Detection of Cadherin-17 in human colon cancer LIM1215 cell secretome and tumour xenograft-derived interstitial fluid and plasma

Oliver K. Bernhard [2,3], David W. Greening [1,2], Thomas W. Barnes [2,4], Hong Ji [1,2], and Richard J. Simpson [1,2*]

[1] *La Trobe Institute for Molecular Science (LIMS), La Trobe University, Bundoora, Australia*

[2] *Ludwig Institute for Cancer Research Ltd (Parkville Branch, Melbourne), Australia*

[3] *Current address: Starpharma Pty Ltd, Prahran, Melbourne, 3004 Australia*

[4] *Current address: Bioanalytical Sciences, Research and Development, CSL Ltd, Parkville, 3052 Australia*

[*]*Corresponding author*
Prof. Richard J. Simpson
La Trobe Institute for Molecular Science (LIMS)
La Trobe University
Bundoora, Victoria, 3083 Australia
Fax: +61 3 9479 1266
Email: Richard.Simpson@latrobe.edu.au

Keywords: animal model, cadherin-17, colorectal cancer, exosome, secretome, tumour interstitial fluid

Title: Colon glyco-secretome

Abbreviations:

BCA, bicinchoninic acid

CDH17, cadherin-17

CEA, carcinoembryonic antigen

CRC, colorectal cancer

CM, culture medium

CCM, concentrated culture media

DTT, dithiothreitol

EDTA, ethylenediaminetetraacetic acid

FOBT, faecal occult blood test

IP, immunoprecipitation

ITS, insulin-transferrin-selenium A

SDS, sodium dodecyl sulphate

TCEP, tris(2-carboxyethyl)phosphine

Tif, tumour interstitial fluid

TFA, trifluoroacetic acid

SUMMARY

Colorectal cancer (CRC), one of the most prevalent cancers in the western world, is treatable if detected early. However, 70 % of CRC is detected at an advanced stage. This is largely due to the inadequacy of current faecal occult blood screening testing and costs involved in conducting population-based colonoscopy, the 'gold standard' for CRC detection. Another biomarker for CRC, carcinoembryonic antigen, while useful for monitoring CRC recurrence, is ineffective, lacking the specificity required early detection of CRC. For these reasons there is a need for more effective blood-based markers for early CRC detection. In this study we targeted glycoproteins secreted from the human colon carcinoma cell line LIM1215 as a source of potential CRC biomarkers. Secreted candidate glycoproteins were confirmed by MS and validated by western blot analysis of tissue/tumour interstitial fluid (Tif) from LIM1215 xenograft tumours grown in immunocompromised mice. Overall, 39 glycoproteins were identified in LIM1215 culture media (CCM) and 5 glycoproteins in LIM1215 tumour xenograft Tif; of these, cadherin-17 (CDH17), galectin-3 binding protein (LGALS3BP), and tyrosine-protein kinase-like 7 (PTK7) were identified in both CM and glycosylation motifs Swiss-Prot was used to annotate Tif. Many of the glycoproteins identified in this study (e.g., AREG, DSG2, EFNA1, EFNA3, EFNA4, EPHB4, ST14, and TIMP1) have been reported to be implicated in CRC biology. Interestingly, the cadherin-17 ectodomain, but not full length cadherin-17, was identified in CM, Tif and plasma derived from mice bearing the LIM1215 xenograft tumor. To our knowledge, this is the first report of the cadherin-17 ectodomain in plasma. In this study, we report for the first time that the presence of full-length cadherin-17 in exosomes released into the CM.

INTRODUCTION

Colorectal cancer (CRC) is the third leading cause of cancer death in The United States (51,690), with one in 19 now diagnosed with the disease (103,170 estimated new diagnoses in 2012) [1]. Early stage detection significantly improves the clinical outcome [2] although this is often made difficult by the lack of specific symptoms [3]. Screening of the general population for CRC is recommended starting at age 50 with colonoscopy or faecal occult blood test (FOBT) at regular intervals [4]. With over 70 % of CRC cases detected at advanced stages screening remains unsatisfactory and non-specific [5]. Problems associated with the FOBT are its low sensitivity and high false-positive rate, however it has been shown to reduce mortality in multiple randomized trials [6]. Colonoscopy is considered the "gold standard" with 90-97 % sensitivity and >98 % specificity [4]. However it is invasive, not without risk and expensive and thus not considered cost-effective for screening [3]. Ideally, measurement of a protein marker in blood would be used for screening for CRC. The lack of specificity and sensitivity preclude the use of all existing serum markers for the early detection of CRC, including carcinoembryonic antigen (CEA), cancer antigen 19.9 (CA19.9), and tissue inhibitor of metalloproteinase 1 (TIMP-1) [7]. To identify novel biomarkers for cancer, genomic changes associated with breast and colon cancers have been identified [8, 9]. Additionally, alterations in mRNA levels during tumour development and progression have been investigated [10], however, such alterations at the DNA and RNA level may not translate into a measurable change of a protein marker [11].

