Analysis of annotated and unannotated long non-coding RNAs from exosome subtypes using next-generation RNA sequencing

Wittaya Suwakulsiri[1], Maoshan Chen[2], David W. Greening[1,3], Rong Xu[1], and Richard J. Simpson[1,*]

[1]*Department of Biochemistry and Genetics, La Trobe Institute for Molecular Science (LIMS), La Trobe University, Melbourne, Victoria 3086, Australia*

[2]*Australian Centre for Blood Diseases, Alfred Hospital, Monash University, Melbourne, Victoria 3004, Australia*

[3]*Baker IDI Heart and Diabetes Institute, Molecular Proteomics, Melbourne, Victoria, 3004, Australia*

*To whom correspondence should be addressed:

Professor Richard J. Simpson

Room 412, La Trobe Institute for Molecular Science (LIMS)

La Trobe University, Bundoora, Victoria 3086, Australia

Tel: +61 03 9479 3099

Fax: +61 03 9479 1226

Email: Richard.Simpson@latrobe.edu.au

Abstract

Long non-coding 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, 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 of 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.

Keywords: Exosomes, Long non-coding RNAs, Transcriptomics, Next-generation RNA sequencing, Bioinformatics

Running title: Next-generation RNA sequencing and long non-coding RNAs from exosome subtypes

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 non-coding RNAs (ncRNAs) [2]. The percentage of transcribed non-coding sequences in whole genome relates to 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 non-coding RNAs 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 heterogenous group of RNA species > 200 nucleotides in length and incapable of encoding proteins [6]. The function of lncRNAs depends on their cell compartmentation, for instance, chromatin modifications, transcriptional control, and post-transcriptional processing occurs in the nuclear compartment while lncRNAs can inhibit of protein synthesis in 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 heterogenous 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. EVs are crucial mediators of

intercellular communication and act by transferring their bioactive cargo (e.g., proteins, RNA species, lipids, metabolites) to recipient cells [18]. Interestingly, the majority of exosomal RNAs maps to intronic regions resulting in the abundance of long intergenic non-coding 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], lncUCA1 [25]) and angiogenesis (e.g., lncPOU3F3 [26], 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 isolation of exosomes secreted from the human colorectal cancer cell line (LIM1863) [28] continuously cultured in Bioreactor classic flasks [21,29]. 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 glycoprotein A33 antigen (Heath et al., 1997) and EpCAM-immunoaffinity capture exosome subtype purification [31] (Fig. 2a, 2b). (Previously, we reported that A33- EpCAM-exosome subtypes derived from LIM1863 cells (A33 and EpCAM) exhibit distinct protein and transcriptomic profiles [31,21]). 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 [32]. Unannotated lncRNA are distinguished from annotated lncRNAs using Stringtie and Gffcompare [33] with cutoffs: (1) nucleotide length more than 200, (2) number of exon less than 2, (3) classification codes “i” (fully contained within a reference intron), “x” (exonic overlap on the opposite strand) and “u” (unknown, intergenic) [34].

Protein coding potential for each unannotated lncRNAs is calculated by Coding-Potential Assessment Tool (CPAT). Coding probability (CP) less than 0.364 (for human samples) indicates noncoding sequence [35] (this tool also allows identifying lncRNAs with protein-coding potential). We further perform raw read count calculation and differential expression analysis using FeatureCounts [36] and DESeq2 [37], respectively. In the final step, significantly enriched lncRNAs are validated using RT-qPCR.

2. Materials

2.1 Cell culture

1. LIM1863 cells [28]

2. 175-cm^2 culture flasks (BD Falcon)

3. RPMI1640

4. Fetal bovine serum (FBS)

5. Insulin-Transferrin-Selenium (ITS, Life Technologies)

6. Penicillin/Streptomycin (P/S, Life Technologies)

7. CELLine CL-1000 Bioreactor classic flask (Integra Biosciences)

8. Incubator, 37 °C capacity for storage of bioreactor flasks

9. Refrigerated centrifuge (to 4 °C), centrifugation capability to 3,000 x *g*

10. 50 ml polypropylene centrifuge tubes (Falcon tubes, Life Technologies)

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. Magnetic stand

10. Tris buffer (50 nM, pH 7.5)

11. Phosphate buffer saline (PBS, pH 7.4)

12. Laboratory rotator

13. 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) with100 mM Dithiothreitol (DTT)

