source of novel and specific cancer biomarkers.

EV DNA fragments

- Double-stranded DNA fragments have been detected in exosomes isolated from diverse human cancer cell lines[207,208].
- The amount of tumour cell-derived exosomal DNA is ~20-fold greater than the amount of fibroblast-derived exosomal DNA[208].
- *EGFR* mutations have been detected in exosomal DNA from non-small-cell lung cancer (NSCLC) cell lines[208].
- High mutation rates in exosomal DNA compared with tumour DNA might reflect tumour heterogeneity that can be detected more effectively using exosomes (owing to high levels of exosome release from tumour cells) than is possible through sequencing of DNA from tissues samples[209].
- The *Myc* oncogene can be amplified in serum EVs from tumour-bearing mice[125].
- Mutated *KRAS* and *TP53* have been detected in the serum exosomal DNA from patients with pancreatic cancer[207].

EV mRNAs

- Exosomal EGFR variant III (EGFRvIII) mRNA promoted tumour growth by transfer to and uptake of this mRNA by endothelial cells[93].
- Exosomal *HRAS* mRNA initiated tumour formation by non-tumorigenic endothelial cells and fibroblasts[210].
- The prostate cancer mRNA biomarkers prostate cancer antigen 3 (PCA3) non-coding RNA and *TMPRSS2–ERG* mRNA have been detected in urinary exosomes from patients with prostate cancer after mild prostate massage[211].
- Oncogenic *KRAS* mutations ($KRAS^{G12D}$ and $KRAS^{G12V}$) mRNAs have been detected in serum exosomes of patients with pancreatic cancer[162].
- Shed microvesicles, glycoprotein A33 $(GPA33)^+$ exosomes, and epithelial cell adhesion molecule $(EPCAM)^+$ exosomes released by human colon cancer (LIM1863) cells contain mRNAs that are upregulated in colorectal cancer (CRC) tumour tissues compared with matched normal tissues[212].
- Combined measurement of exosomal RNA and circulating DNA improved the sensitivity of *EGFR* mutation detection in plasma of patients with NSCLC[213].

EVs provide new sources of neoantigens

- High-resolution profiling of genomic and transcriptomic landscapes identifies plasma exosomes containing copy number variations, point mutations, insertions, deletions, gene fusions, and mutational signatures, which are new sources of neoantigens[209].
- CRC cell line LIM1863-derived EVs contain 268 novel alternative splicing events and 33 fusion genes, many of which have been reported in other cancers — for example, *SH3D19–LRBA* in primary myelofibrosis, *RIPK2–OSGIN2* in primary urethral clear-cell adenocarcinoma, and *GOLT1A–KISS1* in bladder cancer[212].

Diagnostic potential of EV lncRNAs

- Exosomal lncRNAs are stable in plasma; exosomal cytoskeletal regulator RNA (CYTOR) has been evaluated as a biomarker for gastric cancer[214].
- Exosomal lncRNAs, including HOX transcript antisense RNA (HOTAIR), metastasis associated lung adenocarcinoma transcript 1 (MALAT1), and maternally expressed 3 (MEG3), are differentially expressed in exosomes derived from patients with cervical cancer and cancer-free volunteers (human papillomavirus (HPV)-positive or HPV-negative)[215].
- CRC cell line LIM1863-derived EVs are enriched for the lncRNAs small nucleolar RNA host gene 5 (SNGH5)–8, *ZNFX1* antisense RNA 1 (ZFAS1), imprinted maternally expressed transcript H19, and small integral membrane protein 37 (SMIM37), which are upregulated in CRC tumour tissues[212].
- The combination of two mRNAs (*KRTAP5-4* and *MAGEA3*) and one lncRNA (breast cancer anti-oestrogen resistance 4 (BCAR4)) has been evaluated as a potential CRC biomarker[216].

target specificity, and sensitivity in biomarker discovery research have been formulated[142,143].

In contrast to traditional solid biopsies, which are impractical for screening or prognostic assays, liquid biopsies that focus on circulating tumour cells, cell-free tumour DNA, cell-free tumour RNA, and, more recently, exosomes are rapidly gaining recognition in precision or personalized medicine owing to the ease and non-invasive nature of sample collection[144]. The great strength of liquid biopsy is the ability to provide clinical information before and during treatment for therapeutic planning and monitoring. Over the past decade, circulating exosomes have been shown to be a reliable source of cancer-associated molecules (typically miRNAs) with potential as biomarkers for many cancer types[145], including hepatocellular carcinoma[146], lung cancer[147–150], gastrointestinal cancer[151], CRC[152], pancreatic cancer[153], melanoma[154–157], breast cancer[158], ovarian cancer[159], and prostate cancer[160]. In addition to miRNAs, other exosomal cargo molecules, such as oncogenic mRNAs (including fusion gene and splice variant transcripts), double-stranded DNA fragments (including cancer driver mutation genes), lipids, and lncRNAs, are gaining much attention as potential biomarker candidates (BOX 2; TABLE 2). Over the past 20 years, much interest has also been placed on the application of exosome-based cell-free vaccines as alternative approaches to adoptive dendritic cell therapy for suppressing tumour growth[161] (BOX 3).

Two seminal reports of EV-derived biomarkers potentially enabling the detection of pancreatic cancer have generated much interest and discussion[162,163]. The potential implications of these studies are enormous given that pancreatic cancer is currently the third leading cause of cancer-related death in the USA[164] and that early detection of this disease has been hampered by the lack of sufficiently specific and sensitive biomarkers. In both studies[162,163], the authors interrogated exosomal levels of glypican 1 (GPC1), which is expressed on the surface of cancer-associated exosomes circulating in the blood. GPC1 is a membrane-anchored heparan sulfate proteoglycan that is overexpressed in diverse tumour types including glioma[165], breast[166], colorectal[167], and pancreatic[168] tumours. Indeed, exosomal GPC1 and its regulatory miRNAs (miR-96-5p and miR-149) have been reported to be specific biomarkers for the diagnosis of CRC, as well as targets for CRC therapy[167].

In the first report[162], using blood samples from 251 patients divided into discovery and validation cohorts, EV isolation and GPC1 identification enabled the partition of patients with late-stage pancreatic cancer from those with benign pancreatic disease, with a reported accuracy of 100%. In the second seminal study[163], however, exosomal GPC1 was not diagnostic for pancreatic cancer on the basis of an analysis of 12 samples, although an exosomal miRNA signature — comprising high expression of miR-10b, miR-21, miR-30c, and miR-181a and low expression of let-7a — reproducibly differentiated between pancreatic cancer and non-malignant tissue samples. Furthermore, in contrast to exosomal GPC1, within 24 h of pancreatic cancer resection, the levels of these signature miRNAs reverted to levels similar to those in physiological exosomes from nonmalignant tissue from individuals without

Table 2 | **Preclinical evaluation of exosomal cargo as cancer biomarkers**

Cargo	Cancer type	Patient cohorts	Exosome source (isolation method)	Assay used	Outcome and/or utility
DNA					
$KRAS^{G12V}$, $KRAS^{G12D}$, $KRAS^{G12R}$, $KRAS^{G12C}$, $KRAS^{G12S}$, $KRAS^{G12A}$, and $KRAS^{G13D}$ DNA	PDAC	39 patients with PDAC and 82 healthy controls	Plasma (ultracentrifugation)	ddPCR	Exosomal mutant *KRAS* DNA was better for predicting PDAC status than cell-free mutant *KRAS* DNA[240]
$KRAS^{G12D}$ and $TP53^{R273H}$ DNA	Pancreatic cancer	48 patients with PDAC, 7 patients with intraductal papillary mucinous neoplasm, 9 patients with chronic pancreatitis, 12 patients with other pancreatic diseases, and 114 healthy controls	Serum (ultracentrifugation)	ddPCR	Circulating exosomal DNA enabled the detection of $KRAS^{G12D}$ DNA and $TP53^{R273H}$ DNA in 39.6% and 4.2% of patients with PDAC, respectively; however, 2.6% of healthy individuals were positive for the *KRAS* mutation, although none had the $TP53^{R273H}$ mutation[199]
Activating *EGFR* mutant and $EGFR^{T790M}$ mutant DNA	NSCLC	56 patients of the TIGER-X cohort, 50 patients with low levels of $EGFR^{T790M}$ plasma ctDNA, and 21 patients with intrathoracic disease (M0 or M1a)	Plasma (ExoLution Plus)	EXO100 next-generation sequencing of exosomal RNA and DNA, and BEAMing of ctDNA	In comparison with ctDNA analysis alone, combined exosomal RNA–DNA and ctDNA analyses increased the detection sensitivity of activating *EGFR* mutant and $EGFR^{T790M}$ DNA[213]
RNA					
AR-V7 mRNA	CRPC	26 patients receiving abiraterone and 10 patients receiving enzalutamide	Plasma (exoRNeasy kit)	ddPCR	AR-V7 mRNA predicted resistance to hormonal therapy; overall survival was significantly shorter in exosomal AR-V7-mRNA$^+$ participants than in exosomal AR-V7-mRNA$^-$ participants (3 months versus 20 months)[241]
Multiple RNAs	Prostate cancer	148 patients with high-grade prostate cancer and 371 patients with low-grade prostate cancer	Urine (EXOPRO Urine Clinical Sample Concentrator Kit)	RT-PCR	In comparison with the standard of care approach based on clinical risk factors (PSA level, age, race, and family history), urinary exosome gene expression profiling plus standard of care performed better at distinguishing patients with high-grade and low-grade (benign) disease (AUC 0.63 versus 0.73)[242]
Neoantigen transcripts and/or fusion genes	Pancreaticobiliary cancer	3 patients with pancreaticobiliary cancer	Plasma and pleural effusion (ultracentrifugation)	Next-generation sequencing	A wide range of potential cancer-associated RNA and DNA biomarkers were detected, including copy number profiles, point mutations, insertions, deletions, gene fusions, and mutational signatures[209]
CRNDE-h (a long non-coding RNA)	CRC	80 volunteers with normal colonoscopy, 80 patients with hyperplastic polyps, 80 patients with irritable bowel disease, 80 patients with adenoma, and 148 patients with CRC	Serum (ExoQuick)	RT-PCR	CRNDE-h distinguished patients with CRC from patients with benign disease and healthy donors (AUC 0.89) with 70.3% sensitivity and 94.4% specificity, which was superior to carcinoembryonic antigen[243]
Multiple miRNAs	Lung cancer	50 patients with lung adenocarcinoma, 30 patients with lung granuloma, and 25 healthy smokers	Plasma (ExoQuick)	RT-PCR	Multiple miRNAs distinguished patients with lung adenocarcinoma from patients with lung granulomas (AUC 0.76; sensitivity 96% and specificity 60%)[244]
Protein					
GPC1	PDAC	190 patients with PDAC and 100 healthy controls	Serum (ultracentrifugation)	Fluorescence-activated cell sorting	GPC1-positive exosomes distinguished healthy individuals and patients with benign pancreatic disease from patients with early-stage and late-stage PDAC with 100% specificity and 100% sensitivity[162]

Table 2 (cont.) | **Preclinical evaluation of exosomal cargo as cancer biomarkers**

Cargo	Cancer type	Patient cohorts	Exosome source (isolation method)	Assay used	Outcome and/or utility
Protein (cont.)					
MIF	PDAC	12 patients with PDAC with progression to liver metastasis, 10 patients with PDAC without progression, 15 healthy controls, and 18 patients with PDAC and liver metastasis at diagnosis	Plasma (ultracentrifugation)	ELISA	Elevated levels of circulating exosomal MIF predicted development of liver metastasis in patients with stage I PDAC ($P<0.01$)[120]
Lipid					
Phosphatidylserine 18:1/18:1, lactosylceramide (d18:1/16:0), and phosphatidylserine 18:0-18:2	Prostate cancer	15 patients with prostate cancer and 13 healthy controls	Urine (ultracentrifugation)	Mass spectrometry quantitative lipidomics	Combinations of multiple lipid species distinguished patients with prostate cancer from healthy controls (AUC 0.989; sensitivity 93% and specificity 100%)[245]

AR-V7, androgen receptor splice variant 7; AUC, area under the curve; CRC, colorectal cancer; CRNDE-h; colorectal neoplasia differentially expressed variant h; CRPC, castration-resistant prostate cancer; ctDNA, circulating tumour DNA; ddPCR, digital droplet PCR; ELISA, enzyme-linked immunosorbent assay; GPC1, glypican 1; MIF, macrophage migration inhibitory factor; miRNA, microRNA; NSCLC, non-small-cell lung cancer; PDAC, pancreatic ductal adenocarcinoma; PSA, prostate-specific antigen; RT-PCR, reverse transcription PCR.

cancer[163]. Interestingly, although both groups assayed exosomal GPC1 using proteomics platforms, the first study[162] involved an antibody-based assay system, whereas the second study[163] used quantitative mass spectrometry-based methodologies. The latter mass spectrometry method enabled quantitative analysis of exosomal GPC1 levels through liquid chromatography–tandem mass spectrometry (LC–MS/MS) with multiple reaction monitoring (MRM) measurement of the levels of the peptide QIYGAK, which is unique to GPC1 and is not shared with GPC2–GPC6 or with other human proteins. Conceivably, discrepancies in the findings from these important biomarker studies[162,163] could be attributable to differences between the patient cohorts, sample sizes, or reagent specificity; however, given the urgent need for biomarkers for pancreatic cancer surveillance, these studies warrant further research using uniform methodologies. The latter study[163] also highlights the utility of MRM, also referred to as selective reaction monitoring, as a novel LC–MS/MS-based method that is rapid, sensitive, and robust, enabling the efficient high-throughput analysis of clinical samples[169,170].

Perspectives on EV-targeted therapies

As carriers of functional RNA species and oncogenic proteins, EVs have attracted much attention as potential vehicles for the delivery of therapeutic agents, including miRNAs, mRNAs, proteins, peptides, and synthetic drugs[171–175]. Therapeutically active cargo molecules can be loaded into EVs using either passive or active means. The most common in vitro methods use either passive loading through the physical mixing of drugs (for example, curcumin[176], acridine orange[177], doxorubicin[178], or paclitaxel[179]) with isolated EVs or active loading by techniques such as electroporation (for example, with oncogenic *KRAS*^{G12D}-specific small interfering RNA[180]). The direct treatment of EV-secreting donor cells with a therapeutic drug before EV isolation[181] has also been used to generate drug-loaded EVs. Another approach relies on the genetic engineering of EV-producing cells to overexpress proteins (for example, TNF-related apoptosis-inducing ligand (TRAIL; also known as TNFSF10))[182], miRNAs (for example, miR-122 from an expression plasmid[183]), or mRNA[175] cargo molecules of therapeutic interest in order to promote their enrichment in EVs.

EVs can also be obtained from genetically engineered cells to achieve targeted delivery by introducing recipient cell surface recognition molecules[184–186]. An elegant ex vivo study demonstrating the utility of EV-directed drug delivery[187] used multiple superparamagnetic nanoparticles (SPMNs) with conjugated transferrin to first isolate exosomes expressing the transferrin receptor from mouse blood through magnetic separation. These exosomes were passively loaded with doxorubicin before the drug-containing exosome–SPMN conjugates were guided to the tumour using an external magnetic field, thereby suppressing tumour growth in a mouse model of hepatoma[187]. In addition, efficiency of drug delivery using EVs from non-human sources, such as bovine milk[188] and plants (for example, edible exosome-like nanoparticles[189,190]), is currently being evaluated. For example, a phase I clinical trial[191] is underway to assess the use of plant-derived EVs[174] for oral administration of curcumin-loaded EVs to colon tumours.

EVs have many advantages as drug delivery vehicles owing to their biocompatibility, low immunogenicity, and innate ability to interact with target cells, although a number of limitations and challenges associated with this approach require further study in order to bring EV-based therapeutics to the clinic. These include identification of the optimal EV donor cell type, preservation of EV structural integrity during drug loading and large-scale manufacture, storage, and formulation for maintenance of EV potency. EV-based therapeutics have been extensively reviewed

Multiple reaction monitoring (MRM). A targeted mass spectrometry-based proteomics approach for the detection and precise quantification of a predetermined set of proteins or peptides; can also occur via selected reaction monitoring.

Box 3 | **Extracellular vesicle-based vaccines: an update**

Dendritic cell exosomes as EV vaccines

Preclinical study. Dendritic cells (DCs) can secrete exosomes (Dex) that carry functional major histocompatibility complex (MHC) molecules loaded with antigenic peptides and co-stimulatory molecules (such as CD86); Dex induced a better antitumour immune response in mouse tumour models (60% of mice had complete tumour regression at day 60) than that elicited by tumour-peptide-loaded DCs (only 20% of mice were tumour-free at day 60)[161].

Phase I trials: safety and feasibility. Phase I trials using autologous Dex loaded with tumour antigens (first-generation Dex) demonstrated their safety with no serious adverse effects in patients with metastatic melanoma[67] or advanced-stage non-small-cell lung cancer (NSCLC)[68]. First-generation Dex cannot induce a $CD8^+$ T cell response, although they carry natural killer (NK) cell receptor D (NKG2D; also known as KLRK1) ligands to promote NKG2D-dependent NK cell activation[217].

Phase II trial: second-generation Dex. A second-generation autologous Dex with improved immunostimulatory properties was developed for potential peptide-dependent activation of $CD8^+$ T cells[218]. A phase II clinical trial tested the efficacy of second-generation Dex vaccination in patients with unresectable NSCLC ($n = 22$). The median time to progression was 2.2 months, and the median overall survival was 15.0 months. Seven patients (32%) experienced disease stabilization for >4 months. The primary end point (progression-free survival) was not met[219].

Tumour cell-derived exosomes as EV vaccines

Human melanoma-derived exosomes enriched with heat shock 70 kDa protein 1A (HSPA1A) and full-length tumour antigens stimulated DC-dependent activation of tumour-specific $CD8^+$ T cells in vitro; furthermore, mouse tumour cell-derived exosomes induced $CD8^+$ T cell cross-priming and tumour rejection in this preclinical study[220].

Ascites-derived exosomes as EV vaccines

Preclinical study. Exosomes harvested from the ascites (Aex) of melanoma contained the melanoma-associated antigen recognized by T cells (MART-1) tumour antigen, which enabled monocyte-derived DCs (MoDCs) to induce MART-1-specific, HLA-A2-restricted $CD8^+$ T cell responses ex vivo. Furthermore, lymphocytes from seven of nine patients with melanoma expanded into tumour-specific cytotoxic T lymphocytes upon co-culturing with Aex-loaded MoDCs ex vivo[221].

Phase I trials: safety and feasibility. In phase I clinical trials, patients with colorectal cancer received a total of four weekly subcutaneous immunizations of Aex and Aex plus granulocyte–macrophage colony-stimulating factor (GM-CSF). Both therapies were safe and well tolerated; however, only Aex plus GM-CSF, and not Aex alone, induced a tumour-specific cytotoxic T lymphocyte response[222].

Mode of exosome administration

DC-derived exosome-based vaccines have been administered subcutaneously over weekly intervals[67]. Alternatively, autologous tumour cells encapsulated in a biodiffusion chamber and re-implanted in the abdomen of a patient could serve as a slow-release exosome depot to activate an antitumour immune response[223].

elsewhere[171,174,192]. For a more detailed listing of EV-associated drug delivery vehicles and clinical trials involving EVs, see TABLE 3.

Conclusions

With the emergence of new concepts relating to the involvement of EVs in many biological processes, the field of EV research is generating much excitement. Of particular interest for the present discussion is the bidirectional transfer of molecules between tumour cells and the microenvironment, including the role of EVs in establishment of the pre-metastatic niche. Clearly, specific bioactive molecules contained in circulating EVs have great promise as robust surrogates of tumours, thus presenting a new paradigm of using EVs as diagnostic, prognostic, and predictive biomarkers. The increased focus on cancer-derived EVs is, however, accompanied with new challenges, including the standardization of methods for the isolation, quantification, and analysis of EVs from complex tissues, such as blood. Great advances have been made in our ability to accurately determine EV particle numbers[193], notwithstanding the EV size diminution that occurs upon storage at −80°C or freeze–thawing (especially for sMVs), although a pressing need to standardize blood-based EV enumeration procedures remains. This aspect is critical for interlaboratory comparisons of clinical data and for the determination of EV dosages for clinical trials. Another largely unexplored question in the field is how EVs and their contents should best be quantified, with options including vesicle number, protein content, a ratio of the two[193], or classical mAb microarray-based surface profiling. Thus, studies are urgently needed to map the topography of tumour-derived EVs and develop stereospecific mAbs directed against the EV surface[194,195] — particularly those that can enable tumour staging.

A further challenge in the clinical application of EV technologies is that, even for advanced-stage cancers, the percentage of total blood vesicles[196] that are tumour-derived EVs (1%, 0.1%, 0.01%, and so on[197]) is completely unknown. One possible approach to circumventing this issue, in certain settings, is to quantify tumour-derived exosomal mutated DNA, mRNA, RNA species, or fusion gene transcripts — along the lines demonstrated for quantification of circulating tumour DNA[144]; however, the feasibility of applying this strategy to the quantification of tumour-derived exosomes is unclear. Thus, whether any given blood sample will contain levels of tumour-derived EVs within the range of our current technological capabilities for EV detection remains to be determined. Furthermore, our understanding of the constituent and functional differences between EV classes (exosomes and sMVs) and their half-life in biological samples is limited. The answers to these questions will dictate the degree of enrichment required to realistically use EV-containing biomarker candidates as cancer diagnostics. Unlike protein biomarkers, however, RNA markers can be readily amplified and, coupled with high-throughput multiplexed RNA profiling, can offer promising opportunities for cancer diagnostics[198]. A number of rare RNA species[93] and DNA fragments containing cancer driver mutations[199] can be uniquely detected in EVs, offering an exquisite level of biomarker specificity (BOX 2). Importantly, given the likelihood of contamination of blood-derived EV samples with platelet-derived microparticles (historically referred to as platelet dust), any cancer-based EV candidate biomarker or therapeutic target must be evaluated for its specific presence in tumour-derived EVs versus platelet-derived material (BOX 4).

Beyond these practical considerations, current research into the aforementioned roles of EVs in the TME and the pre-metastatic niche will undoubtedly alter our view of cancer biology and present new

Table 3 | **EVs as drug delivery agents for cancer therapy**

Therapeutic cargo	Cancer type	EV source	Recipient cells (mode of EV administration in vivo)	Drug loading techniques	Key findings
Small molecules					
Paclitaxel	Prostate cancer	LNCaP and PC3 prostate cancer cells	LNCaP and PC3 prostate cancer cells	Incubation	Paclitaxel-loaded EVs increased the cytotoxicity of the drug against autologous prostate cancer cells in vitro[246]
	Lung cancer	Mouse RAW 264.7 macrophages	Mouse Lewis 3LL-M27 lung carcinoma cells, wild-type or drug-resistant MDCK cells in vitro, and pulmonary metastases in C57Bl/6 mice (intranasally)	Incubation, electroporation, and sonication	Paclitaxel-loaded EVs increased cytotoxicity against mouse lung carcinoma cells (3LL-M27), wild-type cells, and drug-resistant MDCK cells in vitro and reduced tumour size in vivo[179]
	Lung cancer	Mouse RAW 264.7 macrophages in vitro; primary bone marrow-derived macrophages isolated from C57Bl/6 mice (ex vivo)	Mouse 3LL-M27 lung cancer cells in vitro; mouse lung cancer cell-bearing mice (intravenously)	Engineered hybrid exosomes containing surface AA were generated by fusion with PEG (AA–PEG binds to sigma receptor); hybrid exosomes were loaded with paclitaxel by sonication	Drug-loaded AA-containing exosomes (compared with paclitaxel-loaded non-AA-containing exosomes) were taken up by lung cancer recipient cells in vitro with greater efficiency and showed stronger suppression of metastatic tumour growth in vivo[247]
	Pancreatic cancer	Murine SR4987 MSCs	Human CFPAC-1 pancreatic cancer cell line	Donor cells incubated with paclitaxel	Paclitaxel-loaded EVs inhibited cancer cell proliferation in vitro[181]
Doxorubicin	Lung cancer	Human H1299 and YRC9 lung cancer cells	Human H1299 and A549 lung cancer cells, MRC9 primary lung fibroblasts, and HCASM smooth muscle cells	Gold nanoparticles conjugated with pH-linker-doxorubicin (nanodox) incubated with lung cancer cell-derived exosomes to introduce doxorubicin within the exosome (nanosome)	Nanosomes had greater cytotoxicity than nanodox and selectively reduced growth of cancer cells compared with non-cancerous cells in vitro[178]
	Liver cancer	Mouse serum SMNC-labelled exosomes	Mouse H22 hepatoma cell line in vitro and H22 tumour-bearing mice (intravenously)	Incubation	Doxorubicin-loaded SMNCs exhibited cytotoxicity against H22 cells in vitro and suppressed tumour growth in mice by enabling direction of doxorubicin to the tumour site specifically using an externally positioned magnet[187]
	Breast and ovarian cancers	Human MDA-MB-231 breast cancer cell line and mouse STOSE ovarian cancer cell line	MDA-MB-231 and STOSE cells in vitro and MDA-MB-231 and STOSE cells injected into mice (intraperitoneally)	Electroporation	Doxorubicin-loaded exosomes had cytotoxicity against cancer cells in vitro and reduced tumour volume as well as cardiotoxicity compared with free doxorubicin in vivo[248]
Doxorubicin	Colon cancer	Human U937 lymphoma cells in vitro; mouse RAW 264.7 macrophage -derived extrusion-generated exosome mimetic vesicles in vitro	HUVECs in vitro and mice injected with CT26 cells (intravenously)	Incubation	Adhesion molecules on endothelial cell surfaces were important for targeted delivery of doxorubicin-loaded nanovesicles (drug-loaded exosome mimetic vesicles) in vitro; doxorubicin-containing nanovesicles induced endothelial cell death and suppressed tumour growth in vivo[249]
	Breast cancer	imDC line expressing α_v integrin-specific RGD peptide (CRGDKGPDC)–LAMP2 fusion protein-containing exosomes	α_v integrin-positive human MDA-MB-231 and MCF-7 breast cancer cell lines and B16F10 and HepG2 cancer cell lines in vitro; MDA-MB-231-bearing BALB/c nude mice (intravenously)	Electroporation	Doxorubicin-loaded exosomes expressing LAMP2 on their surface decreased cancer cell proliferation in vitro and caused greater inhibition of tumour growth than free doxorubicin and LAMP2$^-$ exosomes by specifically targeting α_v integrin-positive human breast cancer cell lines[185]

Table 3 (cont.) | **EVs as drug delivery agents for cancer therapy**

Therapeutic cargo	Cancer type	EV source	Recipient cells (mode of EV administration in vivo)	Drug loading techniques	Key findings
Small molecules (cont.)					
Paclitaxel and doxorubicin	Brain cancer	bEND.3 brain endothelial cells	Human brain neuronal glioblastoma–astrocytoma U-87MG xenograft in zebrafish (intravenously)	Incubation	Exosomes delivered anticancer drugs across the blood–brain barrier to xenotransplanted brain cancer cells and significantly reduced expression levels of VEGF RNAs compared with free drugs in vivo[250]
Doxorubicin and anti-inflammatory agents (for example, curcumin)	Breast and colon cancers	Grapefruit and mouse EL4 T cell lymphoma cells in vitro; peripheral blood leukocytes (mouse and human) ex vivo	Mouse 4T1 and 4TO7 breast cancer cell lines, mouse NMuMG mammary gland epithelial cells, CT26 cells, and HUVECs in vitro; mice injected with CT26 cells and 4T1 cells (intravenously)	Sonication and fusion to GNVs coated with inflammatory-related receptor-enriched membranes of activated leukocytes (IGNVs)	Efficient delivery of doxorubicin-loaded IGNVs to inflammatory sites and reduced breast and colon tumour growth in vivo[190]
Withaferin A and paclitaxel	Lung and breast cancers	Bovine milk	Nonmalignant Beas-2B bronchial epithelial cells, MDA-MB-231 and T47D breast cancer cells, and A549 and H1299 lung cancer cells in vitro; A549 cell-injected mice (intraperioneally or by oral gavage)	Incubation	Withaferin A-containing and paclitaxel-containing exosomes had greater cytotoxicity on cancer cells in vitro and caused greater suppression of tumour growth than free drugs in vivo[188]
Curcumin	Colon cancer	Curcumin-loaded plant exosome-like nanovesicles	NA	NA	This is a phase I clinical trial investigating the ability of plant exosomes to deliver curcumin as a dietary supplement[191]
Methotrexate	Lung cancer	MTX-ATMPs	NA	NA	This is a phase II trial testing MTX-ATMPs in the treatment of malignant pleural effusion[251]
	Gastric, colorectal, or ovarian cancer	Autologous erythrocyte-derived microparticles containing methotrexate	NA	NA	This is a clinical study of peritoneal perfusion of erythrocyte-derived microparticles containing methotrexate and systemic therapy combination in the treatment of malignant ascites[252]
Cisplatin	Lung cancer	Chemotherapeutic drug-loaded tumour cells	Treatment-resistant murine H22 hepatocarcinoma cells, human MCF-7 breast cancer, and human A549 lung cancer cells in vitro and patients with lung cancer (by intrathoracic injection)	Incubation	Tumour cell-derived microparticles containing cisplatin reversed drug resistance and induced cell death in drug-resistant tumour-repopulating cells derived from patients with lung cancer[253]
AO	Melanoma	Human blood-derived macrophages ex vivo	Me30966 melanoma cell line in vitro	Incubation	Macrophage-derived exosomes that contain AO induced greater melanoma cell death than free AO in vitro[177]
siRNAs					
Mutant $KRAS^{G12D}$ siRNA or shRNA plasmid	Pancreatic cancer	Human foreskin fibroblasts (BJ), *CD47*-knockout mouse ear fibroblasts, and mesenchymal cells isolated from mice (ex vivo)	MIA PaCa-2, PANC-1, BxPC-3, and T3M4 pancreatic cancer cells in vitro; PANC-1 or KPC689 cells injected into mice; KPC ($Pdx1^{Cre/+}$; $LSL\text{-}Kras^{G12D/+}$; $LSL\text{-}Trp53^{R172H/+}$) mice; and KTC ($Ptf1a^{cre/+}$; $LSL\text{-}Kras^{G12D/+}$; $Tgfbr2^{lox/lox}$) mice (intraperitoneally)	siRNA or shRNA plasmid electroporation	CD47-expressing exosomes escape phagocytotic clearance in vivo; exosomes containing CD47 and mutant *KRAS* siRNA reduced tumour size in mouse pancreatic cancer models compared with exosomes lacking CD47 or exosomes electroporated with mutant *KRAS* siRNA or shRNA[180]

Table 3 (cont.) | **EVs as drug delivery agents for cancer therapy**

Therapeutic cargo	Cancer type	EV source	Recipient cells (mode of EV administration in vivo)	Drug loading techniques	Key findings
siRNAs (cont.)					
VEGF siRNA	Brain cancer	bEND.3 brain endothelial cells	Human brain neuronal glioblastoma–astrocytoma U-87MG xenograft in zebrafish (intravenously)	Lipofectamine-mediated delivery	Exosomes delivered *VEGF* siRNA across the blood–brain barrier to xenotransplanted brain cancer cells to reduce tumour burden in vivo[254]
miRNAs					
miR-122	Liver cancer	Human AMSCs	Human HepG2 cell liver cancer cell line in vitro; HepG2-injected mice in vivo (intratumourally)	Donor cell transfection	miR-122-containing exosomes induced liver cancer cell death in vitro and reduced tumour growth in mice compared with non-miR-122- containing exosomes in vivo[183]
let-7	Breast cancer	Human HCC70, HCC1954, and MCF-7 breast cancer cell lines expressing GE11 (a peptide that targets EGFR)	HCC70 cells in vitro; EGFR-expressing breast cancer xenograft tissue in *Rag2*$^{-/-}$ mice (intravenously)	Donor cell transfection	GE11-containing exosomes specifically targeted EGFR-expressing cancerous tissues and reduced tumour growth compared with non-let-7-containing and non-GE11-containing exosomes in vivo[255]
Anti-miRNAs					
anti-miR-9	Glioblastoma	Human MSCs	Human U87 and T98G glioblastoma cell lines and patient-derived BT145 and BT164 glioma cell lines	Donor cell transfection	Anti-miR-9 from exosomes reversed the resistance of glioblastoma cells to temozolomide compared with non-anti-miR-9-containing exosomes in vitro[256]
mRNAs					
Cytosine deaminase fused to uracil phosphoribosyl transferase (CD–UPRT)	Schwannoma	HEK293T cells	HEI-193 cells in a Schwannoma orthotopic mouse model (intratumourally)	Donor cell transfection	mRNA of CD–UPRT-containing exosomes induced Schwannoma cell death in vitro and reduced tumour growth in mice by increasing 5-fluorocytosine sensitivity compared with non-CD–UPRT-containing exosomes in vivo[175]
Proteins					
T34A-mutated survivin (dominant-negative mutant)	Pancreatic cancer	YUSAC-2 melanoma cell line	MIA PaCa-2 pancreatic cancer cells	Induction of a tetracycline-regulated survivin-T34A expression in melanoma cells, after which survivin-T34A-containing exosomes were collected	Survivin-T34A-containing exosomes induced pancreatic cell death by increasing gemcitabine sensitivity compared with non-survivin-T34A-containing exosomes in vitro[257]
hMUC1	Colon cancer	hMUC1-expressing mouse CT26 and TA3HA cell lines	BMDCs from BALB/c mice in vitro; MUC1-expressing mouse cell lines in BALB/c nude mice (intradermally)	Donor cell transfection	hMUC1-containing exosomes activated splenocytes by promoting IFNγ release in vitro, as well as by suppressing tumour growth compared with non-hMUC1-containing exosomes in vivo[258]
TRAIL	Myeloma	Human K562 chronic myelogenous leukaemia cells	SUDHL, INT12, and KMS11 multiple myeloma cell lines in vitro; mice injected with SUDHL and INT12 cells (intravenously)	Donor cell transfection	TRAIL-containing exosomes were more effective in inducing myeloma apoptosis in vitro and in reducing tumour growth in vivo than free TRAIL[182]

AA, aminoethyl anisamide; AMSC, adipose mesenchymal stem cell; AO, acridine orange; BMDC, bone marrow-derived cell; EV, extracellular vesicle; GNV, grapefruit-derived nanovesicles; hMUC1, human mucin 1; HUVEC, human umbilical vein endothelial cell; imDC, immature mouse dendritic cell; LAMP2, lysosome-associated membrane glycoprotein; miRNA, microRNA; MSC, mesenchymal stem cell; MTX-ATMP, methotrexate-loaded autologous tumour-derived microparticle; NA, not applicable; PEG, polyethylene glycol; RGD, Arg–Gly–Asp; shRNA, short hairpin RNA; siRNA, small interfering RNA; SMNC, superparamagnetic nanoparticle; TRAIL, TNF-related apoptosis-inducing ligand.

Box 4 | **Blood-based EV enumeration in cancer detection**

An increased number of extracellular vesicles (EVs) in human plasma has been suggested as a stand-alone diagnostic marker for cancer. Several reports have shown substantial differences in exosome counts in plasma samples from patients with cancer versus individuals without this disease as determined by either sandwich enzyme-linked immunosorbent assay (ELISA) using anti-CD63 and anti-caveolin 1 antibodies or nanoparticle-tracking analysis[224]. By contrast, other reports showed no statistically significant difference in circulating EV or microparticle levels in plasma or blood from individuals without cancer versus patients with any one of several cancer types[30,225]. Such conflicting results might be due at least in part to differences in sample collection or vesicle enumeration methodologies. Indeed, it has been suggested that the ratio of particle number to protein concentration should be used to address interlaboratory variation in EV enumeration results[193]. A further critical issue for the use of EV enumeration in the clinical setting is the level of platelet products found in human blood and plasma. Platelet-derived EVs constitute the greatest percentage of EVs in blood by far[31], and levels of these platelet particles vary depending on age, stress, and infection status, as well as many other disease states. Also relevant to the use of immunological methods for EV enumeration on the basis of CD63 is the selective presence of CD63 in platelets[12], including both platelet-derived exosomes and platelet-derived microparticles[226,227]. Furthermore, lipoprotein particles, protein complexes, and other microparticles of similar size might, arguably, outnumber EVs in blood[6,193,228–230]. Thus, whether plasma exosome levels alone can clearly distinguish cancer from inflammation-related comorbidities, let alone distinguish one cancer type from another, remains unclear. To address such problems, a shift towards the use of highly specific EV constituent biomarkers has occurred (BOX 2) — as opposed to the sole use of plasma exosome levels for cancer detection[224].

targets for therapeutic intervention. Looking forward, perhaps a future paradigm in EV biology lies in the notion that blocking cancer cell-derived EV release and/or uptake by recipient cells, in conjunction with adjuvant therapies targeting extant cancer, might add another weapon to the armoury for cancer treatment.