From a clinical perspective, focusing on proteins that are secreted by cells of the tumour microenvironment [12] is appealing for diagnostic purposes, as there is a strong likelihood

that proteins may passage into proximal fluids such as the tumour proximal (interstitial) environment [13] and peripheral blood [14-16].

As a model for disease-specific markers, cell culture media includes proteins released from cells via various mechanisms such as classical secretion, non-classical secretory pathway, and extracellular vesicles, such as exosomes [17]. Tissue (or tumour) interstitial fluid (Tif), from xenograft tumours, represents an effective means of enriching the secretome (e.g., the concentration of secreted proteins is reported to be ~1000–1500-fold in Tif relative to cellular concentrations [15]). Recent studies have reported the isolation of Tif *ex vivo* from invasive breast carcinomas [18], ovarian carcinoma tissue [19], head and neck squamous carcinoma [20] and applied this fluid for proteomic analyses, identifying components associated with the extracellular matrix, cell invasion, and inflammation. *In vitro* tumour cell line-derived secretome samples are considered to be a rich source of biomarkers and have the advantage of reproducibility and scalability [21].

Here, we investigated the glyco-secretome as a source of CRC biomarkers using the human colon carcinoma LIM1215 cell line [22] as a model. We employed hydrazide chemistry [23-25] to analyze glycopeptides derived from different secretomes (CM and Tif) to identify various N-glycoproteins implicated in CRC biology. Cadherin-17, an adhesion molecule specific to the intestine [26], was identified in our glyco-secretome study and selected for further evaluation in plasma derived from immuncompromised mice bearing the LIM1215 xenograft tumour.

EXPERIMENTAL PROCEDURES

All buffers and solutions were prepared with water obtained from an A10-Synthesis™ water-polishing system (Millipore, Australia). Urea, SDS and DTT were obtained from Bio-Rad (Australia), tris(2-carboxyethyl)phosphine (TCEP), sodium metaperiodate, sodium deoxycholate and iodoacetamide were from Fluka/Sigma (Australia), methanol and acetonitrile (ChromAR grade) were obtained from Mallinkrodt (Australia), formic acid was from BDH/VCR (PA, USA) and TFA was purchased from Pierce (Rockford, IL, USA). NP-40 was obtained from Calbiochem/Merck (Australia). Polyclonal (AF1032) and monoclonal (MAB1032) antibodies to the extracellular domains of cadherin-17 were from R&D Systems (Minneapolis, MN, USA), the polyclonal antibody (sc-6978) against the C-terminal part of cadherin-17 was from Santa Cruz Biotechnology (Santa Cruz, CA, USA), the secondary anti-goat-Cy5.5 labeled antibody (ab6951) and the mouse anti-human CD9 antibody (ab2215-100) were from Abcam (Cambridge, UK). The secondary anti-mouse-800 antibody was from Licor (Lincoln, NE, USA).

2.1 Cell lines and generation of CCM and detergent lysate

LIM1215 cells [22] (provided by the Ludwig Institute, Parkville, Melbourne) were maintained in RPMI 1640 medium, supplemented with 10 % faetal calf serum (RF-10), 10^{-6} M α-thioglycerol, 25 U/L insulin and 1 mg/mL hydrocortisone. Cells (2×10^6) were plated in 25 mL media per 150 mm diameter culture dish. After reaching 70-80 % confluency, cells were washed twice with RPMI 1640 and twice with phenol red-free RPMI 1640 (Invitrogen). The cells were then cultured in phenol red-free RPMI 1640 supplemented with 1 % insulin-transferrin-selenium-A (ITS, Invitrogen) for 24 h. Cell culture media (CM) was collected, centrifuged at 480 × *g* to remove floating cells followed by another centrifugation at 2,000 ×

g, and concentrated ~100 fold using Amicon Ultra PL-10 device with 5,000 nominal molecular weight limit filter (Millipore). After pelleting membranous vesicles for 1 h at 100,000 × *g* the supernatant (concentrated CM, CCM) was stored at -80 ºC.

For generating detergent lysate, cells were lifted with 50 mM EDTA (avoiding tryptic hydrolysis of membrane proteins) in PBS, washed twice with PBS and resuspended in lysis buffer (10 mM Tris, pH 8.0 containing 140 mM NaCl, 1 mM $MgCl_2$, 0.1 mM EDTA, 0.02 % (w/v) NaN_3, 1 % (v/v) NP-40, 0.5 % (w/v) sodium deoxycholate and 0.1 % (w/v) SDS) at 10^7 cells/mL for 1 h at 4 ºC. Insoluble material was pelleted by centrifugation at 20,000 × *g* for 10 min and the supernatant frozen at –20 ºC until required.