14. NuPAGE™ 4–12% (w/v) Bis-Tris Precast gels (Life Technologies)

15. NuPAGE™ MES running buffer (Life Technologies)

16. Fixing solution (40% (v/v) methanol, 10% (v/v) aqueous acetic acid)

17. SYPRO®Ruby fluorescent stain for protein detection (Life Technologies)

18. Destaining solution (10% (v/v) methanol with 6% (v/v) acetic acid in water)

19. Typhoon 9410 variable mode imager (GMI) or suitable florescent imager

20. ImageQuant TL 8.2 image analysis software (GE Healthcare Life Sciences)

21. BenchMark™ Protein Ladder (Life Technologies)

22. iBlot™2.0 DryBlotting System (Life Technologies)

23. Tris buffered saline (TBS)

24. 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)

25. Anti-EpCAM (1:1,000; Abcam)

26. Anti-TSG101 (1:1,000; BD Biosciences)

27. Anti-ALIX (1:1,000; Cell Signaling)

28. IRDye 800 goat anti-mouse IgG or IRDye 680 goat anti-rabbit IgG (1:15,000; LI-COR Biosciences)

29. Odyssey Infrared Imaging System, v 3.0 (LICOR Biosciences)

30. Supor® Membrane (0.1 µm) (Pall Life Sciences)

31. Amicon® Ultra-15 Ultracel centrifugal filter device with a 3K nominal molecular weight limit (Merck-Millipore)

32. Orbital shaker

33. Dimethyl pimelimidate (DMP)

34. Tween-20

35. Refrigerated centrifuge (to 4 °C), centrifugation capability to 3,000 x *g*

36. Ultracentrifuge (Optima™ XPN-100 and Optima™ MAX-XP, Beckman Coulter) and matched rotor (Type 45Ti and TLA-55)

37. Microcentrifuge polypropylene tube (#357448, Beckman Coulter)

38. Polycarbonate bottle assembly (#355622, Beckman Coulter)

39. Skim milk powder

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, Illumina)

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, Illumina)

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. Magnetic stand

18. A-Tailing control (CTA, Illumina)

19. A-Tailing Mix (ATL, Illumina)

20. Adaptor tubes (Illumina)

21. Index Adaptor plates (Illumina)

22. Ligation Control (CTL, Illumina)

23. Ligation Mix (LIG, Illumina)

24. RNA adaptors (Illumina)

25. PCR Primer Cocktail (PPC, Illumina)

26. PCR Master Mix (PMM, Illumina)

27. A DNA 1000 chip (Agilent Technologies)

28. Agilent Technologies 2100 Bioanalyzer (Agilent Technologies)

29. Tris-HCl (10 mM, pH 8.5)

30. Microseal® 'B' adhesive seal (Bio-Rad)

31. Tween-20

2.5 Bioinformatic analysis pipeline for lncRNA sequencing analysis

1. Computer (64-bit) with either Linux or Mac OS X (10.7 Lion or later), 8GB (recommended) of RAM

2. Software

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/

3. Files

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

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 (Bio-Rad)

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 x 10^7 cells) are collected as floating organoids and centrifuged at 140 x *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 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 x *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 re-seeded 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 section 3.1) is centrifuged at 2,000 x *g*, 4 °C for 10 min to remove cell debris. Pellet is discarded

2. Supernatant is further centrifuged at 10,000 x *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 x 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 1mL 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 3.2.1) are pre-incubated with Dynabeads™ (500 µL, 5 x 10^8 beads) at 4 °C for 2 h with gentle rotation to reduce non-specific binding

7. Dynabeads™ (from 6) are removed by a magnetic stand and supernatant (crude exosomes) is incubated with A33-Dynabeads™ (from 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 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 x *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 sulphate, 125 mM Tris-HCl,pH 6.8, 12.5% (v/v) glycerol, 0.02% (w/v) bromophenol blue) with100 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 1x 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 re-solution.

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 1x NuPAGE™ MES running buffer

3. Proteins on the gel are electro-transferred 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 (1:1,000; 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 (1:1,000; BD Biosciences) and mouse anti-ALIX (1:1,000; 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 sec) and further incubated in RT for 3 min

2. Samples are centrifuged at 12,000 x *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 x *g*, 4 °C for 15 min

5. RNA pellet is washed once with 75% aqueous ethanol, air-dried for 5 min and re-dissolved 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 x *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 airdry 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 x *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 x *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 x *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 x *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 sec then the ethanol is discarded

8. Repeat 7 (*see* Note 2)

9. DFP is placed on a magnetic stand for 15 min to airdry 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 x *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 a Microseal 'B' adhesive seal and centrifuged at 280 x *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 x *g* for 1 min

3.4.5 Adaptor ligation

1. RNA adapters are prepared as followed; Adapter tubes are centrifuged at 600 x *g* for 5 sec, Index Adapter plate is centrifuged at 280 x *g* for 1 min. Ligation Control (CTL) is centrifuged at 600 x *g* for 5 sec 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 x *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 x *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 sec then the ethanol is discarded