Published online 23 May 2018

1. Ahmed, K. A. & Xiang, J. Mechanisms of cellular communication through intercellular protein transfer. *J. Cell. Mol. Med.* **15**, 1458–1473 (2011).
2. Pitt, J. M., Kroemer, G. & Zitvogel, L. Extracellular vesicles: masters of intercellular communication and potential clinical interventions. *J. Clin. Invest.* **126**, 1139–1143 (2016).
3. Peinado, H., Lavotshkin, S. & Lyden, D. The secreted factors responsible for pre-metastatic niche formation: old sayings and new thoughts. *Semin. Cancer Biol.* **21**, 139–146 (2011).
4. Xu, R., Greening, D. W., Zhu, H. J., Takahashi, N. & Simpson, R. J. Extracellular vesicle isolation and characterization: toward clinical application. *J. Clin. Invest.* **126**, 1152–1162 (2016).
5. Colombo, M., Raposo, G. & Thery, C. Biogenesis, secretion, and intercellular interactions of exosomes and other extracellular vesicles. *Annu. Rev. Cell Dev. Biol.* **30**, 255–289 (2014).
6. Witwer, K. W. et al. Standardization of sample collection, isolation and analysis methods in extracellular vesicle research. *J. Extracell. Vesicles* **2**, 20360 (2013).
7. Parolini, I. et al. Microenvironmental pH is a key factor for exosome traffic in tumor cells. *J. Biol. Chem.* **284**, 34211–34222 (2009).
8. Mittelbrunn, M. et al. Unidirectional transfer of microRNA-loaded exosomes from T cells to antigen-presenting cells. *Nat. Commun.* **2**, 282 (2011).
9. Kucharzewska, P. & Belting, M. Emerging roles of extracellular vesicles in the adaptive response of tumour cells to microenvironmental stress. *J. Extracell. Vesicles* **2**, 20304 (2013).
10. An, Q., van Bel, A. J. & Huckelhoven, R. Do plant cells secrete exosomes derived from multivesicular bodies? *Plant Signal. Behav.* **2**, 4–7 (2007).
11. Deatherage, B. L. & Cookson, B. T. Membrane vesicle release in bacteria, eukaryotes, and archaea: a conserved yet underappreciated aspect of microbial life. *Infect. Immun.* **80**, 1948–1957 (2012).
12. Heijnen, H. F., Schiel, A. E., Fijnheer, R., Geuze, H. J. & Sixma, J. J. Activated platelets release two types of membrane vesicles: microvesicles by surface shedding and exosomes derived from exocytosis of multivesicular bodies and alpha-granules. *Blood* **94**, 3791–3799 (1999).
13. Greening, D. W., Gopal, S. K., Xu, R., Simpson, R. J. & Chen, W. Exosomes and their roles in immune regulation and cancer. *Semin. Cell Dev. Biol.* **40**, 72–81 (2015).
14. Robbins, P. D. & Morelli, A. E. Regulation of immune responses by extracellular vesicles. *Nat. Rev. Immunol.* **14**, 195–208 (2014).
15. Nair, R. et al. Extracellular vesicles derived from preosteoblasts influence embryonic stem cell differentiation. *Stem Cells Dev.* **23**, 1625–1635 (2014).
16. Teng, X. et al. Mesenchymal stem cell-derived exosomes improve the microenvironment of infarcted myocardium contributing to angiogenesis and anti-inflammation. *Cell Physiol. Biochem.* **37**, 2415–2424 (2015).
17. Baixauli, F., Lopez-Otin, C. & Mittelbrunn, M. Exosomes and autophagy: coordinated mechanisms for the maintenance of cellular fitness. *Front. Immunol.* **5**, 403 (2014).
18. Greening, D. W., Nguyen, H. P., Elgass, K., Simpson, R. J. & Salamonsen, L. A. Human endometrial exosomes contain hormone-specific cargo modulating trophoblast adhesive capacity: Insights into endometrial-embryo interactions. *Biol. Reprod.* **94**, 38 (2016).
19. Simon, C. et al. Extracellular vesicles in human reproduction in health and disease. *Endocr. Rev.* https://doi.org/10.1210/er.2017-00229 (2018).
20. Mitchell, M. D. et al. Placental exosomes in normal and complicated pregnancy. *Am. J. Obstet. Gynecol.* **213** (Suppl.), S173–S181 (2015).
21. Fruhbeis, C. et al. Neurotransmitter-triggered transfer of exosomes mediates oligodendrocyte-neuron communication. *PLoS Biol.* **11**, e1001604 (2013).
22. Fruhbeis, C., Frohlich, D., Kuo, W. P. & Kramer-Albers, E. M. Extracellular vesicles as mediators of neuron-glia communication. *Front. Cell. Neurosci.* **7**, 182 (2013).
23. Budnik, V., Ruiz-Canada, C. & Wendler, F. Extracellular vesicles round off communication in the nervous system. *Nat. Rev. Neurosci.* **17**, 160–172 (2016).
24. Peinado, H. et al. Pre-metastatic niches: organ-specific homes for metastases. *Nat. Rev. Cancer* **17**, 302–317 (2017).
25. Maas, S. L. N., Breakefield, X. O. & Weaver, A. M. Extracellular vesicles: unique intercellular delivery vehicles. *Trends Cell Biol.* **27**, 172–188 (2017).
26. Gopal, S. K. et al. Extracellular vesicles: their role in cancer biology and epithelial-mesenchymal transition. *Biochem. J.* **474**, 21–45 (2017).
27. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
28. Sadeghipour, S. & Mathias, R. A. Herpesviruses hijack host exosomes for viral pathogenesis. *Semin. Cell Dev. Biol.* **67**, 91–100 (2017).
29. Skog, J. et al. Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nat. Cell Biol.* **10**, 1470–1476 (2008).
30. Peinado, H. et al. Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. *Nat. Med.* **18**, 883–891 (2012).
31. Tao, S. C., Guo, S. C. & Zhang, C. Q. Platelet-derived extracellular vesicles: an emerging therapeutic approach. *Int. J. Biol. Sci.* **13**, 828–834 (2017).
32. Atkin-Smith, G. K. & Poon, I. K. H. Disassembly of the dying: mechanisms and functions. *Trends Cell Biol.* **27**, 151–162 (2017).
33. Xu, R., Greening, D. W., Rai, A., Ji, H. & Simpson, R. J. Highly-purified exosomes and shed microvesicles isolated from the human colon cancer cell line LIM1863 by sequential centrifugal ultrafiltration are biochemically and functionally distinct. *Methods* **87**, 11–21 (2015).
34. Kowal, J. et al. Proteomic comparison defines novel markers to characterize heterogeneous populations of extracellular vesicle subtypes. *Proc. Natl Acad. Sci. USA* **113**, E968–E977 (2016).
35. Tkach, M. et al. Qualitative differences in T cell activation by dendritic cell-derived extracellular vesicle subtypes. *EMBO J.* **36**, 3012–3028 (2017).
36. Tauro, B. J. et al. Two distinct populations of exosomes are released from LIM1863 colon carcinoma cell-derived organoids. *Mol. Cell. Proteom.* **12**, 587–598 (2013).
37. Willms, E. et al. Cells release subpopulations of exosomes with distinct molecular and biological properties. *Sci. Rep.* **6**, 22519 (2016).
38. Zhang, H. et al. Identification of distinct nanoparticles and subsets of extracellular vesicles by asymmetric flow field-flow fractionation. *Nat. Cell Biol.* **20**, 332–343 (2018).
39. Aatonen, M. T. et al. Isolation and characterization of platelet-derived extracellular vesicles. *J. Extracell. Vesicles* **3**, 24692 (2014).
40. Piper, R. C. & Katzmann, D. J. Biogenesis and function of multivesicular bodies. *Annu. Rev. Cell Dev. Biol.* **23**, 519–547 (2007).
41. Appelqvist, H., Waster, P., Kagedal, K. & Ollinger, K. The lysosome: from waste bag to potential therapeutic target. *J. Mol. Cell. Biol.* **5**, 214–226 (2013).
42. Luzio, J. P., Gray, S. R. & Bright, N. A. Endosome-lysosome fusion. *Biochem. Soc. Trans.* **38**, 1413–1416 (2010).
43. Bucci, C., Thomsen, P., Nicoziani, P., McCarthy, J. & van Deurs, B. Rab7: a key to lysosome biogenesis. *Mol. Biol. Cell* **11**, 467–480 (2000).
44. Kowal, J., Tkach, M. & Thery, C. Biogenesis and secretion of exosomes. *Curr. Opin. Cell Biol.* **29**, 116–125 (2014).

45. Bobrie, A., Colombo, M., Raposo, G. & Thery, C. Exosome secretion: molecular mechanisms and roles in immune responses. *Traffic* **12**, 1659–1668 (2011).
46. Katzmann, D. J., Babst, M. & Emr, S. D. Ubiquitin-dependent sorting into the multivesicular body pathway requires the function of a conserved endosomal protein sorting complex. ESCRT-I. *Cell* **106**, 145–155 (2001).
47. Hurley, J. H. ESCRTs are everywhere. *EMBO J.* **34**, 2398–2407 (2015).
48. Henne, W. M., Buchkovich, N. J. & Emr, S. D. The ESCRT pathway. *Dev. Cell* **21**, 77–91 (2011).
49. Colombo, M. et al. Analysis of ESCRT functions in exosome biogenesis, composition and secretion highlights the heterogeneity of extracellular vesicles. *J. Cell Sci.* **126**, 5553–5565 (2013).
50. Buschow, S. I., Liefhebber, J. M., Wubbolts, R. & Stoorvogel, W. Exosomes contain ubiquitinated proteins. *Blood Cells Mol. Dis.* **35**, 398–403 (2005).
51. Clague, M. J., Liu, H. & Urbe, S. Governance of endocytic trafficking and signaling by reversible ubiquitylation. *Dev. Cell* **23**, 457–467 (2012).
52. Hurley, J. H. & Odorizzi, G. Get on the exosome bus with ALIX. *Nat. Cell Biol.* **14**, 654–655 (2012).
53. Trajkovic, K. et al. Ceramide triggers budding of exosome vesicles into multivesicular endosomes. *Science* **319**, 1244–1247 (2008).
54. Ghossoub, R. et al. Syntenin-ALIX exosome biogenesis and budding into multivesicular bodies are controlled by ARF6 and PLD2. *Nat. Commun.* **5**, 3477 (2014).
55. Baietti, M. F. et al. Syndecan-syntenin-ALIX regulates the biogenesis of exosomes. *Nat. Cell Biol.* **14**, 677–685 (2012).
56. Muralidharan-Chari, V. et al. ARF6-regulated shedding of tumor cell-derived plasma membrane microvesicles. *Curr. Biol.* **19**, 1875–1885 (2009).
57. Li, B., Antonyak, M. A., Zhang, J. & Cerione, R. A. RhoA triggers a specific signaling pathway that generates transforming microvesicles in cancer cells. *Oncogene* **31**, 4740–4749 (2012).
58. Nabhan, J. F., Hu, R., Oh, R. S., Cohen, S. N. & Lu, Q. Formation and release of arrestin domain-containing protein 1-mediated microvesicles (ARMMs) at plasma membrane by recruitment of TSG101 protein. *Proc. Natl Acad. Sci. USA* **109**, 4146–4151 (2012).
59. Fujii, K., Hurley, J. H. & Freed, E. O. Beyond Tsg101: the role of Alix in 'ESCRTing' HIV-1. *Nat. Rev. Microbiol.* **5**, 912–916 (2007).
60. Christ, L. et al. ALIX and ESCRT-I/II function as parallel ESCRT-III recruiters in cytokinetic abscission. *J. Cell Biol.* **212**, 499–513 (2016).
61. Morita, E. et al. Human ESCRT and ALIX proteins interact with proteins of the midbody and function in cytokinesis. *EMBO J.* **26**, 4215–4227 (2007).
62. Bianco, F. et al. Acid sphingomyelinase activity triggers microparticle release from glial cells. *EMBO J.* **28**, 1043–1054 (2009).
63. Cocucci, E. & Meldolesi, J. Ectosomes and exosomes: shedding the confusion between extracellular vesicles. *Trends Cell Biol.* **25**, 364–372 (2015).
64. Abels, E. R. & Breakefield, X. O. Introduction to extracellular vesicles: biogenesis, RNA cargo selection, content, release, and uptake. *Cell. Mol. Neurobiol.* **36**, 301–312 (2016).
65. Mulcahy, L. A., Pink, R. C. & Carter, D. R. Routes and mechanisms of extracellular vesicle uptake. *J. Extracell. Vesicles* **3**, 24641 (2014).
66. Tan, A., De La Pena, H. & Seifalian, A. M. The application of exosomes as a nanoscale cancer vaccine. *Int. J. Nanomed.* **5**, 889–900 (2010).
67. Escudier, B. et al. Vaccination of metastatic melanoma patients with autologous dendritic cell (DC) derived-exosomes: results of the first phase I clinical trial. *J. Transl Med.* **3**, 10 (2005).
68. Morse, M. A. et al. A phase I study of dexosome immunotherapy in patients with advanced non-small cell lung cancer. *J. Transl Med.* **3**, 9 (2005).
69. Greening, D. W., Xu, R., Gopal, S. K., Rai, A. & Simpson, R. J. Proteomic insights into extracellular vesicle biology — defining exosomes and shed microvesicles. *Expert Rev. Proteom.* **14**, 69–95 (2017).
70. Nielsen, M., Thomsen, J. L., Primdahl, S., Dyreborg, U. & Andersen, J. A. Breast cancer and atypia among young and middle-aged women: a study of 110 medicolegal autopsies. *Br. J. Cancer* **56**, 814–819 (1987).
71. Harach, H. R., Franssila, K. O. & Wasenius, V. M. Occult papillary carcinoma of the thyroid. A "normal" finding in Finland. A systematic autopsy study. *Cancer* **56**, 531–538 (1985).
72. Manser, R. L., Dodd, M., Byrnes, G., Irving, L. B. & Campbell, D. A. Incidental lung cancers identified at coronial autopsy: implications for overdiagnosis of lung cancer by screening. *Respir. Med.* **99**, 501–507 (2005).
73. Andea, A., Sarkar, F. & Adsay, V. N. Clinicopathological correlates of pancreatic intraepithelial neoplasia: a comparative analysis of 82 cases with and 152 cases without pancreatic ductal adenocarcinoma. *Mod. Pathol.* **16**, 996–1006 (2003).
74. Dolberg, D. S. & Bissell, M. J. Inability of Rous sarcoma virus to cause sarcomas in the avian embryo. *Nature* **309**, 552–556 (1984).
75. Folkman, J. & Kalluri, R. Cancer without disease. *Nature* **427**, 787 (2004).
76. Bissell, M. J. & Hines, W. C. Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. *Nat. Med.* **17**, 320–329 (2011).
77. Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. *Nat. Med.* **19**, 1423–1437 (2013).
78. Barcellos-Hoff, M. H., Lyden, D. & Wang, T. C. The evolution of the cancer niche during multistage carcinogenesis. *Nat. Rev. Cancer* **13**, 511–518 (2013).
79. Hsu, Y. L. et al. Hypoxic lung cancer-secreted exosomal miR-23a increased angiogenesis and vascular permeability by targeting prolyl hydroxylase and tight junction protein ZO-1. *Oncogene* **36**, 4929–4942 (2017).
80. Webber, J., Steadman, R., Mason, M. D., Tabi, Z. & Clayton, A. Cancer exosomes trigger fibroblast to myofibroblast differentiation. *Cancer Res.* **70**, 9621–9630 (2010).
81. Webber, J. P. et al. Differentiation of tumour-promoting stromal myofibroblasts by cancer exosomes. *Oncogene* **34**, 290–302 (2015).
82. Kim, J. W. et al. Fas ligand-positive membranous vesicles isolated from sera of patients with oral cancer induce apoptosis of activated T lymphocytes. *Clin. Cancer Res.* **11**, 1010–1020 (2005).
83. Wieckowski, E. U. et al. Tumor-derived microvesicles promote regulatory T cell expansion and induce apoptosis in tumor-reactive activated CD8+ T lymphocytes. *J. Immunol.* **183**, 3720–3730 (2009).
84. Liu, C. et al. Murine mammary carcinoma exosomes promote tumor growth by suppression of NK cell function. *J. Immunol.* **176**, 1375–1385 (2006).
85. Valadi, H. et al. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat. Cell Biol.* **9**, 654–659 (2007).
86. Zhang, Y. et al. Secreted monocytic miR-150 enhances targeted endothelial cell migration. *Mol. Cell* **39**, 133–144 (2010).
87. Pegtel, D. M. et al. Functional delivery of viral miRNAs via exosomes. *Proc. Natl Acad. Sci. USA* **107**, 6328–6333 (2010).
88. Kosaka, N. et al. Secretory mechanisms and intercellular transfer of microRNAs in living cells. *J. Biol. Chem.* **285**, 17442–17452 (2010).
89. Qu, L. et al. Exosome-transmitted lncARSR promotes sunitinib resistance in renal cancer by acting as a competing endogenous RNA. *Cancer Cell* **29**, 653–668 (2016).
90. Antonyak, M. A. et al. Cancer cell-derived microvesicles induce transformation by transferring tissue transglutaminase and fibronectin to recipient cells. *Proc. Natl Acad. Sci. USA* **108**, 4852–4857 (2011).
91. Lugini, L. et al. Exosomes from human colorectal cancer induce a tumor-like behavior in colonic mesenchymal stromal cells. *Oncotarget* **7**, 50086–50098 (2016).
92. Abd Elmageed, Z. Y. et al. Neoplastic reprogramming of patient-derived adipose stem cells by prostate cancer cell-associated exosomes. *Stem Cells* **32**, 983–997 (2014).
93. Al-Nedawi, K. et al. Intercellular transfer of the oncogenic receptor EGFRvIII by microvesicles derived from tumour cells. *Nat. Cell Biol.* **10**, 619–624 (2008).
94. Melo, S. A. et al. Cancer exosomes perform cell-independent microRNA biogenesis and promote tumorigenesis. *Cancer Cell* **26**, 707–721 (2014).
95. Zhao, H. et al. The key role of extracellular vesicles in the metastatic process. *Biochim. Biophys. Acta* **1869**, 64–77 (2018).
96. Demory Beckler, M. et al. Proteomic analysis of exosomes from mutant KRAS colon cancer cells identifies intercellular transfer of mutant KRAS. *Mol. Cell Proteom.* **12**, 343–355 (2013).
97. Nilsson, R. J. et al. Rearranged EML4-ALK fusion transcripts sequester in circulating blood platelets and enable blood-based crizotinib response monitoring in non-small-cell lung cancer. *Oncotarget* **7**, 1066–1075 (2016).
98. Wei, Y. et al. Exosomal miR-221/222 enhances tamoxifen resistance in recipient ER-positive breast cancer cells. *Breast Cancer Res. Treat.* **147**, 423–431 (2014).
99. Patel, A. P. et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
100. Junttila, M. R. & de Sauvage, F. J. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature* **501**, 346–354 (2013).
101. Luga, V. et al. Exosomes mediate stromal mobilization of autocrine Wnt-PCP signaling in breast cancer cell migration. *Cell* **151**, 1542–1556 (2012).
102. Shimoda, M. et al. Loss of the Timp gene family is sufficient for the acquisition of the CAF-like cell state. *Nat. Cell Biol.* **16**, 889–901 (2014).
103. Zhang, L. et al. Microenvironment-induced PTEN loss by exosomal microRNA primes brain metastasis outgrowth. *Nature* **527**, 100–104 (2015).
104. Nabet, B. Y. et al. Exosome RNA unshielding couples stromal activation to pattern recognition receptor signaling in cancer. *Cell* **170**, 352–366.e13 (2017).
105. Zheng, P. et al. Exosomal transfer of tumor-associated macrophage-derived miR-21 confers cisplatin resistance in gastric cancer cells. *J. Exp. Clin. Cancer Res.* **36**, 53 (2017).
106. Yokoi, A. et al. Malignant extracellular vesicles carrying MMP1 mRNA facilitate peritoneal dissemination in ovarian cancer. *Nat. Commun.* **8**, 14470 (2017).
107. Rahman, M. A. et al. Lung cancer exosomes as drivers of epithelial mesenchymal transition. *Oncotarget* **7**, 54852–54866 (2016).
108. Kalluri, R. The biology and function of exosomes in cancer. *J. Clin. Invest.* **126**, 1208–1215 (2016).
109. Psaila, B. & Lyden, D. The metastatic niche: adapting the foreign soil. *Nat. Rev. Cancer* **9**, 285–293 (2009).
110. Kaplan, R. N. et al. VEGFR1-positive haematopoietic bone marrow progenitors initiate the pre-metastatic niche. *Nature* **438**, 820–827 (2005).
111. Huang, Y. et al. Pulmonary vascular destabilization in the premetastatic phase facilitates lung metastasis. *Cancer Res.* **69**, 7529–7537 (2009).
112. Hiratsuka, S. et al. Primary tumours modulate innate immune signalling to create pre-metastatic vascular hyperpermeability foci. *Nat. Commun.* **4**, 1853 (2013).
113. Gupta, G. P. et al. Mediators of vascular remodelling co-opted for sequential steps in lung metastasis. *Nature* **446**, 765–770 (2007).
114. Liu, Y. et al. Tumor exosomal RNAs promote lung pre-metastatic niche formation by activating alveolar epithelial TLR3 to recruit neutrophils. *Cancer Cell* **30**, 243–256 (2016).
115. Giles, A. J. et al. Activation of hematopoietic stem/progenitor cells promotes immunosuppression within the pre-metastatic niche. *Cancer Res.* **76**, 1335–1347 (2016).
116. Gil-Bernabe, A. M. et al. Recruitment of monocytes/macrophages by tissue factor-mediated coagulation is essential for metastatic cell survival and premetastatic niche establishment in mice. *Blood* **119**, 3164–3175 (2012).
117. Hiratsuka, S. et al. The S100A8-serum amyloid A3-TLR4 paracrine cascade establishes a pre-metastatic phase. *Nat. Cell Biol.* **10**, 1349–1355 (2008).
118. Shojaei, F. et al. G-CSF-initiated myeloid cell mobilization and angiogenesis mediate tumor refractoriness to anti-VEGF therapy in mouse models. *Proc. Natl Acad. Sci. USA* **106**, 6742–6747 (2009).
119. Kaplan, R. N., Rafii, S. & Lyden, D. Preparing the "soil": the premetastatic niche. *Cancer Res.* **66**, 11089–11093 (2006).
120. Costa-Silva, B. et al. Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. *Nat. Cell Biol.* **17**, 816–826 (2015).
121. Hoshino, A. et al. Tumour exosome integrins determine organotropic metastasis. *Nature* **527**, 329–335 (2015).
122. Fong, M. Y. et al. Breast-cancer-secreted miR-122 reprograms glucose metabolism in premetastatic niche to promote metastasis. *Nat. Cell Biol.* **17**, 183–194 (2015).
123. Hood, J. L., San, R. S. & Wickline, S. A. Exosomes released by melanoma cells prepare sentinel lymph nodes for tumor metastasis. *Cancer Res.* **71**, 3792–3801 (2011).

124. Grange, C. et al. Microvesicles released from human renal cancer stem cells stimulate angiogenesis and formation of lung premetastatic niche. *Cancer Res.* **71**, 5346–5356 (2011).
125. Balaj, L. et al. Tumour microvesicles contain retrotransposon elements and amplified oncogene sequences. *Nat. Commun.* **2**, 180 (2011).
126. Carter, C. L., Allen, C. & Henson, D. E. Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. *Cancer* **63**, 181–187 (1989).
127. Pucci, F. et al. SCS macrophages suppress melanoma by restricting tumor-derived vesicle-B cell interactions. *Science* **352**, 242–246 (2016).
128. Otto, B. et al. Molecular changes in pre-metastatic lymph nodes of esophageal cancer patients. *PLoS ONE* **9**, e102552 (2014).
129. Mansfield, A. S. et al. Regional immunity in melanoma: immunosuppressive changes precede nodal metastasis. *Mod. Pathol.* **24**, 487–494 (2011).
130. Matsuura, K. et al. Maturation of dendritic cells and T cell responses in sentinel lymph nodes from patients with breast carcinoma. *Cancer* **106**, 1227–1236 (2006).
131. Sleeman, J. P. The lymph node pre-metastatic niche. *J. Mol. Med.* **93**, 1173–1184 (2015).
132. Jung, T. et al. CD44v6 dependence of premetastatic niche preparation by exosomes. *Neoplasia* **11**, 1093–1105 (2009).
133. Rana, S., Malinowska, K. & Zoller, M. Exosomal tumor microRNA modulates premetastatic organ cells. *Neoplasia* **15**, 281–295 (2013).
134. Kaur, A. et al. sFRP2 in the aged microenvironment drives melanoma metastasis and therapy resistance. *Nature* **532**, 250–254 (2016).
135. Yan, L., Cai, Q. & Xu, Y. The ubiquitin-CXCR4 axis plays an important role in acute lung infection-enhanced lung tumor metastasis. *Clin. Cancer Res.* **19**, 4706–4716 (2013).
136. Bliss, S. A. et al. Mesenchymal stem cell-derived exosomes stimulate cycling quiescence and early breast cancer dormancy in bone marrow. *Cancer Res.* **76**, 5832–5844 (2016).
137. Gao, H. et al. The BMP inhibitor Coco reactivates breast cancer cells at lung metastatic sites. *Cell* **150**, 764–779 (2012).
138. Kobayashi, A. et al. Bone morphogenetic protein 7 in dormancy and metastasis of prostate cancer stem-like cells in bone. *J. Exp. Med.* **208**, 2641–2655 (2011).
139. Ghajar, C. M. et al. The perivascular niche regulates breast tumour dormancy. *Nat. Cell Biol.* **15**, 807–817 (2013).
140. Ono, M. et al. Exosomes from bone marrow mesenchymal stem cells contain a microRNA that promotes dormancy in metastatic breast cancer cells. *Sci. Signal.* **7**, ra63 (2014).
141. Fuzery, A. K., Levin, J., Chan, M. M. & Chan, D. W. Translation of proteomic biomarkers into FDA approved cancer diagnostics: issues and challenges. *Clin. Proteom.* **10**, 13 (2013).
142. Drabovich, A. P., Martinez-Morillo, E. & Diamandis, E. P. Toward an integrated pipeline for protein biomarker development. *Biochim. Biophys. Acta* **1854**, 677–686 (2015).
143. Pepe, M. S. et al. Early-phase studies of biomarkers: what target sensitivity and specificity values might confer clinical utility? *Clin. Chem.* **62**, 737–742 (2016).
144. Siravegna, G., Marsoni, S., Siena, S. & Bardelli, A. Integrating liquid biopsies into the management of cancer. *Nat. Rev. Clin. Oncol.* **14**, 531–548 (2017).
145. Urabe, F., Kosaka, N., Yoshioka, Y., Egawa, S. & Ochiya, T. The small vesicular culprits: the investigation of extracellular vesicles as new targets for cancer treatment. *Clin. Transl Med.* **6**, 45 (2017).
146. Sugimachi, K. et al. Identification of a bona fide microRNA biomarker in serum exosomes that predicts hepatocellular carcinoma recurrence after liver transplantation. *Br. J. Cancer* **112**, 532–538 (2015).
147. Jin, X. et al. Evaluation of tumor-derived exosomal miRNA as potential diagnostic biomarkers for early-stage non-small cell lung cancer using next-generation sequencing. *Clin. Cancer Res.* **23**, 5311–5319 (2017).
148. Huang, S. H., Li, Y., Zhang, J., Rong, J. & Ye, S. Epidermal growth factor receptor-containing exosomes induce tumor-specific regulatory T cells. *Cancer Invest.* **31**, 330–335 (2013).
149. Reclusa, P. et al. Exosomes as diagnostic and predictive biomarkers in lung cancer. *J. Thorac. Dis.* **9**, S1373–S1382 (2017).
150. Liu, Q. et al. Circulating exosomal microRNAs as prognostic biomarkers for non-small-cell lung cancer. *Oncotarget* **8**, 13048–13058 (2017).
151. Nedaeinia, R. et al. Circulating exosomes and exosomal microRNAs as biomarkers in gastrointestinal cancer. *Cancer Gene Ther.* **24**, 48–56 (2017).
152. Ogata-Kawata, H. et al. Circulating exosomal microRNAs as biomarkers of colon cancer. *PLoS ONE* **9**, e92921 (2014).
153. Madhavan, B. et al. Combined evaluation of a panel of protein and miRNA serum-exosome biomarkers for pancreatic cancer diagnosis increases sensitivity and specificity. *Int. J. Cancer* **136**, 2616–2627 (2015).
154. Alegre, E. et al. Study of circulating microRNA-125b levels in serum exosomes in advanced melanoma. *Arch. Pathol. Lab. Med.* **138**, 828–832 (2014).
155. Ragusa, M. et al. miRNA profiling in vitreous humor, vitreal exosomes and serum from uveal melanoma patients: pathological and diagnostic implications. *Cancer Biol. Ther.* **16**, 1387–1396 (2015).
156. Fleming, N. H. et al. Serum-based miRNAs in the prediction and detection of recurrence in melanoma patients. *Cancer* **121**, 51–59 (2015).
157. Logozzi, M. et al. High levels of exosomes expressing CD63 and caveolin-1 in plasma of melanoma patients. *PLoS ONE* **4**, e5219 (2009).
158. Chen, I. H. et al. Phosphoproteins in extracellular vesicles as candidate markers for breast cancer. *Proc. Natl Acad. Sci. USA* **114**, 3175–3180 (2017).
159. Zhou, J. et al. Urinary microRNA-30a-5p is a potential biomarker for ovarian serous adenocarcinoma. *Oncol. Rep.* **33**, 2915–2923 (2015).
160. Valentino, A. et al. Exosomal microRNAs in liquid biopsies: future biomarkers for prostate cancer. *Clin. Transl Oncol.* **19**, 651–657 (2017).
161. Zitvogel, L. et al. Eradication of established murine tumors using a novel cell-free vaccine: dendritic cell-derived exosomes. *Nat. Med.* **4**, 594–600 (1998).
162. Melo, S. A. et al. Glypican-1 identifies cancer exosomes and detects early pancreatic cancer. *Nature* **523**, 177–182 (2015).
163. Lai, X. et al. A microRNA signature in circulating exosomes is superior to exosomal glypican-1 levels for diagnosing pancreatic cancer. *Cancer Lett.* **393**, 86–93 (2017).
164. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2016. *CA Cancer J. Clin.* **66**, 7–30 (2016).
165. Su, G. et al. Glypican-1 is frequently overexpressed in human gliomas and enhances FGF-2 signaling in glioma cells. *Am. J. Pathol.* **168**, 2014–2026 (2006).
166. Matsuda, K. et al. Glypican-1 is overexpressed in human breast cancer and modulates the mitogenic effects of multiple heparin-binding growth factors in breast cancer cells. *Cancer Res.* **61**, 5562–5569 (2001).
167. Li, J. et al. GPC1 exosome and its regulatory miRNAs are specific markers for the detection and target therapy of colorectal cancer. *J. Cell. Mol. Med.* **21**, 838–847 (2017).
168. Kleeff, J. et al. The cell-surface heparan sulfate proteoglycan glypican-1 regulates growth factor action in pancreatic carcinoma cells and is overexpressed in human pancreatic cancer. *J. Clin. Invest.* **102**, 1662–1673 (1998).
169. Picotti, P. & Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* **9**, 555–566 (2012).
170. Wang, Q. et al. Selected reaction monitoring approach for validating peptide biomarkers. *Proc. Natl Acad. Sci. USA* **114**, 13519–13524 (2017).
171. Ha, D., Yang, N. & Nadithe, V. Exosomes as therapeutic drug carriers and delivery vehicles across biological membranes: current perspectives and future challenges. *Acta Pharm. Sin. B* **6**, 287–296 (2016).
172. Barile, L. & Vassalli, G. Exosomes: therapy delivery tools and biomarkers of diseases. *Pharmacol. Ther.* **174**, 63–78 (2017).
173. Wang, J., Zheng, Y. & Zhao, M. Exosome-based cancer therapy: implication for targeting cancer stem cells. *Front. Pharmacol.* **7**, 533 (2016).
174. Luan, X. et al. Engineering exosomes as refined biological nanoplatforms for drug delivery. *Acta Pharmacol. Sin.* **38**, 754–763 (2017).
175. Mizrak, A. et al. Genetically engineered microvesicles carrying suicide mRNA/protein inhibit schwannoma tumor growth. *Mol. Ther.* **21**, 101–108 (2013).
176. Sun, D. et al. A novel nanoparticle drug delivery system: the anti-inflammatory activity of curcumin is enhanced when encapsulated in exosomes. *Mol. Ther.* **18**, 1606–1614 (2010).
177. Iessi, E. et al. Acridine orange/exosomes increase the delivery and the effectiveness of acridine orange in human melanoma cells: a new prototype for theranostics of tumors. *J. Enzyme Inhib Med. Chem.* **32**, 648–657 (2017).
178. Srivastava, A. et al. Nanosomes carrying doxorubicin exhibit potent anticancer activity against human lung cancer cells. *Sci. Rep.* **6**, 38541 (2016).
179. Kim, M. S. et al. Development of exosome-encapsulated paclitaxel to overcome MDR in cancer cells. *Nanomedicine* **12**, 655–664 (2016).
180. Kamerkar, S. et al. Exosomes facilitate therapeutic targeting of oncogenic KRAS in pancreatic cancer. *Nature* **546**, 498–503 (2017).
181. Pascucci, L. et al. Paclitaxel is incorporated by mesenchymal stromal cells and released in exosomes that inhibit in vitro tumor growth: a new approach for drug delivery. *J. Control. Release* **192**, 262–270 (2014).
182. Rivoltini, L. et al. TNF-related apoptosis-inducing ligand (TRAIL)-armed exosomes deliver proapoptotic signals to tumor site. *Clin. Cancer Res.* **22**, 3499–3512 (2016).
183. Lou, G. et al. Exosomes derived from miR-122-modified adipose tissue-derived MSCs increase chemosensitivity of hepatocellular carcinoma. *J. Hematol. Oncol.* **8**, 122 (2015).
184. Alvarez-Erviti, L. et al. Delivery of siRNA to the mouse brain by systemic injection of targeted exosomes. *Nat. Biotechnol.* **29**, 341–345 (2011).
185. Tian, Y. et al. A doxorubicin delivery platform using engineered natural membrane vesicle exosomes for targeted tumor therapy. *Biomaterials* **35**, 2383–2390 (2014).
186. Kooijmans, S. A. et al. Display of GPI-anchored anti-EGFR nanobodies on extracellular vesicles promotes tumour cell targeting. *J. Extracell. Vesicles* **5**, 31053 (2016).
187. Qi, H. et al. Blood exosomes endowed with magnetic and targeting properties for cancer therapy. *ACS Nano* **10**, 3323–3333 (2016).
188. Munagala, R., Aqil, F., Jeyabalan, J. & Gupta, R. C. Bovine milk-derived exosomes for drug delivery. *Cancer Lett.* **371**, 48–61 (2016).
189. Mu, J. et al. Interspecies communication between plant and mouse gut host cells through edible plant derived exosome-like nanoparticles. *Mol. Nutr. Food Res.* **58**, 1561–1573 (2014).
190. Wang, Q. et al. Grapefruit-derived nanovectors use an activated leukocyte trafficking pathway to deliver therapeutic agents to inflammatory tumor sites. *Cancer Res.* **75**, 2520–2529 (2015).
191. US National Library of Medicine. *ClinicalTrials.gov* https://clinicaltrials.gov/ct2/show/NCT01939899 (2016).
192. Lener, T. et al. Applying extracellular vesicles based therapeutics in clinical trials — an ISEV position paper. *J. Extracell. Vesicles* **4**, 30087 (2015).
193. Webber, J. & Clayton, A. How pure are your vesicles? *J. Extracell. Vesicles* **2**, 19861 (2013).
194. Belov, L. et al. Surface profiling of extracellular vesicles from plasma or ascites fluid using DotScan antibody microarrays. *Methods Mol. Biol.* **1619**, 263–301 (2017).
195. Baek, R. & Jorgensen, M. M. Multiplexed phenotyping of small extracellular vesicles using protein microarray (EV array). *Methods Mol. Biol.* **1545**, 117–127 (2017).
196. Arraud, N. et al. Extracellular vesicles from blood plasma: determination of their morphology, size, phenotype and concentration. *J. Thromb. Haemost.* **12**, 614–627 (2014).
197. Liang, K. et al. Nanoplasmonic quantification of tumor-derived extracellular vesicles in plasma microsamples for diagnosis and treatment monitoring. *Nat. Biomed. Eng.* **1**, 0021 (2017).
198. Pritchard, C. C., Cheng, H. H. & Tewari, M. MicroRNA profiling: approaches and considerations. *Nat. Rev. Genet.* **13**, 358–369 (2012).
199. Yang, S. et al. Detection of mutant KRAS and TP53 DNA in circulating exosomes from healthy individuals and patients with pancreatic cancer. *Cancer Biol. Ther.* **18**, 158–165 (2017).
200. Li, P., Kaslan, M., Lee, S. H., Yao, J. & Gao, Z. Progress in exosome isolation techniques. *Theranostics* **7**, 789–804 (2017).
201. Tauro, B. J. et al. Comparison of ultracentrifugation, density gradient separation, and immunoaffinity capture methods for isolating human colon cancer cell line LIM1863-derived exosomes. *Methods* **56**, 293–304 (2012).
202. Ghosh, A. et al. Rapid isolation of extracellular vesicles from cell culture and biological fluids using a synthetic peptide with specific affinity for heat shock proteins. *PLoS ONE* **9**, e110443 (2014).