*2.2 Xenograft establishment and collection of plasma and tumour interstitial fluid (*Tif*)*

Animal experiments were approved by The Ludwig Institute for Cancer Research/Department of Surgery Animal Ethics Committee and conducted in accordance with the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes. SCID mice and Balb/c nude/nude mice (Animal Resources Centre, Australia) were maintained in microisolator cages with access to food and water *ad libitum*. SCID mice were irradiated with 200 rad 24h prior to injection. No adverse reaction was observed in mice following administration of this dose. For SCID mice (n=8), xenografts were established by the s.c. injection of 5×10^6 LIM1215 cells (washed twice with PBS) in 0.1 mL PBS into both ventral flanks, anterior to the hind leg. Xenograft volume was determined 3× per week by measurement with standard calipers using the formula V = (small diameter)^2 × (large diameter) × 0.5 [27]. Animals were sacrificed when the combined tumour volume reached 1.0 cm^3 or after weight loss exceeding 20 % of body weight. Blood was collected by post-mortem cardiac puncture into 4 mL-Greiner EDTA Vacuette® blood collection tubes

(CenMed, East Brunswick, NJ, USA), centrifuged at 850 × *g* for 20 min at 19 ºC, and the supernatant aspirated and re-centrifuged using the same parameters. The supernatant (plasma) was stored at -80 ºC (n=3). Xenografts were excised, rinsed with PBS and intermediately stored on ice. Xenografts were gently diced and incubated in 500 µL PBS for 1 h at 37 ºC in a humidified incubator with 10 % CO_2 as described by Celis *et al*., [18, 28]. The sample was then subjected to two subsequent centrifugations, the first at 1,000 × *g* for 2 min (after which ~ 500 µL supernatant was aspirated (preTif)) and again following 5,000 × *g* for 20 min (after which ~50 µL supernatant (Tif) was collected). Both preTif and Tif samples were combined stored at -80 ºC for further use (n=3, matched with plasma samples described above). Only 3 matched plasma-Tif samples were obtained due to xenograft tumour size >1.0 cm^3.

2.3 Glycoprotein capture and isolation of glycopeptides

Concentrated CM (CCM) (n=2) containing 1.0 mg protein (determined by bicinchoninic acid (BCA) protein assay [29]) was thawed and buffer-exchanged into coupling buffer (0.1 M sodium acetate, pH 5.5, containing 0.15 M NaCl) via centrifugal ultrafiltration using an Amicon Ultra PL-10 device with a 10,000 nominal molecular weight cut-off (Millipore) and adjusted to a final concentration of 2 mg/mL. For analyzing Tif, 1.0 mg of Tif (n=2) was used (determined by BCA assay) and diluted with coupling buffer to 2 mg/mL. Carbohydrate side chains were oxidized by adding an equal volume of a freshly prepared 30 mM solution of sodium metaperiodate in coupling buffer and incubation for 1 h in the dark. Residual oxidizing agent was removed by gel filtration into coupling buffer using a NAP-10 column (GE Healthcare, UK) and 100–200 µL hydrazide beads (Pierce) that had been washed with coupling buffer were added and the coupling reaction performed overnight. Unbound proteins were removed by washing 5 × with urea wash buffer (8 M urea, 0.4 M NH_4HCO_3,

pH 8.3) followed by reduction of disulfide bonds with TCEP and alkylation with 0.5 M iodoacetamide as described in detail in our previous study [23]. Following another 5 × washes in urea wash buffer, proteins were digested with trypsin (Promega, Australia) over night and then unbound peptides were removed as described [23]. Tethered glycopeptides were released by incubating with 1 µL PNGase F over night at 37 ºC.

2.4 Peptide purification for LC-MS/MS analysis

Released glycopeptides were extracted from the beads with 0.1 % aqueous TFA (total sample volume of 450 µL). The sample was split into 3 × 150 µL aliquots and each aliquot subjected to two subsequent ZipTip® (Millipore) purifications. Briefly, the ZipTip® resin was pretreated with 60 % (v/v) aqueous acetonitrile containing 0.1 % (v/v) TFA in water and washed (2 ×) with 0.1 % (v/v) aqueous TFA. The sample was then applied to the resin followed by 5 washes with 0.1 % (v/v) aqueous TFA. The purified peptides were eluted with 10 µL 60 % (v/v) aqueous acetonitrile containing 0.1 % (v/v) TFA. The 6 × 10 µL elution fractions were pooled and acetonitrile was evaporated using centrifugal lyophilization and aqueous 0.1 % TFA was added to a total volume of 30 µL.