8. Repeat 7 (*see* Note 2)

9. ALP is placed on a magnetic stand for 15 min to airdry 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 x *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 4-12 with 50 µL AMPure XP beads and 22.5 µL RSB

14. Supernatant (20 µL) from 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 x *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 sec, 15 cycles of 98 °C for 10 sec, 60 °C for 30 sec, 72 °C for 30 sec (end of the 15 cycles), and finish with 72 °C for 5 min, hold at 4 °C)

5. PCR plate is centrifuged at 280 x *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 x *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 sec then the ethanol is discarded

10. Repeat 9 (*see* Note 2)

11. PCR plate is placed on a magnetic stand for 15 min to airdry 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 x *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 x *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 x *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

3.5.1 Download files for RNA sequence alignment

1. Establish a work 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. 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, 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

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 3.5.1.4) using `stringtie` with options `-p 8 -G -o SRR1662176.gtf (SRR1662177.gtf) -l` (*see* Note 10, Note 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 3.5.1.4) 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 non-coding 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 -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 non-coding 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 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	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

*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 (2x, 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 sec and placed on PCR machine (95 °C for 2 min, 30-45 cycles of denaturation at 95 °C for 5-15 sec; annealing/extension at 60 °C for 10-30 sec)

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 x *g* for 5 sec before use

4. FSA is centrifuged at 600 x *g* for 5 sec before use

5. CTE is centrifuged at 600 x *g* for 5 sec and diluted to 1:50 in RSB before use

6. SMM is centrifuged at 600 x *g* for 5 sec before use

7. CTA is centrifuged at 600 x *g* for 5 sec and diluted to 1:100 in RSB before use

8. ATL is centrifuged at 600 x *g* for 5 sec 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 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, gene biotype, and so on (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 different R packages and one of the packages is "DESeq2". It requires 2 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)).

5. Conclusion

Long non-coding RNAs are regulatory molecules in transcription and translation processes. Using next-generation RNA sequencing, several studies have shown that lncRNAs are present in exosomes. The lncRNA analysis method described here utilizes RNA sequencing data and stringent parameters to discover annotated and unannotated lncRNAs in two exosome subtypes 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.

Figure legend

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 isolation 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. 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 [36] and DESeq2 [37], 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 $A33^{+}$, $ALIX^{+}$, $TSG101^{+}$, $EpCAM^{-}$ and EpCAM-exosomes are $EpCAM^{+}$, $ALIX^{+}$, $TSG101^{+}$, $A33^{-}$. Electron microscopy analysis shows A33- and EpCAM-exosomes are cupped shape structure with diameter around 40-60 nm *see* [31]. (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.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

The authors are supported, in part, by the National Health and Medical Research Council of Australia project grants #1057741 and #1139489 (D.W.G) and La Trobe University (R.J.S). W.S is supported by La Trobe University Postgraduate Scholarship.

References

1. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X (2001) The sequence of the human genome. Science 291 (5507):1304-1351. doi:10.1126/science.1058040

2. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermuller J, Hofacker IL, Bell I, Cheung E, Drenkow J, Dumais E, Patel S, Helt G, Ganesh M, Ghosh S, Piccolboni A, Sementchenko V, Tammana H, Gingeras TR (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 316 (5830):1484-1488. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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. doi:10.1038/nrg2521
9. Anderson DM, Anderson KM, Chang CL, Makarewich CA, Nelson BR, McAnally JR, Kasaragod P, Shelton JM, Liou J, Bassel-Duby R, Olson EN (2015) A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. Cell 160 (4):595-606. doi:10.1016/j.cell.2015.01.009
10. Stein CS, Jadiya P, Zhang X, McLendon JM, Abouassaly GM, Witmer NH, Anderson EJ, Elrod JW, Boudreau RL (2018) Mitoregulin: A lncRNA-Encoded Microprotein that Supports Mitochondrial Supercomplexes and Respiratory Efficiency. Cell Rep 23 (13):3710-3720 e3718. doi:10.1016/j.celrep.2018.06.002
11. Cai B, Li Z, Ma M, Wang Z, Han P, Abdalla BA, Nie Q, Zhang X (2017) LncRNA-Six1 Encodes a Micropeptide to Activate Six1 in Cis and Is Involved in Cell Proliferation and Muscle Growth. Front Physiol 8:230. doi:10.3389/fphys.2017.00230
12. Pan Y, Chen H, Shen X, Wang X, Ju S, Lu M, Cong H (2018) Serum level of long noncoding RNA H19 as a diagnostic biomarker of multiple myeloma. Clin Chim Acta 480:199-205. doi:10.1016/j.cca.2018.02.019
13. Dong H, Wang W, Chen R, Zhang Y, Zou K, Ye M, He X, Zhang F, Han J (2018) Exosome-mediated transfer of lncRNASNHG14 promotes trastuzumab chemoresistance in breast cancer. Int J Oncol 53 (3):1013-1026. doi:10.3892/ijo.2018.4467
14. Karlsson O, Rodosthenous RS, Jara C, Brennan KJ, Wright RO, Baccarelli AA, Wright RJ (2016) Detection of long non-coding RNAs in human breastmilk extracellular vesicles: Implications for early child development. Epigenetics 11 (10):721-729. doi: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. doi:10.1007/s00264-018-4093-6
16. Berrondo C, Flax J, Kucherov V, Siebert A, Osinski T, Rosenberg A, Fucile C, Richheimer S, Beckham CJ (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. doi:10.1371/journal.pone.0147236
17. Xu R, Greening DW, Zhu HJ, Takahashi N, Simpson RJ (2016) Extracellular vesicle isolation and characterization: toward clinical application. J Clin Invest 126 (4):1152-1162. doi:10.1172/JCI81129
18. Xu R, Rai A, Chen M, Suwakulsiri W, Greening DW, Simpson RJ (2018) Extracellular vesicles in cancer - implications for future improvements in cancer care. Nat Rev Clin Oncol 15 (10):617-638. doi: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. doi:10.1080/14789450.2018.1537788
20. Jeppesen DK, Fenix AM, Franklin JL, Higginbotham JN, Zhang Q, Zimmerman LJ, Liebler DC, Ping J, Liu Q, Evans R, Fissell WH, Patton JG, Rome LH, Burnette DT, Coffey RJ (2019) Reassessment of Exosome Composition. Cell 177 (2):428-445 e418. doi:10.1016/j.cell.2019.02.029

21. Chen M, Xu R, Ji H, Greening DW, Rai A, Izumikawa K, Ishikawa H, Takahashi N, Simpson RJ (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. doi:10.1038/srep38397
22. Qu L, Ding J, Chen C, Wu ZJ, Liu B, Gao Y, Chen W, Liu F, Sun W, Li XF, Wang X, Wang Y, Xu ZY, Gao L, Yang Q, Xu B, Li YM, Fang ZY, Xu ZP, Bao Y, Wu DS, Miao X, Sun HY, Sun YH, Wang HY, Wang LH (2016) Exosome-Transmitted lncARSR Promotes Sunitinib Resistance in Renal Cancer by Acting as a Competing Endogenous RNA. Cancer Cell 29 (5):653-668. doi:10.1016/j.ccell.2016.03.004
23. Takahashi K, Yan IK, Kogure T, Haga H, Patel T (2014) Extracellular vesicle-mediated transfer of long non-coding RNA ROR modulates chemosensitivity in human hepatocellular cancer. FEBS Open Bio 4:458-467. doi:10.1016/j.fob.2014.04.007
24. Takahashi K, Yan IK, Wood J, Haga H, Patel T (2014) Involvement of extracellular vesicle long noncoding RNA (linc-VLDLR) in tumor cell responses to chemotherapy. Mol Cancer Res 12 (10):1377-1387. doi:10.1158/1541-7786.MCR-13-0636
25. Xu CG, Yang MF, Ren YQ, Wu CH, Wang LQ (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, Liu Y, Tu W, Lu YM, Wu L, Xu GH (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, Kuang W, Chen Y, Wu L, Xu GH (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. doi:10.3892/or.2017.5742
28. Whitehead RH, Jones JK, Gabriel A, Lukies RE (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, He W, Rai A, Zhang W, Simpson RJ (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. doi:10.1371/journal.pone.0110314
30. Suwakulsiri W, Rai A, Xu R, Chen M, Greening DW, Simpson RJ (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. doi:10.1016/j.bbapap.2018.11.008
31. Tauro BJ, Greening DW, Mathias RA, Mathivanan S, Ji H, Simpson RJ (2013) Two distinct populations of exosomes are released from LIM1863 colon carcinoma cell-derived organoids. Mol Cell Proteomics 12 (3):587-598. doi:10.1074/mcp.M112.021303
32. Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12 (4):357-360. doi:10.1038/nmeth.3317
33. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 33 (3):290-295. doi:10.1038/nbt.3122
34. Zhao Q, Sun Y, Wang D, Zhang H, Yu K, Zheng J, Zuo Z (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. doi:10.1016/j.jgg.2018.06.005

35. Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W (2013) CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res 41 (6):e74. doi:10.1093/nar/gkt006
36. 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. doi:10.1093/bioinformatics/btt656
37. 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. doi:10.1186/s13059-014-0550-8