124. Grange, C. et al. Microvesicles released from human renal cancer stem cells stimulate angiogenesis and formation of lung premetastatic niche. *Cancer Res.* **71**, 5346–5356 (2011).
125. Balaj, L. et al. Tumour microvesicles contain retrotransposon elements and amplified oncogene sequences. *Nat. Commun.* **2**, 180 (2011).
126. Carter, C. L., Allen, C. & Henson, D. E. Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. *Cancer* **63**, 181–187 (1989).
127. Pucci, F. et al. SCS macrophages suppress melanoma by restricting tumor-derived vesicle-B cell interactions. *Science* **352**, 242–246 (2016).
128. Otto, B. et al. Molecular changes in pre-metastatic lymph nodes of esophageal cancer patients. *PLoS ONE* **9**, e102552 (2014).
129. Mansfield, A. S. et al. Regional immunity in melanoma: immunosuppressive changes precede nodal metastasis. *Mod. Pathol.* **24**, 487–494 (2011).
130. Matsuura, K. et al. Maturation of dendritic cells and T cell responses in sentinel lymph nodes from patients with breast carcinoma. *Cancer* **106**, 1227–1236 (2006).
131. Sleeman, J. P. The lymph node pre-metastatic niche. *J. Mol. Med.* **93**, 1173–1184 (2015).
132. Jung, T. et al. CD44v6 dependence of premetastatic niche preparation by exosomes. *Neoplasia* **11**, 1093–1105 (2009).
133. Rana, S., Malinowska, K. & Zoller, M. Exosomal tumor microRNA modulates premetastatic organ cells. *Neoplasia* **15**, 281–295 (2013).
134. Kaur, A. et al. sFRP2 in the aged microenvironment drives melanoma metastasis and therapy resistance. *Nature* **532**, 250–254 (2016).
135. Yan, L., Cai, Q. & Xu, Y. The ubiquitin-CXCR4 axis plays an important role in acute lung infection-enhanced lung tumor metastasis. *Clin. Cancer Res.* **19**, 4706–4716 (2013).
136. Bliss, S. A. et al. Mesenchymal stem cell-derived exosomes stimulate cycling quiescence and early breast cancer dormancy in bone marrow. *Cancer Res.* **76**, 5832–5844 (2016).
137. Gao, H. et al. The BMP inhibitor Coco reactivates breast cancer cells at lung metastatic sites. *Cell* **150**, 764–779 (2012).
138. Kobayashi, A. et al. Bone morphogenetic protein 7 in dormancy and metastasis of prostate cancer stem-like cells in bone. *J. Exp. Med.* **208**, 2641–2655 (2011).
139. Ghajar, C. M. et al. The perivascular niche regulates breast tumour dormancy. *Nat. Cell Biol.* **15**, 807–817 (2013).
140. Ono, M. et al. Exosomes from bone marrow mesenchymal stem cells contain a microRNA that promotes dormancy in metastatic breast cancer cells. *Sci. Signal.* **7**, ra63 (2014).
141. Fuzery, A. K., Levin, J., Chan, M. M. & Chan, D. W. Translation of proteomic biomarkers into FDA approved cancer diagnostics: issues and challenges. *Clin. Proteom.* **10**, 13 (2013).
142. Drabovich, A. P., Martinez-Morillo, E. & Diamandis, E. P. Toward an integrated pipeline for protein biomarker development. *Biochim. Biophys. Acta* **1854**, 677–686 (2015).
143. Pepe, M. S. et al. Early-phase studies of biomarkers: what target sensitivity and specificity values might confer clinical utility? *Clin. Chem.* **62**, 737–742 (2016).
144. Siravegna, G., Marsoni, S., Siena, S. & Bardelli, A. Integrating liquid biopsies into the management of cancer. *Nat. Rev. Clin. Oncol.* **14**, 531–548 (2017).
145. Urabe, F., Kosaka, N., Yoshioka, Y., Egawa, S. & Ochiya, T. The small vesicular culprits: the investigation of extracellular vesicles as new targets for cancer treatment. *Clin. Transl Med.* **6**, 45 (2017).
146. Sugimachi, K. et al. Identification of a bona fide microRNA biomarker in serum exosomes that predicts hepatocellular carcinoma recurrence after liver transplantation. *Br. J. Cancer* **112**, 532–538 (2015).
147. Jin, X. et al. Evaluation of tumor-derived exosomal miRNA as potential diagnostic biomarkers for early-stage non-small cell lung cancer using next-generation sequencing. *Clin. Cancer Res.* **23**, 5311–5319 (2017).
148. Huang, S. H., Li, Y., Zhang, J., Rong, J. & Ye, S. Epidermal growth factor receptor-containing exosomes induce tumor-specific regulatory T cells. *Cancer Invest.* **31**, 330–335 (2013).
149. Reclusa, P. et al. Exosomes as diagnostic and predictive biomarkers in lung cancer. *J. Thorac. Dis.* **9**, S1373–S1382 (2017).
150. Liu, Q. et al. Circulating exosomal microRNAs as prognostic biomarkers for non-small-cell lung cancer. *Oncotarget* **8**, 13048–13058 (2017).
151. Nedaeinia, R. et al. Circulating exosomes and exosomal microRNAs as biomarkers in gastrointestinal cancer. *Cancer Gene Ther.* **24**, 48–56 (2017).
152. Ogata-Kawata, H. et al. Circulating exosomal microRNAs as biomarkers of colon cancer. *PLoS ONE* **9**, e92921 (2014).
153. Madhavan, B. et al. Combined evaluation of a panel of protein and miRNA serum-exosome biomarkers for pancreatic cancer diagnosis increases sensitivity and specificity. *Int. J. Cancer* **136**, 2616–2627 (2015).
154. Alegre, E. et al. Study of circulating microRNA-125b levels in serum exosomes in advanced melanoma. *Arch. Pathol. Lab. Med.* **138**, 828–832 (2014).
155. Ragusa, M. et al. miRNA profiling in vitreous humor, vitreal exosomes and serum from uveal melanoma patients: pathological and diagnostic implications. *Cancer Biol. Ther.* **16**, 1387–1396 (2015).
156. Fleming, N. H. et al. Serum-based miRNAs in the prediction and detection of recurrence in melanoma patients. *Cancer* **121**, 51–59 (2015).
157. Logozzi, M. et al. High levels of exosomes expressing CD63 and caveolin-1 in plasma of melanoma patients. *PLoS ONE* **4**, e5219 (2009).
158. Chen, I. H. et al. Phosphoproteins in extracellular vesicles as candidate markers for breast cancer. *Proc. Natl Acad. Sci. USA* **114**, 3175–3180 (2017).
159. Zhou, J. et al. Urinary microRNA-30a-5p is a potential biomarker for ovarian serous adenocarcinoma. *Oncol. Rep.* **33**, 2915–2923 (2015).
160. Valentino, A. et al. Exosomal microRNAs in liquid biopsies: future biomarkers for prostate cancer. *Clin. Transl Oncol.* **19**, 651–657 (2017).
161. Zitvogel, L. et al. Eradication of established murine tumors using a novel cell-free vaccine: dendritic cell-derived exosomes. *Nat. Med.* **4**, 594–600 (1998).
162. Melo, S. A. et al. Glypican-1 identifies cancer exosomes and detects early pancreatic cancer. *Nature* **523**, 177–182 (2015).
163. Lai, X. et al. A microRNA signature in circulating exosomes is superior to exosomal glypican-1 levels for diagnosing pancreatic cancer. *Cancer Lett.* **393**, 86–93 (2017).
164. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2016. *CA Cancer J. Clin.* **66**, 7–30 (2016).
165. Su, G. et al. Glypican-1 is frequently overexpressed in human gliomas and enhances FGF-2 signaling in glioma cells. *Am. J. Pathol.* **168**, 2014–2026 (2006).
166. Matsuda, K. et al. Glypican-1 is overexpressed in human breast cancer and modulates the mitogenic effects of multiple heparin-binding growth factors in breast cancer cells. *Cancer Res.* **61**, 5562–5569 (2001).
167. Li, J. et al. GPC1 exosome and its regulatory miRNAs are specific markers for the detection and target therapy of colorectal cancer. *J. Cell. Mol. Med.* **21**, 838–847 (2017).
168. Kleeff, J. et al. The cell-surface heparan sulfate proteoglycan glypican-1 regulates growth factor action in pancreatic carcinoma cells and is overexpressed in human pancreatic cancer. *J. Clin. Invest.* **102**, 1662–1673 (1998).
169. Picotti, P. & Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* **9**, 555–566 (2012).
170. Wang, Q. et al. Selected reaction monitoring approach for validating peptide biomarkers. *Proc. Natl Acad. Sci. USA* **114**, 13519–13524 (2017).
171. Ha, D., Yang, N. & Nadithe, V. Exosomes as therapeutic drug carriers and delivery vehicles across biological membranes: current perspectives and future challenges. *Acta Pharm. Sin. B* **6**, 287–296 (2016).
172. Barile, L. & Vassalli, G. Exosomes: therapy delivery tools and biomarkers of diseases. *Pharmacol. Ther.* **174**, 63–78 (2017).
173. Wang, J., Zheng, Y. & Zhao, M. Exosome-based cancer therapy: implication for targeting cancer stem cells. *Front. Pharmacol.* **7**, 533 (2016).
174. Luan, X. et al. Engineering exosomes as refined biological nanoplatforms for drug delivery. *Acta Pharmacol. Sin.* **38**, 754–763 (2017).
175. Mizrak, A. et al. Genetically engineered microvesicles carrying suicide mRNA/protein inhibit schwannoma tumor growth. *Mol. Ther.* **21**, 101–108 (2013).
176. Sun, D. et al. A novel nanoparticle drug delivery system: the anti-inflammatory activity of curcumin is enhanced when encapsulated in exosomes. *Mol. Ther.* **18**, 1606–1614 (2010).
177. Iessi, E. et al. Acridine orange/exosomes increase the delivery and the effectiveness of acridine orange in human melanoma cells: a new prototype for theranostics of tumors. *J. Enzyme Inhib Med. Chem.* **32**, 648–657 (2017).
178. Srivastava, A. et al. Nanosomes carrying doxorubicin exhibit potent anticancer activity against human lung cancer cells. *Sci. Rep.* **6**, 38541 (2016).
179. Kim, M. S. et al. Development of exosome-encapsulated paclitaxel to overcome MDR in cancer cells. *Nanomedicine* **12**, 655–664 (2016).
180. Kamerkar, S. et al. Exosomes facilitate therapeutic targeting of oncogenic KRAS in pancreatic cancer. *Nature* **546**, 498–503 (2017).
181. Pascucci, L. et al. Paclitaxel is incorporated by mesenchymal stromal cells and released in exosomes that inhibit in vitro tumor growth: a new approach for drug delivery. *J. Control. Release* **192**, 262–270 (2014).
182. Rivoltini, L. et al. TNF-related apoptosis-inducing ligand (TRAIL)-armed exosomes deliver proapoptotic signals to tumor site. *Clin. Cancer Res.* **22**, 3499–3512 (2016).
183. Lou, G. et al. Exosomes derived from miR-122-modified adipose tissue-derived MSCs increase chemosensitivity of hepatocellular carcinoma. *J. Hematol. Oncol.* **8**, 122 (2015).
184. Alvarez-Erviti, L. et al. Delivery of siRNA to the mouse brain by systemic injection of targeted exosomes. *Nat. Biotechnol.* **29**, 341–345 (2011).
185. Tian, Y. et al. A doxorubicin delivery platform using engineered natural membrane vesicle exosomes for targeted tumor therapy. *Biomaterials* **35**, 2383–2390 (2014).
186. Kooijmans, S. A. et al. Display of GPI-anchored anti-EGFR nanobodies on extracellular vesicles promotes tumour cell targeting. *J. Extracell. Vesicles* **5**, 31053 (2016).
187. Qi, H. et al. Blood exosomes endowed with magnetic and targeting properties for cancer therapy. *ACS Nano* **10**, 3323–3333 (2016).
188. Munagala, R., Aqil, F., Jeyabalan, J. & Gupta, R. C. Bovine milk-derived exosomes for drug delivery. *Cancer Lett.* **371**, 48–61 (2016).
189. Mu, J. et al. Interspecies communication between plant and mouse gut host cells through edible plant derived exosome-like nanoparticles. *Mol. Nutr. Food Res.* **58**, 1561–1573 (2014).
190. Wang, Q. et al. Grapefruit-derived nanovectors use an activated leukocyte trafficking pathway to deliver therapeutic agents to inflammatory tumor sites. *Cancer Res.* **75**, 2520–2529 (2015).
191. US National Library of Medicine. *ClinicalTrials.gov* https://clinicaltrials.gov/ct2/show/NCT01939899 (2016).
192. Lener, T. et al. Applying extracellular vesicles based therapeutics in clinical trials — an ISEV position paper. *J. Extracell. Vesicles* **4**, 30087 (2015).
193. Webber, J. & Clayton, A. How pure are your vesicles? *J. Extracell. Vesicles* **2**, 19861 (2013).
194. Belov, L. et al. Surface profiling of extracellular vesicles from plasma or ascites fluid using DotScan antibody microarrays. *Methods Mol. Biol.* **1619**, 263–301 (2017).
195. Baek, R. & Jorgensen, M. M. Multiplexed phenotyping of small extracellular vesicles using protein microarray (EV array). *Methods Mol. Biol.* **1545**, 117–127 (2017).
196. Arraud, N. et al. Extracellular vesicles from blood plasma: determination of their morphology, size, phenotype and concentration. *J. Thromb. Haemost.* **12**, 614–627 (2014).
197. Liang, K. et al. Nanoplasmonic quantification of tumor-derived extracellular vesicles in plasma microsamples for diagnosis and treatment monitoring. *Nat. Biomed. Eng.* **1**, 0021 (2017).
198. Pritchard, C. C., Cheng, H. H. & Tewari, M. MicroRNA profiling: approaches and considerations. *Nat. Rev. Genet.* **13**, 358–369 (2012).
199. Yang, S. et al. Detection of mutant KRAS and TP53 DNA in circulating exosomes from healthy individuals and patients with pancreatic cancer. *Cancer Biol. Ther.* **18**, 158–165 (2017).
200. Li, P., Kaslan, M., Lee, S. H., Yao, J. & Gao, Z. Progress in exosome isolation techniques. *Theranostics* **7**, 789–804 (2017).
201. Tauro, B. J. et al. Comparison of ultracentrifugation, density gradient separation, and immunoaffinity capture methods for isolating human colon cancer cell line LIM1863-derived exosomes. *Methods* **56**, 293–304 (2012).
202. Ghosh, A. et al. Rapid isolation of extracellular vesicles from cell culture and biological fluids using a synthetic peptide with specific affinity for heat shock proteins. *PLoS ONE* **9**, e110443 (2014).

Appendix Article 3

Chapter 12

Analysis of Annotated and Unannotated Long Noncoding RNAs from Exosome Subtypes Using Next-Generation RNA Sequencing

Wittaya Suwakulsiri, Maoshan Chen, David W. Greening, Rong Xu, and Richard J. Simpson

Abstract

Long noncoding RNAs (lncRNAs) contain >200 nucleotides and act as regulatory molecules in transcription and translation processes in both normal and pathological conditions. LncRNAs have been reported to localize in nuclei, cytoplasm, and, more recently, extracellular vesicles such as exosomes. Exosomal lncRNAs have gained much attention as exosomes secreted from one cell type can transfer their cargo (e.g., protein, RNA species, and lipids) to recipient cells and mediate phenotypic changes in the recipient cell. In recent years, many exosomal lncRNAs have been discovered and annotated and are attracting much attention as potential markers for disease diagnosis and prognosis. It is expected that many exosomal lncRNAs are yet to be identified. However, characterization of unannotated exosomal RNAs with non–protein-coding sequences from massive RNA sequencing data is technically challenging. Here, we describe a method for the discovery of annotated and unannotated exosomal lncRNA. This method includes a large-scale isolation and purification strategy for exosome subtypes, using the human colorectal cancer cell line (LIM1863) as a model. The method inputs RNA sequencing clean reads and performs transcript assembly to identify annotated and unannotated exosomal lncRNAs. Cutoffs (length, number of exon, classification code, and human protein-coding probability) are used to identify potentially novel exosomal lncRNAs. Raw read count calculation and differential expression analysis are also introduced for downstream analysis and candidate selection. Exosomal lncRNA candidates are validated using RT-qPCR. This method provides a template for exosomal lncRNA discovery and analysis from next-generation RNA sequencing.

Key words Exosomes, Long noncoding RNAs, Transcriptomics, Next-generation RNA sequencing, Bioinformatics

1 Introduction

It is now well recognized that <2% of genomic DNA is transcribed and translated into functional proteins [1], and that genomic DNA transcribes a broad spectrum of RNA species including protein-coding and noncoding RNAs (ncRNAs) [2]. The percentage of transcribed noncoding sequences in whole genome relates to

Haiming Cao (ed.), *Functional Analysis of Long Non-Coding RNAs: Methods and Protocols*, Methods in Molecular Biology, vol. 2254, https://doi.org/10.1007/978-1-0716-1158-6_12,

developmental complexity which is 57% in human, whereas it is 33%, 10%, and 1% in nematode, yeast, and bacteria, respectively [3, 4]. While protein-coding RNAs are further processed for protein synthesis, non–protein-coding RNAs were considered as dysfunctional RNAs [5]. Recently, however, it has become evident that ncRNAs can crucially function in both normal and pathological conditions by acting as regulatory molecules in transcription, RNA processing and translation [6, 7].

Long non-coding RNAs (lncRNAs) are defined as a heterogeneous group of RNA species >200 nucleotides in length and incapable of encoding proteins [6]. The functions of lncRNAs depends on their cell compartmentation, for instance, chromatin modifications, transcriptional control, and posttranscriptional processing occurs in the nuclear compartment while lncRNAs can inhibit protein synthesis in the cytoplasmic compartment [8]. Moreover, some lncRNAs containing small open reading frames (sORFs) such as LINC00948 [9], LINC00116 [10] and Six1 [11], have been shown to encode functional micropeptides (typically, ~50 amino acids in length). Interestingly, lncRNAs have been detected in extracellular vesicles (EVs) isolated from biofluids such as plasma [12], serum [13], breast milk [14], synovial fluid [15], and urine [16].

EVs are a heterogeneous population of lipid bilayer-membrane vesicles derived from diverse cell types [17–19]. Based upon their mechanism of biogenesis, EVs comprise two major classes—exosomes (of endosomal origin) and shed microvesicles (also referred to as microparticles and ectomeres), which originate by blebbing of the plasma membrane [19]. EVs are crucial mediators of intercellular communication and act by transferring their bioactive cargo (e.g., proteins, RNA species, lipids, and metabolites) to recipient cells [18]. Interestingly, the majority of exosomal RNAs map to intronic regions resulting in the abundance of long intergenic noncoding RNAs (lincRNAs) when compared to parental cells [20, 21]. Exosomal lncRNAs have been implicated in the onset of cancer drug resistance (e.g., lncARSR [22], linc-ROR [23], linc-VLDLR [24], and lncUCA1 [25]) and angiogenesis (e.g., lncPOU3F3 [26] and lncCCAT2 [27]).

Here, we report a method used to discover annotated and unannotated exosomal lncRNAs from next-generation RNA sequencing data, focusing on human exosomal lncRNAs (Fig. 1). The method starts from large-scale generation of exosomes secreted from the human colorectal cancer cell line, LIM1863 [28] continuously cultured in Bioreactor classic flasks [21, 29]. To isolate exosomes, we next performed differential centrifugation on the cell culture medium (CM) to remove cells, cell debris, and larger EVs known as shed microvesicles [30] followed by sequential glycoprotein A33 antigen (A33) [31] and EpCAM-immunoaffinity capture exosome subtype purification

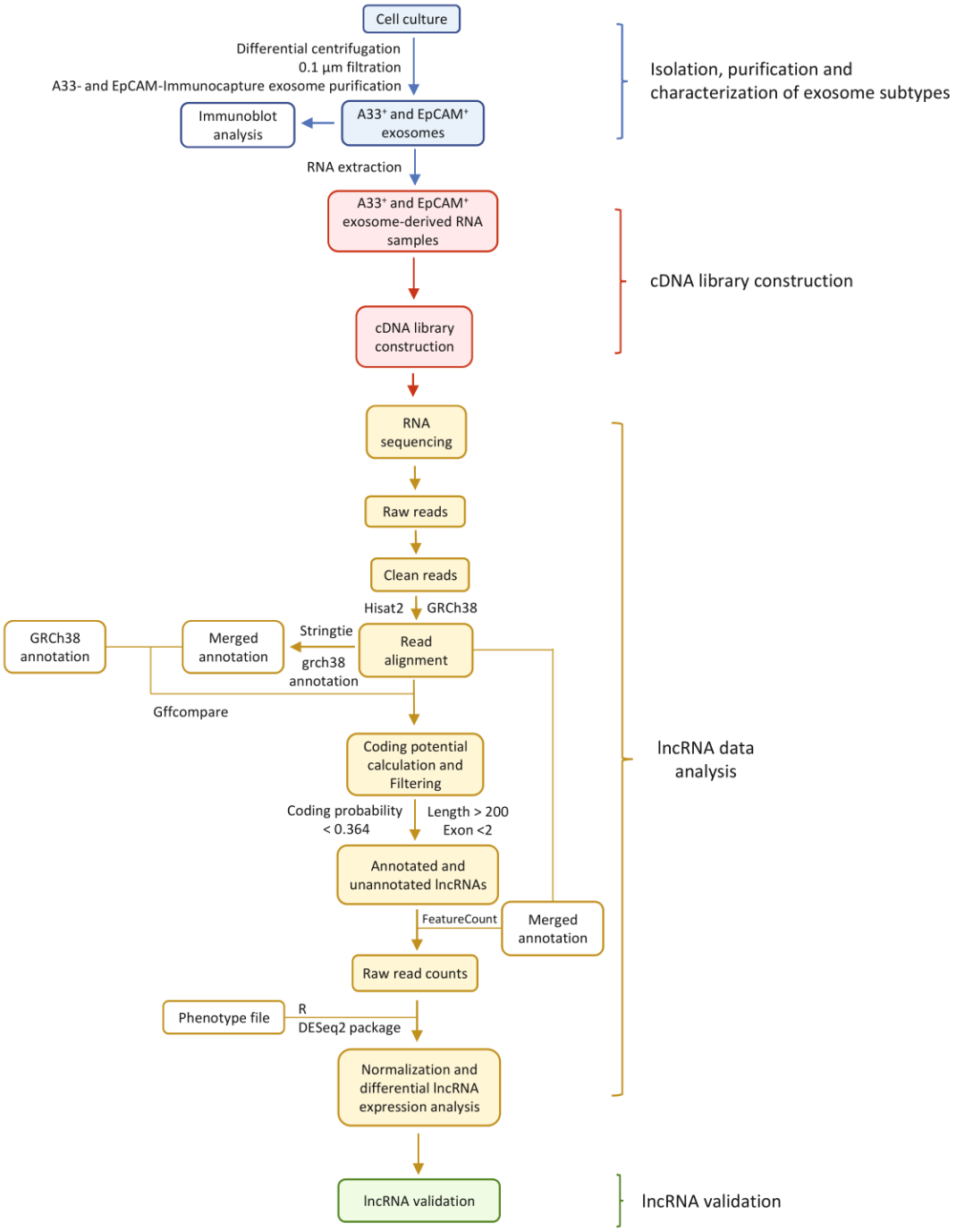


Fig. 1 Exosomal lncRNA analysis workflow. The workflow is designed for discovery of annotated and unannotated exosomal lncRNAs. The workflow starts from LIM1863 culture in Bioreactor classic flasks for a large-scale exosome generation followed by isolation and characterization of exosome subtypes using differential centrifugations, filtration and immunoaffinity capture exosome purification ($A33^+$ and $EpCAM^+$ exosomes), respectively. Total RNA is isolated from $A33^+$ and $EpCAM^+$ exosomes for cDNA library construction and next-generation RNA sequencing. Clean reads are aligned with human genome (GRCh38) using Hisat2.

[32] (Fig. 2a, b). (Previously, we reported that two exosome subtypes derived from LIM1863 cells ($A33^+$ and $EpCAM^+$) exhibit distinct protein and transcriptomic profiles [21, 32]). Then total RNA is isolated from $A33^+$ and $EpCAM^+$ exosomes for cDNA library construction (Fig. 3). We also provide paired-ended clean reads of $A33^+$ and $EpCAM^+$ exosomes for lncRNA data analysis. The clean reads can be downloaded from NCBI using SRA Toolkit (SRR1662176 for $A33^+$ exosome clean reads and SRR1662177 for $EpCAM^+$ exosome clean reads) [21]. Clean reads are mapped to human genome (Ensembl, GRCh38) using Hisat2 [33]. Unannotated lncRNA are distinguished from annotated lncRNAs using Stringtie and Gffcompare [34] with cutoffs: (1) nucleotide length more than 200, (2) number of exon <2, (3) classification codes "i" (fully contained within a reference intron), "x" (exonic overlap on the opposite strand), and "u" (unknown, intergenic) [35]. Protein coding potential for each unannotated lncRNAs is calculated by Coding-Potential Assessment Tool (CPAT). Coding probability (CP) < 0.364 (for human samples) indicates noncoding sequence [36] (this tool also allows identifying lncRNAs with protein-coding potential). We further perform raw read count calculation and differential expression analysis using FeatureCounts [37] and DESeq2 [38], respectively. In the final step, significantly enriched lncRNAs are validated using RT-qPCR.

The lncRNA analysis method described here utilizes RNA sequencing data and stringent parameters to discover annotated and unannotated lncRNAs in two exosome subtypes sequentially isolated from the human colon cell line, LIM1863. The large-scale isolation, purification, and characterization of exosomes are described. Our method permits novel exosomal lncRNA discovery using RNA sequencing coupled with stringent bioinformatic approaches. This protocol facilitates the discovery of novel exosomal lncRNA candidates for future cell biology studies and potential disease biomarkers.

2 Materials

2.1 Cell Culture

1. LIM1863 cells [28].
2. RPMI1640.
3. EV-depleted fetal bovine serum (FBS).

Fig. 1 (continued) Unannotated lncRNAs are identified using Stringtie coupled with Gffcompare. Cutoffs (nucleotide length, number of exon, classification codes, and protein coding probability) are applied to identify unannotated lncRNAs. Raw read count calculation and differential expression analysis are performed using FeatureCounts [37] and DESeq2 [38], respectively. RT-qPCR is used to validate significantly enriched lncRNAs between the samples

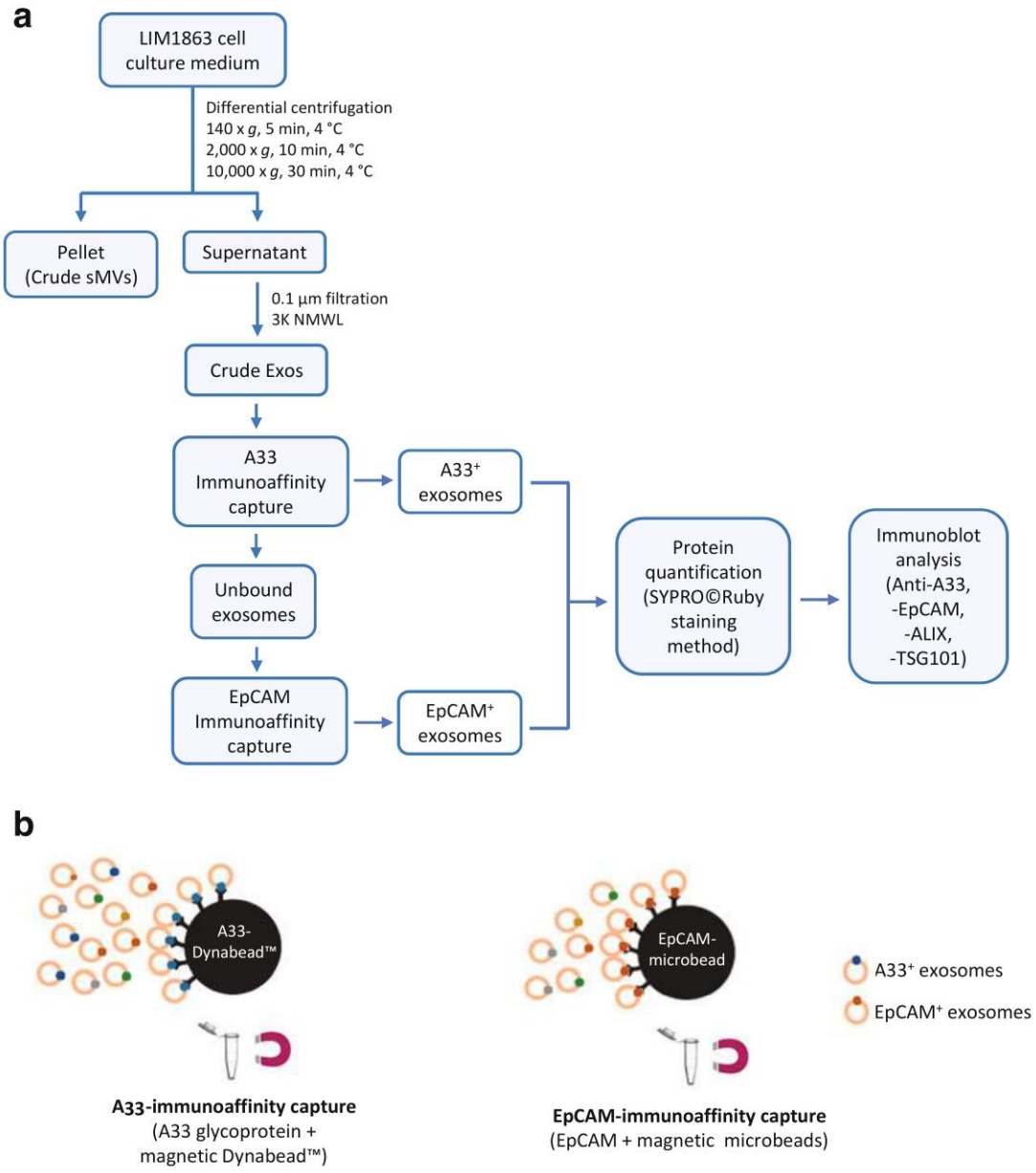


Fig. 2 Isolation, purification and characterization of exosomes. (**a**) LIM1863 cells are cultured in Bioreactor classic flasks. Exosomes are isolated and purified using differential centrifugation, filtration and A33-, EpCAM-immunoaffinity captures. Proteins in $A33^+$ and $EpCAM^+$ exosomes are quantified using SYPRO® Ruby staining method. A33, EpCAM, ALIX, and TSG101 abundance is analyzed by immunoblot. $A33^+$ exosomes are $A33^+$, $ALIX^+$, $TSG101^+$, and $EpCAM^-$, and $EpCAM^+$ exosomes are $EpCAM^+$, $ALIX^+$, $TSG101^+$, $A33^-$. Transmission electron microscopy analysis shows $A33^+$ and $EpCAM^+$ exosomes are cupped shape structure with diameter around 40–60 nm *see* [32] (sMVs = Shed microvesicles, Exos = Exosomes). (**b**) A diagram shows purification of $A33^+$ and $EpCAM^+$ exosomes using immunoaffinity capture technique

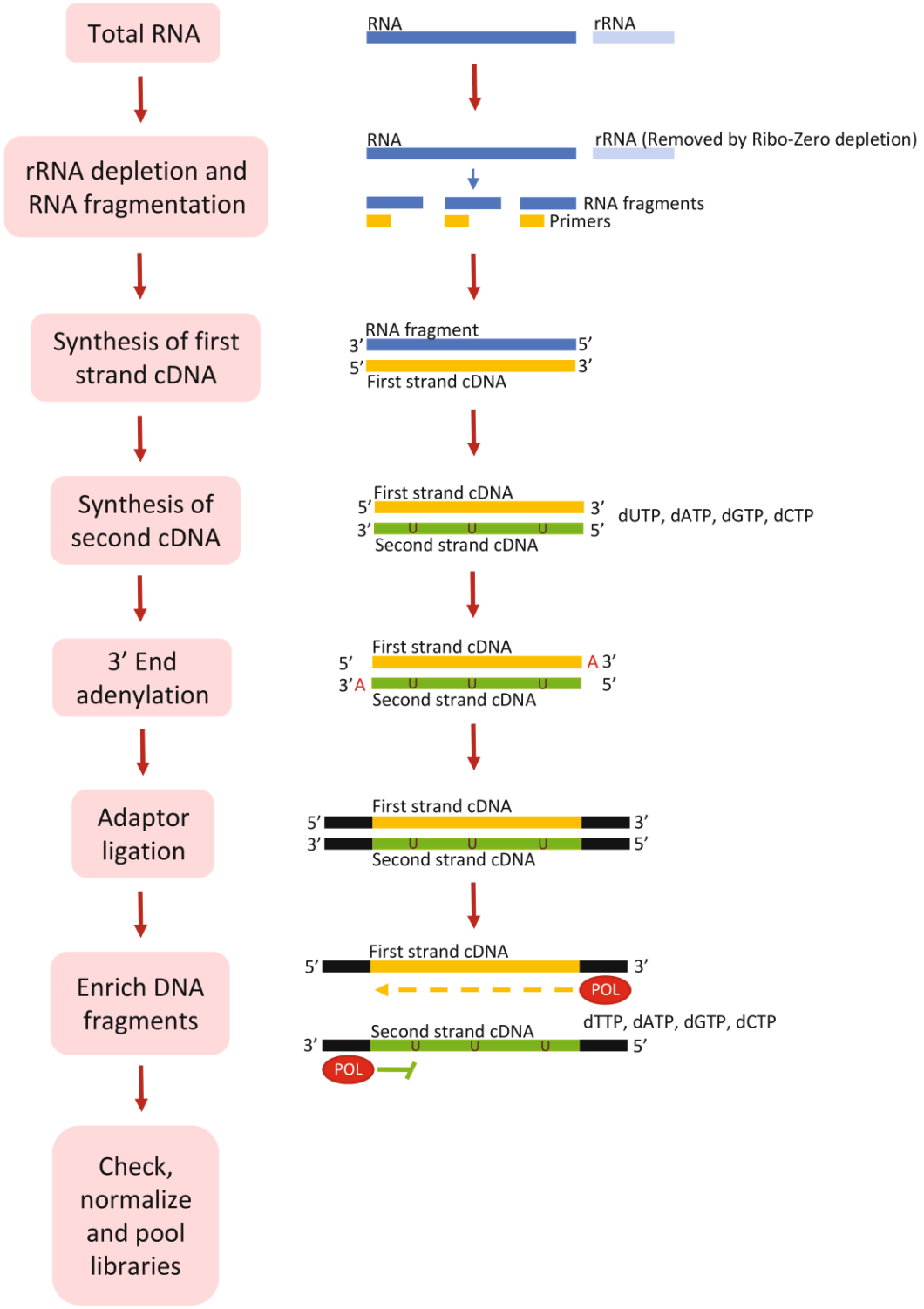


Fig. 3 cDNA library construction. This method captures both poly-A tail and non-poly-A tail RNAs and depletes rRNAs in samples. Starting from rRNA depletion/RNA fragmentation using divalent cations under elevated temperature and synthesis of first and second strands. Then double-stranded cDNAs are conjugated with adenine at 3′ and adapter sequence. cDNAs are enriched using PCR technique and measured quality using Agilent 2100 Bioanalyzer

4. Insulin–transferrin–selenium (ITS).
5. Penicillin/Streptomycin (P/S).
6. CELLine CL-1000 Bioreactor classic flask.

2.2 Isolation, Purification and Characterization of Exosomes

1. Protein G Dynabeads™ (Invitrogen).
2. Citrate phosphate buffer (pH 5.0).
3. Triethanolamine (0.2 M, pH 7.5).
4. Phosphate buffer saline (PBS, pH 7.4).
5. Glycine (0.2 M, pH 2.8).
6. EpCAM magnetic microbeads (Miltenyi Biotec).
7. Rinsing solution (MACS® BSA Stock Solution diluted 1:20 with autoMACS® Rinsing Solution; Miltenyi Biotec).
8. LS microcolumn (Miltenyi Biotec).
9. Tris buffer (50 nM, pH 7.5).
10. Phosphate buffer saline (PBS, pH 7.4).
11. SDS sample buffer (2% (w/v) sodium dodecyl sulfate, 125 mM Tris–HCl, pH 6.8, 12.5% (v/v) glycerol, 0.02% (w/v) bromophenol blue) with 100 mM dithiothreitol (DTT).
12. NuPAGE™ 4–12% (w/v) Bis-Tris Precast gels (Life Technologies).
13. NuPAGE™ MES running buffer (Life Technologies).
14. Fixing solution (40% (v/v) methanol, 10% (v/v) aqueous acetic acid).
15. SYPRO®Ruby fluorescent stain.
16. Destaining solution (10% (v/v) methanol with 6% (v/v) acetic acid in water).
17. Typhoon 9410 variable mode imager (GMI) or suitable florescent imager.
18. ImageQuant TL 8.2 image analysis software (GE Healthcare Life Sciences).
19. BenchMark™ Protein Ladder (Life Technologies).
20. iBlot™2.0 DryBlotting System (Life Technologies),
21. Tris buffered saline (TBS).
22. Anti-A33 (human, 1 μg/mL, a gift from Andrew Scott, Ludwig Institute for Cancer Research Ltd.—Austin Campus, Melbourne, AU) or suitable sourced Anti-GPA33 antibody (reactivity: human, IgG isotype) (#MA5-24139, Invitrogen).
23. Anti-EpCAM (1:1000; Abcam).
24. Anti-TSG101 (1:1000; BD Biosciences).
25. Anti-ALIX (1:1000; Cell Signaling).

26. IRDye 800 goat anti-mouse IgG or IRDye 680 goat anti-rabbit IgG (1:15,000; LI-COR Biosciences).
27. Odyssey Infrared Imaging System, v 3.0 (LICOR Biosciences).
28. Supor® Membrane (0.1 μm) (Pall Life Sciences).
29. Amicon® Ultra-15 Ultracel centrifugal filter device with a 3K nominal molecular weight limit (Merck-Millipore).
30. Dimethyl pimelimidate (DMP).
31. Tween 20.
32. Ultracentrifuge (performance speeds of up to 100,000 × *g*).
33. Tabletop ultracentrifuge (performance speeds of up to 400,000 × *g*, *TLA100 rotor*).
34. Microcentrifuge polypropylene tube.
35. Polycarbonate bottle assembly.
36. Skim milk powder (5% w/v).

2.3 Isolation of Exosomal RNA

1. TRIzol™ reagent (Invitrogen).
2. Chloroform.
3. Glycogen (Invitrogen).
4. Isopropyl alcohol.
5. RNase-free water.
6. Agilent 2100 Bioanalyzer (Agilent Technologies).

2.4 cDNA Library Construction

1. Nuclease-free water.
2. 96-well PCR plates.
3. rRNA Binding Buffer (RBB).
4. Removal Mix (Illumina).
5. PCR machine (Bio-Rad).
6. rRNA Removal Beads (Illumina).
7. RNAclean XP Beads (Illumina).
8. 70% aqueous ethanol and 80% aqueous ethanol.
9. Elution Buffer (ELB).
10. Elute, Prime, Fragment High Mix (EPH, Illumina).
11. First Strand Synthesis Act D mix (FSA, Illumina).
12. Super Script II (Invitrogen).
13. Diluted End Pair Control (CTE, Illumina).
14. Resuspension Buffer (RSB, Illumina).
15. Second Strand Marking Master Mix (SMM, Illumina).
16. AMPure XP beads (Illumina).
17. A-Tailing control (CTA, Illumina).

18. A-Tailing Mix (ATL, Illumina).
19. Adaptor tubes.
20. Index Adaptor plates.
21. Ligation Control (CTL, Illumina).
22. Ligation Mix (LIG, Illumina).
23. RNA adaptors (Illumina).
24. PCR Primer Cocktail (PPC, Illumina).
25. PCR Master Mix (PMM, Illumina).
26. A DNA 1000 chip (Agilent Technologies).
27. Agilent Technologies 2100 Bioanalyzer (Agilent Technologies).
28. Tris-HCl (10 mM, pH 8.5).
29. Adhesive PCR plate seals.
30. Tween 20.

2.5 Bioinformatic Analysis Pipeline for lncRNA Sequencing Analysis

1. A computer (64-bit) with either Linux or Mac OS X (10.7 Lion or later), 8GB (recommended) of RAM.
2. Software (*see* Table 1).
3. Files (*see* Table 2).

2.6 LncRNA Validation

1. SuperScript™ IV VILO™ Master Mix (Invitrogen).
2. SSoAdvanced universal probes Supermix (Bio-Rad).
3. Forward and reverse primers.
4. PCR machine.
5. CFX Manager™ Software (v 3.1) (Bio-Rad).

3 Methods

3.1 Cell Culture

1. LIM1863 cells are cultured in a 175-cm^2 flask in RPMI1640 supplemented with 5% (v/v) fetal bovine serum (FBS), 0.1% (v/v) insulin–transferrin–selenium (ITS), and 1% (v/v) Penicillin/Streptomycin (P/S) at 37 °C and 5% CO_2.
2. LIM1863 cells (~3×10^7 cells) are collected as floating organoids and centrifuged at 140 × *g*, 4 °C for 5 min.
3. Supernatant is removed, and pellet (LIM1863 cells) are resuspended in 15 mL RPMI1640 medium containing 0.5% (v/v) ITS and 1% (v/v) P/S.
4. Cell solution (15 mL) is transferred into the Cultivation chamber (lower chamber) of a CELLine CL-1000 Bioreactor classic flask. The Nutrient Supply chamber (upper chamber) contains

Table 1
Software required for exosomal lncRNA discovery

File	URL for downloading
SRAtoolkit	https://www.ncbi.nlm.nih.gov/sra/docs/toolkitsoft/
Hisat2	https://ccb.jhu.edu/software/hisat2/index.shtml
SAMtools	http://www.htslib.org/
Stringtie	http://ccb.jhu.edu/software/stringtie/#install
GffCompare	http://ccb.jhu.edu/software/stringtie/gffcompare.shtml
Gffread	https://github.com/gpertea/gffread
CPAT	http://rna-cpat.sourceforge.net/
FeatureCounts	http://bioinf.wehi.edu.au/subread/
R	https://www.r-project.org/
biomaRt	https://bioconductor.org/packages/release/bioc/html/biomaRt.html
DESeq2	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
FileZilla	https://filezilla-project.org/

Table 2
Files required for exosomal lncRNA discovery

File	Source	URL for downloading
Human genome indexes	Ensembl, *H. sapiens*, GRCh38	ftp://ftp.ccb.jhu.edu/pub/infphilo/hisat2/data/grch38.tar.gz
Human annotation file	Ensembl, *H. sapiens*, GRCh38	ftp://ftp.ensembl.org/pub/release-96/gtf/homo_sapiens
Human DNA data	Ensembl, *H. sapiens*	ftp://ftp.ensembl.org/pub/release-96/fasta/homo_sapiens/dna

500 mL of RPMI-1640 supplemented with 5% (v/v) FBS and 1% (v/v) P/S and cultured in 37 °C, 5% CO_2.

5. After a 5-day seeding timeframe, the medium in the Cultivation chamber is harvested every 48 h and then centrifuged at 140 × *g*, 4 °C for 5 min.
6. The cell pellet (LIM1863 cells) is resuspended in 15 mL RPMI-1640 supplemented with 0.5% (v/v) ITS and 1% (v/v) P/S and reseeded back to the Cultivation chamber cultured in 37 °C and 5% CO_2. Supernatant is prepared for sequential centrifugation.

   (The medium in Nutrient Supply chamber is replaced with 500 mL RPMI-1640 containing 5% (v/v) FBS and 1% (v/v) P/S twice a week.)

3.2 Isolation, Purification and Characterization of Exosomes

3.2.1 Sequential Centrifugation and Filtration

1. Supernatant (from Subheading 3.1) is centrifuged at 2000 × *g*, 4 °C for 10 min to remove cell debris. Pellet is discarded.
2. Supernatant is further centrifuged at 10,000 × *g*, 4 °C for 30 min to remove shed microvesicles (sMVs). Pellet (sMVs) is discarded.
3. Supernatant is filtered using a VacuCap® 60 filter unit fitted with a 0.1 μm Supor® Membrane.
4. Supernatant is further concentrated to 500 μL using an Amicon® Ultra-15 Ultracel centrifugal filter device with a 3K nominal molecular weight limit (3K NMWL).

3.2.2 A33- and EpCAM-Immunoaffinity Capture for $A33^+$ and $EpCAM^+$ Exosome Purification

1. Protein G Dynabead™ (5×10^8 beads) in citrate phosphate buffer (pH 5.0, 500 μL) is mixed with A33 antibody (100 μL, 300 μg) and incubated with gentle rotation at 25 °C for 40 min.
2. A33 conjugated with Dynabeads™ (A33-Dynabeads™) are placed on a magnetic stand for 2 min (supernatant is discarded) and washed twice with 1 mL citrate-phosphate buffer (pH 5.0) and subsequently 0.2 M triethanolamine (pH 8.2).
3. A33-Dynabeads™ are resuspended in 1 mL of freshly prepared 20 mM dimethyl pimelimidate in 0.2 M triethanolamine (pH 8.2) with gentle rotation for 30 min.
4. A33-Dynabeads™ are placed on a magnetic stand for 2 min (supernatant is discarded). The beads are resuspended in 1 mL Tris buffer (50 nM, pH 7.5) with gentle rotation for 15 min.
5. A33-Dynabeads™ are place on a magnetic stand for 2 min (supernatant is discarded) and washed with PBS containing 0.05% (v/v) Tween 20 for three times.
6. Crude exosomes (from Subheading 3.2.1) are preincubated with Dynabeads™ (500 μL, 5×10^8 beads) at 4 °C for 2 h with gentle rotation to reduce nonspecific binding.
7. Dynabeads™ (from **step 6**) are removed by a magnetic stand and supernatant (crude exosomes) is incubated with A33-Dynabeads™ (from **step 4**) at 4 °C for 2 h with gentle rotation.
8. $A33^+$ exosomes are collected using a magnetic stand and washed three times with 1 mL PBS for 5 min. Supernatant is collected for $EpCAM^+$ exosome purification.
9. A33-Dynabeads™ are eluted from $A33^+$ exosome complex by 100 μL glycine (0.2 M, pH 2.8).
10. A33-depleted exosomes (from **step 8**) is incubated with 100 μg EpCAM (CD326) magnetic microbeads at 4 °C for 4 h.
11. $EpCAM^+$ exosome-bound microbeads are subjected on to a 3 mL LS microcolumn and washed three times with 1 mL

Rinsing solution using a magnetic stand (LS column is rinsed three times with Rinsing solution before use).