2.5 LC-MS/MS analysis

RP-HPLC was performed (CCM n=2, Tif n=2) with a capillary HPLC system (model 1100; Agilent Technologies, Santa Clara, CA, USA). Samples from each vial (8 µL) were loaded onto a capillary RP-HPLC column (150 µm internal diameter × 100 mm; 1.7 µm particle size; BEH C18, Waters-Micromass, Altrincham, UK) at a flow rate of 2 µL/min for 10 min. The column was developed at a flow rate of 0.8 µL/min using a linear 60 min gradient from 0-100 % B, where solvent A was aqueous 0.1 % (v/v) formic acid, and solvent B was 60 % (v/v) acetonitrile in aqueous 0.1 % (v/v) formic acid. RP-HPLC was coupled to an ESI-IT-

MS (LCQ Deca, Thermo-Finnigan, San Jose, USA). The ESI parameters were set as follows: spray voltage, 1.9 kV; capillary temperature, 175 ºC; capillary voltage, 34 V; tube lens offset voltage, 2 V; electron multiplier voltage, -1.16 kV. A constant number of ions (1×10^8 ions for full MS scans; 5×10^6 ions for 'zoom' scans (see below); 1×10^8 ions for MS/MS scans) was injected into the trap by allowing it to remain open for a maximum of 200 ms for full MS and 'zoom' scans, and 500 ms for MS/MS scans. The IT automatic gain control parameter was also activated for all experiments in order to maintain a constant number of ions within the trap. A 'triple scan' method was used to collect MS data in centroid mode as follows: (1) A full MS scan was acquired (400-2000 *m/z*), (2) the most abundant peak (above a preset threshold of 2×10^5 counts) was subjected to a higher resolution 'zoom' scan to allow determination of the charge state, and (3) the monoisotopic peak was selected for collision-induced dissociation with maximal collision energy of 35 %. The *m/z* range acquired for MS/MS experiments was varied according to the MW of the parent ion as determined from the 'zoom' scan (minimum 50 *m/z*; maximum: calculated parent ion MW + 10 % (upper limit 2000 *m/z*)).

2.6 Data analysis

Each raw MS/MS data file was searched with the Mascot™ algorithm (v2.2.04, Matrix Science) using Mascot Demon (v2.2.2, Matrix Science) and a ThermoFinnigan LCQ/DECA data import filter for peak list generation with the following settings: minimum mass, 700; maximum mass; 5,000; grouping tolerance, 0.01; intermediate scans, 1; min. scans/group, 1. Taxonomic restriction to mammalian entries was used to increase coverage of proteins that may have only a few correct peptide hits and acceptable because all assignments were manually validated. The search parameters were as follows: database, Swiss-Prot (v04/2009, 64,367 mammalian sequences); enzyme, semi-trypsin; fixed modifications,

carbamidomethylation (C); variable modifications, deamidation (N), oxidation (M), pyroglutamic acid (N-terminal Q); missed cleavages, two; peptide tolerance, 1500 ppm; MS/MS tolerance, 0.6 Da. Search results were exported to an Excel spreadsheet and then filtered for peptides that have an associated E-value ≤ 1 or which are fully tryptic and contain a deamidation in a glycosylation motif. Filtered peptides and additional peptides not passing the filtering criteria, but belonging to already inferred proteins were then manually validated using the Xcalibur™ (v1.4, Thermo-Finnigan) software tool to display MS/MS spectra. The masses of the product ions of the peptide were calculated with Protein Prospector (http://prospector.ucsf.edu/). Peptide verification was conducted according to previously published guidelines [30] confirming *b*- and *y*-ion continuity, significant signal-to-noise ratio (typically, 2 or more) and assignment of the most abundant product ions taking into account preferred peptide bond cleavage or precursor ion neutral loss [23, 30]. All identified peptides were then BLAST-searched (http://au.expasy.org/tools/blast/) against the Swiss-Prot database (v04/2009) with restriction to mammalian entries. Peptide sequence, name and accession number of the inferred protein were added to the list of identified peptides and proteins. If more than one protein entry matched the identified peptide, one representative protein was listed.

2.7 SDS-PAGE and Western blotting

Samples were prepared in 1 × Lämmli sample buffer (0.06 M Tris-HCl, 2 % (w/v) SDS, 10 % (v/v) glycerol, 0.01 % bromophenol blue, pH 6.8) containing 50 mM DTT and heated for 5 min at 95 ºC. Proteins were separated on a 4-12 % NuPAGE® Novex Bis-Tris Gel (Invitrogen) at constant current and transferred to a nitrocellulose membrane using an iBlot device (Invitrogen). The membrane was incubated in 1 % (w/v non-fat dry skim milk

powder, 0.02 % (w/v) sodium azide in water for 1 h at RT on an orbital shaker. Proteins were probed with the respective antibodies and detected using the Odyssey system (LI-COR).

2.8 Cadherin-17 immunoprecipitation (IP)

Xenograft-matched plasma (75 µL diluted with 155 µL PBS containing 0.02 % (v/v) Tween-20 (TPBS)) or 1.0 mL detergent lysate was precipitated by incubation at 4 ºC for 1 h with 1 µg monoclonal anti-human cadherin-17 antibody bound to 20 µL Protein G sepharose (GE Healthcare). The beads were washed with 1.0 mL TPBS (5 ×), eluted by heating in 20 µL 1 × Lämmli sample buffer at 95 ºC for 5 min and subjected to SDS-PAGE and Western detection as described above.

2.9 Purification of LIM1215-derived exosomes

LIM1215 cells were seeded into the cultivation chamber of a CELLine AD 1000 flask (Integra Biosciences, Switzerland) at 1×10^7 cells/15 ml of RPMI 1640 containing 0.8 % ITS, 100 U/mL penicillin and 100 µg/mL streptomycin. The nutrient supply chamber of the CELLine AD 1000 flask was filled with 900 ml of RPMI1640 containing 5 % FCS, 100 U/mL penicillin and 100 µg/mL streptomycin. The CM was collected, centrifuged at 480 × *g* for 5 min to sediment floating cells and again at 2,000 × *g* to remove cell debris. Resultant supernatant was filtered through a 0.1 µm membrane filter and concentrated to 0.2 mL using a 100,000-M_r cut-off filter (Millipore). The CCM was adjusted to 500 µl containing 250 mM sucrose and 10 mM Tris-HCl, pH 7.4 and then loaded onto a density cushion composed of 5, 10, 20 and 40 % of OptiPrep™ density gradient medium (Sigma) as described [31]. After centrifuging at 100,000 × g for 16 h, 12 fractions were collected. Each fraction was diluted three times with PBS and exosomes pelleted by centrifugation for 3 h at 100,000 × g. The

pellets were then washed once with PBS and resuspended in 1 × SDS-PAGE sample buffer for analysis.

RESULTS

Analysis of CCM from LIM1215 cells

To identify glycoproteins secreted by LIM1215 cells, CCM containing 1.0 mg of protein was used per analysis (two biological replicates). For LIM1215 cells, typically, protein yields for recovered CCM was approximately 20–30 mg per 3×10^8 cells (10 × 150-mm^2 dishes, 150 mL) per 24 h. Cell proliferation and viability of LIM1215 cells cultured using serum-free conditions over 24 h was ~95-97 % based on MTT assay (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) and Trypan Blue dye exclusion assay (data not shown), which is in accordance with our previous studies [32]. Glycopeptide CCM samples were prepared as outlined in Experimental Procedures with each sample subjected to three LC-MS/MS analyses to reduce analytical incompleteness. Following validation, a total of 39 proteins were identified from 59 detected peptides (discarding identifications of the media supplements insulin and transferrin). All peptides that mapped to transmembrane proteins were part of the extracellular or lumenal domains (based on Swiss-Prot annotation). Table 1 contains all glycoproteins identified in LIM1215 CCM and Tif, with all glycopeptides information available in Supplementary Tables S1-2.

Analysis of proteins secreted in vivo tumour xenograft-derived interstitial fluid (Tif)

In order to confirm proteins detected *in vitro* in an *in vivo* environment, we established LIM1215 xenografts in Balb/c nu/nu and SCID mice. LIM1215 only formed large tumours in SCID mice (n=8) after approximately 7 weeks whilst tumours in nude mice (n=8) remained

small. To obtain Tif (n=3), we used a centrifugal method [18, 33] for harvesting Tif from each tumour xenograft. Protein concentration of Tif preparations varied significantly (7-17 mg/mL, with no correlation to xenograft size or centrifugation speed.

Tif (~1.0 mg protein) was subjected to glycopeptide capture (n=2) and despite predominance of mouse plasma proteins (see Supplementary Table S3), 5 human proteins from 7 glycopeptides in LIM1215 Tif (Table 2, Supplementary Table S3) were detected. Two proteins detected in Tif (polymeric-immunoglobulin receptor (P01833, PIGR) and apolipoprotein A-I (P02647, APOA1) were not detected in the CCM. Interestingly, cadherin-17 was identified in both *in vitro* (CCM) and *in vivo* (Tif). Because cadherin-17 exhibits intestine-specific expression and its expression is dysregulated in cancer [34, 35] it was selected for further characterization.

Western detection of cadherin-17 in Tif, CCM and mouse plasma

The presence of cadherin-17 was analyzed by Western blot in LIM1215-derived samples. Fig. 1A shows that cadherin-17 could be detected as a 97 kDa band in Tif, CCM and cell lysate derived from LIM1215 cells. Furthermore a 75 kDa form of the protein could be seen in Tif and CCM with varying intensity but not in the detergent cell lysate. Cadherin-17 could not be detected by western blotting directly in plasma from mice carrying LIM1215 xenografts. For this reason, immunoprecipitation (IP) of cadherin-17 from plasma was performed using a monoclonal antibody. IP proteins were resolved by SDS-PAGE and Western blotting performed with a polyclonal antibody to detect plasma-derived cadherin-17. Fig. 1B shows the detection of cadherin-17 precipitated from mice plasma (n=3) bearing a LIM1215 xenograft with the monoclonal anti-cadherin-17 antibody. Control IP using plasma

from mice bearing no xenograft showed no cadherin-17 expression, confirming mouse cadherin-17 is not reacting with the monoclonal antibody.