12. $EpCAM^+$ exosome-bound microbeads are washed with 1 mL PBS twice and centrifuged at 100,000 × *g* at 4 °C for 1 h.
13. $EpCAM^+$ exosome (pellet) are eluted from the microbeads by 100 μL glycine (0.2 M, pH 2.8).

3.2.3 Characterization of Exosomes (Protein Quantification)

1. $A33^+$, $EpCAM^+$, and unbound exosome samples (5 μL) are mixed with 5 μL SDS sample buffer (2% (w/v) sodium dodecyl sulfate, 125 mM Tris–HCl, pH 6.8, 12.5% (v/v) glycerol, and 0.02% (w/v) bromophenol blue) with 100 mM dithiothreitol (DTT) and loaded into a 1 mm, 10-well NuPAGE™ 4–12% (w/v) Bis-Tris Precast gel.
2. Electrophoresis is performed at 150 V for 1 h in 1× NuPAGE™ MES running buffer.
3. The gel is fixed in 50 mL fixing solution (40% (v/v) methanol, 10% (v/v) acetic acid in water) for 30 min on an orbital shaker.
4. The gel is stained with SYPRO® Ruby overnight, followed by destaining solution twice in 50 mL of 10% (v/v) methanol with 6% (v/v) acetic acid in water for 2 h.
5. The gel is imaged on a Typhoon 9410 variable mode imager (Molecular Dynamics, Sunnyvale, USA), using a green (532 nm) excitation laser and a 610BP30 emission filter at 100 μm resolution.
6. Densitometry quantitation is performed using ImageQuant TL 8.1 image analysis software to determine protein concentration relative to a BenchMark™ Protein Ladder standard of known protein concentration (1.7 μg/μL) according to the manufacturer's instructions (GE Healthcare Life Sciences, USA, ImageQuant TL User's Guide).

3.2.4 Characterization of Exosomes (Immunoblot Analysis)

1. $A33^+$, $EpCAM^+$ and unbound exosome samples in SDS sample buffer (20 μg total protein) are loaded into a 1 mm, 10-well NuPAGE™ 4–12% (w/v) Bis-Tris Precast gel.
2. Electrophoresis is performed at 150 V for 1 h in 1× NuPAGE™ MES running buffer.
3. Proteins on the gel are electrotransferred onto nitrocellulose membranes using the iBlot™ 2.0 DryBlotting System.
4. Membranes are incubated with 5% (w/v) skim milk powder in TTBS at 25 °C for 1 h.
5. Membranes are probed with primary antibodies in TTBS on an orbital shaker at 4 °C for overnight (rabbit anti-EpCAM (1:1000; Abcam), mouse anti-A33 (1 μg/mL, Andrew Scott, Ludwig Institute for Cancer Research Ltd.—Austin Campus,

Melbourne, AU or suitable sourced Anti-GPA33 antibody (reactivity: human, IgG isotype)), mouse anti-TSG101 (1:1000; BD Biosciences), and mouse anti-ALIX (1:1000; Cell Signaling).

6. Membranes are washed three times with 20 mL TTBS at 25 °C for 10 min on an orbital shaker.
7. Membranes are incubated with the secondary antibody, IRDye 800 goat anti-mouse IgG or IRDye 680 goat anti-rabbit IgG (1:15,000; LI-COR Biosciences, USA), at 25 °C for 1 h in darkness on an orbital shaker.
8. Membranes are washed three times with 20 mL TTBS at 25 °C for 10 min on an orbital shaker.
9. Membranes are visualized using the Odyssey Infrared Imaging System, v 3.0 according to the manufacturer's instructions (LI-COR Biosciences, USA, Odyssey Infrared Imaging System Operator's Manual v 3.0).

3.3 Isolation of Exosomal RNA

1. Exosome samples are incubated with 1 mL TRizol™ reagent for 5 min at 25 °C then 0.2 mL chloroform is added. (samples are vortexed vigorously for 15 s) and further incubated in RT for 3 min.
2. Samples are centrifuged at 12,000 × *g*, 4 °C for 15 min.
3. Aqueous phase is collected, mixed with 5–10 μg of glycogen (20 mg/mL aqueous glycogen, Invitrogen) and isopropyl alcohol (0.5 mL isopropyl alcohol/1 mL aqueous phase) incubated at 25 °C for 10 min.
4. Total RNA is centrifuged at 12,000 × *g*, 4 °C for 15 min.
5. RNA pellet is washed once with 75% aqueous ethanol, air-dried for 5 min, and redissolved in RNase-free water (the quality and composition of RNA samples are evaluated using an Agilent 2100 Bioanalyzer according to the manufacturer's instructions (Agilent Technologies, USA, 2100 Expert Software User's Guide).

3.4 cDNA Library Construction

3.4.1 rRNA Depletion and RNA Fragmentation

1. Total RNA is diluted in nuclease-free water to final volume of 10 μL in each well of the Bind rRNA plate (BRP).
2. rRNA binding buffer (5 μL) is added into each well.
3. Removal Mix (5 μL) is added into each well and mixed by gentle pipetting.
4. BRP is centrifuged at 280 × *g* for 1 min.
5. BRP is placed on PCR machine and run RNA denaturation program then further incubated at 25 °C for 1 min.

6. Solution (~20 μL) from BRP is transferred to rRNA Removal Plate (RRP) where rRNA Removal Beads (35 μL) is placed in then further incubated at 25 °C for 1 min.
7. RRP is placed in a magnetic stand for 1 min.
8. Solution in each well is transferred into each well of RNA Clean up Plate (RCP) and placed in a magnetic stand for 1 min.
9. RNAClean XP Beads (99 μL) are added into each well and mixed with gentle pipetting (*see* **Note 1**) then incubated at 25 °C for 15 min.
10. RCP is placed in a magnetic stand until supernatant is clear then supernatant is discarded from each well.
11. Freshly prepared 70% ethanol (200 μL) is added into each well.
12. RCP is placed in a magnetic stand until supernatant is clear then supernatant is discarded from each well (*see* **Note 2**).
13. RCP is still placed on a magnetic stand for 15 min to air-dry ethanol.
14. Elution Buffer (ELB, 11 μL) is added into each well and mixed with gentle pipetting (*see* **Note 3**).
15. RCP is incubated at 25 °C for 2 min then centrifuged at 280 × *g* for 1 min.
16. RCP is placed in a magnetic stand until supernatant is clear then supernatant (8.5 μL) is transferred to the corresponding well of the Depleted RNA Fragmentation Plate (DFP).
17. Elute, Prime, Fragment High Mix (EPH, 8.5 μL) is added into each well of DFP and mixed with gentle pipetting.
18. DFP is placed on PCR machine and run Elution 2-Frag-Prime program (preheat lid option and set to 100 °C, 94 °C for 8 min, hold at 4 °C).

3.4.2 Synthesis of First Strand cDNA

1. First Strand Synthesis Act D mix (FSA) and Super Script II are mixed at the ratio of 9 μL FSA and 1 μL Super Script II (*see* **Note 4**).
2. FSA and Super Script II mixture (8 μL) is added into each well of DFP and mixed with gentle pipetting.
3. DFP is centrifuged at 280 × *g* for 1 min then placed to a PCR machine and run the Synthesis of first strand program (preheat lid option and set to 100 °C, 25 °C for 10 min, 42 °C for 15 min, 70 °C for 15 min, hold at 4 °C).

3.4.3 Synthesis of Second Strand cDNA

1. Diluted End Pair Control (CTE, 5 μL) and Resuspension Buffer (RSB, 5 μL) are added into each well of DFP and mixed with gentle pipetting (*see* **Note 5** for diluted CTE preparation).

2. Second Strand Marking Master Mix (SMM, 20 μL) is added into each well and mixed with gentle pipetting (*see* **Note 6**) then centrifuged at 280 × *g* for 1 min.
3. DFP is placed in PCR machine and incubated at 16 °C for 1 h then moved to 25 °C.
4. AMPure XP beads (90 μL) is added into each well of the DFP and mixed with gentle pipetting then incubated at 25 °C for 15 min.
5. DFP is centrifuged at 280 × *g* for 1 min.
6. DFP is placed on a magnetic stand until the supernatant is clear then 135 μL supernatant is discarded from each well.
7. Freshly prepared 80% ethanol (200 μL) is added into each well of DFP and incubated on the magnetic stand for 30 s then the ethanol is discarded.
8. Repeat **step 7** (*see* **Note 2**).
9. DFP is placed on a magnetic stand for 15 min to air-dry ethanol then removed from the magnetic stand.
10. RSB (17.5 μL) is added into each well of DFP and mixed with gentle pipetting then incubated at 25 °C for 2 min.
11. DFP is centrifuged at 280 × *g* for 1 min.
12. DFP is placed on a magnetic stand until the supernatant is clear then 15 μL supernatant is transferred to the corresponding well of the Adaptor Ligation Plate (ALP).

3.4.4 3′ End Adenylation

1. Diluted A-Tailing control (CTA, 2.5 μL) is added into each well of ALP and mixed with gentle pipetting (*see* **Note 7** for diluted CTA preparation).
2. A-Tailing Mix (ATL, 12.5 μL) is added into each well of ALP and mixed with gentle pipetting (*see* **Note 8**).
3. ALP is sealed with an adhesive PCR plate seal and centrifuged at 280 × *g* for 1 min.
4. ALP is placed on PCR machine and run ATAIL70 program (preheat lid option and set to 100 °C, 37 °C for 30 min, 70 °C for 5 min, hold at 4 °C) then centrifuged at 280 × *g* for 1 min.

3.4.5 Adaptor Ligation

1. RNA adapters are prepared as followed; Adapter tubes are centrifuged at 600 × *g* for 5 s, Index Adapter plate is centrifuged at 280 × *g* for 1 min. Ligation Control (CTL) is centrifuged at 600 × *g* for 5 s then diluted to 1:100 in RSB.
2. Diluted CTL (2.5 μL), Ligation Mix (LIG, 2.5 μL), and RNA adapters (2.5 μL) are added in the order into each well of ALP and mixed with gentle pipetting.
3. ALP is centrifuged at 280 × *g* for 1 min.

4. AMPure XP beads (42 μL) are added into each well of ALP and mixed with gentle pipetting then incubated at 25 °C for 15 min.
5. ALP is centrifuged at 280 × *g* for 1 min.
6. ALP is placed on a magnetic stand until the supernatant is clear then supernatant is discarded from each well.
7. Freshly prepared 80% ethanol (200 μL) is added into each well and incubated on the magnetic stand for 30 s then the ethanol is discarded.
8. Repeat **step 7** (*see* **Note 2**).
9. ALP is placed on a magnetic stand for 15 min to air-dry ethanol then removed from the magnetic stand.
10. RSB (52.5 μL) is added into each well of ALP and mixed with gentle pipetting.
11. ALP is incubated at 25 °C for 2 min and further centrifuged at 280 × *g* for 1 min.
12. ALP is placed on a magnetic stand until the supernatant is clear then 50 μL supernatant is transferred into the corresponding well of the Clean Up ALP (CAP).
13. Repeat **steps 4–12** with 50 μL AMPure XP beads and 22.5 μL RSB.
14. Supernatant (20 μL) from **step 13** is transferred into the corresponding well of the PCR plate.

3.4.6 Enrich DNA Fragments

1. PCR plate is placed on ice and 5 μL PCR Primer Cocktail (PPC) is added into each well.
2. PCR Master Mix (PMM, 25 μL) is added into each well and mixed with gentle pipetting.
3. PCR plate is centrifuged at 280 × *g* for 1 min.
4. PCR plate is placed in PCR machine and run PCR program (preheat lid option and set to 100 °C, 98 °C for 30 s, 15 cycles of 98 °C for 10 s, 60 °C for 30 s, 72 °C for 30 s (end of the 15 cycles), and finish with 72 °C for 5 min, hold at 4 °C).
5. PCR plate is centrifuged at 280 × *g* for 1 min.
6. AMPure beads (50 μL for Adaptor tubes, 47.5 μL for Index Adaptor Plate) are added into each well and mixed with gentle pipetting.
7. PCR plate is incubated at 25 °C for 15 min then further centrifuged at 280 × *g* for 1 min.
8. PCR plate is placed on a magnetic stand until the supernatant is clear then supernatant is discarded from each well.

9. Freshly prepared 80% ethanol (200 μL) is added into each well and incubated on the magnetic stand for 30 s then the ethanol is discarded.
10. Repeat **step 9** (*see* **Note 2**).
11. PCR plate is placed on a magnetic stand for 15 min to air-dry ethanol then removed from the magnetic stand.
12. RSB (32.5 μL) is added into each well and mixed with gentle pipetting.
13. PCR plate is centrifuged at 280 × *g* for 1 min.
14. PCR plate is placed on a magnetic stand until the supernatant is clear, then 30 μL supernatant is transferred to the corresponding well of the Target Sample Plate1 (TSP1).

3.4.7 Check Libraries

1. DNA library from TSP1 (1 μL) is added to a DNA 1000 chip on an Agilent Technologies 2100 Bioanalyzer according to the manufacturer's instructions (Agilent Technologies, USA, 2100 Expert Software User's Guide).
2. Check size (~260 bp) and purity of samples.

3.4.8 Normalize and Pool Libraries

1. DNA library (10 μL) is transferred to Diluted Cluster Template plate (DCT).
2. DNA library is normalized to 10 nm by adding Tris–HCl 10 mM, pH 8.5 with 0.1% (v/v) Tween 20.
3. DCT plate is centrifuged at 280 × *g* for 1 min.
4. Each normalized library (10 μL) is transferred to a single well of the Pooled DCT plate (PDP) and centrifuged at 280 × *g* for 1 min.
5. Process to cluster generation according to the manufacturer's instructions (Illumina Inc., USA, HiSeq 3000/4000 PE Cluster Kit).

3.5 Bioinformatic Analysis Pipeline for lncRNA Sequencing Analysis

1. Establish a working directory either in Linux or Mac OS X.
2. Download human genome indexes (*H. sapiens*, GRCh38) from https://ccb.jhu.edu/software/hisat2/index.shtml using `wget` command following `tar xvzf`.

3.5.1 Download Files for RNA Sequence Alignment

3. Download pair-ended RNA sequencing clean reads from NCBI using SRA Toolkit: `prefetch -v SRR1662176 (and SRR1662177)` then separate two pair-ended reads using `fasta-dump --split-files SRR1662176 (and SRR1662177).`

   *SRR1662176 is $A33^+$ exosome clean reads, and SRR1662177 is $EpCAM^+$ exosome clean reads.

4. Download human gene annotation file from ftp://ftp.ensembl.org/pub/release-95/gtf/homo_sapiens using `wget` command.
5. Download human DNA data file from http://ensembl.org/info/data/ftp/index.html using `wget` command and rename the file as fullsequence.fa.

3.5.2 RNA Sequence Alignment

1. Align pair-ended RNA sequencing clean reads using `hisat2` with options `-q --dta -x` and `--summary-file` to obtain an align-read summary result (*see* **Note 9**).
2. Convert SAM files to BAM files using `samtools` with options `sort -@ 8 -o SRR1662176.bam (SRR1662177.bam).`

3.5.3 Transcript Assembly and Annotation Merging

1. Assemble transcripts from SRR1662176.bam and SRR1662177.bam files with human gene annotation (from Subheading 3.5.1, **step 4**) using `stringtie` with options `-p 8 -G -o SRR1662176.gtf (SRR1662177.gtf) -l` (*see* **Notes 10** and **11** for alternatively direct annotated lncRNA discovery).
2. Merge assembled transcript files from both GTF files (SRR1662176.gtf and SRR1662177.gtf) using `Stringtie` with options `--merge -p 8 -G -o merged.gtf` (*see* **Note 12**).

3.5.4 Discovery of Unannotated lncRNA

1. Compare merged.gtf with human annotation file (from Subheading 3.5.1, **step 4**) using `gffcompare` with options `-r -G -o merged` (*see* **Note 13**).
2. `Gffcompare` output files (.stats, .combined.gtf, .refmap, .tracking, and .tmap) are generated.
3. Extract novel transcripts by selecting classification codes "i" (fully contained within a reference intron), "x" (exonic overlap on the opposite strand) and "u" (unknown, intergenic) from .tmap file.
4. Transcripts that contain more than 2 exons are removed.
5. Unannotated long noncoding transcript list is generated.

3.5.5 Protein Coding Potential Calculation

1. Obtain full sequence of each transcript to calculate protein-coding potential by using `gffread` with options `-w fullsequence.fa` (from Subreading 3.5.1, **step 5**) `-g merged.gtf` (*see* **Note 14**).
2. Calculate protein-coding potential of each transcript using `cpat.py` with options `-g fullsequence.fa -d -x -o` (*see* **Note 15**).
3. Filter out transcripts that have human protein-coding probability more than 0.364 and length less than 200 nucleotides.
4. Long noncoding transcript list is generated.

3.5.6 Raw Read Counting

1. Calculate raw read counts using `featureCounts` with options `-p -t exon -g gene_id -a -o counts.txt` (*see* **Note 16**).
2. Compare transcripts from counts.txt with lists from Subheadings 3.5.4 and 3.5.5 to obtain unannotated lncRNAs.
3. RNA biotypes of annotated lncRNAs are obtained using R package "biomaRt." Sense-intronic, lincRNA and macro lncRNA are considered as annotated lncRNAs (Antisense can be a either sncRNA or lncRNA) (*see* **Note 17**).

3.5.7 Differential Gene Expression Analysis

1. A raw read count file containing raw read counts of lncRNA is used to perform differential gene expression analysis using R package "DESeq2."
2. DESeq2 package is called by `library(DESeq2)`.
3. `countData` is imported using `read.csv` following by `as.matrix`.
4. Raw read counts that are equal 0 and 1 are removed by `[rowSums(countData)>1,]`.
5. `colData` that contains phenotypes (condition) of samples is imported using `read.csv` (*see* **Note 18**).
6. DESeqDataSet is calculated using `DESeqDataSetFromMatrix(countData,colData,design = ~condition)` following by `DESeq`.
7. Differential gene expression result is exported using `results` for downstream analysis. Annotated and unannotated lncRNAs derived from differential gene expression analysis are shown in Table 3 (*see* **Note 19** for normalized count extraction, **Note 20** for Fragments Per Kilobase of transcript per Million mapped read (FPKM) calculation for down steam analysis, and **Note 21** for transferring files to Windows OS).

Table 3
Selected annotated and unannotated exosomal lncRNAs in LIM1863 A33$^+$ and EpCAM$^+$ exosomes

LncRNA annotation	Gene ID	Gene name	RNA type	Log2 Foldchange (A33$^+$ Exo vs EpCAM$^+$ Exo)	*P*-value
Annotated lncRNAs	ENSG00000268713	AC005261.3	lincRNA	−8.9	1.58E-02
	ENSG00000224417	AL606970.1	sense_intronic	−8.2	2.93E-02
	ENSG00000254854	AP003390.1	Antisense	−8.1	3.21E-02
	ENSG00000260804	LINC01963	lincRNA	−8.0	3.54E-02
Unannotated lncRNAs[a]	MSTRG.10899	–	–	−9.8	7.07E-03
	MSTRG.40911	–	–	−9.3	1.09E-02
	MSTRG.92003	–	–	−8.8	1.67E-02
	MSTRG.97444	–	–	−8.8	1.67E-02

[a]Unannotated lncRNAs based on the version of Ensembl human GRCh38 annotation file (March, 2019)

3.6 lncRNA Validation Using RT-qPCR

3.6.1 Reverse Transcription (cDNA Synthesis)

1. Template lncRNA (100 ng to 2.5 μg) and SuperScript™ IV VILO™ Master Mix (4 μL) are mixed and topped up with nuclease-free water to 20 μL.
2. Samples are incubated at 25 °C for 10 min.
3. Samples are incubated at 50 °C for 10 min and 85 °C for 5 min, respectively (cDNAs are generated).
4. Primers are designed using OligoArchitec™ Online.

3.6.2 Quantitative PCR

1. SSoAdvanced universal probes Supermix (2×, 10 μL), forward and reverse primers (250–900 nM) and cDNA template (~100 ng) are mixed and topped up with nuclease-free water to 20 μL.
2. Sample is vortexed for 30 s and placed on PCR machine (95 °C for 2 min, 30–45 cycles of denaturation at 95 °C for 5–15 s; annealing/extension at 60 °C for 10–30 s).
3. LncRNA expression levels between samples are analyzed and calculated using The CFX Manager™ Software (v3.1) according to the manufacturer's instructions (Bio-Rad, USA, CFX96™ and CFX384™ Real-Time PCR Detection Systems Instruction Manual).

4 Notes

1. RNAClean XP Beads (193 μL) are added into each well if starting with degraded total RNA.
2. Remove the rest of ethanol by using a 20 μL pipette.
3. ELB is centrifuged at 600 × *g* for 5 s before use.
4. FSA is centrifuged at 600 × *g* for 5 s before use.
5. CTE is centrifuged at 600 × *g* for 5 s and diluted to 1:50 in RSB before use.
6. SMM is centrifuged at 600 × *g* for 5 s before use.
7. CTA is centrifuged at 600 × *g* for 5 s and diluted to 1:100 in RSB before use.
8. ATL is centrifuged at 600 × *g* for 5 s before use.
9. `hisat2` inputs either single-ended or pair-ended reads. For pair-ended reads, use `-1` and `-2` for forward and reverse reads, respectively. Use options, `–x` to identify indexes, `-S` following filename to output alignread files in `.sam` file, `--summary-file` following `filename.txt` to output summary alignread.
10. `stringtie` inputs `.bam` file(s) and uses option `-G` following the directory of a human annotation file.

11. Input `.bam` file(s) and use option `-G` following the directory of a human lncRNA annotation file downloaded from GENCODE by using `wget ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_30/gencode.v30.long_noncoding_RNAs.gtf.gz`.
12. `--merge` option from `stringtie` inputs all transcript assembly files from previous step. Use options, `-G` following the directory of human annotation file, `-o` following the new annotation file. This new annotation file will be used for raw read count calculation.
13. `gffcompare` compares transcripts between the human annotation file and the new annotation file. Use options, `-r` following the directory of human annotation file, `-o` merged following the directory of the new annotation file.
14. `gffread` inputs files in FASTA format. Use option `-g` following the annotation file (.gtf) and outputs a full sequence file using option `-w`.
15. CPAT inputs the full sequence file (from Subreading 3.5.1, **step 5**) from the previous step using option `-g`. CPAT requires logitModel.Rdata and Hexamer.tsv files using options `-d` and `-x`, respectively. The logitModel.Rdata and Hexamer.tsv files are organism-specific. All human files associated with CPAT are provided at http://rna-cpat.sourceforge.net/.
16. `featureCounts` inputs all align read files from `hisat2` (.bam files). Use options, `-a` following the directory of the new annotation file and `-o` following an output filename (.txt).
17. `biomaRt` in `R` is used to extract information of annotated transcripts such as gene name, gene description, and gene biotype (for more information use option `attribute`). To perform `biomaRt` in `R`, use.

```
library("biomaRt")
ensembl=useMart("ensembl")
datasets <- listDatasets(ensembl)
ensembl = useMart("ensembl",dataset="hsapiens_gene_ensembl")
```

Then input Ensembl gene id as `value`.

After that perform the command below to obtain Ensembl gene id, Ensembl gene name, and gene biotype.

```
getBM(attributes=c('ensembl_gene_id','external_gene_name','description',
'gene_biotype'), filter='ensembl_gene_id', value=value, mart=ensembl).
```

Finally, the information is exported using `write.csv`.

18. An example of phenotype file is shown in Table 4.

Table 4
An example of phenotype file

ID	Condition
SRR1662176	A33
SRR1662177	EpCAM

19. Normalized counts are calculation from raw read counts (from `FeatureCounts`) by using R package "DESeq2" followed by these commands.

```
raw_read_count_matrix = as.matrix(raw_read_counts)
dds = DESeq(raw_read_count_matrix)
dds = estimateSizeFactor(dds)
normalized_counts = counts(dds, normalized=TRUE)
normalized_counts then are exported using write.csv command
```

20. FPKM can be calculated by various R packages and one of the packages is "DESeq2." It requires two input files: 1. `raw_read_counts` file generated by `FeatureCounts` and 2. `Length_file` that contains length of each transcript in a column format named "`Length`."

```
raw_read_count_matrix = as.matrix(raw_read_counts)
se = SummarizedExperiment(list(counts = raw_read_count_matrix))
dds = DESeqDataSet(se, ~1)
Then import a length_file
mcols(dds)$basepairs = length_file[, "Length"]
fpkm = fpkm(dds)
```

Then `fpkm` are exported using `write.csv` command.

21. FileZilla is an open source software used to connect different servers. For instance, it connects High Performance Computing (HPC) Linux clusters with Windows OS then files can be managed and transferred between these servers. In this case, hostname, port, username, and password used for HPC login are required for connection in FileZilla (FileZilla tutorial available at https://wiki.filezilla-project.org/FileZilla_Client_Tutorial_(en)).

Acknowledgments

W.S. is supported by a LaTrobe University Postgraduate Scholarship. R.J.S. and R.X are supported by La Trobe University and D. W.G, in part, by the National Health and Medical Research Council of Australia project grants #1057741 and #1139489.

References

1. Venter JC, Adams MD, Myers EW et al (2001) The sequence of the human genome. Science 291(5507):1304–1351. https://doi.org/10.1126/science.1058040
2. Kapranov P, Cheng J, Dike S et al (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 316(5830):1484–1488. https://doi.org/10.1126/science.1138341
3. Frith MC, Pheasant M, Mattick JS (2005) The amazing complexity of the human transcriptome. Eur J Hum Genet 13(8):894–897. https://doi.org/10.1038/sj.ejhg.5201459
4. Shabalina SA, Spiridonov NA (2004) The mammalian transcriptome and the function of non-coding DNA sequences. Genome Biol 5 (4):105. https://doi.org/10.1186/gb-2004-5-4-105
5. Palazzo AF, Lee ES (2015) Non-coding RNA: what is functional and what is junk? Front Genet 6:2. https://doi.org/10.3389/fgene.2015.00002
6. Quinn JJ, Chang HY (2016) Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet 17(1):47–62. https://doi.org/10.1038/nrg.2015.10
7. Fatica A, Bozzoni I (2014) Long non-coding RNAs: new players in cell differentiation and development. Nat Rev Genet 15(1):7–21. https://doi.org/10.1038/nrg3606
8. Mercer TR, Dinger ME, Mattick JS (2009) Long non-coding RNAs: insights into functions. Nat Rev Genet 10(3):155–159. https://doi.org/10.1038/nrg2521
9. Anderson DM, Anderson KM, Chang CL et al (2015) A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. Cell 160(4):595–606. https://doi.org/10.1016/j.cell.2015.01.009
10. Stein CS, Jadiya P, Zhang X et al (2018) Mitoregulin: a lncRNA-encoded microprotein that supports mitochondrial Supercomplexes and respiratory efficiency. Cell Rep 23 (13):3710–3720. e3718. https://doi.org/10.1016/j.celrep.2018.06.002
11. Cai B, Li Z, Ma M et al (2017) LncRNA-Six1 encodes a micropeptide to activate Six1 in cis and is involved in cell proliferation and muscle growth. Front Physiol 8:230. https://doi.org/10.3389/fphys.2017.00230
12. Pan Y, Chen H, Shen X et al (2018) Serum level of long noncoding RNA H19 as a diagnostic biomarker of multiple myeloma. Clin Chim Acta 480:199–205. https://doi.org/10.1016/j.cca.2018.02.019
13. Dong H, Wang W, Chen R et al (2018) Exosome-mediated transfer of lncRNASNHG14 promotes trastuzumab chemoresistance in breast cancer. Int J Oncol 53 (3):1013–1026. https://doi.org/10.3892/ijo.2018.4467
14. Karlsson O, Rodosthenous RS, Jara C et al (2016) Detection of long non-coding RNAs in human breastmilk extracellular vesicles: implications for early child development. Epigenetics 11(10):721–729. https://doi.org/10.1080/15592294.2016.1216285
15. Zhao Y, Xu J (2018) Synovial fluid-derived exosomal lncRNA PCGEM1 as biomarker for the different stages of osteoarthritis. Int Orthop 42(12):2865–2872. https://doi.org/10.1007/s00264-018-4093-6
16. Berrondo C, Flax J, Kucherov V et al (2016) Expression of the long non-coding RNA HOTAIR correlates with disease progression in bladder cancer and is contained in bladder cancer patient urinary exosomes. PLoS One 11 (1):e0147236. https://doi.org/10.1371/journal.pone.0147236
17. Xu R, Greening DW, Zhu HJ et al (2016) Extracellular vesicle isolation and characterization: toward clinical application. J Clin Invest 126(4):1152–1162. https://doi.org/10.1172/JCI81129
18. Xu R, Rai A, Chen M et al (2018) Extracellular vesicles in cancer - implications for future improvements in cancer care. Nat Rev Clin Oncol 15(10):617–638. https://doi.org/10.1038/s41571-018-0036-9
19. Greening DW, Simpson RJ (2018) Understanding extracellular vesicle diversity - current

status. Expert Rev Proteomics 15 (11):887–910. https://doi.org/10.1080/14789450.2018.1537788

20. Jeppesen DK, Fenix AM, Franklin JL et al (2019) Reassessment of exosome composition. Cell 177(2):428–445. e418. https://doi.org/10.1016/j.cell.2019.02.029
21. Chen M, Xu R, Ji H et al (2016) Transcriptome and long noncoding RNA sequencing of three extracellular vesicle subtypes released from the human colon cancer LIM1863 cell line. Sci Rep 6:38397. https://doi.org/10.1038/srep38397
22. Qu L, Ding J, Chen C et al (2016) Exosome-transmitted lncARSR promotes Sunitinib resistance in renal cancer by acting as a competing endogenous RNA. Cancer Cell 29 (5):653–668. https://doi.org/10.1016/j.ccell.2016.03.004
23. Takahashi K, Yan IK, Kogure T et al (2014) Extracellular vesicle-mediated transfer of long non-coding RNA ROR modulates chemosensitivity in human hepatocellular cancer. FEBS Open Biol 4:458–467. https://doi.org/10.1016/j.fob.2014.04.007
24. Takahashi K, Yan IK, Wood J et al (2014) Involvement of extracellular vesicle long noncoding RNA (linc-VLDLR) in tumor cell responses to chemotherapy. Mol Cancer Res 12(10):1377–1387. https://doi.org/10.1158/1541-7786.MCR-13-0636
25. Xu CG, Yang MF, Ren YQ et al (2016) Exosomes mediated transfer of lncRNA UCA1 results in increased tamoxifen resistance in breast cancer cells. Eur Rev Med Pharmacol Sci 20(20):4362–4368
26. Lang HL, Hu GW, Chen Y et al (2017) Glioma cells promote angiogenesis through the release of exosomes containing long non-coding RNA POU3F3. Eur Rev Med Pharmacol Sci 21 (5):959–972
27. Lang HL, Hu GW, Zhang B et al (2017) Glioma cells enhance angiogenesis and inhibit endothelial cell apoptosis through the release of exosomes that contain long non-coding RNA CCAT2. Oncol Rep 38(2):785–798. https://doi.org/10.3892/or.2017.5742
28. Whitehead RH, Jones JK, Gabriel A et al (1987) A new colon carcinoma cell line (LIM1863) that grows as organoids with spontaneous differentiation into crypt-like structures in vitro. Cancer Res 47(10):2683–2689
29. Ji H, Chen M, Greening DW et al (2014) Deep sequencing of RNA from three different extracellular vesicle (EV) subtypes released from the human LIM1863 colon cancer cell line uncovers distinct miRNA-enrichment signatures. PLoS One 9(10):e110314. https://doi.org/10.1371/journal.pone.0110314
30. Suwakulsiri W, Rai A, Xu R et al (2018) Proteomic profiling reveals key cancer progression modulators in shed microvesicles released from isogenic human primary and metastatic colorectal cancer cell lines. Biochim Biophys Acta Proteins Proteom 1867(12):140171. https://doi.org/10.1016/j.bbapap.2018.11.008
31. The human A33 antigen is a transmembrane glycoprotein and a novel member of the immunoglobulin superfamily
32. Tauro BJ, Greening DW, Mathias RA et al (2013) Two distinct populations of exosomes are released from LIM1863 colon carcinoma cell-derived organoids. Mol Cell Proteomics 12(3):587–598. https://doi.org/10.1074/mcp.M112.021303
33. Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12(4):357–360. https://doi.org/10.1038/nmeth.3317
34. Pertea M, Pertea GM, Antonescu CM et al (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 33(3):290–295. https://doi.org/10.1038/nbt.3122
35. Zhao Q, Sun Y, Wang D et al (2018) LncPipe: a Nextflow-based pipeline for identification and analysis of long non-coding RNAs from RNA-Seq data. J Genet Genomics 45 (7):399–401. https://doi.org/10.1016/j.jgg.2018.06.005
36. Wang L, Park HJ, Dasari S et al (2013) CPAT: coding-potential assessment tool using an alignment-free logistic regression model. Nucleic Acids Res 41(6):e74. https://doi.org/10.1093/nar/gkt006
37. Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30(7):923–930. https://doi.org/10.1093/bioinformatics/btt656
38. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15(12):550. https://doi.org/10.1186/s13059-014-0550-8

Appendix Article 4

PLOS ONE

RESEARCH ARTICLE

Distinct shed microvesicle and exosome microRNA signatures reveal diagnostic markers for colorectal cancer

Maoshan Chen[1]*, Rong Xu[1], Alin Rai[1], Wittaya Suwakulsiri[1], Keiichi Izumikawa[2], Hideaki Ishikawa[2], David W. Greening[1], Nobuhiro Takahashi[2], Richard J. Simpson[1]*

1 Department of Biochemistry and Genetics, La Trobe Institute for Molecular Science (LIMS), La Trobe University, Melbourne, Australia, **2** Department of Applied Biological Science, Graduate School of Agriculture, and Global Innovation Research Organisation, Tokyo University of Agriculture and Technology, Tokyo Japan

* m.chen@latrobe.edu.au (MC); Richard.simpson@latrobe.edu.au (RJS)

OPEN ACCESS

Citation: Chen M, Xu R, Rai A, Suwakulsiri W, Izumikawa K, Ishikawa H, et al. (2019) Distinct shed microvesicle and exosome microRNA signatures reveal diagnostic markers for colorectal cancer. PLoS ONE 14(1): e0210003. https://doi.org/10.1371/journal.pone.0210003

Editor: Aamir Ahmad, University of South Alabama Mitchell Cancer Institute, UNITED STATES

Received: August 20, 2018

Accepted: December 14, 2018

Published: January 4, 2019

Data Availability Statement: Raw data (FASTQ formatted files) can be accessed in NCBI SRA database under the accession numbers SRA440609 and SRA448517.

Funding: MC, RX, AR, WS were supported by La Trobe University Post-Graduate Research Scholarships. RS is supported by a Distinguished Professorship (La Trobe University) and DG, a Stone Molecular Biology Fellowship (La Trobe University) and Australian National Health and Medical Research Council (Project: 1139489 and

Abstract

Extracellular vesicle (EV) microRNAs are of major interest as potential diagnostic biomarkers in all cancer types. This study aims to identify miRNA profiles of shed microvesicles (sMVs) and exosomes (Exos) secreted from the isogenic colorectal cancer (CRC) cell lines SW480 and SW620 and evaluate their ability to predict CRC. Deep sequencing of miRNAs in parental cell lysates (CLs) and highly-purified sMVs and Exos was performed. We focused on miRNAs enriched in EVs and dysregulated miRNAs in metastatic cells (SW620) relative to primary cancer cells (SW480). We investigated the ability of EV miRNA signatures to predict CRC tumours using 594 tumours (representing different pathological stages) and 11 normal samples obtained from TCGA. In SW480 and SW620 cells we identified 345 miRNAs, of which 61 and 73 were upregulated and downregulated in SW620-CLs compared to SW480-CLs, respectively. Selective distribution of cellular miRNAs into EVs results in distinct miRNA signatures for sMVs and Exos in each cell line. Cross cell line comparisons of EV miRNA profiles reveal a subset of miRNAs critical in CRC progression from primary carcinoma to metastasis. Many miRNAs non-detectable (<5 TPM) in CLs were significantly enriched (>1000 TPM) in secreted EVs. Strikingly, *miR-7641* which is non-detectable in SW480-CL but upregulated in SW620-CL is highly enriched in EVs secreted from both cell lines. Pearson correlation analysis demonstrated that EV miRNA profiles can be used to predict CRC tumours with ~96% accuracy. Our findings suggest that EV miRNA profiles from CRC cell lines may allow prediction of CRC tumours, and that *miR-7641* may serve as an attractive candidate for the specific, non-invasive diagnosis and prognosis of CRC.

Introduction

Extracellular vesicles (EVs) are nano-membranous particles (30–2000 nm) released by most cell types and function in cell-cell communications [1, 2]. According to particle size, two major EV subtypes have been reported: shed microvesicles (sMVs), also referred to as

1141946). NT, KI, HI were supported by Core Research for Evolutional Science and Technology (CREST) from Japan Science and Technology Agency (JST).

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: 3'-UTR, 3'untranslated region; CEA, carcinoembryonic antigen; CLs, cell lysates; CRC, colorectal cancer; EV, extracellular vesicles; Exos, exosomes; miRNA, microRNA; qRT-PCR, quantitative real-time PCR; SEER, Surveillance, Epidemiology, and End Results; sMVs, shed microvesicles; TCGA, The Cancer Genome Atlas; TMUGS, Tumour Marker Utility Grading System; TPM, transcripts per million reads.

microparticles or microvesicles that range in size from 400–1500 nm and exosomes (Exos) that range in size from 50–150 nm [3]. Exosomes are released as intraluminal vesicles (ILVs) from the multivesicular bodies (MVBs), which are formed by budding of the limiting membrane of late endosomes [1]. By contrast, sMVs are generated by the direct budding and fission of the plasma membrane [4]. To date, EVs have been found to be secreted not only by eukaryotic cells but also from plant cells and pathogens (e.g., bacteria, mycobacteria, archaea, and fungi) [5, 6].

During EV biogenesis, cellular bioactive cargo molecules such as protein, lipid, DNA and RNA are selectively packaged into the EVs, [1]. Comprehensive studies have identified ALIX, TSG101, RAB27A, RAB11B, CD9, CD63, CD81 and CD82 as stereotypic markers for exosomes and KIF23 as for sMVs [7, 8]. Since Valadi and colleagues reported ~121 microRNAs (miRNAs), small noncoding RNAs (~ 22 nt) that function in targeting mRNAs for cleavage or translational repression in animals and plants [9, 10], EV-associated miRNAs have attracted much attention because of their multiple functions in cellular activities and, importantly, their potential as diagnostic and prognostic biomarkers for cancer [11, 12]. For example, miR-200 promotes epithelial-to-mesenchymal transition (EMT) and breast cancer cell metastasis through the exosomal transfer into non-metastatic cells [13]. Lung and pancreatic tumour-derived EVs have been reported to transfer miR-21, activate TLR7 receptor on murine myoblasts and promote apoptosis through c-Jun N-terminal kinase (JNK) activity in C2C12 immortalized myoblasts or primary myoblasts [14]. Because EVs from a variety of cancer types are found in circulating body fluids such as blood their miRNA signatures are being extensively studied as potential diagnostic/ prognostic markers of disease [15].

Colorectal cancer (CRC), a leading cause of cancer death in the Western world [16], is the third most commonly diagnosed cancer in males and the second in females, with an estimated 1.4 million new cases and 50,630 deaths occurring in 2018 [17]. Metastatic CRC is the principal cause of death and only early diagnosis can improve the treatment. Some groups have demonstrated that circulating miRNAs can be a powerful tool to predict early CRC. Chen *et al.* reported that miRNAs, such as miR-134, miR-146a, miR-221, miR-222 and miR-23a, were deregulated in the serum of CRC patients compared to healthy controls [18]. Huang and colleagues identified that plasma miR-29a and miR-92a have significant diagnostic value for advanced colorectal neoplasia [19]. Some studies have also identified miRNA candidates for early CRC diagnosis in circulating EVs. For example, seven miRNAs (let-7a, miR-1229, miR-1246, miR-150, miR-21, miR-223, and miR-23a) were significantly higher in serum exosomes of primary CRC patients, compared to healthy controls [20]. Plasma extracellular RNA profiles identified that six EV RNAs including five miRNAs (miR-1343-3p, miR-125a-5p, miR-708-5p, miR-381-3p and miR-543) can differentiate CRC patients from healthy controls with an area under curve (AUC) of 0.76 (sensitivity = 0.72 and specificity = 0.80) [21]. Further, plasma exosomal miR-125a-3p and miR-320c were up-regulated in early stage CRC patients (stage I and II) and they showed significant correlation with nerve infiltration ($P < 0.01$), but not with tumour size, infiltration depth, and differentiation degree ($P > 0.05$) [22].