The secreted form of cadherin-17 (ectodomain) is truncated at the C-terminus

Because cadherin-17 is a type I transmembrane protein [26] and such proteins (ectodomains) are often shed from cells by proteases [36], we investigated whether cadherin-17 present in Tif and CCM is truncated. Visual examination of the cadherin-17 bands showed no obvious difference in the relative mobility of the molecule from either secreted or cell-associated samples. As the cytoplasmic and the transmembrane domain of cadherin-17 consist of 45 amino acids altogether, a loss of these domains by membrane-proximal cleavage was not separated from the full length form by SDS-PAGE gel used in our current study. Hence we probed samples containing secreted cadherin-17 (CCM and Tif) and samples containing cell-associated cadherin-17 (detergent cell lysate) with an antibody against the C-terminus of the molecule. Fig. 1C shows that the antibody against the C-terminus detected cell-associated cadherin-17 but not cadherin-17 from Tif or CCM. The antibody against the N-terminal part of the molecule detected the protein in all samples, suggesting that cadherin-17 is truncated at the C-terminus producing a soluble secreted form (ectodomain) of the protein.

To examine whether plasma-cadherin-17 is also truncated we precipitated the molecule with a monoclonal antibody from LIM1215 detergent cell lysates and plasma from mice carrying a LIM1215 or no xenograft. Fig. 1D shows that the cadherin-17 ectodomain-specific antibody detected the protein in detergent lysate and plasma samples. The antibody against the cadherin-17 C-terminus only detected cadherin-17 precipitated from detergent cell lysates, providing evidence that plasma-based cadherin-17 is also C-terminally truncated.

Non-classical (exosomal) secretion of Cadherin-17

We further isolated exosomes released from LIM1215 *in vitro* to determine whether they contained cadherin-17 and, if so, in what form - full-length or truncated. Exosomes are vesicles of 30-100 nm diameters which are secreted by a variety of cells, including intestinal enterocytes [37, 38]. Exosomal fractions from an OptiPrep™ density gradient were analyzed for full-length/ ectodomain cadherin-17 by western blot analysis. Fig. 2 reveals the presence of exosomes in fractions 7-9 (buoyant density ~1.11 g/mL, and presence of typical exosomal marker CD9 [31, 39, 40]). Interestingly, fractions 7-8 contain full-length cadherin-17.

DISCUSSION

This study has identified and characterized the glyco-proteins secreted by colon tumour cells into the extracellular environment, utilizing both an *in vitro* secretome model and an *in vivo* xenograft tissue interstitial fluid approach. This strategy identified secreted or ectodomain-derived membrane proteins as a source of potential biomarkers for the detection of human CRC (Table 1). We analyzed soluble-secreted glycoproteins from CRC LIM1215 cells (CCM) because secreted or shed plasma membrane proteins are hypothesized to be a source of potential biomarkers [41]. N-glycosylation is often associated with proteins classically secreted or cleaved from the plasma membrane [42] consequently non-glycosylated intracellular proteins, often released via cell death in the culture medium (estimated at 6-7 % [41]), do not interfere. Further, N-glycans regulate cell-cell communication, modulating adhesion, molecular and cellular homeostasis, receptor activation and signal transduction, endocytosis, and associated with cancer progression [43]. We detected 39 glycoproteins (Table 1), none of them cytoplasmic or nuclear demonstrating that targeting the glyco-secretome detects mainly secreted and membrane proteins (*bona fide* secreted proteins).

Furthermore the reduction to glycopeptides provided a less complex analytical sample for MS analysis.

Cell lines are artificial tumour models because *in vivo* the tissue surrounding the tumour plays an active part in tumour development. Especially expression of proteases and their inhibitors can potentially alter the tumour secretome [44-46]. To determine the effect of the microenvironment on the LIM1215 secretome we grew LIM1215 cells as xenograft tumours in immunocompromised mice. Tif, a model for the *in vivo* secretome, was isolated using a differential centrifugation strategy [18, 33]. The results confirmed three of the *in vitro* secreted proteins (Table 2). However the dominance of mouse plasma proteins (Supplementary Table S3) hindered detection of tumour-derived proteins in Tif and confounded a direct comparison between CCM and Tif. Only five tumour-derived proteins were identified in Tif. Two proteins, the polymeric immunoglobulin receptor and apolipoprotein A-I were only seen in Tif.

Collectively, 41 proteins were identified in CCM and Tif which is few compared to other studies (i.e., 262 proteins in a prostate cancer secretome [47]). However, detection of known biomarkers is thought to be critical for the validity of biomarker discovery approaches [48]. Unfortunately, known markers such as CEA are often heterogeneously glycosylated hence difficult to identify, but a reductionist approach like our glycopeptide-based strategy could improve this [48]. We identified proteins which have been evaluated as biomarkers such as the galectin-3 binding protein [49] and TIMP-1 which has been investigated as a CRC biomarker in combination with CEA [50] confirming the validity of our enrichment approach. Although we did not identify CEA in the LIM1215 secretome, it was detected in other CRC secretomes using the same approach (Simpson *et al.,* manuscript in preparation).