Recently, we described the isolation of two populations of exosomes (A33-Exos and EpCAM-Exos) as well as sMVs from the CRC cell line LIM1863 and identified 254 cellular miRNAs [23] of which 63 selectively distribute to the EVs–thereby, providing miRNA signatures for these EV subtypes. Interestingly, of 32 miRNAs that selectively distribute to A33-Exos, 13 have not been previously reported in publicly-accessible miRNA databases for human CRC tissue/ biofluids [23]. In addition, we identified 2,389 mRNAs, 317 pseudogene transcripts, 1,028 long noncoding RNAs and 206 short non-coding RNAs selectively distributed to LIM1863 EVs, of which several RNA species are up-regulated in tumour biopsies [24]. Here, using RNA-Seq analysis we compared the miRNA profiles of EVs (Exos and sMVs) secreted from two isogenic

CRC cell lines–primary carcinoma-derived SW480 cells and its lymph node metastatic variant SW620 cells [25]. We sought to compare the miRNA profiles of EVs secreted from SW480/SW620 cells and ascertain whether they reveal a subset of miRNAs critical in the aetiology of CRC and provide potential markers of the disease. Because there is increasing evidence that cancer cell EVs are present in biofluids such as blood, we also sought to investigate whether CRC cell line derived EV-miRNA profiles might provide a useful tool to predict CRC tumours.

Material and methods

Cell culture and exosome isolation

Human CRC cell lines SW480 (CCL-228) and SW620 (CCL-227) were from ATCC (Manassas, VA, USA). Cells were initially cultured in 175-cm^2 flasks in RPMI-1640 medium (Life Technologies, CA) supplemented with 10% (v/v) foetal bovine serum (Thermo Fisher Scientific, CA) and 1% (v/v) Penicillin Streptomycin (Pen/Strep, Thermo Fisher Scientific) at 37˚C and 10% CO_2 and then transferred to CELLine AD-1000 Bioreactor Flasks (Integra Biosciences, NH) and continuous culture performed over a period of several weeks, as described [26]. Cell culture medium (CM) from two separate bioreactor flasks (150 ml) was collected for EV isolation (in duplicate, biological replicates), as described [26]. Briefly, low-speed centrifugation removed cells and cell debris. Then the supernatant was centrifuged at 10,000 × g for 30 min to obtain shed microvesicle (sMVs). The sMV-depleted supernatant was further centrifuged at 100,000 × g for 1 h to isolate crude exosomes. OptiPrep density gradient (100,000 × g 18 h) was employed to obtain purified exosomes. Protein quantification was performed by protein staining (SYPRO Ruby) densitometry following 1D-SDS-PAGE [26].

Cell lysate preparation

Cell lysates were prepared as previously described [26]. Briefly, SW480 and SW620 cells were cultured in 10-cm dishes to 70% confluency. The growth medium was removed and TRIzol reagent (1 mL) added to lyse the cells and the lysate subjected to RNA extraction.

Cryo-electron microscopy

Cryo-EM imaging of EV preparations was performed essentially as described [27, 28], with minor modifications. Briefly, EV preparations (~2 μg protein, non-frozen samples prepared within 2 days of analysis) were transferred onto glow-discharged C-flat holey carbon grids (ProSciTech Pty Ltd). Excess liquid was blotted and grids were plunge-frozen in liquid ethane. Grids were mounted in a Gatan cryoholder (Gatan, Inc., Warrendale, PA, USA) in liquid nitrogen. Images were acquired at 300 kV using a Tecnai G2 F30 (FEI, Eidhoven, NL) in low dose mode. Size distribution of vesicles (range 30–>1000 nm) was calculated using ImageJ for the 12 fields of view (~200 different vesicles for each EV preparation).

Nanoparticle Tracking analysis

EV diameter (size) and concentration was determined using the NanoSight NS300 system (Malvern, UK) equipped with a blue laser (488 nm). Briefly, EVs were diluted in water (~8×10^8 particles/ml) and loaded into a flow-cell top plate using a syringe pump. Three videos (1 min) were recorded for each sample and analysed by NTA software (Build 3.1.45).

Western blotting

Western blot analyses of cell/EV samples (10 μg protein) were performed as described [26]. Briefly, membranes were probed with primary mouse anti-TSG101 (1:500, BD Biosciences, San Jose, CA), mouse anti-Alix (1:1000, Cell Signalling Technology Danvers, MA), mouse anti-CD9 antibody (1:1000, Abcam, Cambridge, MA), mouse anti-KIF23 (1:1000, Thermo Fisher Scientific, CA), and rabbit anti-GAPDH (1:1000, Abcam, Cambridge, UK) in TTBS (Tris buffered-saline containing 0.1% Tween-20) for 1 h. After washing with TTBS (3 × 10 min), membranes were probed with secondary antibody (IRDye 800 goat anti-mouse or IRDye 700 goat anti-rabbit IgG, 1: 15,000). The fluorescent signals were detected using the Odyssey Infrared Imaging System, v3.0 (Li-COR Biosciences, Nebraska, USA).

Total RNA isolation

Total RNA extraction from samples was performed (in duplicate), as described previously [24].

Small RNA library construction and deep sequencing

Small RNA (18~30 nt) libraries for SW480/SW620 cells and derived EVs (sMVs and exosomes) were constructed using Illumina TruSeq Small RNA Sample Preparation Kit v2 and sequenced on a HiSeq 2000 platform (Illumina), according to the manufacturer's protocols. Briefly, small RNAs (18~30 nt) were fractioned on a 15% Tris-borate-EDTA (TBE) polyacrylamide gel (Life Technologies, CA) from total RNA (200 ng), purified by centrifugation, and ligated with adaptors. Small RNAs were then reverse transcribed into cDNAs and amplified using the adaptor primers for 14 cycles. The cDNA fragments (~150 bp) were isolated from a 6% TBE PAGE-gel and directly used for cluster generation by using TruSeq PE Cluster Kit v3 (Illumina). Using TruSeq SBS Kit v3 (Illumina) biological replicates (n = 2) were sequenced for each sample.

Small RNA annotation

Clean reads were obtained for subsequent analyses by removing low-quality, adapter, and short (<18 nt) reads from the raw reads. Clean reads were first aligned to the human reference genome (GRCh38, http://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/) by SOAP2 [29] without mismatch. Then, miRNAs were identified and profiled by mapping the clean reads to human miRNA precursors obtained from miRBase (v21, http://www.mirbase.org). Other noncoding RNA types (rRNA, tRNA, snRNA, snoRNA, srcRNA, srpRNA) in the clean reads were annotated using Rfam and GenBank databases by BLAST software [30]. Repeat-associated small RNAs and mRNA degraded fragments were characterized by mapping clean reads to repeat, exon and intron regions on the human genome. Target prediction and pathway analyses were performed by using Ingenuity Pathway Analysis (IPA) [31].

qRT-PCR validation

Quantitative real-time PCR (qRT-PCR) was used to validate expression levels of 11 candidate miRNAs (let-7a-5p, miR-10a-5p, miR-182-5p, miR-183-5p, miR-192-5p, miR-193b-3p, miR-19b-3p, miR-222-5p, miR-28-3p, miR-339-5p and miR-486-5p) and a housekeeping gene RNU43 was used as the internal control. Briefly, total RNA (100 ng) from cells/ derived-EVs was obtained as described above, transcribed into cDNAs by reverse transcription (RT) primers using iScript Reverse Transcription Supermix for RT-qPCR (Bio-Rad Laboratories Inc., USA). Then, cDNA templates (2 μl), forward and universal reverse primers (2 μl),

1×SsoAdvanced Universal SYBR Green Supermix (10 μl, Bio-Rad Laboratories Inc., USA) and ddH_2O (6 μl) were mixed together to make a super qRT-PCR mix (final concentration of primers: 0.3 μM). Three amplifications for every candidate miRNA in each sample were run on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad Laboratories Inc., USA). All miRNA expression levels were normalized to RNU43 using CFX Manager Software v3.1 (Bio-Rad Laboratories Inc., USA).

Clinical sample miRNA data

Clinical sample miRNA data used in this study were obtained from The Cancer Genome Atlas (TCGA). miRNA expression profile data for normal and CRC tumour tissues were obtained from two CRC projects (TCGA-COAD and TCGA-READ) of TCGA, which recruited miRNA expression profiles of 11 normal colon tissues, and CRC tumours of known pathological stage —103 stage-I, 223 stage-II, 179 stage-III, and 89 stage-IV tumour tissues. Original miRNA counts, and normalized miRNA expression files were downloaded for subsequent analysis.

We also obtained serum and plasma exosomal miRNA profiles of CRC patients from two publicly-available datasets [20, 21]. The serum exosomal miRNA study screened 88 primary CRC patients and 11 healthy controls (HCs) using microarray technology and identified 69 miRNAs that were up-regulated in serum-derived exosomes from CRC patients relative to serum-derived exosomes from healthy controls. The plasma exosomal miRNA study recruited a total of 159 CRC patients from different pathological stages (58 stage I, 46 stage II, 28 stage III and 27 stage IV) and HCs (n = 50).

Statistical analysis

edgeR [32] was used to identify miRNAs enriched in the EVs and dysregulated in the tumour samples in comparison with normal tissues. We first filtered miRNAs expressed no more than 5 transcripts per million reads (TPM). Raw miRNA read counts were then used in edgeR as input for differential expression analysis. Strict criteria were used to select differentially-expressed miRNAs as follows: log2 fold change (log2FC) >1 (for up-regulated miRNAs) or log2FC <-1 (for down-regulated miRNAs), p-value <0.05, and false discovery rate (FDR) <0.05. The Pearson correlation coefficient calculated by 'cor' (an R package) was used to evaluate the correlation between cell/EVs and clinical samples.

Results

Extracellular vesicle isolation, purification and characterization

SW480 and SW620 cells were characterized as previously described [26]. Cells were maintained and cultured in two CELLine AD-1000 Bioreactor classic flasks and 10 collections of culture media (total volume 150 ml) were made over a period of 5 days. As shown in Fig 1A, shed microvesicles (sMVs) and crude exosomes were isolated from the CM by sequential centrifugation. The presence of sMVs in the 10,000 × g pellet was confirmed by the stereotypic sMV marker KIF23 [8] (Fig 1B). The yields of SW480- and SW620-derived sMVs were 1.85 mg and 1.41 mg protein, respectively. Next, the crude exosomes (reconstituted in 500 μl PBS) were further purified by density gradient centrifugation using a 5–40% iodixanol (OptiPrep) density gradient. Exosomes were enriched in factions 6 and 7 (buoyant density 1.08–1.12 μg/ml) based on western blot analysis of the stereotypic exosomal marker proteins Alix, TSG101 and CD9 (Fig 1B). The yields of SW480- and SW620-derived exosomes were 0.83 mg and 0.69 mg protein, respectively. Both cryo-EM and NTA analyses (Fig 1C and 1D) showed the particle size of SW480 and SW620 exosomes to be in the range 40–200 nm (mean size: SW480–119

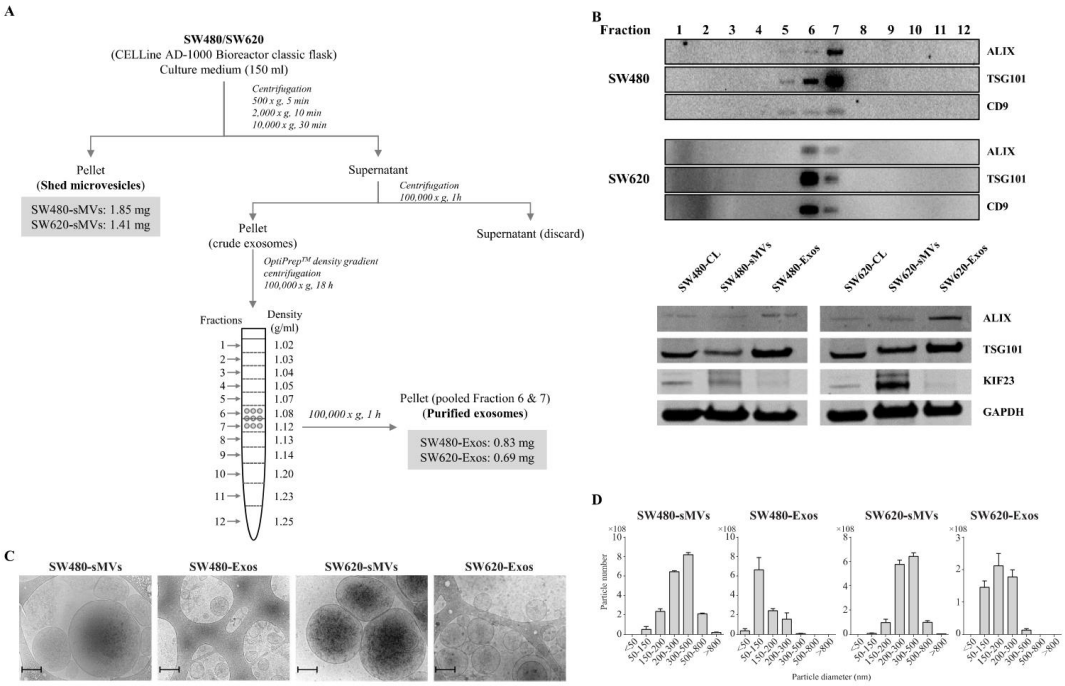


Fig 1. Purification and characterization of sMVs and exosomes secreted from human colorectal cancer cell lines SW480 and SW620. A, A combination of differential centrifugation and density gradient ultracentrifugation (OptiPrep) was used to purify shed microvesicles (sMVs) and exosomes (Exos) in high yield. **B**, Western blot analysis of EV markers. Antibodies for stereotypic exosomal markers (ALIX, TSG101, and CD9) revealed enrichment of exosomes in fractions 6–7 (buoyant density, 1.08–1.12 g/mL) (up panel); antibodies of TSG101, ALIX, KIF23 and GAPDH showed KIF23 enriched in sMVs (below panel). **C**, Cryo-electron microscopy was used to determine median particle diameters of purified sMVs and Exos: SW480-sMVs 266 nm, SW480-Exos—119 nm, SW620-sMVs—486 nm, and SW620-Exos—127 nm (scale bar: 200 nm). Images representative of 2 biological replicates with n = 12 images obtained for each replicate; vesicle diameters determined using ImageJ software. **D**, NanoSight Tracking Analysis was used to calculate size distribution of sMVs and exosomes: SW480-sMVs, 339 nm; SW480-Exos, 138 nm; SW620-sMVs, 327 nm; SW620-Exos, 201 nm (representative of a single biological replicate, with 3 technical replicates performed, and data averaged and merged).

https://doi.org/10.1371/journal.pone.0210003.g001

nm and SW620–127 nm), while those of sMVs were in the range 50–1500 nm (mean size: SW480–266 nm and SW620–486 nm). Total RNA was extracted from SW480 and SW620 cell lysates and their purified EVs using Trizol reagent (Life Technologies, CA). RNA quality analysis, evaluated using an Agilent 2100 Bioanalyzer (S1 Fig), revealed that small RNAs (<50 nt) were enriched in the EVs.

Differential expression of miRNAs in SW480 and SW620 cell lysates

Small RNA sequencing was performed to profile miRNA expression in SW480 and SW620 cell lysates (CLs) and their released EVs. Initially, a total of 160.46 million raw reads were generated by an Illumina HiSeq-2000 system for all the samples, yielding an average 12.77 million clean reads. Clean reads were aligned to miRBase (v21) to profile known human miRNAs in each sample. We used total miRNA mapped reads for normalization purposes, and then filtered lowly-expressed miRNAs (<5 TPM) to obtain a total of 345 miRNAs in SW480 and

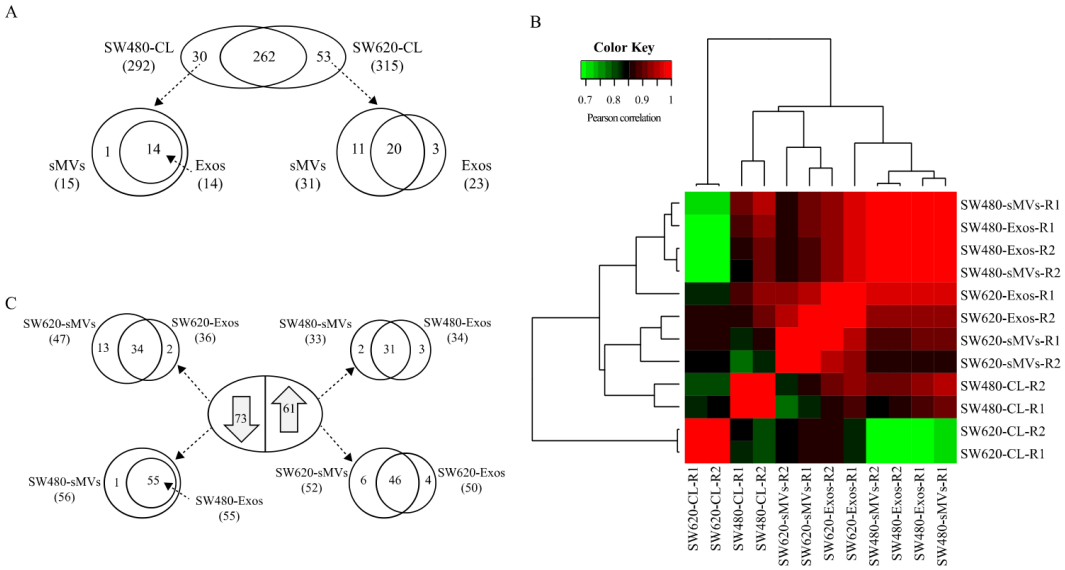


Fig 2. Analysis of cellular miRNAs in human colon cancer cell models. A, Venn diagrams of miRNAs identified in SW480/ SW620 cell lysates (biological replicates combined) and for each cell line the number of unique miRNAs that selectively distribute to secreted EVs: sMVs and Exos. **B**, Correlation coefficient heat map shows limited variability between biological replicates (R1-2) and illustrates differences between miRNA profiles of SW480/SW620-CLs and their secreted EV subtypes. **C**, Differential miRNA expression analysis (log_2 FC >1 or <-1, p-value <0.05, FDR <0.05) in SW480/SW620 cell lines by edgeR revealed 134 dysregulated miRNAs (S2 Table). Of these, 73 are miRNAs down-regulated in SW620-CL and 61 up-regulated, relative to SW480-CL; the Venn diagrams show the number of dysregulated miRNAs that selectively distribute to SW480/SW620 secreted EVs.

https://doi.org/10.1371/journal.pone.0210003.g002

SW620 CLs (S1 Table); 292 and 315 being identified in SW480 and SW620 CLs, respectively. Among these, 102 (29.6%) were identified as passenger (star) miRNAs and 262 cellular miRNAs are common to both cell lines (Fig 2A). A heat map (Fig 2B) of sample correlations, based on their miRNA profiles, indicated sMVs, exosomes and cells contain distinct miRNA signatures (clusters); 30 and 53 miRNAs were found to be present exclusively in SW480 and SW620 CLs, respectively. Of these, the most abundant miRNAs found uniquely in SW480 CLs included *miR-371a-5p*, *miR372-3p*, *miR-373-3p*, and *miR-509-3p*. In the case of SW620 CLs, *miR-1224-5p* and *miR-125b-5p* were the most abundant.

Next, the statistical software package edgeR was used to identify miRNAs differentially expressed in SW480/ SW620 CLs. We identified 134 dysregulated miRNAs (Fig 2C, S2 Table). Of these, 61 SW480 cellular miRNAs were found to be upregulated in SW620 CLs, and 73 downregulated. The most highly-dysregulated miRNAs in SW480/SW620 CLs are listed in Table 1. Most prominent of these are four miRNAs in SW480 CLs (*miR-371a-5p*, *miR-509-3p*, *miR-200c-3p*, and *miR-141-3p*) whose levels are >28-fold higher when compared with SW620 CLs and 6 members of the miR-17~92a cluster (*miR-17-5p*, *miR-18a-5p*, *miR-19a/b-3p*, *miR-20a-5p*, and *miR-92a-1-5p*) that are significantly upregulated in SW620 relative to SW480 CLs (Table 1, S2 Table). Using qRT-PCR, and RNU43 as the internal control, we validated 11 miRNAs in all samples. Then, we normalized the miRNA expression by using SW480-CL as the control. Normalized expression by both qRT-PCR and miRNA sequencing revealed 69.7% of the comparisons to be consistent (S2 Fig). IPA analysis revealed cellular dysregulated miRNAs

Table 1. Selected miRNAs highly expressed in SW480/SW620 cell lines.

Cell line	miRNA[a]	SW480[b]	SW620[b]	Log_2FC[c]	TCGA[d]
SW480	miR-371a-5p*	247.74	0.13	-11.277	
	miR-34c-5p	27.3135	0	-11.13	Down
	miR-342-5p	12.7025	0	-10.021	Down
	miR-342-3p	12.3445	0	-9.977	Down
	miR-372-3p	100.0815	0.134	-9.966	
	miR-509-3p	317.6	0.71	-9.346	
	miR-373-3p	119.141	0.2615	-9.319	
	miR-200c-3p	205.48	9.88	-4.953	Down
	miR-141-3p	1697.01	88.54	-4.842	Up
	miR-203a-3p	619.21	100.72	-3.198	Up
	miR-424-3p*	558.24	133.38	-2.639	
	miR-450b-5p	211.02	60.68	-2.373	Up
	let-7e-5p	4865.66	2220.8	-1.701	Down
	miR-10a-5p	322652.89	161525.68	-1.585	Up
	miR-146b-5p	292.59	146.26	-1.581	Up
	miR-125a-5p	1741.36	881.99	-1.568	Down
	miR-26a-5p	20625.27	10740.42	-1.524	Up
	miR-29a-3p	2196.02	1204.33	-1.445	Up
	miR-423-3p	6145.78	3559.41	-1.366	Down
	miR-21-5p	109349.32	65249.43	-1.327	Up
	miR-96-5p	207.37	131.89	-1.238	Up
	miR-22-3p	15892.02	10135.87	-1.23	Up
	miR-182-5p	82988.6	53721.39	-1.214	Up
	miR-378a-3p	6214.12	4003.12	-1.214	Down
	miR-500a-3p*	979.22	673.06	-1.125	Down
SW620	miR-142-5p	37.23	3579.56	6.006	Up
	miR-142-3p	7.89	498.91	5.395	Up
	miR-181a-3p	17.33	986.11	5.251	Up
	miR-375	626.96	29844.77	4.989	Down
	miR-192-5p	3291.59	111573.13	4.502	Up
	miR-194-5p	50.2	1538.3	4.358	Up
	miR-1224-5p	3.2065	60.325	3.648	
	miR-100-5p	125	2267.91	3.601	
	miR-1247-3p*	30.66	450.09	3.296	Down
	miR-125b-5p	4.009	56.979	3.246	Down
	miR-181a-5p	6098.54	60778.94	2.735	Down
	miR-140-3p	145.57	1190.55	2.45	Down
	miR-181b-5p	1108.63	8695.7	2.387	Down
	miR-151a-5p	474.61	2362.55	1.735	Up
	miR-19a-3p	141.13	652.83	1.629	Up
	miR-19b-3p	481.3	2171.18	1.591	Up
	miR-92a-1-5p	110.98	460.75	1.469	Down
	miR-582-3p	144.06	593.63	1.459	Up
	miR-151a-3p	1035.92	4226.32	1.447	Up
	miR-196a-5p	212.3	858.05	1.431	Up
	miR-486-5p	199.2	676.53	1.182	Down
	miR-17-5p	190.82	643.33	1.172	Up

* are passenger miRNAs.

[a]miRNAs annotated with superscript

[b]Average of normalized expression values from both biological replicates, >5 TPM observed in at least one replicate.

[c]p-value<0.05 and FDR <0.05.

[d]miRNAs dysregulated in CRC tumors compared to normal colon tissues (source, TCGA) according to dbDEMC 2.0; UP–upregulated, DOWN–down regulated.

https://doi.org/10.1371/journal.pone.0210003.t001

were involved in the pathways associated with CRC progression and metastasis, such as 'colorectal cancer metastasis signaling', 'p53 signaling' and 'Wnt/β-catenin signaling' (S3 Fig). Experimentally validated target genes of the cellular dysregulated miRNAs revealed known oncogenes and tumour suppressors from these three pathways, including APC, TP53, KRAS, CD44 and TGFBR2 (S3 Fig).

Cellular miRNAs selectively traffic into sMVs and exosomes

Next, we asked how many of the unique, as well as dysregulated, miRNAs found in SW480/SW620 CLs sort into their cognate EVs. Of the 30 unique miRNAs found in SW480-CLs, 14 distribute to both sMVs and Exos while an additional miRNA (*miR-1262*) uniquely distributes to sMVs. In the case of the 53 unique SW620 CL miRNAs, 31 and 23 were found to distribute into sMVs and Exos, respectively; 20 of these sorts to both EV subtypes (Fig 2A). Concerning those miRNAs in SW480-CL whose expression levels are dysregulated (relative to their expression in SW620-CL), 55/73 of the up-regulated miRNAs are common to both SW480-derived sMVs /Exos, while 46/61 of the down-regulated miRNAs in SW480-CL distribute to both SW620-derived EV subtypes; interestingly, 6 and 4 of the up-regulated miRNAs in SW620-CL were observed to uniquely distribute to SW620-derived sMVs and Exos, respectively (Fig 2C).

Cross cell line comparison of miRNA profiles in sMVs and Exos

In addition to comparing miRNA profiles of SW480/SW620 CLs, a further goal of this study was to compare the miRNA profiles of EVs derived from these isogenic cell lines to ascertain whether they reveal a subset of miRNAs critical in CRC progression from primary carcinoma (SW480 cell line surrogate) to metastasis (SW620 cell line surrogate)–information that might be used as markers to assist in the clinical staging of the disease. To address this question, we performed a cross cell-line comparison of miRNA profiles of the two major EV subtypes isolated from SW480/SW620 cell culture media.

Of the 292 miRNAs (expression levels >5 TPM) in SW480-CL, 227 and 222 were detected in derived sMVs and Exos, respectively (Fig 3A, S3 and S4 Tables). 214 of these EV miRNAs are common to both EV subtypes while 13 and 8 are exclusively detected in derived sMVs and Exos, respectively. edgeR analysis of the 214 common miRNAs revealed 32 /214 to be significantly (log2FC>1) enriched in both EV subtypes; one (*miR-210-3p*) and 15 miRNAs (including the passenger miRNA of *miR-210*, miR-210-5p) are selectively enriched in SW480-sMVs and -Exos, respectively. Next, we applied this analysis to the SW620 cell model. Of the identified 315 SW620-CL miRNAs, a combined total of 276 miRNAs are found in SW620-sMVs (261, S5 Table) and SW620-Exos (242, S6 Table). 227/276 of these miRNAs are common to both EV subtypes; 34 and 15 are unique to SW620-sMVs and -Exos, respectively (Fig 3B). 16/227 miRNAs were significantly enriched in both SW620-sMVs/ -Exos (middle panel of bar chart); while one (left panel of bar chart) and 19 (right panel of bar chart) were enriched only in SW620-sMVs and SW620-Exos, respectively.

Next, we performed a cross cell-line (SW480/SW620) analysis of dysregulated sMVs and Exos miRNAs (see 4-way Venn diagram Fig 3C). It can be seen that a total of 186 miRNAs are common to all sMVs and Exos from these two cell lines (S7 Table). Interestingly, with the exception of miR-365a-3p and miR-365b-3p, 17 miRNAs observed in SW480 EVs are down-regulated in SW620-CLs, relative to SW480-CLs. In the case of the 33 miRNAs detected in SW620 EVs (S8 Table), including 17 SW620-CL up-regulated miRNAs; in addition, 1 (miR-1262), 18 and 8 miRNAs are found exclusively in SW620-sMVs and SW620-Exos, respectively. 6 miRNAs (miR-15b-3p, miR-29c-3p, miR-34a-5p, miR-449c-5p, miR-9-3p, and miR-95-3p)

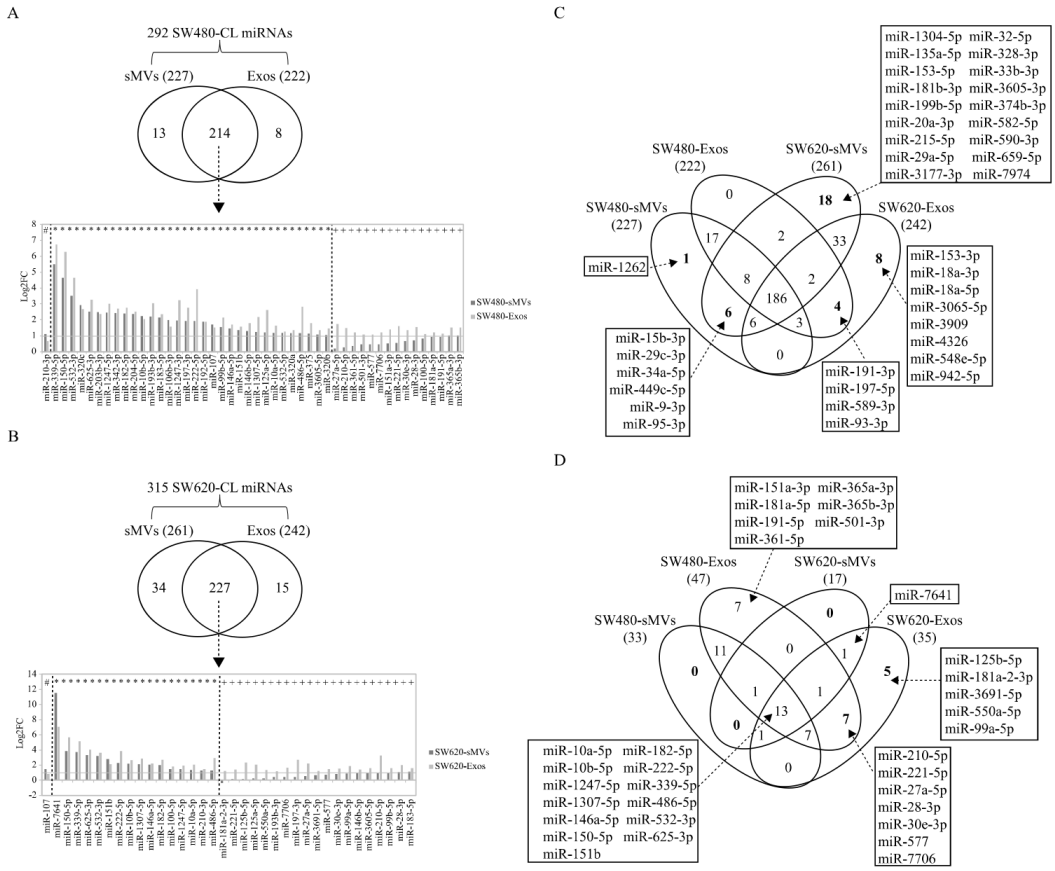


Fig 3. Selective packaging of SW480/SW620 cellular miRNAs into sMVs and exosomes. A, Of the 292 total SW480 cellular miRNAs detected, 227 and 222 distribute to sMVs and Exos, respectively (based on normalized expression >5 TPM); 214 miRNAs are common to both SW480-sMVs and–Exos, of which 48 are enriched according to differential miRNA expression analysis by edgeR (log2FC > 1, p-value <0.05, FDR <0.05); the bar plot below lists 32 common to sMVs and Exos (*), and 1 (#) and 15 (+) miRNAs enriched in sMVs and Exos, respectively. **B**, Of the 315 detected SW620 cellular miRNAs, 261 and 242 were distributed into sMVs and exosomes, respectively (based on normalized expression >5 TPM). 227 cellular miRNAs are common to both sMVs and exosomes. Differential miRNA expression analysis by edgeR (log2FC > 1, p-value <0.05, FDR <0.05) revealed 36 cellular miRNAs are selectively enriched in both EV subtypes; the bar plot below lists 16 miRNAs common to both EV types (*), and 1 (#) and 19 (+) miRNAs selectively enriched in sMVs and Exos, respectively. **C**, Venn diagram showing distribution of 186 miRNAs (S7 Table) that are common to all SW480/SW620-secreted EVs into their respective cell line-derived sMVs and Exos (S3–S6 Tables) and miRNAs identified specifically in each EV subtype. **D**, Venn diagram showing distribution of 13 significantly enriched (log2FC > 1, p-value <0.05, FDR <0.05) miRNAs in all SW480/SW620 EV types (S3–S6 Tables) and miRNAs enriched specifically in each EV subtype.

https://doi.org/10.1371/journal.pone.0210003.g003

are found exclusively in sMVs and 4 miRNAs (miR-191-3p, miR-197-5p, miR-589-3p, miR-93-3p) are unique to Exos.

In another analysis, we examined miRNAs enriched in the four EV subtypes relative to their respective parental CLs (Fig 3D). It can be seen that 13 miRNAs are commonly enriched in all four EV subtypes. Interestingly, only one miRNA (*miR-7641*) is specifically enriched in

both SW620 EV subtypes. Also, there are 7 EV-enriched miRNAs common to exosomes released by the two cell lines, including the only SW480-CL up-regulated miRNA (*miR-210-5p*). Enrichment of *miR-210-3p* (mature miRNA of *miR-210*) is common to SW480-sMVs, SW620-sMVs and SW620-Exos.

Low-abundance cellular SW480/SW620 miRNAs are significantly enriched in EVs

A salient finding in this study was the significant enrichment of many low-abundance cellular miRNAs (<5 TPM) in secreted EVs. In some cases, these non-detectable cellular miRNAs were found in EVs at levels in excess of 1,000 TPM (S4 Fig and S9 Table). In the case of SW480 cells, 35 non-detectable cellular miRNAs were significantly enriched in SW480-derived EVs (9 exclusively found in sMVs, 7 in Exos and 19 common to both EVs). For SW620 cells, 73 non-detectable cellular miRNAs were found to be enriched in EVs from these cells (55 in sMVs, 4 in Exos and 14 common to both EV subtypes). Further examination of these data reveals striking miRNA distribution patterns. For example, 7 miRNAs non-detectable in SW480-CL (miR-1224-*5p*, *miR-125b-5p*, *miR-7641*, *miR-99a-5p*, *miR-1266-5p*, *miR-194-3p*, and *miR-125b-2-3p*), but enriched in SW480-derived EVs (log2FC in the range 1.36 to 7.94), were found to be up-regulated in SW620-CL (log2FC in the range 1.25 to 4.80); of these, *miR-125b-5p*, *miR-7641* and *miR-99a-5p* are also enriched in SW620 EVs (relative to SW620-CL). Another 7 miRNAs that are highly abundant in SW480-CL yet non-detectable in SW620-CL (*miR-6716-3p*, *miR-26a-1-3p*, *miR-203b-3p*, *miR-891a-5p*, *miR-143-3p*, *miR-371a-5p*, and *miR-509-3p*) are significantly enriched in SW620-derived EVs (log2FC in the range 1.03 to 5.40).

Correlation of miRNA profiles of SW480/SW620 cells and derived EVs with colorectal normal/tumour tissues

To investigate the relationship of SW480/SW620 cellular miRNA expression patterns and those reported for primary CRC tumours, we performed a correlation analysis of our cell line miRNA data with CRC tumour miRNA profiles obtained from The Cancer Genome Atlas, TCGA [33]–594 tumours representing 11 normal colon tissues and 103 CRC stage I tumours, 223 stage II, 179 stage III, and 89 stage IV tumours. First, we examined correlations between SW480-CL and SW620-CL and CRC tumours based on 292 and 315 miRNAs, respectively, observed in these cell lines. The mean Pearson correlation coefficients (PCC) for SW480-CL range from 0.05 (normal colon tissue) to 0.41 (for CRC tumours) and, in the case of SW620-CL 0.22 (for normal tissue) to 0.55 for tumours. Based on these correlation coefficients, the miRNA profiles for SW480-CL can be used to predict 592/594 CRC tumours and SW620-CL miRNA profiles, 564/594 (94.9%) of CRC tumours. These predictions improve marginally if the correlations are analysed using the 134 dysregulated miRNAs observed in SW480-/SW620-CLs (see Fig 4A, Table 2)–for example, 593/594 and 572/594 (96.3%) for SW480-/SW620-CLs, respectively. Correlations between SW480/SW620 cell lines and CRC tumours are further illustrated in Fig 4A.

Next, in line with the recognition that functional miRNAs in EVs are potential candidates as stable blood-based biomarkers, we performed PCC analyses between SW480/SW620 cell-derived EVs (sMVs and Exos) and CRC tumours. A total of 97 miRNAs significantly enriched in SW480/SW620 cell-secreted EVs (sMVs and Exos) were used for these analyses (see S3–S6 Tables). As shown in the volcano plots in Fig 4B more than half of the EV-enriched miRNAs were dysregulated in different CRC tumour stages compared to normal colon tissue. The number of miRNAs up-regulated and down-regulated revealed in different CRC tumour stages is given in the Venn diagrams shown in Fig 4C. For the EV-enriched miRNAs, we next calculated their PCC values between normal colon tissue and different tumour stage tissues

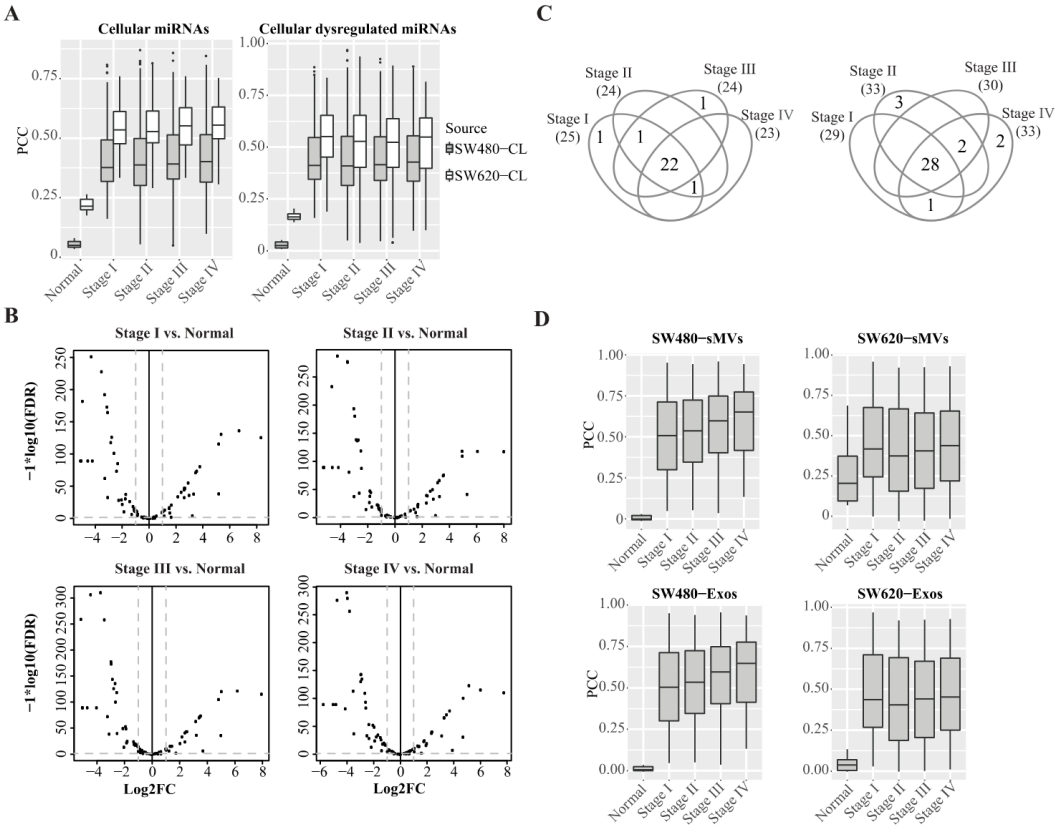


Fig 4. Correlation of miRNAs in SW480/SW620 cell lines and their secreted EVs with miRNAs expressed in non-tumoral colon tissues and CRC tumours. A, Pearson correlation coefficients of SW480/SW620 cellular miRNAs and dysregulated cellular miRNAs with miRNA expression profiles of normal colon tissue (11 samples) and 594 tumor samples from different pathological stages; tissue samples were obtained from TCGA. **B**, Volcano plot for comparing 97 EV-enriched miRNAs (from this study) in different CRC tumor stages with normal colon tissue; x-axis: log2FC, y-axis: -log10 (FDR). The vertical dash lines represent the cut-off levels of log2FCs <-1 (down-regulated miRNAs in tumours, relative to normal colon tissue) or >1 (up-regulated miRNAs in tumours, relative to normal colon tissues), respectfully. The horizontal dash line shows the cut-off for FDR <0.05. **C**, Venn diagrams of the number of up-regulated (left) /down-regulated (right) miRNAs identified in CRC tumours of different pathological stages. **D**, Pearson correlation coefficients of SW480/SW620-EV miRNA profiles with miRNA expression profile of different stage colon tumours and normal colon samples.

https://doi.org/10.1371/journal.pone.0210003.g004

(Table 2). A salient finding of this study is that the mean PCC values between EV miRNA profiles and normal colon tissue miRNA profiles are striking lower than the mean PCC values between cellular miRNA profiles and normal colon tissue miRNA profiles, thereby enhancing CRC tumour prediction accuracy (Fig 4D, Table 2).

Correlation of miRNA profiles of SW480/SW620 cells and derived EVs with CRC patient serum/plasma exosomal miRNA profiles

To further evaluate the potential of SW480/SW620 EV-enriched miRNAs as CRC biomarkers, we compared the 97 miRNAs with previously reported serum/plasma exosomal miRNA

Table 2. Correlation of SW480/SW620 miRNA profiles with normal colon tissue and CRC tumors of different pathological stages.

	Normal colon tissue	Stage I CRC tumors	Stage II CRC tumors	Stage III CRC tumors	Stage IV CRC tumors
SW480-CL[a]	0.034~0.080 (0.054)	0.162~0.808 (0.408)	0.054~0.870 (0.410)	0.050~0.858 (0.412)	0.098~0.845 (0.421)
SW620-CL[a]	0.175~0.265 (0.219)	0.333~0.760 (0.539)	0.290~0.815 (0.543)	0.333~0.760 (0.549)	0.307~0.753 (0.559)
SW480-CL[b]	0.007~0.051 (0.027)	0.158~0.886 (0.447)	0.049~0.969 (0.443)	0.045~0.926 (0.441)	0.096~0.890 (0.453)
SW620-CL[b]	0.138~0.204 (0.167)	0.188~0.836 (0.546)	0.038~0.938 (0.526)	0.040~0.893 (0.514)	0.099~0.816 (0.527)
SW480-sMVs[c]	-0.012~0.028 (0.006)	0.047~0.952 (0.497)	0.052~0.943 (0.521)	0.035~0.959 (0.568)	0.134~0.943 (0.591)
SW480-Exos[c]	-0.006~0.035 (0.011)	0.046~0.950 (0.497)	0.050~0.940 (0.521)	0.036~0.956 (0.568)	0.133~0.936 (0.592)
SW620-sMVs[c]	0.068~0.686 (0.246)	-0.000~0.958 (0.444)	-0.030~0.921 (0.410)	-0.027~0.924 (0.415)	-0.016~0.930 (0.431)
SW620-Exos[c]	-0.004~0.134 (0.045)	0.028~0.970 (0.473)	-0.003~0.920 (0.440)	0.000~0.925 (0.443)	0.010~0.930 (0.461)

[a]PCC values are calculated using miRNAs identified in SW480-CL or SW620-CL; values in brackets are mean values.

[b]PCC values are calculated using miRNAs dysregulated in SW620-CL relative to SW480-CL; values in brackets are mean values.

[c]PCC values are calculated using the 97 miRNAs significantly enriched in the EVs compared to the cell lysates; values in brackets are mean values.

https://doi.org/10.1371/journal.pone.0210003.t002

profiles of CRC patients [20, 21]. The serum exosomal miRNA study screened 88 primary CRC patients and 11 HCs using microarray technology and identified 69 miRNAs were up-regulated in CRC patients than healthy controls. Among them, miR-1280 and miR-1308 were removed from miRBase because they are tRNA fragments of a. Of the remaining, 67 miRNAs, 10, 7, 7 and 6 were selectively enriched in the SW480-Exos, SW480-sMVs, SW620-Exos and SW620-sMVs, respectively (Fig 5A). Four miRNAs (miR-150-5p, miR-1246, miR-766-3p and miR-10b-5p) were commonly enriched in the EVs. In addition, we found that miR-1268a was specifically enriched in SW480-Exos and miR-107 was enriched in only SW480-CL released EVs.

Next, we compared the 97 miRNAs against plasma exosomal miRNA profiles of different CRC stage patients (58 stage I, 46 stage II, 28 stage III and 27 stage IV) and HCs (n = 50). Differential expression analysis (fold change > 1.5 or < -1.5, p-value < 0.05 and FDR < 0.05) showed that 4, 9, 35 and 64 miRNAs were differentially expressed in stage I, II, III and IV, respectively, relative to the HCs (Fig 5B, S10 Table). It is clear that in advanced CRC stage patients more miRNAs were dysregulated in the plasma exosomes. Because 7 miRNAs (miR-1266-5p, miR-3620-3p, miR-4640-3p, miR-4758-3p, miR-5096, miR-6720-5p and miR-6840-3p) were expressed less than 5 TPM in the plasma exosomes, we compared the 92 EV-enriched miRNAs with the 84 differentially expressed miRNAs in CRC plasma exosomes (Table 3). It is shown that among the EV-enriched miRNAs only miR-342-3p, commonly enriched in SW480-CL released EVs, was down-regulated in the plasma exosomes of stage III CRC patients (S10 Table). We found that miR-483-5p was the only miRNA enriched in SW480 EVs and up-regulated in the plasma exosomes of stage I CRC patients (Table 3). The up-regulation of miR-483-5p was also found in stage II and IV CRC patients. While miR-1246 was the only miRNA which was abundant in all four EV subtypes and up-regulated in stage II, III and IV CRC patients, which maybe a broad biomarker for CRC cancer. Several EV-enriched miRNAs were found to be up-regulated only in advanced CRC patients, such as let-7b-3p, miR-27a-5p, miR-182-5p, miR-192-5p and miR-486-5p. The Pearson correlation model was not suitable

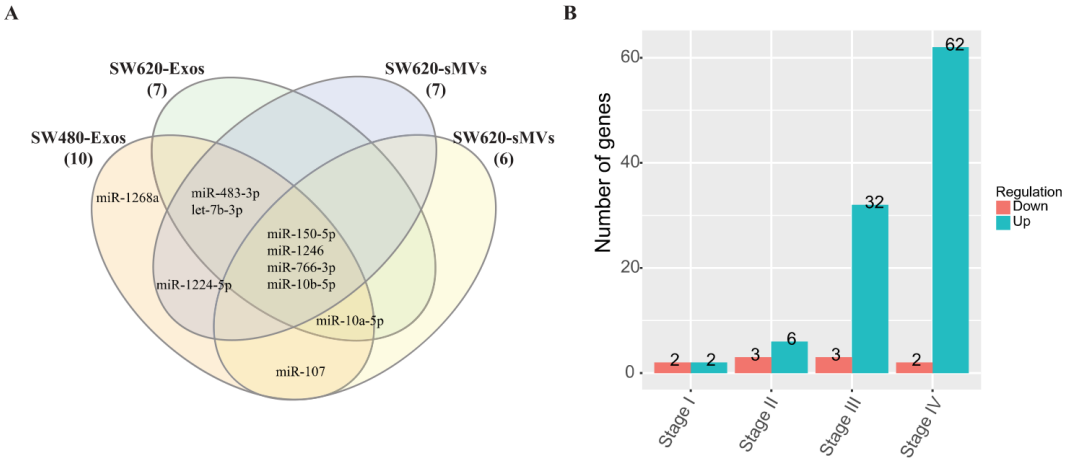


Fig 5. Comparison of miRNA profiles of SW480/SW620 EVs and circulating EVs. A. Venn diagram of miRNAs enriched in SW480/S620 EVs and miRNAs up-regulated in the serum EVs of CRC patients. B. Number of miRNAs enriched in SW480/SW620 EVs that are differentially expressed in the plasma EVs of different CRC stage patients.

https://doi.org/10.1371/journal.pone.0210003.g005

for these two datasets as the linear correlation was very low (< 0.1) between SW480/SW620 EVs and serum/plasma exosomal miRNA profiles.

Discussion

This study was designed to compare miRNA expression profiles of isogenic CRC cell lines SW480 (from primary carcinoma) and SW620 (from lymph node metastasis) [25] and their secreted EVs (sMVs and Exos), with the aim understanding the role of miRNAs in the aetiology of CRC. Further, we examined the suitability of miRNA profiles of SW480/SW620 cell-line-secreted EVs as clinical indicators for predicting CRC.

Table 3. Number of EV-enriched miRNAs up-regulated in exosomes isolated from CRC-patient plasma compared to exosomes from healthy individuals*.

CRC stage	SW480-Exos	SW620-Exos	SW480-sMVs	SW620-sMVs
Stage I	miR-483-5p		miR-483-5p	
Stage II	miR-1246 miR-483-5p	miR-1246	miR-1246 miR-483-5p	miR-1246
Stage III	miR-182-5p miR-1246 miR-486-5p miR-27a-5p	miR-486-5p miR-182-5p miR-1246 miR-27a-5p miR-550a-5p	miR-1246 miR-182-5p	miR-1246 miR-182-5p miR-486-5p
Stage IV	miR-1246 miR-486-5p let-7b-3p miR-192-5p miR-483-5p	miR-486-5p miR-1246 let-7b-3p miR-550a-5p miR-181a-2-3p	miR-1246 let-7b-3p miR-192-5p miR-483-5p	miR-1246 miR-486-5p

*Data from Ogata-Kawata et al [19].

https://doi.org/10.1371/journal.pone.0210003.t003

First, we compared miRNA expression profiles of SW480 and SW620 cells. This study revealed 292 miRNAs in SW480-CLs and 315 miRNAs in SW620-CLs. Of these, 30 and 53 miRNAs were found to be exclusively expressed in SW480- and SW620-CLs, respectively. Using edgeR, a total of 134 cellular SW480/SW620 miRNAs were found to be significantly dysregulated (log2FC>1 or log2FC<-1, p-value <0.05, FDR <0.05) with 61 SW480-CL miRNAs up-regulated in SW620-CL and 73 miRNAs down-regulated (S2 Table). KEGG pathway analysis of the down-regulated miRNAs identified significant (p<0.0.1) enrichment in "Wnt signalling pathway" (ko04310) terms, while the up-regulated miRNAs were enriched in "TGF-beta signalling pathway" (ko04350) terms. These observations are consistent with the biological role of Wnt/TGF-β signalling (involving genes APC, AXIN1/2, GSK3β, the TCF gene family and TP53) in CRC [34, 35]. For example, dysregulation of Wnt signalling is an early event during CRC initiation [35] and dysregulation of TGF-β signalling pathway plays a critical role in cancer metastasis [36]. According to dbDEMC 2.0 [37], a database of differentially expressed miRNAs in all human cancers, all 134 dysregulated miRNAs identified in our study are listed. While *miR-486-3p* and *miR-561-5p* have been previously shown to be dysregulated in lung and kidney cancers (dbDEMC 2.0), the up-regulation of *miR-486-3p* and down-regulation of *miR-561-5p* in SW620-CL is the first report, to our knowledge, of their dysregulation during CRC progression. Interestingly, the known CRC metastasis suppressor miRNAs *miR-141* [38] and *miR-203* [39] were found to be significantly down-regulated in SW620-CL compared to SW480-CL. Notably, another two anti-oncomiRNAs, *miR-34c* and *miR-342*, that operate upstream of Wnt were found to be highly expressed in SW480-CL but not detectable in SW620-CL using a <5 TPM threshold cut off. It has been previously reported that the expression of these two miRNAs in CRC primary tumours is silenced by CpG island methylation of their encoding DNA sequences [40, 41] suggesting that depletion of *miR-34c* and *miR-342* levels might contribute to CRC metastasis. Other miRNAs down-regulated in SW620-CL (compared to SW480-CL) such as *miR-125a* and *miR-10a* also have been reported to suppress metastasis of lung cancer [42] and gastric cancer [43]. Of those oncomiRNAs up-regulated in SW620-CL (relative to SW480-CL), *miR-181a*, *miR17~92a* cluster (*miR-19a/b* and *miR-92a*, *miR-17*, *miR-18a* and *miR-20a*) and *miR-1246* are implicated in CRC development [44–49]. Paradoxically, in addition to these oncomiRNAs, we see several anti-oncomiRNAs up-regulated in SW620-CL (e.g., *miR-100*, *miR-142*, *miR-192* and *miR-194*); controversially, these miRNAs are frequently down-regulated in CRC [50–52].

Second, and in line with the observation that EVs released from cells can localise to blood [19, 53–55], we characterized the miRNA profiles of SW480/SW620-derived sMV and Exos EV subtypes. For both CRC cell lines, many cellular oncomiRNAs /anti-oncomiRNAs (e.g., *miR-10b*, *miR-100*, *miR-192*, *miR-200c*, and *miR-222*) were selectively sorted into EVs (Fig 3A and 3B). In all, 186 miRNAs were detected in SW480/SW620-derived EV subtypes, many being in common and several exclusively distributing to each EV subtype (Fig 3C). Similar to our previous miRNA profiling of three CRC LIM1863 cell line-derived EV populations [23], each SW480/SW620 cell-derived EV subtype has a distinct miRNA portrait (Fig 3C) containing many potential candidate biomarkers that can be explored for CRC diagnosis and pursued in future clinical studies.

Third, our intriguing observation that miRNAs non-detectable (i.e., <5 TPM) in SW480/SW620-CLs can be selectively sorted into sMVs and Exos and actively packaged in high concentration (e.g., > 1000 TPM) (S4 Fig and S9 Table) represents, to the best of our knowledge, the first documentation of such a paradigm. Whether the elimination of certain clusters of cellular miRNAs (e.g., miRNAs that modulate Wnt/β-catenin and TGF-β activities) via EVs represents a vehicle for lowering the cellular concentration of these miRNA clusters, thereby impacting on biological function, is not known. Counterpoint to this speculation is whether

these same EV-containing miRNAs cleared from the parental cell act synergistically to regulate Wnt/β-catenin and TGF-β activities in recipient cell(s) upon EV uptake. Needless to say, the underlying mechanism of this phenomenon warrants investigation. We have shown that miRNA expression patterns vary between parental SW480/SW620 cells and their cognate EVs. These data suggest that specific structural motifs must exist to sort distinct miRNAs into EVs. We therefore performed a motif search in all 371 EV miRNAs identified in this study (S3–S6 Tables, derived from 334 miRNA precursors) that might serve as targeting signals. No global enrichment of specific sequences or motifs was found, including GGAG which is reported to be recognized by hnRNP A2B1 [56]; this motif was identified in 108 miRNAs. However, an interesting finding was a 52.7% enrichment (176/334) of the pre-miRNA motif CUGU reported for the RNA-binding protein adenamotous polyposis coli (APC) [57] which also plays a pivotal role in the canonical Wnt/β-catenin signalling pathway [58].

Fourth, the observation that *miR-7641* is not only up-regulated in SW620-CL, compared to SW480-CL, but also significantly enriched (log2FC in the range 6.1 to 11.5) in both SW480 and SW620 EVs (Fig 3D, S3–S6 Tables) is striking. There are two possible regions in the human genome that can encode *miR-7641*, one the 5S rRNA gene (RNA5SP387), the other in an intergenic region. It has been previously reported that *miR-7641* is a potential biomarker of aging [59] while, in another report, it has been shown to modulate the expression of CXCL1 during endothelial differentiation derived from human embryonic stem cells [60]. According to the dbDEMC database, miR-7641 was down-regulated in blood samples from pancreatic, biliary tract, colon, gastric, oesophageal and liver cancer patients. While the biological function of *miR-7641* remains largely unknown, its significant enrichment (log2FC in the range 6.1–11.5) in SW480- and SW620-EVs makes it an attractive candidate for the diagnosis and prognosis of CRC.

Finally, because of the stability of EV-miRNAs in the circulatory system [61], we reasoned that miRNA profiles of EVs (sMVs and Exos) from the SW480/SW620 cell lines might contain attractive candidate biomarkers for the diagnosis of CRC. Given that there are many differences in mRNA profiling of cancer cell lines and tumours [62, 63] due to the absence of stromal and immune components *in vitro*, we undertook a correlation analysis between miRNA profiles of SW480/SW620 cell line-derived EVs and CRC tumour tissues to gauge the extent that EV miRNA signatures mirror pathological changes in CRC tumours. Our study shows that the mean PCC values between EV miRNA signatures and normal colon miRNA signatures are significantly lower than the mean PCC values between cellular miRNA signatures and normal colon tissue (presumably, due to the striking enrichment in EVs of miRNAs non-detectable in SW480-/SW620-CLs), resulting in improved tumour prediction accuracy using EV miRNA signatures. Our results show that the Pearson correlation coefficient approach based on EV (sMVs and Exos) miRNA signatures has a high sensitivity (94.9%) and specificity (100%) to predict CRC tumours of different stages listed in the TCGA dataset. More CRC cell line-derived EV profiles are required to assess the correlation of miRNA expression in tumours according to pathological stage. Also, some EV-enriched miRNAs are up-regulated in serum/plasma exosomes from CRC patients, compared to HCs, especially in advanced CRC pathological stages (Fig 5). This not only indicates that EVs can transfer regulatory miRNAs that promote cancer progression but also shows EV-containing miRNAs have demonstrated potential CRC biomarkers. The low linear correlation of miRNA profiles in cell-derived EVs and circulating EVs may be due to the different technologies used for miRNA profiling and/or the predicting model needs to be improved.

In summary, there is emerging recognition that EV miRNAs represent promising candidates as stable biomarkers for CRC detection [20], and as prognostic biomarkers for recurrence of CRC [64]. Moreover, the unique properties of EVs provide a potential tool for the

development of drug delivery systems for cancer therapy [65, 66]. It is anticipated that many of the EV-associated miRNAs identified in the present study might represent attractive candidate biomarkers for the diagnosis of CRC and be useful prognostic indicators of this disease.

Supporting information

S1 Fig. Total RNA profile analysis using Agilent 2100 Bioanalyzer.
(PDF)

S2 Fig. qRT-PCR.
(PDF)

S3 Fig. IPA analysis for cellular dysregulated miRNAs.
(PDF)

S4 Fig. Venn diagram of low-abundance cellular miRNAs in the EVs.
(PDF)

S1 Table. 345 miRNAs identified in SW480 and SW620 cell lysates (> 5 TPM).
(XLSX)

S2 Table. Differentially expressed miRNAs between SW480 and SW620 cell lysates.
(XLSX)

S3 Table. miRNAs detected in sMVs of SW480.
(XLSX)

S4 Table. miRNAs detected in exosomes of SW480.
(XLSX)

S5 Table. miRNAs detected in sMVs of SW620.
(XLSX)

S6 Table. miRNAs detected in exosomes of SW620.
(XLSX)

S7 Table. miRNAs detected in all EV subtypes of two CRC cell lines.
(XLSX)

S8 Table. miRNAs specifically distributed into the EVs released by SW620 cells.
(XLSX)

S9 Table. miRNAs > 5 TPM in EVs but < 5 TPM in the cell lysates.
(XLSX)

S10 Table. Differentially expressed genes in plasma exosomes of CRC patients.
(XLSX)

Author Contributions

Conceptualization: Maoshan Chen, Richard J. Simpson.

Data curation: Maoshan Chen, Rong Xu, Richard J. Simpson.

Formal analysis: Maoshan Chen, Wittaya Suwakulsiri, Richard J. Simpson.

Funding acquisition: Richard J. Simpson.

Investigation: Richard J. Simpson.

Methodology: Rong Xu, Alin Rai, Wittaya Suwakulsiri, Keiichi Izumikawa, Hideaki Ishikawa, David W. Greening, Nobuhiro Takahashi, Richard J. Simpson.

Project administration: Richard J. Simpson.

Resources: Richard J. Simpson.

Software: Maoshan Chen.

Supervision: Maoshan Chen, Richard J. Simpson.

Visualization: Maoshan Chen.

Writing – original draft: Maoshan Chen, Richard J. Simpson.

Writing – review & editing: Maoshan Chen, Alin Rai, Richard J. Simpson.

References

1. Xu R, Rai A, Chen M, Suwakulsiri W, Greening DW, Simpson RJ. Extracellular vesicles in cancer—implications for future improvements in cancer care. Nat Rev Clin Oncol. 2018. Epub 2018/05/26. https://doi.org/10.1038/s41571-018-0036-9 PMID: 29795272.
2. Xu R, Greening DW, Zhu HJ, Takahashi N, Simpson RJ. Extracellular vesicle isolation and characterization: toward clinical application. J Clin Invest. 2016; 126(4):1152–62. https://doi.org/10.1172/JCI81129 PMID: 27035807
3. Greening DW, Simpson RJ. Understanding extracellular vesicle diversity—current status. Expert Rev Proteomics. 2018; 15(11):887–910. Epub 2018/10/18. https://doi.org/10.1080/14789450.2018.1537788 PMID: 30326765.
4. Greening DW, Xu R, Gopal SK, Rai A, Simpson RJ. Proteomic insights into extracellular vesicle biology—defining exosomes and shed microvesicles. Expert Rev Proteomics. 2017; 14(1):69–95. 10.1080/14789450.2017.1260450. https://doi.org/10.1080/14789450.2017.1260450 PMID: 27838931.
5. An Q, van Bel AJ, Huckelhoven R. Do plant cells secrete exosomes derived from multivesicular bodies? Plant Signal Behav. 2007; 2(1):4–7. Epub 2007/01/01. PMID: 19704795
6. Deatherage BL, Cookson BT. Membrane vesicle release in bacteria, eukaryotes, and archaea: a conserved yet underappreciated aspect of microbial life. Infect Immun. 2012; 80(6):1948–57. Epub 2012/03/14. https://doi.org/10.1128/IAI.06014-11 PMID: 22409932
7. Simpson RJ, Lim JW, Moritz RL, Mathivanan S. Exosomes: proteomic insights and diagnostic potential. Expert review of proteomics. 2009; 6(3):267–83. Epub 2009/06/06. https://doi.org/10.1586/epr.09.17 PMID: 19489699.
8. Xu R, Greening DW, Rai A, Ji H, Simpson RJ. Highly-purified exosomes and shed microvesicles isolated from the human colon cancer cell line LIM1863 by sequential centrifugal ultrafiltration are biochemically and functionally distinct. Methods. 2015; 87:11–21. https://doi.org/10.1016/j.ymeth.2015.04.008 PMID: 25890246.
9. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004; 116(2):281–97. PMID: 14744438.
10. Ha M, Kim VN. Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol. 2014; 15(8):509–24. https://doi.org/10.1038/nrm3838 PMID: 25027649.
11. Rabinowits G, Gercel-Taylor C, Day JM, Taylor DD, Kloecker GH. Exosomal microRNA: a diagnostic marker for lung cancer. Clin Lung Cancer. 2009; 10(1):42–6. https://doi.org/10.3816/CLC.2009.n.006 PMID: 19289371.
12. Pegtel DM, Cosmopoulos K, Thorley-Lawson DA, van Eijndhoven MA, Hopmans ES, Lindenberg JL, et al. Functional delivery of viral miRNAs via exosomes. Proc Natl Acad Sci U S A. 2010; 107(14):6328–33. https://doi.org/10.1073/pnas.0914843107 PMID: 20304794
13. Le MT, Hamar P, Guo C, Basar E, Perdigao-Henriques R, Balaj L, et al. miR-200-containing extracellular vesicles promote breast cancer cell metastasis. J Clin Invest. 2014; 124(12):5109–28. https://doi.org/10.1172/JCI75695 PMID: 25401471
14. He WA, Calore F, Londhe P, Canella A, Guttridge DC, Croce CM. Microvesicles containing miRNAs promote muscle cell death in cancer cachexia via TLR7. Proc Natl Acad Sci U S A. 2014; 111(12):4525–9. Epub 2014/03/13. https://doi.org/10.1073/pnas.1402714111 PMID: 24616506
15. Xu R, Rai A, Chen M, Suwakulsiri W, Green DR, Simpson RJ. Extracellular vesicles in cancer—implications for future improvements in cancer care. Nature Reviews Clinical Oncology. 018.

16. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin. 2016; 66(1):7–30. https://doi.org/10.3322/caac.21332 PMID: 26742998.
17. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin. 2015; 65(2):87–108. https://doi.org/10.3322/caac.21262 PMID: 25651787.
18. Chen X, Ba Y, Ma L, Cai X, Yin Y, Wang K, et al. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. Cell Res. 2008; 18(10):997–1006. https://doi.org/10.1038/cr.2008.282 PMID: 18766170.
19. Huang Z, Huang D, Ni S, Peng Z, Sheng W, Du X. Plasma microRNAs are promising novel biomarkers for early detection of colorectal cancer. International journal of cancer Journal international du cancer. 2010; 127(1):118–26. https://doi.org/10.1002/ijc.25007 PMID: 19876917.
20. Ogata-Kawata H, Izumiya M, Kurioka D, Honma Y, Yamada Y, Furuta K, et al. Circulating exosomal microRNAs as biomarkers of colon cancer. PLoS One. 2014; 9(4):e92921. https://doi.org/10.1371/journal.pone.0092921 PMID: 24705249
21. Yuan T, Huang X, Woodcock M, Du M, Dittmar R, Wang Y, et al. Plasma extracellular RNA profiles in healthy and cancer patients. Sci Rep. 2016; 6:19413. Epub 2016/01/21. https://doi.org/10.1038/srep19413 PMID: 26786760
22. Wang J, Yan F, Zhao Q, Zhan F, Wang R, Wang L, et al. Circulating exosomal miR-125a-3p as a novel biomarker for early-stage colon cancer. Sci Rep. 2017; 7(1):4150. Epub 2017/06/25. https://doi.org/10.1038/s41598-017-04386-1 PMID: 28646161
23. Ji H, Chen M, Greening DW, He W, Rai A, Zhang W, et al. Deep sequencing of RNA from three different extracellular vesicle (EV) subtypes released from the human LIM1863 colon cancer cell line uncovers distinct miRNA-enrichment signatures. PLoS One. 2014; 9(10):e110314. https://doi.org/10.1371/journal.pone.0110314 PMID: 25330373
24. Chen M, Xu R, Ji H, Greening DW, Rai A, Izumikawa K, et al. Transcriptome and long noncoding RNA sequencing of three extracellular vesicle subtypes released from the human colon cancer LIM1863 cell line. Sci Rep. 2016; 6:38397. https://doi.org/10.1038/srep38397 PMID: 27917920.
25. Leibovitz A, Stinson JC, McCombs WB 3rd, McCoy CE, Mazur KC, Mabry ND. Classification of human colorectal adenocarcinoma cell lines. Cancer Res. 1976; 36(12):4562–9. PMID: 1000501.
26. Ji H, Greening DW, Barnes TW, Lim JW, Tauro BJ, Rai A, et al. Proteome profiling of exosomes derived from human primary and metastatic colorectal cancer cells reveal differential expression of key metastatic factors and signal transduction components. Proteomics. 2013; 13(10–11):1672–86. Epub 2013/04/16. https://doi.org/10.1002/pmic.201200562 PMID: 23585443.
27. Tauro BJ, Mathias RA, Greening DW, Gopal SK, Ji H, Kapp EA, et al. Oncogenic H-ras reprograms Madin-Darby canine kidney (MDCK) cell-derived exosomal proteins following epithelial-mesenchymal transition. Mol Cell Proteomics. 2013; 12(8):2148–59. https://doi.org/10.1074/mcp.M112.027086 PMID: 23645497
28. Tauro BJ, Greening DW, Mathias RA, Mathivanan S, Ji H, Simpson RJ. Two distinct populations of exosomes are released from LIM1863 colon carcinoma cell-derived organoids. Mol Cell Proteomics. 2013; 12(3):587–98. https://doi.org/10.1074/mcp.M112.021303 PMID: 23230278
29. Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, et al. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics. 2009; 25(15):1966–7. https://doi.org/10.1093/bioinformatics/btp336 PMID: 19497933.
30. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25(17):3389–402. PMID: 9254694
31. Kramer A, Green J, Pollard J Jr., Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics. 2014; 30(4):523–30. https://doi.org/10.1093/bioinformatics/btt703 PMID: 24336805
32. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26(1):139–40. https://doi.org/10.1093/bioinformatics/btp616 PMID: 19910308
33. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012; 487(7407):330–7. https://doi.org/10.1038/nature11252 PMID: 22810696
34. Bienz M, Clevers H. Linking colorectal cancer to Wnt signaling. Cell. 2000; 103(2):311–20. PMID: 11057903.
35. Markowitz SD, Bertagnolli MM. Molecular origins of cancer: Molecular basis of colorectal cancer. N Engl J Med. 2009; 361(25):2449–60. https://doi.org/10.1056/NEJMra0804588 PMID: 20018966
36. Padua D, Massague J. Roles of TGFbeta in metastasis. Cell Res. 2009; 19(1):89–102. https://doi.org/10.1038/cr.2008.316 PMID: 19050696.

37. Yang Z, Wu L, Wang A, Tang W, Zhao Y, Zhao H, et al. dbDEMC 2.0: updated database of differentially expressed miRNAs in human cancers. Nucleic Acids Res. 2017; 45(D1):D812–D8. https://doi.org/10.1093/nar/gkw1079 PMID: 27899556

38. Feng L, Ma H, Chang L, Zhou X, Wang N, Zhao L, et al. Role of microRNA-141 in colorectal cancer with lymph node metastasis. Exp Ther Med. 2016; 12(5):3405–10. https://doi.org/10.3892/etm.2016.3751 PMID: 27882171

39. Deng B, Wang B, Fang J, Zhu X, Cao Z, Lin Q, et al. MiRNA-203 suppresses cell proliferation, migration and invasion in colorectal cancer via targeting of EIF5A2. Sci Rep. 2016; 6:28301. https://doi.org/10.1038/srep28301 PMID: 27376958

40. Deng G, Kakar S, Kim YS. MicroRNA-124a and microRNA-34b/c are frequently methylated in all histological types of colorectal cancer and polyps, and in the adjacent normal mucosa. Oncol Lett. 2011; 2 (1):175–80. https://doi.org/10.3892/ol.2010.222 PMID: 22870149

41. Grady WM, Parkin RK, Mitchell PS, Lee JH, Kim YH, Tsuchiya KD, et al. Epigenetic silencing of the intronic microRNA hsa-miR-342 and its host gene EVL in colorectal cancer. Oncogene. 2008; 27 (27):3880–8. Epub 2008/02/12. https://doi.org/10.1038/onc.2008.10 PMID: 18264139.

42. Wang G, Mao W, Zheng S, Ye J. Epidermal growth factor receptor-regulated miR-125a-5p—a metastatic inhibitor of lung cancer. FEBS J. 2009; 276(19):5571–8. https://doi.org/10.1111/j.1742-4658.2009.07238.x PMID: 19702827

43. Chen W, Tang Z, Sun Y, Zhang Y, Wang X, Shen Z, et al. miRNA expression profile in primary gastric cancers and paired lymph node metastases indicates that miR-10a plays a role in metastasis from primary gastric cancer to lymph nodes. Exp Ther Med. 2012; 3(2):351–6. https://doi.org/10.3892/etm.2011.411 PMID: 22969895

44. Ji D, Chen Z, Li M, Zhan T, Yao Y, Zhang Z, et al. MicroRNA-181a promotes tumor growth and liver metastasis in colorectal cancer by targeting the tumor suppressor WIF-1. Mol Cancer. 2014; 13:86. https://doi.org/10.1186/1476-4598-13-86 PMID: 24755295

45. Zirvi KA, Keogh JP, Slomiany A, Slomiany BL. Transglutaminase activity in human colorectal carcinomas of differing metastatic potential. Cancer Lett. 1991; 60(1):85–92. PMID: 1680545.

46. Miyoshi N, Ishii H, Mimori K, Tanaka F, Hitora T, Tei M, et al. TGM2 is a novel marker for prognosis and therapeutic target in colorectal cancer. Ann Surg Oncol. 2010; 17(4):967–72. https://doi.org/10.1245/s10434-009-0865-y PMID: 20033322.

47. Cellura D, Pickard K, Quaratino S, Parker H, Strefford JC, Thomas GJ, et al. miR-19-Mediated Inhibition of Transglutaminase-2 Leads to Enhanced Invasion and Metastasis in Colorectal Cancer. Mol Cancer Res. 2015; 13(7):1095–105. https://doi.org/10.1158/1541-7786.MCR-14-0466 PMID: 25934693

48. Zhou T, Zhang G, Liu Z, Xia S, Tian H. Overexpression of miR-92a correlates with tumor metastasis and poor prognosis in patients with colorectal cancer. Int J Colorectal Dis. 2013; 28(1):19–24. https://doi.org/10.1007/s00384-012-1528-1 PMID: 22772712.

49. Wang S, Zeng Y, Zhou JM, Nie SL, Peng Q, Gong J, et al. MicroRNA-1246 promotes growth and metastasis of colorectal cancer cells involving CCNG2 reduction. Mol Med Rep. 2016; 13(1):273–80. https://doi.org/10.3892/mmr.2015.4557 PMID: 26573378.

50. Chiang Y, Song Y, Wang Z, Liu Z, Gao P, Liang J, et al. microRNA-192, -194 and -215 are frequently downregulated in colorectal cancer. Exp Ther Med. 2012; 3(3):560–6. https://doi.org/10.3892/etm.2011.436 PMID: 22969930

51. Shi D, Zhai B, Zheng Y, Ren R, Han M, Wang X. Transcatheter arterial infusion chemotherapy increases expression level of miR-142-5p in stage III colorectal cancer. Indian J Cancer. 2015; 52 Suppl 2:e47–55. https://doi.org/10.4103/0019-509X.172513 PMID: 26728674.

52. Peng H, Luo J, Hao H, Hu J, Xie SK, Ren D, et al. MicroRNA-100 regulates SW620 colorectal cancer cell proliferation and invasion by targeting RAP1B. Oncol Rep. 2014; 31(5):2055–62. https://doi.org/10.3892/or.2014.3075 PMID: 24626817.

53. Wang S, Xiang J, Li Z, Lu S, Hu J, Gao X, et al. A plasma microRNA panel for early detection of colorectal cancer. International journal of cancer Journal international du cancer. 2015; 136(1):152–61. https://doi.org/10.1002/ijc.28136 PMID: 23456911.

54. Hofsli E, Sjursen W, Prestvik WS, Johansen J, Rye M, Trano G, et al. Identification of serum microRNA profiles in colon cancer. British journal of cancer. 2013; 108(8):1712–9. https://doi.org/10.1038/bjc.2013.121 PMID: 23558896

55. Luo X, Stock C, Burwinkel B, Brenner H. Identification and evaluation of plasma microRNAs for early detection of colorectal cancer. PLoS One. 2013; 8(5):e62880. https://doi.org/10.1371/journal.pone.0062880 PMID: 23690963

56. Villarroya-Beltri C, Gutierrez-Vazquez C, Sanchez-Cabo F, Perez-Hernandez D, Vazquez J, Martin-Cofreces N, et al. Sumoylated hnRNPA2B1 controls the sorting of miRNAs into exosomes through

binding to specific motifs. Nature communications. 2013; 4:2980. https://doi.org/10.1038/ncomms3980 PMID: 24356509

57. Preitner N, Quan J, Nowakowski DW, Hancock ML, Shi J, Tcherkezian J, et al. APC is an RNA-binding protein, and its interactome provides a link to neural development and microtubule assembly. Cell. 2014; 158(2):368–82. https://doi.org/10.1016/j.cell.2014.05.042 PMID: 25036633

58. Clevers H, Nusse R. Wnt/beta-catenin signaling and disease. Cell. 2012; 149(6):1192–205. https://doi.org/10.1016/j.cell.2012.05.012 PMID: 22682243.

59. Machida T, Tomofuji T, Ekuni D, Maruyama T, Yoneda T, Kawabata Y, et al. MicroRNAs in Salivary Exosome as Potential Biomarkers of Aging. Int J Mol Sci. 2015; 16(9):21294–309. https://doi.org/10.3390/ijms160921294 PMID: 26370963

60. Yoo JK, Jung HY, Kim CH, Son WS, Kim JK. miR-7641 modulates the expression of CXCL1 during endothelial differentiation derived from human embryonic stem cells. Arch Pharm Res. 2013; 36 (3):353–8. https://doi.org/10.1007/s12272-013-0067-9 PMID: 23444042.

61. Ge Q, Zhou Y, Lu J, Bai Y, Xie X, Lu Z. miRNA in plasma exosome is stable under different storage conditions. Molecules. 2014; 19(2):1568–75. https://doi.org/10.3390/molecules19021568 PMID: 24473213.

62. Domcke S, Sinha R, Levine DA, Sander C, Schultz N. Evaluating cell lines as tumour models by comparison of genomic profiles. Nature communications. 2013; 4:2126. https://doi.org/10.1038/ncomms3126 PMID: 23839242

63. Jiang G, Zhang S, Yazdanparast A, Li M, Pawar AV, Liu Y, et al. Comprehensive comparison of molecular portraits between cell lines and tumors in breast cancer. BMC Genomics. 2016; 17 Suppl 7:525. https://doi.org/10.1186/s12864-016-2911-z PMID: 27556158

64. Matsumura T, Sugimachi K, Iinuma H, Takahashi Y, Kurashige J, Sawada G, et al. Exosomal microRNA in serum is a novel biomarker of recurrence in human colorectal cancer. British journal of cancer. 2015; 113(2):275–81. https://doi.org/10.1038/bjc.2015.201 PMID: 26057451

65. Tominaga N, Yoshioka Y, Ochiya T. A novel platform for cancer therapy using extracellular vesicles. Adv Drug Deliv Rev. 2015; 95:50–5. https://doi.org/10.1016/j.addr.2015.10.002 PMID: 26482189.

66. Vader P, Breakefield XO, Wood MJ. Extracellular vesicles: emerging targets for cancer therapy. Trends in molecular medicine. 2014; 20(7):385–93. https://doi.org/10.1016/j.molmed.2014.03.002 PMID: 24703619

Appendix Article 5

Proteomics
Proteomics and Systems Biology
www.proteomics-journal.com

Exosomes Derived from the Human Primary Colorectal Cancer Cell Line SW480 Orchestrate Fibroblast-Led Cancer Invasion

Alin Rai, David W. Greening, Rong Xu, Wittaya Suwakulsiri, and Richard J. Simpson**

In localized tumors, basement membrane (BM) prevents invasive outgrowth of tumor cells into surrounding tissues. When carcinomas become invasive, cancer cells either degrade BM or reprogram stromal fibroblasts to breach BM barrier and lead invasion of cancer cells into surrounding tissues in a process called fibroblast-led invasion. However, tumor-derived factors orchestrating fibroblast-led invasion remain poorly understood. Here it is shown that although early-stage primary colorectal adenocarcinoma (SW480) cells are themselves unable to invade Matrigel matrix, they secrete exosomes that reprogram normal fibroblasts to acquire de novo capacity to invade matrix and lead invasion of SW480 cells. Strikingly, cancer cells follow leading fibroblasts as collective epithelial-clusters, thereby circumventing need for epithelial to mesenchymal transition, a key event associated with invasion. Moreover, acquisition of pro-invasive phenotype by fibroblasts treated with SW480-derived exosomes relied on exosome-mediated MAPK pathway activation. Mass spectrometry-based protein profiling reveals that cancer exosomes upregulate fibroblasts proteins implicated in focal adhesion (ITGA2/A6/AV, ITGB1/B4/B5, EGFR, CRK), regulators of actin cytoskeleton (RAC1, ARF1, ARPC3, CYFIP1, NCKAP1, ICAM1, ERM complex), and signalling pathways (MAPK, Rap1, RAC1, Ras) important in pro-invasive remodeling of extracellular matrix. Blocking tumor exosome-mediated signaling to fibroblasts therefore represents an attractive therapeutic strategy in restraining tumors by perturbing stroma-driven invasive outgrowth.