Table 1 lists the subcellular localisation and reported function of proteins identified in this study.

Further, we focused on proteins identified *ex vivo* as well as because not all secreted proteins (secretome proteins) necessarily passage from Tif (extracellular environment) to plasma. Of these, cadherin-17 was of interest because of its narrow GI tissue distribution and its evaluation as a plasma biomarker has not been described in CRC context. Cadherin-17 was first described as an adhesion molecule of the cadherin subfamily in rat liver and intestinal tissue and alternatively named "liver intestinal" (LI) cadherin [51]. In humans, cadherin-17 is only expressed in intestinal enterocytes and distributed on the basolateral surface but excluded from adherens junctions [26, 51]. Cell adhesion is mediated by calcium-dependent binding of two cadherin-17 dimers on adjacent cells [52] not involving the very short cytoplasmic tail [53]. The protein is regulated by the transcription factor CDX2 which is responsible for intestinal-type gene expression [54]. Cadherin-17 expression in human CRC is preserved in well-differentiated but lost in dedifferentiated carcinomas [35, 54]. Cadherin-17 has also been detected in pancreatic cancer [55], hepatocellular cancer [56] and gastric cancer [57-60] where its expression is strongly correlated with the intestinal phenotype [57, 58], gastrointestinal adenocarcinoma [61, 62], poor prognosis [58, 59], dedifferentiated phenotype [63] and lymph node metastasis [60].

These characteristics prompted us to further analyze cadherin-17 by Western blotting. Cadherin-17 was detected in LIM1215 whole cell lysate (detergent lysis), in Tif derived from LIM1215 cells xenograft tumours, and in plasma from these mice (n=3). We also showed that secretion of cadherin-17, involves loss of the C-terminus (i.e., as an ectodomain). Besides the 97K form of cadherin-17, a 75K species was also detected, with varying intensities, in CCM

and Tif. This truncated version of the molecule may be a result of extracellular degradation as it was not observed in the detergent lysate.

Exosomes are 30-100 nm in size and released by a variety of cell lines including intestinal enterocytes [38, 40, 64] via inward budding into multivesicular bodies (MVB) and fusion of the MVB with the plasma membrane allowing release of the exosomes. Exosomes contain both membrane proteins on their surface as well as cytoplasmic proteins in their interior. Their function depends on the cell of origin – e.g., dendritic cell-derived exosomes contain co-stimulatory proteins required for T-cell activation [65]. Tumour cells also secrete exosomes and recently Skog *et. al.* demonstrated that glioblastoma-derived exosomes contain both RNA and proteins which are taken up by target cells [66]. They also showed that tumour-derived exosomes, present in patient plasma, can be used as biomarkers for glioblastoma [66]. Similarly, colon cancer cell-derived exosomes secreted into plasma may have biomarker potential, hence a thorough investigation of their composition is required. Recently, exosomes derived from LIM1215 cells were characterized and 22 proteins were reported differentially expressed in exosomes upon treatment with the NSAID drug sulindac, but cadherin-17 expression was not affected [32]. Further, expression of cadherin-17 was confirmed in LIM1215 [37] and LIM1863 [31, 67] exosomes from our laboratory. Here we confirmed that cadherin-17 is present in LIM1215-derived exosomes. Due to its narrow tissue distribution, cadherin-17 could be used as a marker for intestinal-derived exosomes allowing selective targeting of intestinal exosomes among plasma exosomes facilitating their characterization.

The identification of cadherin-17 ectodomain shed by LIM1215 cells into plasma raises important questions. Firstly, if secretion of cadherin-17 unique to cancer tissue then the

protein could be a potential selective marker for CRC. Or is secretion also present in normal colonic tissue then should be present in normal human plasma through consistent shedding from intestinal enterocytes. We attempted detection of cadherin-17 in human plasma but could not obtain conclusive data confirming its presence (data not shown) - we are currently establishing a bead-based assay [68] for high sensitivity plasma detection of cadherin-17.

Another interesting aspect is the expression of cadherin-17 in gastric cancer [57-60, 63]. The data suggests that the protein is a potential biomarker for intestinal-type gastric cancer [57, 58], for dedifferentiated cancer [63], for lymph node metastasis [60] and for poor prognosis [58, 59]. Hence, in addition to its application within CRC, plasma cadherin-17 may also be useful as a potential marker for these sub-types of gastric carcinoma.

In summary, we show the versatility of analyzing the glyco-secretome of the CRC cell line LIM1215 for identifying potential biomarkers. The identification of known markers in our study, such as TIMP-1, confirms the glyco-secretome strategy reported here. Using cadherin-17 as an example, we confirm, using a mouse xenograft tumour model, that this secretome protein (from CM) is also present in Tif and plasma Further, we show that cadherin-17 is secreted in CM, Tif and plasma as an ectodomain (i.e., C-terminally truncated) (ectodomain) and as such may be a potential blood-based disease biomarker [69] for CRC. Additionally, we confirm the exosomal secretion of full length cadherin-17 where it may serve as a potential exosomal surface protein for immunocapture of specific CRC-derived exosomes from blood for the purpose of diagnostic marker development.