1. Introduction

The basal surface of epithelium is lined by a thin layer of extracellular matrix called the basement membrane (BM) that plays an essential role in structuring, polarizing, and compartmentalizing cells.[1,2] Breach of the BM in developing carcinoma in situ is imperative for escape of localized tumor cells into surrounding tissues.[3] Cancer cells can perforate the BM by proteolytic or physical remodeling of the BM matrix.[4] However, accumulating evidence suggests remodeling of the BM can also be mediated by stromal cells.[5–11] An important class of stromal cells that support transition from carcinoma in situ to an invasive stage are cancer-associated fibroblasts (CAFs).[12] For instance, in colon carcinoma in situ CAFs have been shown to physically remodel the BM matrix to generate tracks through which cancer cells could migrate.[11] In squamous cell carcinoma, generation of such tracks involved proteolytic- and force-mediated matrix remodeling, which were sufficient to enable collective cell invasion.[5] Additionally, CAFs have been shown to exert a physical force on cancer cells via heterotypic cell–cell interaction that stimulates their invasion.[13] This process whereby CAFs lead invasion of cancer cells has been termed as "fibroblast-led invasion." Although underlying molecular mechanisms occurring in fibroblasts and cancer cells during fibroblast-led invasion are beginning to emerge,[5,11,13] tumor-derived factors that trigger and orchestrate this process remains poorly understood.

Tumor-derived exosomes are important mediators of intercellular communication in the tumor microenvironment (TME).[14] They contain oncogenic proteins,[15] RNA species,[16,17] mutated DNA fragments,[18,19] and signaling molecules,[20] which can be functionally transferred to non-cancer cells within the TME to support tumor growth, invasion. and metastasis.[14] For instance, tumor-derived exosomes can transfer TGF-β to resting fibroblasts triggering their differentiation into CAF-like state that supported angiogenesis, and tumor growth.[21,22] Recently, exosomes from mutant p53 expressing tumor cells were shown to regulate integrin trafficking in normal fibroblasts resulting in increased deposition of pro-invasive extracellular matrix (ECM) that supported tumor cell migration and invasion.[23] We previously reported that exosomes from early stage colon adenocarcinoma were enriched in signal transduction molecules that can potentially remodel the TME.[24] Recently, we extended this

Dr. A. Rai, Dr. D. W. Greening, Dr. R. Xu, W. Suwakulsiri, Prof. R. J. Simpson
Department of Biochemistry and Genetics
La Trobe Institute for Molecular Science
La Trobe University
Melbourne, Victoria 3086, Australia
E-mail: alin.rai@baker.edu.au; richard.simpson@latrobe.edu.au
Dr. A. Rai, Dr. D. W. Greening
Baker Heart and Diabetes Institute
Melbourne, Victoria 3004, Australia

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/pmic.202000016

DOI: 10.1002/pmic.202000016

observation by showing that exosomes from early stage colon adenocarcinoma cells (SW480) and its lymph node metastatic variant (SW620 cells) were capable of activating quiescent fibroblasts toward heterogeneous population of CAF-like state that displayed anchorage independent growth, pro-proliferative, pro-angiogenic, and invasive capacities.[25]

Here, we show normal colon fibroblasts (CRL1541) when treated with exosomes derived from SW480 cells acquire an invasive phenotype. These fibroblasts in turn led invasion of SW480 cells into Matrigel matrix as collective epithelial cell clusters, thereby circumventing the need for epithelial to mesenchymal transition, a process often associated with acquisition of invasive phenotype.[26,27] Acquisition of this pro-invasive phenotype relied on MAPK activation. Moreover, mass spectrometry-based protein profiling of SW480-exosome treated CRL1541 fibroblasts identified upregulation of key ECM remodeling pathway proteins. Collectively, our in vitro findings support the role exosomes play in orchestrating fibroblast-led invasion of cancer cells.

Significance Statement

Understanding how localized tumors breach the basement membrane and invade into surrounding tissues is critical in controlling tumor outgrowth and metastasis. Here we show that cancer cells release exosomes that reprogram fibroblast to lead invasion of cancer cells across extracellular matrix. Using mass spectrometry-based proteomics, we provide molecular basis underlying pro-invasive reprogramming of the fibroblasts by cancer exosomes during this process. These findings will provide molecular targets that will enable design of novel ways to restrain tumors by targeting fibroblast-driven invasion.

2. Results

2.1. SW480 Exosome Isolation and Characterization

SW480 cells grown in culture display typical epithelial characteristics[28,29] as evidenced by their tight cell–cell junction (**Figure 1**A), expression of the junction protein E-cadherin (Figure 1B), and inability to invade across extracellular matrix as evidenced in the transwell-Matrigel assay (Figure 1C). Large-scale production of highly-purified exosomes (SW480-Exos) was obtained from SW480 cells grown in continuous culture using the strategy described previously[24,25] (Figure 1D). Consistent with our previous reports, purified exosomes expressed stereotypic marker proteins ALIX and TSG101 (Figure 1E), were spherical in shape (Figure 1F) and ≈150 nm in size (Figure 1G).

2.2. Exosomes from Early-Stage Cancer Cells (SW480) Are Taken Up by Colon Fibroblasts

In this study we first questioned whether SW480-Exos could be taken up by fibroblasts isolated from colon tissue of healthy individual (CRL1541). SW480-Exos labeled with lipophilic dye DiI were taken up by CRL1541 fibroblasts (data not shown). To rule out the possibility of dye transfer without exosome uptake per se, we incubated CRL1541 fibroblasts with GFP-labeled exosomes derived from SW480 cells stably expressing GFP-GAP fusion protein (SW480-GAP-GFP)[30] that tags the plasma membrane; this construct enabled GFP tagging of SW480-derived exosomes (SW480-GFP-Exos). SW480-GFP-Exos were isolated from SW480-GAP-GFP cells and purified as described in Section 3.1 (**Figure 2**A). Fluorescence microscopy revealed that GFP-labeled SW480 exosomes were taken up by fibroblasts (Figure 2B).

2.3. Exosomes from Early-Stage Cancer Cells (SW480-Exos) Promote Fibroblast Invasion into Matrigel Matrix

We next assayed normal human fibroblasts (CRL1541) that had been exposed to SW480-Exos for their ability to invade Matrigel matrix using the Iuvo 3D invasion slide (Figure 2C) comprising three ports: a central port containing solidified Matrigel matrix, a right port containing test fibroblasts, and a left port containing medium. CRL1541 cells (stained with DiI) were treated with SW480-Exos for 2 h and then seeded into the right port. Transmigration of SW40-Exos treated CRL1541 fibroblasts into Matrigel (central port) was monitored using fluorescence microscopy. While control CRL1541 fibroblasts (vehicle-treated) remained confined to the matrix boundary (indicated by white broken line), SW480-Exos treated CRL1541 fibroblasts migrated into the Matrigel matrix in a time-dependent manner (Figure 2D,E). In an orthogonal invasive capacity assay, we demonstrate that SW480 exosome-treated fibroblasts (CRL1541 as well as NIH3T3 cells) also acquire invasive capacity (Figure S1, Supporting Information). Moreover, exosomes derived from non-transformed cells (such as fibroblasts or kidney cells (MDCK cells) does not support pro-invasive fibroblast-phenotype (data not shown). This shows that exosomes from early stage human colon cancer cells (SW480) can trigger fibroblast invasion across the ECM.

2.4. SW480 Exosomes Promote Fibroblast-Led Collective Invasion of SW480 Cells

We next questioned whether quiescent fibroblasts treated with cancer exosomes were capable of facilitating SW480 cell invasion into the ECM. To test this in our in vitro Iuvo 3D cellular-based invasion assay, we allowed DiI-labeled CRL1541 cells treated with SW480-Exos to transmigrate into the Matrigel (central port). After 120 h, we washed the right port and then added fluorescently-labeled SW480 GAP-GFP cells (**Figure 3**A). After 48 h it was apparent that SW480 GAP-GFP cells had entered into the matrix "tailing" behind the CRL1541 fibroblasts (Figure 3B). In contrast, CRL1541 fibroblasts treated with PBS vehicle did not support invasion of SW480 GAP-GFP cells into the matrix. Moreover, we also demonstrate that exosomes orchestrate fibroblast-led invasion of SW480 cells using exosomes from CRC cell line SW620 cells (a metastatic variant of SW480 cells) (Figure S2, Supporting Information). Interestingly, SW480 cells that followed the leading CRL1541 fibroblasts displayed cell–cell junction and were entering matrix as an epithelial collective cluster (Figure 3C). This fibroblast-led collective invasion potentially

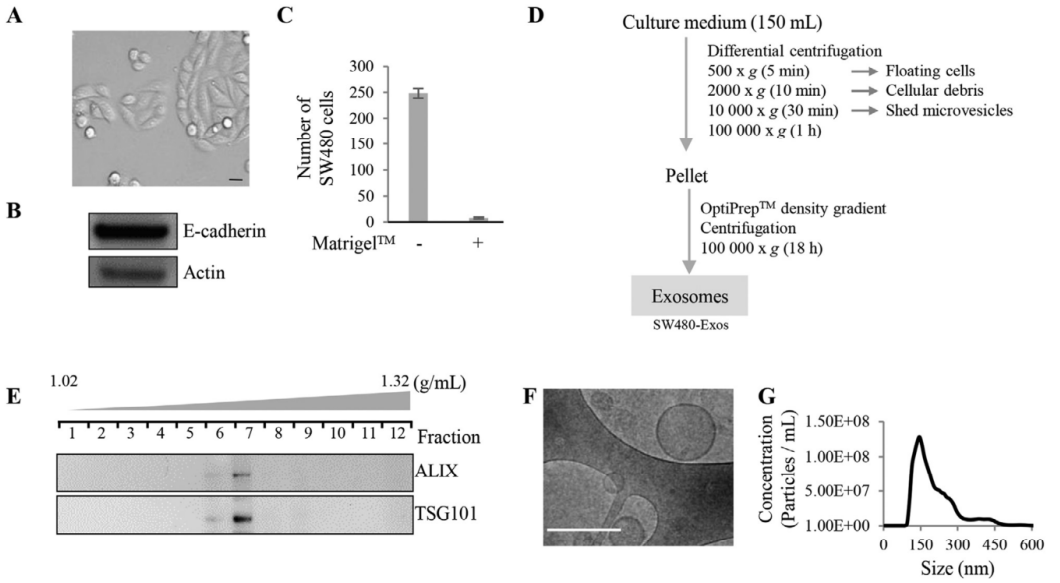


Figure 1. Isolation and characterization of exosomes from SW480 cells. A) Bright-field image of SW480 cells. Scale bar, 10 μm. B) Western blot analysis of SW480 cells using indicated antibodies. C) Transwell-Matrigel matrix invasion assay of SW480 cells ($n > 3$, with five fields of view per insert). Data are presented as average ± SEM (error bars). D) Workflow for isolation of exosomes from culture medium of SW480 cells using a combination of differential centrifugation and buoyant density gradient (iodixanol/OptiPrep) centrifugation. E) Western blot analysis of twelve 1-mL OptiPrep fractions (from Figure 1D) using indicated antibodies for stereotypic exosome marker proteins ALIX and TSG101 ($n > 3$). F) Cryo electron microscopy image SW480-Exos from pooled fractions 6 and 7 (Figure 1E); representative of >10 images, scale bar = 200 nm, and G) Nanoparticle tracking analysis of SW480-Exos, mean size ≈ 150 nm) ($n > 3$).

requires direct local remodeling of matrix (compared to diffusible paracrine signals) by fibroblasts because SW480 GAP-GFP cells added to the left port were unable to invade into the central port Matrigel matrix (data not shown). Of note, DiI-membrane signals from leading CRL1541 fibroblasts could be occasionally observed inside "tailing" SW480-GAP-GFP cells, reminiscent of migrasomes[31] that are potentially released by leading fibroblasts (Figure S3, Supporting Information). Importantly, SW480-Exos alone does not promote SW480 cell invasion across the Matrigel matrix, highlighting the need for stromal-intervention to enable invasive outgrowth (Figure S4, Supporting Information). Thus, our data show that exosomes orchestrate fibroblast-led invasion of SW480 cells that is potentially important in stromal-mediated invasive outgrowth during early stages of cancer.

2.5. Exosome-Mediated Pro-Invasive Fibroblast Phenotype Is Not Dependent on α-SMA Activation

Typically, CAFs at invasive fronts are identified by expression of α-SMA and associated with aggressive disease, greater metastatic burden, and poor clinical outcome.[32,33] However, the role of α-SMA-dependent signaling in pro-invasive phenotype of fibroblasts remains vexed.[34] Interestingly, in our in vitro fibroblast-led invasion model we found the basal level of α-SMA in CRL1541 cells remained unchanged following 16 h exposure to SW480-Exos (Figure S5A, Supporting Information); a finding in contrast to their de novo acquisition of invasive capacity (Figure S5B, Supporting Information). In another experiment NIH3T3 fibroblasts treated with SW480-Exos lack expression of α-SMA but were still able to acquire invasive capacity (Figure S5C,D, Supporting Information). These data suggest that acquisition of pro-invasive capacity by fibroblasts may be uncoupled to α-SMA expression.

2.6. MAPK Signaling Implicated in SW480-Exos Fibroblast-Led Invasion

To gain an insight into the molecular changes in CRL1541 fibroblasts during fibroblast-led invasion, we subjected CRL1541 cells treated with SW480-Exos to GeLC-MS/MS based protein profiling (Table S1, Supporting Information). CRL1541 cells treated with PBS vehicle served as control fibroblasts. Based on stringent informatics, a total of 1387 proteins were identified. For control CRL1541 cells, 1172 proteins were identified, while 1171 proteins were identified in CRL1541 cells treated with SW480-Exos (see **Figure 4**A). A cursory inspection of the data set revealed that 95/1171 proteins were upregulated proteins in CRL1541 treated with SW480 exosomes, which include those that were identified in both biological replicates either uniquely

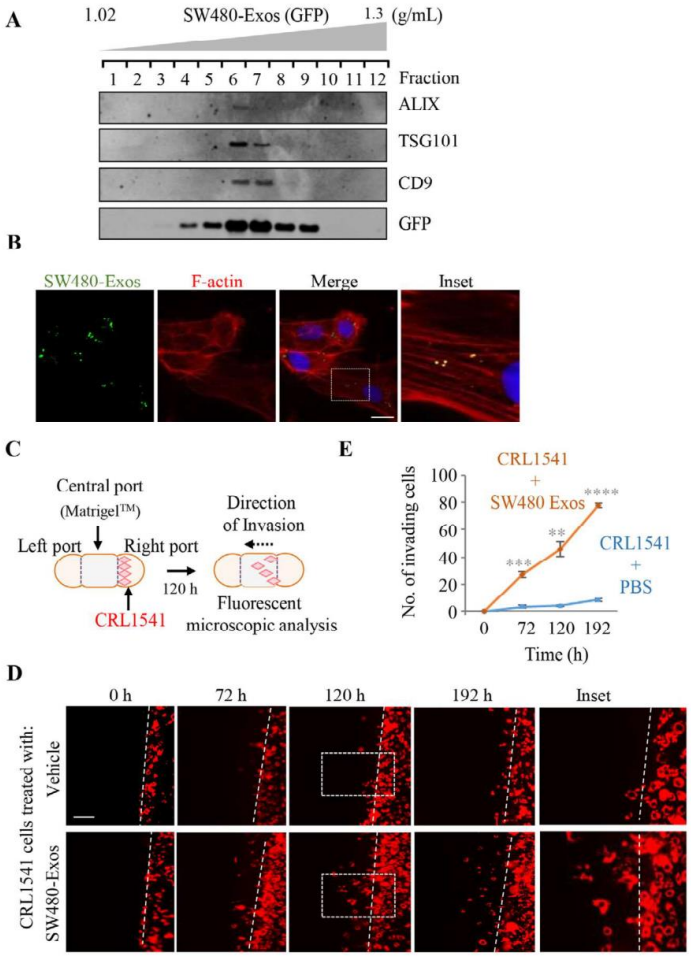


Figure 2. SW480-Exosomes promote normal human colon fibroblast (CRL1541) invasion into Matrigel matrix. A) Western blot analysis of OptiPrep density gradient fractions of SW480 GAP-GFP cell-derived exosomes (SW480-GFP-Exos) ($n > 3$). B) CRL1541 fibroblasts were incubated with SW480-GFP-Exos (green). After 2 h, cells were washed, subjected to immunofluorescence assay (actin filaments immunolabeled with Alexa Fluor 555 Phalloidin (red)) and analyzed by fluorescence microscopy. Hoechst nuclear stain (blue). Scale bar, 10 μm. Right panel: Magnification of boxed region in merged panel ($n = 3$). C) Schematic representation of Transwell Matrigel invasion assay using the Iuvo 3D cellular assay device (50 μm) that consists of three ports. The central port of Iuvo slide device that separates the left and right ports was filled with Matrigel matrix and allowed to solidify for 30 min at 37 °C. DiI-labeled CRL1541 fibroblasts treated with SW480-Exos were added to the right port while the left port was filled with culture medium (DMEM containing 5% FBS, 1% Pen/Strep and supplemented with 30 μg mL^{-1} SW480-Exos or PBS. Fibroblasts were allowed to invade into the Matrigel matrix for 120 h. Cells were then imaged by fluorescence microscopy at indicated time points. D) Fluorescence microscopy analysis of SW480-Exos or PBS treated CRL1541 fibroblast invasion (red) over 0–192 h time period (120 h inset). Broken white lines represent Matrigel boundary ($n > 3$). Scale bar, 100 μm. E) Number of CRL1541 fibroblasts treated with SW480-Exos invading into Matrigel over time ($n > 3$). Data are presented as average ± SEM (error bars, $^{**}p < 0.002$, $^{***}p < 0.0005$, $^{****}p < 0.0001$).

or whose fold change (log 2) ≥ 1.0 compared to CRL1541 control (Table S2, Supporting Information).

We next performed enrichment map analysis of these 95 upregulated proteins to gain insight into enriched pathways and functions (Gene Ontology [GO], KEGG, and Reactome). Enrichment map analysis revealed that these upregulated proteins were implicated in "Regulation of actin cytoskeleton," "Focal adhesion/ECM-receptor interaction," "ERK 1 and ERK2 cascade," and "RAP1 and RAC1 signaling" amongst others (Figure 4B). Importantly, STRING protein–protein interaction network analysis of proteins identified in these individual pathways revealed involvement of components associated with MAPK signaling (Figure 4C; Figure S6, Supporting Information). These include upstream effectors of MAPK signaling

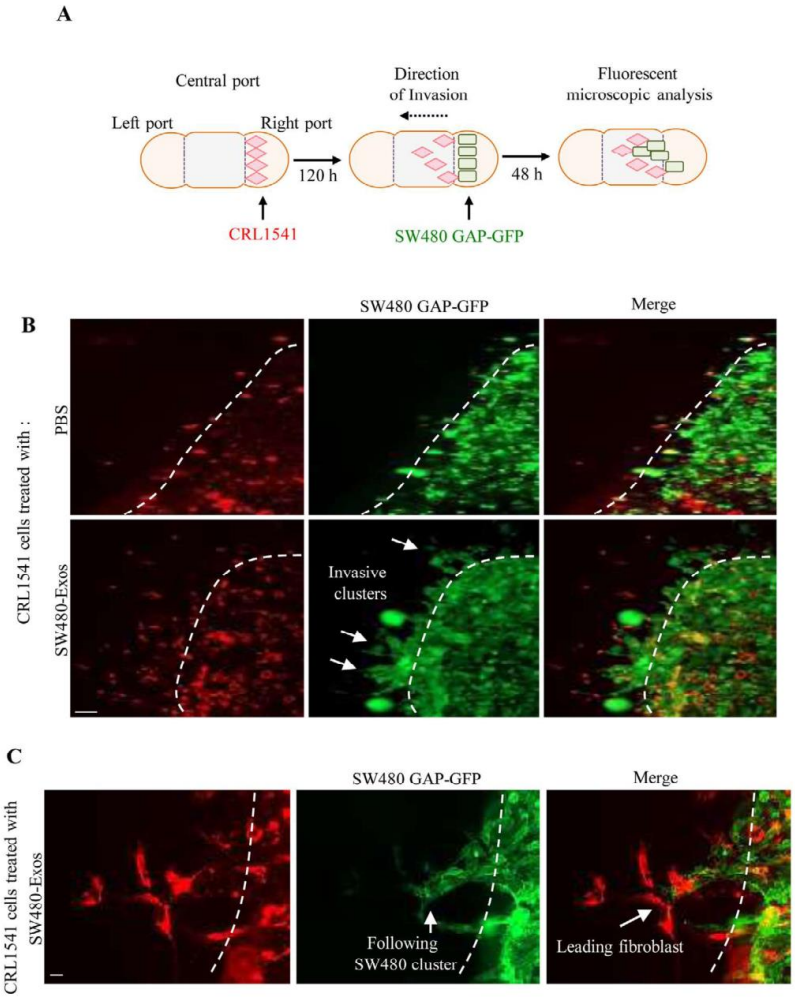


Figure 3. SW480-Exosomes promote fibroblast-led invasion of SW480 cells. A) Schematic representation of fibroblast-led invasion assay using Iuvo 3D cellular assay slide device. CRL1541 fibroblasts were allowed to invade into the Matrigel matrix for 120 h as described in (A). SW480 GAP-GFP cells (green) were then added and the assay continued for a further 48 h and imaged using fluorescence microscopy. B) Broken white lines represent the Matrigel boundary. White arrows indicate regions of SW480 GAP-GFP cell clusters invading into the Matrigel matrix ($n = 3$). Scale bar, 100 μm. C) Fluorescence microscopic analysis reveal that CRL1541 fibroblasts (red) stimulated with SW480-Exos lead invasion of SW480 GAP-GFP as collective epithelial cell clusters (green) ($n = 3$). Scale bar, 10 μm.

which include HRAS and RRAS,[35] EGFR,[36] and EPHA2.[37] Moreover, EGFR[38] can be functionally transferred by exosomes between cells to activate downstream MAPK signaling pathway. The involvement of MAPK signaling in our in vitro fibroblast-led invasion model was further highlighted in global STRING based protein–protein interaction network analysis of all 95 upregulated proteins (Figure 4D). Moreover, invasion of cells through ECM also requires extensive remodeling of cytoskeleton network and integrin rich focal adhesions that cells use to cells anchor to ECM.[39] Given that several proteins implicated in "MAPK signaling" are also implicated in "focal adhesion" and "regulation of actin cytoskeleton," these processes are not mutually exclusive.

Because MAPK pathway has been reported to regulate invasion and metastasis,[40] we questioned whether exposure to SW480-Exos might activate MAPK signaling in CRL1541 fibroblasts and importantly, promote their invasive capacity. Indeed, SW480-Exos activated MAPK signaling in CRL1541 fibroblasts (Figure 4E; Figure S7, Supporting Information). We then questioned whether abrogating MAPK signaling in CRL1541 fibroblasts treated with SW480-Exos interferes with invasive

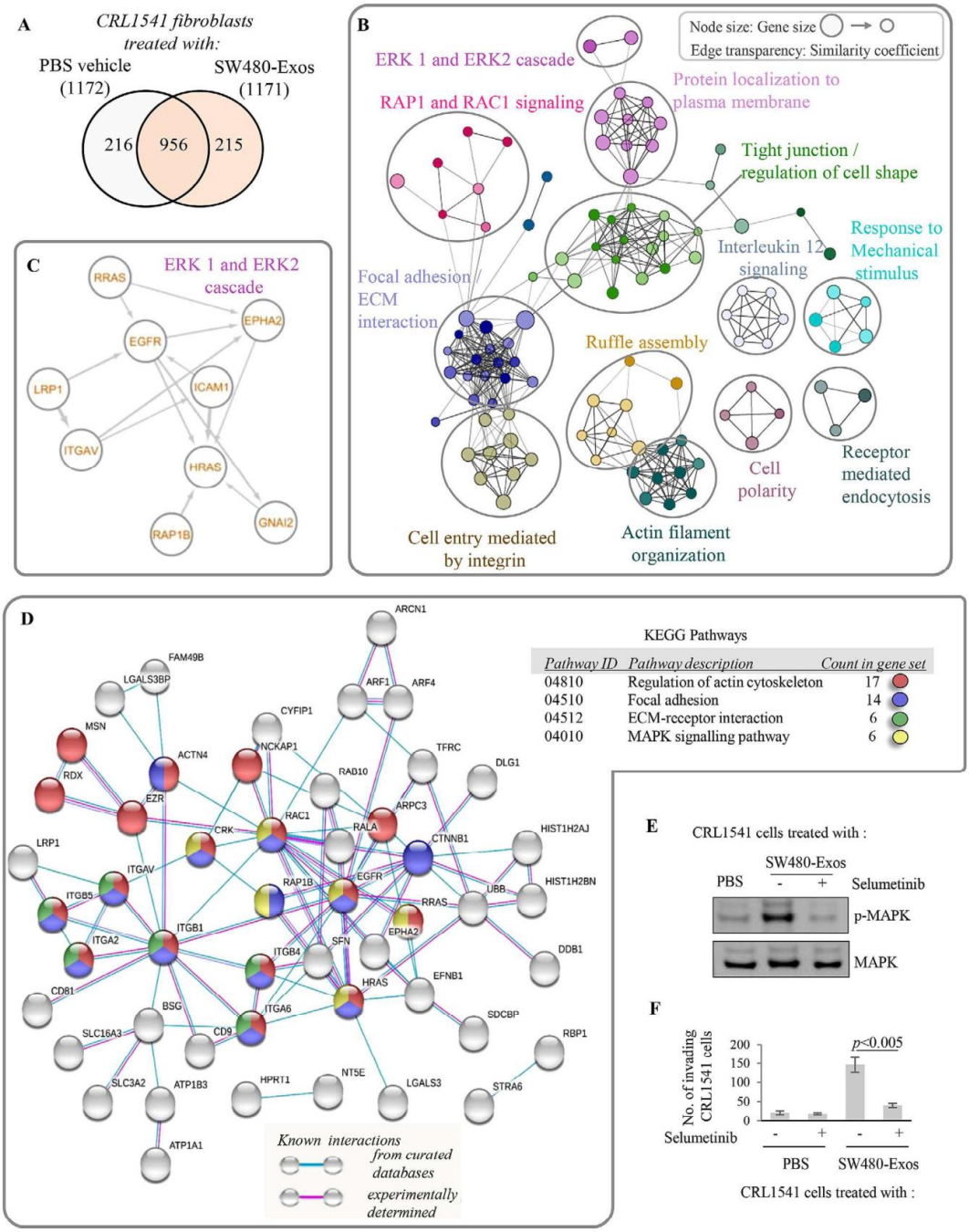


Figure 4. Proteomic analysis of CRL1541 fibroblasts stimulated with SW480-Exos. A) Venn diagram of proteins identified using mass spectrometry analysis in control CRL1541 cells and CRL1541 cells stimulated with SW480-Exos. B) Network visualization of enriched pathways in 95 abundant proteins in CRL1541 stimulated with SW480-Exos compared to control CRL1541 cells using was generated using Enrichment MAP Cytoscape. Functional enrich-

phenotype. For this, we first treated CRL1541 fibroblasts with Selumetinib,[41] a pharmacological inhibition of MAPK signaling pathway, and then stimulated them with SW480-Exos and tested their invasive capacity in Transwell invasion assay (Figure 4F). Selumetinib significantly reduced SW480-Exos-dependent invasive phenotype in CRL1541 fibroblasts (Figure 4F). Thus, our data suggest that exosomes support pro-invasive fibroblast phenotype through MAPK activation.

3. Discussion

Several lines of evidence indicate the presence of CAFs at tumor fronts[42–46] where they lead regional invasion of cancer cells into surrounding tissues.[5,6,8,47] Cancer cells have been shown to trigger activation of different cell types into CAF-like cells through secretion of exosomes.[48–51] In this in vitro model study, we show that exosomes from an early stage colon adenocarcinoma (SW480-Exos) regulate fibroblast-led invasion of cancer cells (SW480) and that this process involves activation of the MAPK pathway and upregulation of key ECM remodeling pathways (see schemata in **Figure 5**). Strikingly, SW480 cells follow the invading fibroblasts as collective clusters (Figure 3) typical of fibroblast-led invasion of cancer cells observed in tumor tissues.[5] Histologically, several carcinomas retain epithelial characteristics but display invasive tumor fronts[5,52] that are spearheaded by CAFs.[45,46] Protein and functional dissection revealed the role of MAPK signaling in acquisition of invasive phenotype in leading fibroblasts (Figure 4). Supporting this observation, MAPK activation was also observed in CAFs in breast cancer tissues.[53] Moreover, MAPK activation at the invasive front was significantly associated with recurrent hepatic colorectal metastasis.[54] CAFs have been shown to display MAPK activation in response to tumor-derived signals in ovarian cancer to support invasion and metastasis[55] and pre-neoplastic keratinocyte cell growth in xenograft experiments.[56] Although the molecular player involved is not known, MAPK activation in fibroblasts could be induced by TGF-β to promote CRC invasion and metastasis.[57] Moreover, TGF-β signaling in CAFs promotes outgrowth of tumor-initiating cells and defines poor prognosis for CRC.[58]

In this study, proteome profiling of SW480-Exos treated CRL1541 fibroblasts also revealed upregulation of proteins implicated in physical remodeling of the matrix (see GO analysis, Figure 4), suggesting physical ECM remodeling might be the primary mode by which fibroblasts support invasion. Cells anchor to the ECM through highly specialized plasma membrane domains consisting of cluster of integrins (called focal adhesion).[59] We found that SW480-Exos treated CRL1541 fibroblasts showed elevated expression of several integrins

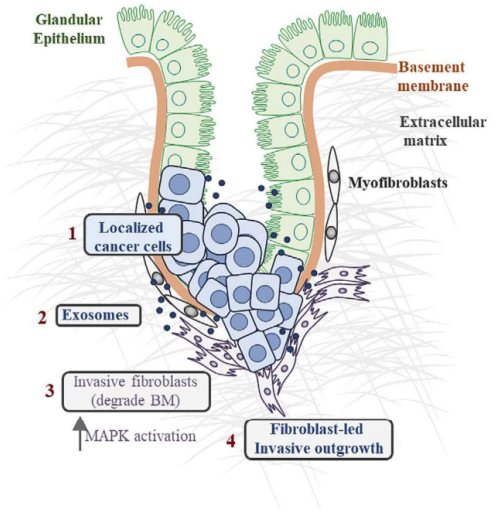


Figure 5. Schemata of cancer exosomes orchestrating fibroblast-led invasion of cancer cells. 1) Cancer cells localized to the confines of ECM 2) release exosomes that are taken up by neighboring fibroblasts resulting in 3) acquisition of invasive capacity that is MAPK-dependent. 4) These fibroblasts in turn lead invasion of cancer cells through ECM to support invasive outgrowth.

(e.g., integrin B1, integrin A2). It has been reported elsewhere that depletion of integrin B1 attenuates matrix remodeling by leading CAFs with reduced number of holes being formed in the matrix.[5] Moreover, assembly of signaling molecules at focal adhesions then regulate Rho GTPases, which has been shown to be important in re-organizing actin cytoskeleton and generation of force required to contract matrix by CAFs when generating tracks in the matrix.[5] Consistent with this observation, CAFs isolated from human colon cancers also expressed proteins implicated in physically remodeling of ECM during fibroblast-led invasion.[11] Moreover, these CAFs were shown to generate holes in the matrix in a metalloproteinase-independent process and rely exclusively on contractile force exerted by CAFs.[11] We also noted that EVs secreted by leading fibroblasts might be taken up by following cancer cells (Figure 3C; Figure S3, Supporting Information). Exosome secretion by cancer cells have been shown to be critical for persistent directional migration, most likely by stabilizing leading edge protrusions of cancer cells.[60] It is possible that leading fibroblasts secrete EVs during their migration—these deposited exosomes might be also used by trailing cancer cells to enable their directional migration through complex tissue

ment analyses (Gene Ontology (GO), KEGG and Reactome pathways) were performed g:Profiler. C) Protein–protein interaction network for proteins in "ERK 1 and ERK2 Cascade" shown in Figure S6, Supporting Information. D) STRING-based protein–protein interaction network analysis of 95 abundant proteins in CRL1541 stimulated with SW480-Exos compared to control CRL1541 cells. The interactions were based on "experiments" and "curated databases" as active interaction source and interaction threshold set at 0.700 (high confidence). Disconnected nodes in the network are hidden. Proteins identified under KEGG pathways terms are indicated. E) Western blot analysis of CRL1541 fibroblasts that were stimulated with FBS or SW480-Exos for 30 min ($n > 3$). F) Transwell-Matrigel invasion assay was used to quantify the effect of Selumetinib (14 nM) on SW480-Exos (30 μg mL^{-1}) ($n \geq 3$). Data are presented as average ± SEM (error bars).

environments. Indeed, stromal fibroblasts are reported to release EVs that promote directional cancer cell migration.[61] These EVs are also reminiscent of migrasomes that are released by migrating cells[62] and are known to be rich in chemokines that provide regional cues during organ morphogenesis.[63]

We currently do not understand how EVs from localized tumors cross the BM barrier and reach the underlying fibroblasts. High-resolution microscopy has revealed that EVs can indeed penetrate dense network of fibrous ECM,[64] however the underlying mechanism of transport is not yet understood. In our previous study we reported that matrix metalloproteinases (MMPs, e.g., membrane-type MMP14/ MT1-MMP), cell surface-anchored sheddases including ADAMS (a disintegrin and metalloproteinases) (e.g., ADAM17 and ADAM10), and an upstream regulator of MMPs (e.g., ECM-inducer EMMPRIN—also known as Basigin or CD147) are packaged in SW480-Exos,[24] it is conceivable that these ECM remodeling enzymes/ modulators, in conjunction to soluble secreted factors, may contribute to localized but limited degradation of ECM that would allow exosomes to pass through the BM. Several studies have also reported that tumor-derived EVs contain ECM-degrading enzymes[65–68] and play an important role in ECM remodeling and tumor progression.[69] However, it should be noted that SW480-Exos themselves were unable to promote SW480 cell invasion; this further highlights the need for stromal intervention. We speculate that due to inherent matrix remodeling capacity of fibroblasts, EVs are able to push them to pro-invasive phenotype, whereas SW480 cells that are of epithelial origin, would potentially require additional reprogramming to drive toward invasive phenotype. Moreover, we have previously shown SW480 soluble secretome also contains various pro-invasive molecules, which include fibronectin, cathepsin B1, niban-like protein 1, serpin A5, and kallikrein-6.[24] Expectedly, soluble secretome from SW480 also conferred fibroblasts with invasive capacity (data not shown). Hence, in vivo, the overall process of invasive outgrowth is highly complex with many alternative compensatory pathways in place, reason why clinically controlling cancer spread remains a challenge.

Of note, we observed that fibroblasts acquire invasive capacity independent of α-SMA expression (Figure S5, Supporting Information), findings consistent with other groups.[34] Although we have recently demonstrated that fibroblasts treated with SW480-Exos for longer periods resulted in eventual activation of α-SMA expression,[25] it appears that acquisition of invasive phenotype by fibroblasts is not totally dependent on this event (Figure S5, Supporting Information). Accumulating evidence shows that within the TME, CAFs are highly heterogeneous and that α-SMA is not expressed by all CAFs.[70–72]

Cancer cells potentially invade through the ECM by employing several mechanisms[73] which can operate exclusively or in concordance, interplay that is often dictated by accumulating oncogenic mutations (due to genomic instability) and evolving tumor niche. The complexity of invasive process is further exacerbated by intra tumor heterogeneity, which can contribute to both individual-cell migration as well as collective invasion within the same tumor.[74] In this regard, in early stages, stromal-intervention potentially plays a crucial role in escaping the BM and contribute to expansion of neoplastic cells.[56,58] However, as the primary tumor (e.g., SW480 cell) evolves and mutations accumulate, new clonal subpopulation (e.g., SW620 cells) that arise may no longer display strict dependency on stromal-led invasion for invasive outgrowth. Indeed, we have shown that SW620 cells derived exosomes also promote pro-invasive phenotype to fibroblasts, albeit to a higher degree compared to SW480-Exos.[25] Moreover, these fibroblasts are also able to lead SW480 cell invasion (Figure S2, Supporting Information). Thus, in SW620 cells which display mesenchymal phenotype and possess matrix-invading capacity, it is conceivable that fibroblast-intervention is not crucial but enhances invasion to support overall invasive outgrowth. Instead, these co-invading CAFs may serve SW620 cells with alternative purpose in cancer spread, for instance, enhancing metastatic colonization through co-dissemination.[75] Unique reprograming of fibroblast proteome by SW620 exosomes compared to SW480 exosomes further highlights this.[25]

Currently, although surgical resection of operable CRC is highly effective treatment modality,[76] accumulating evidence suggests that metastasis is an early event in CRC, preceding clinical detectability.[77] This highlights the need for patient stratification and design of preventive measures.[77] Attempt to restrain tumors in early stages of invasive outgrowth, for example by targeting exosome mediated cancer-stromal signaling (e.g., stromal MAPK activation) in high risk individuals could present an attractive preventive measure.

In conclusion, we demonstrate that exosomes released from primary colon cancer cells orchestrate fibroblast-led invasion of cancer cells through activation of MAPK signaling pathway in fibroblasts. Interfering with tumor exosome-mediated signaling in stromal fibroblasts or abrogating exosome uptake by stromal fibroblasts might represent an attractive therapeutic strategy in restraining tumors to the confines of BM boundaries.

4. Experimental Section

Cell Culture: The human colorectal cancer (CRC) cell line SW480, established from primary colon adenocarcinoma of epithelial origin,[28,29] was commercially obtained from ATTC (CCL-228). Human colon fibroblast cell line derived from a healthy individual (CRL1541) and mouse NIH3T3 fibroblasts were obtained from ATCC. SW480 cells were cultured in RPMI-1640 media (Invitrogen) supplemented with 10% v/v fetal bovine serum (FBS) (Invitrogen), 1% v/v penicillin streptomycin (Pen/Strep, Life Technologies), and maintained at 37 °C with 10% CO_2. Fibroblast cell lines were routinely cultured in DMEM with high-glucose (Invitrogen) at 37 °C with 10% CO_2. Cells were passaged using Trypsin-EDTA (Gibco).

Exosome Isolation and Purification: Exosomes were harvested from SW480 cells cultured in CELLine AD-1000 Bioreactor classic flasks (Integra Biosciences), as described.[78] Cell culture media (CM) were subjected to differential centrifugation (500 × *g*, 5 min; 2000 × *g*, 10 min; 10 000 × *g*, 30 min; and 100 000 × *g*, 1 h) to obtain crude exosomes, which were further purified using isopycnic (iodixanol-density) ultracentrifugation,[24,79] as previously described.[24,80]

Cryo-Electron Microscopy: Exosome morphology was analyzed using cryo-electron microscopy (cryo-EM).[81] SW480-Exos (2 μg) were applied to flow-discharged C-flat holey carbon grids (ProSciTech), and snap-frozen in liquid ethane. Girds were

mounted onto Gatan cryoholder (Gatan Inc) in liquid nitrogen and imaged using Tecnai G2 F30 (FEI, Eidhoven) at 300 kV.

Nanoparticle Tracking Analysis: Particle size distribution of SW480-Exos was determined by nanoparticle tracking analysis (NanoSight NS300, Malvern), as previously described.[82] SW480-Exos sample (1 μg μL^{-1} in PBS (1:10 000 dilution) was analyzed with detection threshold = 10, flowrate = 100, and temperature = 25 °C. Data were analyzed using NTA software 3.0 (ATA Scientific).

Protein Quantification and Western Blotting: Protein quantification was performed using 1D SDS-PAGE / SYPRO Ruby protein staining-based densitometry, as previously described.[83] Western blotting was performed using iBlot Dry 2.0 blotting system (Life Technologies), with primary antibodies (1:1000 dilution); rabbit polyclonal against GAPDH (#2118, Cell Signaling), β-actin (#ab8227, Abcam), GFP (#ab290, Abcam), p44/42 MAPK (#9102, Cell Signaling), phospho-p44/42 MAPK (#9215, Cell Signaling), and mouse monoclonal against E-cadherin (#610181, BD Transduction Laboratories), ALIX (#2171, Cell Signaling), TSG101 (#612696, BD Biosciences), CD9 (#sc-13118, Santa Cruz Biotechnology), and alpha(α)-SMA (#ab7817, Abcam). Secondary antibodies; IRDye 800 goat anti-mouse IgG (#926-32210) or IRDye 680 goat anti-rabbit IgG (#926-68071) (1:15 000, LI-COR Biosciences).

Immunofluorescence Assay: Cells were cultured on glass coverslip to 70% confluency. Immunofluorescence was performed as previously described.[25] Primary antibodies (1:200) (mouse anti-α-SMA (Abcam); secondary antibodies (1:200) Alexa Fluor 488-conjugated goat anti-mouse IgG (Invitrogen). F-actin was labeled with Alexa Fluor 555 Phalloidin (Cell Signaling). Cells were imaged using Zeiss AxioObserver Z1 microscope and images were analyzed using Zen 2011 (Blue edition, Zeiss).

3D Iuvo Invasion Assay: Invasion assays were performed using Iuvo Slide-3D ICC (Bellbrook Labs). Briefly, 2.5 μL of phenol red-free growth-factor-reduced Matrigel matrix (Corning) was added to middle input port and allowed to polymerize at 37 °C for 30 min. CRL1541 cells (3250 cells) stained with DiI were stimulated with SW480-Exos (100 μg mL^{-1}) or vehicle (PBS) alone for 2 h at 37 °C and loaded into the right port. DMEM medium (5% FBS, 1% Pen/Strep, 6.5 μL) supplemented with SW480-Exos (100 μg mL^{-1}) or PBS added to the left port. The Iuvo slide was incubated at 37 °C with 10% CO_2, and maintained at a 30° angle, with medium changed every 24 h. Where indicated, 120 h later, SW480 GAP-GFP cells (650) were added in the right port. Slides were further incubated for 48 h with medium changed every 24 h. Cells were imaged using Zeiss AxioObserver Z1 microscope. Images were quantified using Image J software v1.49e. Where indicated, instead of SW480-Exos, SW620[28,29] cell-derived exosomes (100 μg mL^{-1}), isolated as described for SW480, were used.

Transwell-Matrigel Invasion Assay: Transwell-Matrigel invasion assays were performed using growth factor-reduced Matrigel matrix (Corning), as previously described.[25] Briefly, 8 μm Transwell inserts (Corning) coated with Matrigel were overlaid with fibroblasts (5×10^4) that were pre-treated SW480-Exos (30 μg mL^{-1}) or PBS alone for 2 h at 37 °C. The inserts were nested onto 24-well plate companion plate (Corning) containing DMEM (5% FCS, 1% Pen/Strep) supplemented with SW480-Exos (30 μg mL^{-1}) or PBS alone and incubated overnight (≈16 h) at 37 °C. Inserts were then fixed with formaldehyde, permeabilized (0.2% TX-100), and nuclei stained with Hoechst (10 μg mL^{-1}). Non-invading cells were removed (cotton swab) and nuclei of invasive fibroblasts were imaged using Zeiss AxioObserver Z1 microscope. Where indicated, fibroblasts were treated 14 nM Selumetinib (Sellekchem, AZD6244) or DMSO alone for 30 min before stimulation with SW480-Exos with lower chamber media also supplemented with Selumetinib or DMSO. Data are presented as average ± SEM, with unpaired Student's *t*-test for statistical analyses.

Proteomics: Sample Preparation: Proteomic experiments were performed in biological duplicate (with technical duplicate) using GeLC-MS/MS as previously described.[84,85] Samples (whole cell lysates, protein amount 10 μg) were lysed in SDS sample buffer (SDS, 20% v/v glycerol and 0.01% v/v bromophenol blue, 0.125 M Tris-Hydrochloride (Tris-HCl), pH 6.8), denatured at 95 °C (5 min), and run on SDS-PAGE (150 V, 8 min). Proteins bands (2) were excised, reduced with 10 mM dithiothreitol (DTT) for 30 min at 37 °C followed by alkylation with 50 mM iodoacetamide for 30 min at 37 °C in the dark. Proteins were digested with trypsin (Promega, V5111) at a 1:50 enzyme-to-substrate ratio for 16 h at 37 °C. The peptide mixture was acidified to a final concentration of 2% formic acid, 0.1% trifluoroacetic acid (TFA), and peptides desalted using in-house packed styrene divinylbenzene-reversed-phase sulfonate (SDB-RPS, Empore, Agilent Technologies) microcolumns and eluted with 80% acetonitrile in 2% ammonium hydroxide followed by vacuum concentration. For proteomic analysis, peptides were resuspended in 0.5% acetonitrile, 0.1% TFA.

Liquid Chromatography–Tandem Mass Spectrometry: Peptides were analyzed on a Dionex 3000 nanoUHPLC coupled to a Q-Exactive Orbitrap Elite mass spectrometer equipped with a nanospray ion source in positive mode. Peptides were loaded (Acclaim PepMap100 C18 5 µm beads with 100 Å pore-size, Thermo Fisher Scientific) and separated by a linear gradient for 120 min of 0–100% v/v phase B (0.1% v/v FA in 80% v/v acetonitrile) at a flow rate of 250 nL min^{-1} at 45 °C in a 50-cm fused-silica emitter reversed-phase PepMapRSLC C18, 75 µm inner diameter, 2 µm resin with 100 Å pore-size (Thermo Fisher Scientific). An MS1 scan was acquired from 350–1500 m/z (120 000 resolution, 3×10^6 automatic gain control [AGC], 20 ms injection time) followed by MS/MS data-dependent acquisition (top 20) with collision-induced dissociation and detection in the ion trap (30 000 resolution, 1×10^5 AGC, 60-ms injection time, using a normalized collision energy of 25, and isolation window of 3 Th).[86] Unassigned precursor ions charge states and slightly charged species were rejected and peptide match disabled. Selected sequenced ions were dynamically excluded for 90 s. The mass spectrometry proteomics data deposited to the PeptideAtlas repository with the data set identifier PASS01496; ftp://PASS01496:MD227x@ftp.peptideatlas.org/.

Database Searching and Protein Identification: Peptide identification and quantification were performed using MaxQuant (v1.5.3.30) with its built-in search engine Andromeda.[87] Tandem mass spectra were searched against human-only (UniProt #71785 entries) sequence database (February, 2018) supplemented with common contaminants as described.[88] Search parameters included parent tolerance of 10 ppm, fragment tolerance of 0.5 Da, and minimum peptide length 7, tryptic digestion with up to two missed cleavages, cysteine carbamidomethylation

as fixed modification, and methionine oxidation and protein N-terminal acetylation as variable modifications, and data analyzed and normalized with label-free quantitation (maxLFQ).[89] All data were searched as a single batch and the peptide spectral matches of each database search filtered to 1% FDR with reported proteins containing at least two unique peptide identifications. Contaminants and reverse identifications were excluded from further data analysis. Resulting *p*-values were adjusted by the Benjamini–Hochberg multi-test adjustment method for a high number of comparisons[90] and statistics were performed as previously described.[83]

Data Analysis: Data analysis was done using Microsoft Office Excel and Perseus[91] (Max-Planck Institute of Biochemistry, Martinsried) software. Hierarchical clustering was performed in Perseus using Euclidian distance and average linkage clustering. Proteins upregulated in CRL1541 treated with SW480 exosomes include those that were identified in both biological replicates either uniquely or whose fold change (log 2) $\geq$ 1.0 compared to CRL1541 control. Functional enrichment analyses (GO, KEGG, and Reactome pathways) were performed using g:Profiler.[92] Pathway enrichment map analysis was performed using Cytoscape (v3.7.1).[93] Protein–protein interaction networks (as well as KEGG pathway enrichment) were generated using STRING (v11.0).[94]

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

A.R., D.W.G., R.X., W.S., and R.J.S. acknowledge funding support from La Trobe University, Australia. A.R, R.X, and W.S. acknowledge support by La Trobe University Postgraduate Scholarships. This work was supported in part by National Health and Medical Research Council Project grants #1057741 and #1139489 (D.W.G.) and Helen Amelia Hanis Fellowship (to D.W.G.). The authors acknowledge the La Trobe University Comprehensive Proteomics Platform and Bioimaging Platform for providing infrastructure.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

cancer, cancer-associated fibroblasts, exosomes, fibroblast-led invasion, tumor microenvironment

Received: January 19, 2020
Revised: April 14, 2020

[1] R. G. Rowe, S. J. Weiss, *Trends Cell Biol.* **2008**, *18*, 560.
[2] A. Glentis, V. Gurchenkov, D. M. Vignjevic, *Cell Adhes. Migr.* **2014**, *8*, 236.
[3] L. C. Kelley, L. L. Lohmer, E. J. Hagedorn, D. R. Sherwood, *J. Cell Biol.* **2014**, *204*, 291.
[4] S. Linder, C. Wiesner, M. Himmel, *Annu. Rev. Cell Dev. Biol.* **2011**, *27*, 185.
[5] C. Gaggioli, S. Hooper, C. Hidalgo-Carcedo, R. Grosse, J. F. Marshall, K. Harrington, E. Sahai, *Nat. Cell Biol.* **2007**, *9*, 1392.
[6] S. Neri, G. Ishii, H. Hashimoto, T. Kuwata, K. Nagai, H. Date, A. Ochiai, *Int. J. Cancer* **2015**, *137*, 784.
[7] H. Yamaguchi, N. Yoshida, M. Takanashi, Y. Ito, K. Fukami, K. Yanagihara, M. Yashiro, R. Sakai, *PLoS One* **2014**, *9*, e85485.
[8] S. M. Goicoechea, R. GarcÃ-a-Mata, J. Staub, A. Valdivia, L. Sharek, C. G. McCulloch, R. F. Hwang, R. Urrutia, J. J. Yeh, H. J. Kim, C. A. Otey, *Oncogene* **2014**, *33*, 1265.
[9] J. Li, Z. Jia, J. Kong, F. Zhang, S. Fang, X. Li, W. Li, X. Yang, Y. Luo, B. Lin, T. Liu, *PLoS One* **2016**, *11*, e0150247.
[10] S. Neri, H. Hashimoto, H. Kii, H. Watanabe, K. Masutomi, T. Kuwata, H. Date, M. Tsuboi, K. Goto, A. Ochiai, G. Ishii, *J. Cancer Res. Clin. Oncol.* **2016**, *142*, 437.
[11] A. Glentis, P. Oertle, P. Mariani, A. Chikina, F. E. Marjou, Y. Attieh, F. Zaccarini, M. Lae, D. Loew, F. Dingli, P. Sirven, M. Schoumacher, B. G. Gurchenkov, M. Plodinec, D. M. Vignjevic, *Nat. Commun.* **2017**, *8*, 924.
[12] R. Kalluri, *Nat. Rev. Cancer* **2016**, *16*, 582.
[13] A. Labernadie, T. Kato, A. Brugués, X. Serra-Picamal, S. Derzsi, E. Arwert, A. Weston, V. González-Tarragó, A. Elosegui-Artola, L. Albertazzi, J. Alcaraz, P. Roca-Cusachs, E. Sahai, X. Trepat, *Nat. Cell Biol.* **2017**, *19*, 224.
[14] R. Xu, A. Rai, M. Chen, W. Suwakulsiri, D. W. Greening, R. J. Simpson, *Nat. Rev. Clin. Oncol.* **2018**, *15*, 617.
[15] K. Al-Nedawi, B. Meehan, J. Micallef, V. Lhotak, L. May, A. Guha, J. Rak, *Nat. Cell Biol.* **2008**, *10*, 619.
[16] R. J. A. Nilsson, N. Karachaliou, J. Berenguer, A. Gimenez-Capitan, P. Schellen, C. Teixido, J. Tannous, J. L. Kuiper, E. Drees, M. Grabowska, M. van Keulen, D. A. M. Heideman, E. Thunnissen, A.-M. C. Dingemans, S. Viteri, B. A. Tannous, A. Drozdowskyj, R. Rosell, E. F. Smit, T. Wurdinger, *Oncotarget* **2016**, *7*, 1066.
[17] L. Qu, J. Ding, C. Chen, Z.-J. Wu, B. Liu, Y. Gao, W. Chen, F. Liu, W. Sun, X.-F. Li, X. Wang, Y. Wang, Z.-Y. Xu, L. Gao, Q. Yang, B. Xu, Y.-M. Li, Z.-Y. Fang, Z.-P. Xu, Y. Bao, D.-S. Wu, X. Miao, H.-Y. Sun, Y.-H. Sun, H.-Y. Wang, L.-H. Wang, *Cancer Cell* **2016**, *29*, 653.
[18] K. Allenson, J. Castillo, F. A. S. Lucas, G. Scelo, D. U. Kim, V. Bernard, G. Davis, T. Kumar, M. Katz, M. J. Overman, L. Foretova, E. Fabianova, I. Holcatova, V. Janout, F. Meric-Bernstam, P. Gascoyne, I. Wistuba, G. Varadhachary, P. Brennan, S. Hanash, D. Li, A. Maitra, H. Alvarez, *Ann. Oncol.* **2017**, *28*, 741.
[19] C. Kahlert, S. A. Melo, A. Protopopov, J. Tang, S. Seth, M. Koch, J. Zhang, J. Weitz, L. Chin, A. Futreal, R. Kalluri, *J. Biol. Chem.* **2014**, *289*, 3869.
[20] H. Peinado, M. Alečković, S. Lavotshkin, I. Matei, B. Costa-Silva, G. Moreno-Bueno, M. Hergueta-Redondo, C. Williams, G. García-Santos, C. M. Ghajar, A. Nitadori-Hoshino, C. Hoffman, K. Badal, B. A. Garcia, M. K. Callahan, J. Yuan, V. R. Martins, J. Skog, R. N. Kaplan, M. S. Brady, J. D. Wolchok, P. B. Chapman, Y. Kang, J. Bromberg, D. Lyden, *Nat. Med.* **2012**, *18*, 883.
[21] W. G. Jiang, R. Steadman, J. Wymant, A. T. Jones, H. Kynaston, M. D. Mason, Z. Tabi, A. Clayton, *Oncogene* **2014**, *34*, 290.
[22] J. Webber, R. Steadman, M. D. Mason, Z. Tabi, A. Clayton, *Cancer Res.* **2010**, *70*, 9621.
[23] D. Novo, N. Heath, L. Mitchell, G. Caligiuri, A. MacFarlane, D. Reijmer, L. Charlton, J. Knight, M. Calka, E. McGhee, E. Dornier, D. Sumpton, S. Mason, A. Echard, K. Klinkert, J. Secklehner, F. Kruiswijk, K. Vousden, I. R. Macpherson, K. Blyth, P. Bailey, H. Yin, L. M. Carlin, J. Morton, S. Zanivan, J. C. Norman, *Nat. Commun.* **2018**, *9*, 5069.

[24] H. Ji, D. W. Greening, T. W. Barnes, J. W. Lim, B. J. Tauro, A. Rai, R. Xu, C. Adda, S. Mathivanan, W. Zhao, Y. Xue, T. Xu, H.-J. Zhu, R. J. Simpson, *Proteomics* **2013**, *13*, 1672.
[25] A. Rai, D. W. Greening, M. Chen, R. Xu, H. Ji, R. J. Simpson, *Proteomics* **2019**, *19*, 1800148.
[26] M. Â. A. Nieto, R. Y.-J. Huang, R. A. Jackson, J. P. Thiery, *Cell* **2016**, *166*, 21.
[27] S. Lamouille, J. Xu, R. Derynck, *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 178.
[28] R. E. Hewitt, A. McMarlin, D. Kleiner, R. Wersto, P. Martin, M. Tsoskas, G. W. H. Stamp, W. G. Stetler-Stevenson, *J. Pathol.* **2000**, *192*, 446.
[29] A. Leibovitz, J. C. Stinson, W. B. McCombs, 3rd, C. E. McCoy, K. C. Mazur, N. D. Mabry, *Cancer Res.* **1976**, *36*, 4562.
[30] R. Xu, D. W. Greening, M. Chen, A. Rai, H. Ji, N. Takahashi, R. J. Simpson, *Proteomics* **2019**, *19*, 1700453.
[31] B. da Rocha-Azevedo, S. L. Schmid, *Cell Res.* **2015**, *25*, 1.
[32] S. Madar, I. Goldstein, V. Rotter, *Trends Mol. Med.* **2013**, *19*, 447.
[33] J. G. Parikh, A. Kulkarni, C. Johns, *Oncol. Lett.* **2014**, *7*, 573.
[34] J. Albrengues, I. Bourget, C. Pons, V. Butet, P. Hofman, S. Tartare-Deckert, C. C. Feral, G. Meneguzzi, C. Gaggioli, *Cell Rep.* **2014**, *7*, 1664.
[35] D. K. Simanshu, D. V. Nissley, F. McCormick, *Cell* **2017**, *170*, 17.
[36] N. E. Hynes, H. A. Lane, *Nat. Rev. Cancer* **2005**, *5*, 341.
[37] Y. Hamaoka, M. Negishi, H. Katoh, *Cell. Signal.* **2016**, *28*, 97.
[38] K. Al-Nedawi, B. Meehan, R. S. Kerbel, A. C. Allison, J. Rak, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 3794.
[39] C. T. Mierke, T. Fischer, S. Puder, T. Kunschmann, B. Soetje, W. H. Ziegler, *Sci. Rep.* **2017**, *7*, 42780.
[40] A. S. Dhillon, S. Hagan, O. Rath, W. Kolch, *Oncogene* **2007**, *26*, 3279.
[41] H. Kalra, R. J. Simpson, H. Ji, E. Aikawa, P. Altevogt, P. Askenase, V. C. Bond, F. E. Borràs, X. Breakefield, V. Budnik, E. Buzas, G. Camussi, A. Clayton, E. Cocucci, J. M. Falcon-Perez, S. Gabrielsson, Y. S. Gho, D. Gupta, H. C. Harsha, A. Hendrix, A. F. Hill, J. M. Inal, G. Jenster, E.-M. Krämer-Albers, S. K. Lim, A. Llorente, J. Lötvall, A. Marcilla, L. Mincheva-Nilsson, I. Nazarenko, et al., *PLoS Biol.* **2012**, *10*, e1001450.
[42] C. Pyke, P. Kristensen, E. Ralfkiaer, J. Grondahl-Hansen, J. Eriksen, F. Blasi, K. Dan, *Am. J. Pathol.* **1991**, *138*, 1059.
[43] B. S. Nielsen, F. Rank, J. M. Lopez, M. Balbin, F. Vizoso, L. R. Lund, K. Danø, C. López-Otín, *Cancer Res.* **2001**, *61*, 7091.
[44] F. Xing, *Front. Biosci.* **2010**, *15*, 166.
[45] H. Nakayama, H. Enzan, E. Miyazaki, K. Naruse, H. Kiyoku, M. Hiroi, *Jpn. J. Clin. Oncol.* **1998**, *28*, 615.
[46] M. Martin, P. Pujuguet, F. Martin, *Pathol., Res. Pract.* **1996**, *192*, 712.
[47] R. W. Scott, S. Hooper, D. Crighton, A. Li, I. König, J. Munro, E. Trivier, G. Wickman, P. Morin, D. R. Croft, J. Dawson, L. Machesky, K. I. Anderson, E. A. Sahai, M. F. Olson, *J. Cell Biol.* **2010**, *191*, 169.
[48] J. A. Cho, H. Park, E. H. Lim, K. H. Kim, J. S. Choi, J. H. Lee, J. W. Shin, K. W. Lee, *Gynecol. Oncol.* **2011**, *123*, 379.
[49] J. A. Cho, H. Park, E. H. Lim, K. W. Lee, *Int. J. Oncol.* **2012**, *40*, 1040.
[50] R. Chowdhury, J. P. Webber, M. Gurney, M. D. Mason, Z. Tabi, A. Clayton, *Oncotarget* **2015**, *6*, 715.
[51] J. Gu, H. Qian, L. Shen, X. Zhang, W. Zhu, L. Huang, Y. Yan, F. Mao, C. Zhao, Y. Shi, W. Xu, *PLoS One* **2012**, *7*, e52465.
[52] I. R. Macpherson, S. Hooper, A. Serrels, L. McGarry, B. W. Ozanne, K. Harrington, M. C. Frame, E. Sahai, V. G. Brunton, *Oncogene* **2007**, *26*, 5214.
[53] S. Busch, L. Rydén, O. Stål, K. Jirström, G. Landberg, *PLoS One* **2012**, *7*, e45669.
[54] C.-J. Tai, C.-H. Lee, H.-C. Chen, H.-K. Wang, M.-C. Jiang, T.-C. Su, K.-H. Shen, S.-H. Lin, C.-M. Yeh, C.-J. Chen, K.-T. Yeh, C.-C. Chang, *Ann. Diagn. Pathol.* **2013**, *17*, 165.
[55] M. Curtis, H. A. Kenny, B. Ashcroft, A. Mukherjee, A. Johnson, Y. Zhang, Y. Helou, R. Batlle, X. Liu, N. Gutierrez, X. Gao, S. D Yamada, R. Lastra, A. Montag, N. Ahsan, J. W. Locasale, A. R. Salomon, A. R. Nebreda, E. Lengyel, *Cell Metab.* **2019**, *29*, 141.
[56] E. Alspach, K. C. Flanagan, X. Luo, M. K. Ruhland, H. Huang, E. Pazolli, M. J. Donlin, T. Marsh, D. Piwnica-Worms, J. Monahan, D. V. Novack, S. S. McAllister, S. A. Stewart, *Cancer Discovery* **2014**, *4*, 716.
[57] H. R. Cheruku, A. Mohamedali, D. I. Cantor, S. H. Tan, E. C. Nice, M. S. Baker, *EuPA Open Proteomics* **2015**, *8*, 104.
[58] G. Finak, N. Bertos, F. Pepin, S. Sadekova, M. Souleimanova, H. Zhao, H. Chen, G. Omeroglu, S. Meterissian, A. Omeroglu, M. Hallett, M. Park, *Nat. Med.* **2008**, *14*, 518.
[59] M. A. Schwartz, *Cold Spring Harbor Perspect. Biol.* **2010**, *2*, a005066.
[60] B. H. Sung, T. Ketova, D. Hoshino, A. Zijlstra, A. M. Weaver, *Nat. Commun.* **2015**, *6*, 7164.
[61] V. Luga, L. Zhang, A. M. Viloria-Petit, A. A. Ogunjimi, M. R. Inanlou, E. Chiu, M. Buchanan, A. N. Hosein, M. Basik, J. L. Wrana, *Cell* **2012**, *151*, 1542.
[62] L. Ma, Y. Li, J. Peng, D. Wu, X. Zhao, Y. Cui, L. Chen, X. Yan, Y. Du, Li Yu, *Cell Res.* **2015**, *25*, 24.
[63] D. Jiang, Z. Jiang, Di Lu, X. Wang, H. Liang, J. Zhang, Y. Meng, Y. Li, D. Wu, Y. Huang, Y. Chen, H. Deng, Q. Wu, J. Xiong, A. Meng, Li Yu, *Nat. Cell Biol.* **2019**, *21*, 966.
[64] Y. Chen, A. George, *Front. Physiol.* **2018**, *9*, 1092.
[65] J. Hakulinen, L. Sankkila, N. Sugiyama, K. Lehti, J. Keski-Oja, *J. Cell. Biochem.* **2008**, *105*, 1211.
[66] M. Shimoda, R. Khokha, *Biochim. Biophys. Acta,– Mol. Cell Res.* **2017**, *1864*, 1989.
[67] K. Rilla, A.-M. Mustonen, U. T. Arasu, K. Härkönen, J. Matilainen, P. Nieminen, *Matrix Biol.* **2019**, *75-76*, 201.
[68] R. D. Sanderson, S. K. Bandari, I. Vlodavsky, *Matrix Biol.* **2019**, *75-76*, 160.
[69] M. Nawaz, N. Shah, B. Zanetti, M. Maugeri, R. Silvestre, F. Fatima, L. Neder, H. Valadi, *Cells* 167, **2018**, *7*.
[70] H. Sugimoto, T. M. Mundel, M. W. Kieran, R. Kalluri, *Cancer Biol. Therapy* **2006**, *5*, 1640.
[71] D. Öhlund, A. Handly-Santana, G. Biffi, E. Elyada, A. S. Almeida, M. Ponz-Sarvise, V. Corbo, T. E. Oni, S. A. Hearn, E. J. Lee, I. I. C. Chio, C.-I. Hwang, H. Tiriac, L. A. Baker, D. D. Engle, C. Feig, A. Kultti, M. Egeblad, D. T. Fearon, J. M. Crawford, H. Clevers, Y. Park, D. A. Tuveson, *J. Exp. Med.* **2017**, *214*, 579.
[72] A.-M. Givel, Y. Kieffer, A. Scholer-Dahirel, P. Sirven, M. Cardon, F. Pelon, I. Magagna, G. Gentric, A. Costa, C. Bonneau, V. Mieulet, A. Vincent-Salomon, F. Mechta-Grigoriou, *Nat. Commun.* **2018**, *9*, 1056.
[73] P. Friedl, J. Locker, E. Sahai, J. E. Segall, *Nat. Cell Biol.* **2012**, *14*, 777.
[74] A. Hallou, J. Jennings, A. J. Kabla, *R. Soc. Open Sci.* **2017**, *4*, 161007.
[75] I. Gonzalez-Zubeldia, J. Dotor, M. Redrado, A.-M. Bleau, I. Manrique, A. L. de Aberasturi, M. Villalba, A. Calvo, *Cell Tissue Res.* **2015**, *359*, 829.
[76] E. Huang, M. McGee, *Clinical Algorithms in General Surgery* (Eds: S. Docimo Jr., E. M. Pauli), Springer, Switzerland **2019**, pp. 255–259.
[77] Z. Hu, J. Ding, Z. Ma, R. Sun, J. A. Seoane, J. S. Shaffer, C. J. Suarez, A. S. Berghoff, C. Cremolini, A. Falcone, F. Loupakis, P. Birner, M. Preusser, H.-J. Lenz, C. Curtis, *Nat. Genet.* **2019**, *51*, 1113.
[78] H. Ji, M. Chen, D. W. Greening, W. He, A. Rai, W. Zhang, R. J. Simpson, *PLoS One* **2014**, *9*, e110314.
[79] J. Graham, T. Ford, D. Rickwood, *Anal. Biochem.* **1994**, *220*, 367.
[80] D. W. Greening, R. Xu, H. Ji, B. J. Tauro, R. J. Simpson, *Methods Mol. Biol.* **2015**, *1295*, 179.
[81] R. Xu, D. W. Greening, A. Rai, H. Ji, R. J. Simpson, *Methods* **2015**, *87*, 11.
[82] W. Suwakulsiri, A. Rai, R. Xu, M. Chen, D. W. Greening, R. J. Simpson, *Biochim. Biophys. Acta, Proteins Proteomics* **2019**, *1867*, 140171.

[83] B. J. Tauro, D. W. Greening, R. A. Mathias, S. Mathivanan, H. Ji, R. J. Simpson, *Mol. Cell. Proteomics* **2013**, *12*, 587.
[84] S. K. Gopal, D. W. Greening, R. A. Mathias, H. Ji, A. Rai, M. Chen, H.-J. Zhu, R. J. Simpson, *Oncotarget* **2015**, *6*, 13718.
[85] D. W. Greening, M. Notaras, M. Chen, R. Xu, J. D. Smith, L. Cheng, R. J. Simpson, A. F. Hill, M. van den Buuse, *Mol. Psychiatry* **2019**, In Press.
[86] D. W. Greening, H. P. T. Nguyen, K. Elgass, R. J. Simpson, L. A. Salamonsen, *Biol. Reprod.* **2016**, *94*, 38.
[87] J. Cox, M. Mann, *Nat. Biotechnol.* **2008**, *26*, 1367.
[88] J. Evans, A. Rai, H. P. T. Nguyen, Q. H. Poh, K. Elglass, R. J. Simpson, L. A. Salamonsen, D. W. Greening, *Proteomics* **2019**, *19*, 1800423.
[89] C. A. Luber, J. Cox, H. Lauterbach, B. Fancke, M. Selbach, J. Tschopp, S. Akira, M. Wiegand, H. Hochrein, M. O'Keeffe, M. Mann, *Immunity* **2010**, *32*, 279.
[90] Y. Benjamini, Y. Hochberg, *J. R. Stat. Soc. B* **1995**, *57*, 289.
[91] S. Tyanova, T. Temu, P. Sinitcyn, A. Carlson, M. Y. Hein, T. Geiger, M. Mann, J. Cox, *Nat. Methods* **2016**, *13*, 731.
[92] U. Raudvere, L. Kolberg, I. Kuzmin, T. Arak, P. Adler, H. Peterson, J. Vilo, *Nucleic Acids Res.* **2019**, *47*, W191.
[93] P. Shannon, *Genome Res.* **2003**, *13*, 2498.
[94] D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork, L. J. Jensen, C. V. Mering, *Nucleic Acids Res.* **2019**, *47*, D607.

Copyright permission for reproduced figures in this thesis

Extracellular vesicles in cancer — implications for future improvements in cancer care

Author: Rong Xu et al

Publication: Nature Reviews Clinical Oncology

Publisher: Springer Nature

Date: May 23, 2018

Author Request

Proteomic profiling reveals key cancer progression modulators in shed microvesicles released from isogenic human primary and metastatic colorectal cancer cell lines

Author:
Wittaya Suwakulsiri,Alin Rai,Rong Xu,Maoshan Chen,David W. Greening,Richard J. Simpson

Publication: Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics

Publisher: Elsevier

Date: December 2019

Distinct shed microvesicle and exosome microRNA signatures reveal diagnostic markers for colorectal cancer

Maoshan Chen , Rong Xu, Alin Rai, Wittaya Suwakulsiri, Keiichi Izumikawa, Hideaki Ishikawa, David W. Greening, Nobuhiro Takahashi, Richard J. Simpson

Published: January 4, 2019 • https://doi.org/10.1371/journal.pone.0210003

Abstract

Extracellular vesicle (EV) microRNAs are of major interest as potential diagnostic biomarkers in all cancer types. This study aims to identify miRNA profiles of shed microvesicles (sMVs) and exosomes (Exos) secreted from the isogenic colorectal cancer (CRC) cell lines SW480 and SW620 and evaluate their ability to predict CRC. Deep sequencing of miRNAs in parental cell lysates (CLs) and highly-purified sMVs and Exos was performed. We focused on miRNAs enriched in EVs and dysregulated miRNAs in metastatic cells (SW620) relative to primary cancer cells (SW480). We investigated the ability of EV miRNA signatures to predict CRC tumours using 594 tumours (representing different pathological stages) and 11 normal samples obtained from TCGA. In SW480 and SW620 cells we identified 345 miRNAs, of which 61 and 73 were upregulated and downregulated in SW620-CLs compared to SW480-CLs, respectively. Selective distribution of cellular miRNAs into EVs results in distinct miRNA signatures for sMVs and Exos in each cell line. Cross cell line comparisons of EV miRNA profiles reveal a subset of miRNAs critical in CRC progression from primary carcinoma to metastasis. Many miRNAs non-detectable (<5 TPM) in CLs were significantly enriched (>1000 TPM) in secreted EVs. Strikingly, *miR-7641* which is non-detectable in SW480-CL but upregulated in SW620-CL is highly enriched in EVs secreted from both cell lines. Pearson correlation analysis demonstrated that EV miRNA profiles can be used to predict CRC tumours with ~96% accuracy. Our findings suggest that EV miRNA profiles from CRC cell lines may allow prediction of CRC tumours, and that *miR-7641* may serve as an attractive candidate for the specific, non-invasive diagnosis and prognosis of CRC.

Citation: Chen M, Xu R, Rai A, Suwakulsiri W, Izumikawa K, Ishikawa H, et al. (2019) Distinct shed microvesicle and exosome microRNA signatures reveal diagnostic markers for colorectal cancer. PLoS ONE 14(1): e0210003. https://doi.org/10.1371/journal.pone.0210003

Editor: Aamir Ahmad, University of South Alabama Mitchell Cancer Institute, UNITED STATES

Received: August 20, 2018; **Accepted:** December 14, 2018; **Published:** January 4, 2019

Data Availability: Raw data (FASTQ formatted files) can be accessed in NCBI SRA database under the accession numbers SRA440609 and SRA448517.

Funding: MC, RX, AR, WS were supported by La Trobe University Post-Graduate Research Scholarships. RS is supported by a Distinguished Professorship (La Trobe University) and DG, a Stone Molecular Biology Fellowship (La Trobe University) and Australian National Health and Medical Research Council (Project: 1139489 and 1141946). NT, KI, HI were supported by Core Research for Evolutional Science and Technology (CREST) from Japan Science and Technology Agency (JST).

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: 3'-UTR, 3'untranslated region; CEA, carcinoembryonic antigen; CLs, cell lysates; CRC, colorectal cancer; EV, extracellular vesicles; Exos, exosomes; miRNA, microRNA; qRT-PCR, quantitative real-time PCR; SEER, Surveillance, Epidemiology, and End Results; sMVs, shed microvesicles; TCGA, The Cancer Genome Atlas; TMUGS, Tumour Marker Utility Grading System; TPM, transcripts per million reads

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Jan 27, 2021

This Agreement between Mr. Wittaya Suwakulsiri ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4997080258623
License date	Jan 27, 2021
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Springer eBook
Licensed Content Title	Analysis of Annotated and Unannotated Long Noncoding RNAs from Exosome Subtypes Using Next-Generation RNA Sequencing
Licensed Content Author	Wittaya Suwakulsiri, Maoshan Chen, David W. Greening et al
Licensed Content Date	Jan 1, 2021
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	full article/chapter
Will you be	no

translating?	
Circulation/distribution	1 - 29
Author of this Springer Nature content	yes
Title	Analysis of Annotated and Unannotated Long Noncoding RNAs from Exosome Subtypes Using Next-Generation RNA Sequencing
Institution name	La Trobe University
Expected presentation date	Jan 2021

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Feb 05, 2021

This Agreement between Mr. Wittaya Suwakulsiri ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number 5002790414956

License date	Feb 05, 2021
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Proteomics
Licensed Content Title	Exosomes Derived from the Human Primary Colorectal Cancer Cell Line SW480 Orchestrate Fibroblast-Led Cancer Invasion
Licensed Content Author	Richard J. Simpson, Wittaya Suwakulsiri, Rong Xu, et al
Licensed Content Date	Jul 20, 2020
Licensed Content Volume	20
Licensed Content Issue	14
Licensed Content Pages	12

Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Electronic
Portion	Full article
Will you be translating?	No
Title	Analysis of Annotated and Unannotated Long Noncoding RNAs from Exosome Subtypes Using Next-Generation RNA Sequencing
Institution name	La Trobe University
Expected presentation date	Feb 2021

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Jan 30, 2021

This Agreement between Mr. Wittaya Suwakulsiri ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4999041260122
License date	Jan 30, 2021
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Cancer
Licensed Content Title	APC, Signal transduction and genetic instability in colorectal cancer
Licensed Content Author	Riccardo Fodde et al
Licensed Content Date	Oct 1, 2001
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1

Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	Analysis of Annotated and Unannotated Long Noncoding RNAs from Exosome Subtypes Using Next-Generation RNA Sequencing
Institution name	La Trobe University
Expected presentation date	Jan 2021
Portions	A figure in Box1

Permissions

About NEJM

Submit Your Permission Request Using RightsLink

If you are seeking permission to copy, reproduce, or republish content from the *New England Journal of Medicine* (NEJM) and you are not the author of the content, you may use the Copyright Clearance Center's RightsLink® service. Simply visit NEJM.org and locate the article or video from which you seek to reuse content.

Once you have located and accessed the article you are looking for on NEJM.org:

- Click on the Permissions link © in the toolbar to the left of the article.
- The RightsLink window will pop up with information about the content you have selected.
- Follow the prompts to select the options that best suit your needs.
- The system will either provide you with an immediate price quote or direct your request to the NEJM

Permissions Department for further processing through the RightsLink interface.
- When prompted to continue, either sign in to your existing RightsLink account, or create a new account if you do not already have one, and follow the additional prompts to secure permission.

Confirmation of your permission and related terms and conditions will be sent to you instantly via email for most requests. Some requests may require Publisher review.

If you have questions about using the RightsLink service, please contact:

RightsLink® Customer Support
+1-877-622-5543 (toll free)
+1-978-777-9929
Email: **customercare@copyright.com**

For general questions about NEJM Permissions, email **permissions@nejm.org**.

Permission for Authors

If you are the author of an article that has been published in NEJM, see **Author Permissions**.

Reuse of Content Within a Thesis or Dissertation

Content (full-text or portions thereof) may be used in print and electronic versions of a dissertation or thesis without formal permission from the Massachusetts Medical Society (MMS), Publisher of the *New England Journal of Medicine*.

The following credit line must be printed along with the copyrighted material:

> Reproduced with permission from (scientific reference citation), Copyright Massachusetts Medical Society.

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Feb 06, 2021

This Agreement between Mr. Wittaya Suwakulsiri ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5003331370796
License date	Feb 06, 2021
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Gastroenterology & Hepatology
Licensed Content Title	Epigenetics and colorectal cancer
Licensed Content Author	Victoria Valinluck Lao et al
Licensed Content Date	Oct 18, 2011
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Will you be translating?	no

Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	Analysis of Annotated and Unannotated Long Noncoding RNAs from Exosome Subtypes Using Next-Generation RNA Sequencing
Institution name	La Trobe University
Expected presentation date	Feb 2021
Portions	Figure 1

AMERICAN ASSOCIATION FOR CANCER RESEARCH LICENSE
TERMS AND CONDITIONS

Feb 06, 2021

This Agreement between Mr. Wittaya Suwakulsiri ("You") and American Association for Cancer Research ("American Association for Cancer Research") consists of your license details and the terms and conditions provided by American Association for Cancer Research and Copyright Clearance Center.

License Number	4997480174036
License date	Jan 28, 2021
Licensed Content Publisher	American Association for Cancer Research
Licensed Content Publication	Clinical Cancer Research
Licensed Content Title	Colon Cancer: It's CIN or CIMP
Licensed Content Author	Jean-Pierre Issa
Licensed Content Date	Oct 1, 2008
Licensed Content Volume	14
Licensed Content Issue	19
Type of Use	Thesis/Dissertation
Requestor type	academic/educational
Format	electronic

Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Will you be translating?	no
Circulation	1
Territory of distribution	Worldwide
Title	Analysis of Annotated and Unannotated Long Noncoding RNAs from Exosome Subtypes Using Next-Generation RNA Sequencing
Institution name	La Trobe University
Expected presentation date	Jan 2021
Portions	Figure 1

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Jan 30, 2021

This Agreement between Mr. Wittaya Suwakulsiri ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4999051483232
License date	Jan 30, 2021
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Cancer
Licensed Content Title	Microenvironmental regulation of metastasis
Licensed Content Author	Johanna A. Joyce et al
Licensed Content Date	Mar 12, 2008
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1

High-res required	no
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	Analysis of Annotated and Unannotated Long Noncoding RNAs from Exosome Subtypes Using Next-Generation RNA Sequencing
Institution name	La Trobe University
Expected presentation date	Jan 2021
Portions	Figure1

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Jan 30, 2021

This Agreement between Mr. Wittaya Suwakulsiri ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4999060867365
License date	Jan 30, 2021
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Oncogene
Licensed Content Title	The tumor microenvironment and its role in promoting tumor growth
Licensed Content Author	T L Whiteside
Licensed Content Date	Oct 6, 2008
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1

High-res required	no
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	Analysis of Annotated and Unannotated Long Noncoding RNAs from Exosome Subtypes Using Next-Generation RNA Sequencing
Institution name	La Trobe University
Expected presentation date	Jan 2021
Portions	Figure 1

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Jan 30, 2021

This Agreement between Mr. Wittaya Suwakulsiri ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4999061284138
License date	Jan 30, 2021
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Journal of Neuro-Oncology
Licensed Content Title	Biogenesis of extracellular vesicles (EV): exosomes, microvesicles, retrovirus-like vesicles, and apoptotic bodies
Licensed Content Author	Johnny C. Akers et al
Licensed Content Date	Mar 2, 2013
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1

Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	Analysis of Annotated and Unannotated Long Noncoding RNAs from Exosome Subtypes Using Next-Generation RNA Sequencing
Institution name	La Trobe University
Expected presentation date	Jan 2021
Portions	Figure 4

Secreted midbody remnants are a class of extracellular vesicles molecularly distinct from exosomes and microparticles

Author: Alin Rai et al

Publication: Communications Biology

Publisher: Springer Nature

Date: Mar 25, 2021