Acknowledgements

This work was supported by the National Health & Medical Research Council of Australia (NH&MRC) program grant #487922 (RJS)

The authors declare that there are no commercial or financial conflicts of interest.

References

1. Siegel, R., Naishadham, D., Jemal, A. *CA Cancer J Clin*. **2012**. 62 (1), 10-29.
2. Wong Kee Song, L.M. and Wilson, B.C. *Best Pract Res Clin Gastroenterol*. **2005**. 19 (6), 833-856.
3. Steele, R.J. *Surgeon*. **2006**. 4 (5), 285-291.
4. Weinberg, D.S. *Ann Intern Med*. **2008**. 148 (3), ITC2-1-ITC2-16.
5. Habermann, J.K., Bader, F.G., Franke, C., Zimmermann, K., Gemoll, T., Fritzsche, B., Ried, T., Auer, G., Bruch, H.P., Roblick, U.J. *Langenbecks Arch Surg*. **2008**. 393 (1), 93-104.
6. Whitlock, E.P., Lin, J.S., Liles, E., Beil, T.L., Fu, R. *Ann Intern Med*. **2008**. 149 (9), 638-658.
7. Duffy, M.J., van Dalen, A., Haglund, C., Hansson, L., Holinski-Feder, E., Klapdor, R., Lamerz, R., Peltomaki, P., Sturgeon, C., Topolcan, O. *Eur J Cancer*. **2007**. 43 (9), 1348-1360.
8. Wood, L.D., Parsons, D.W., Jones, S., Lin, J., Sjoblom, T., Leary, R.J., Shen, D., Boca, S.M., Barber, T., Ptak, J., Silliman, N., Szabo, S., Dezso, Z., Ustyanksky, V., Nikolskaya, T., Nikolsky, Y., Karchin, R., Wilson, P.A., Kaminker, J.S., Zhang, Z., Croshaw, R., Willis, J., Dawson, D., Shipitsin, M., Willson, J.K., Sukumar, S., Polyak, K., Park, B.H., Pethiyagoda, C.L., Pant, P.V., Ballinger, D.G., Sparks, A.B., Hartigan, J., Smith, D.R., Suh, E., Papadopoulos, N., Buckhaults, P., Markowitz, S.D., Parmigiani, G., Kinzler, K.W., Velculescu, V.E. and Vogelstein, B. *Science (New York, N.Y.* **2007**. 318 (5853), 1108-1113.
9. Sjoblom, T., Jones, S., Wood, L.D., Parsons, D.W., Lin, J., Barber, T.D., Mandelker, D., Leary, R.J., Ptak, J., Silliman, N., Szabo, S., Buckhaults, P., Farrell, C., Meeh, P., Markowitz, S.D., Willis, J., Dawson, D., Willson, J.K., Gazdar, A.F., Hartigan, J., Wu, L., Liu, C., Parmigiani, G., Park, B.H., Bachman, K.E., Papadopoulos, N., Vogelstein, B., Kinzler, K.W. and Velculescu, V.E. *Science (New York, N.Y.* **2006**. 314 (5797), 268-274.
10. Ross, D.T., Scherf, U., Eisen, M.B., Perou, C.M., Rees, C., Spellman, P., Iyer, V., Jeffrey, S.S., Van de Rijn, M., Waltham, M., Pergamenschikov, A., Lee, J.C., Lashkari, D., Shalon, D., Myers, T.G., Weinstein, J.N., Botstein, D. and Brown, P.O. *Nature genetics*. **2000**. 24 (3), 227-235.
11. Simpson, R.J., Bernhard, O.K., Greening, D.W., Moritz, R.L. *Curr Opin Chem Biol*. **2008**. 12 (1), 72-77.
12. Dowling, P. and Clynes, M. *Proteomics*. **2011**. 11 (4), 794-804.
13. Wiig, H., Tenstad, O., Iversen, P.O., Kalluri, R., Bjerkvig, R. *Fibrogenesis Tissue Repair*. **2010**. 3 12.
14. Clark, H.F., Gurney, A.L., Abaya, E., Baker, K., Baldwin, D., Brush, J., Chen, J., Chow, B., Chui, C., Crowley, C., Currell, B., Deuel, B., Dowd, P., Eaton, D., Foster, J., Grimaldi, C., Gu, Q., Hass, P.E., Heldens, S., Huang, A., Kim, H.S., Klimowski, L., Jin, Y., Johnson, S., Lee, J., Lewis, L., Liao, D., Mark, M., Robbie, E., Sanchez, C., Schoenfeld, J., Seshagiri, S., Simmons, L., Singh, J., Smith, V., Stinson, J., Vagts, A., Vandlen, R., Watanabe, C., Wieand, D., Woods, K., Xie, M.H., Yansura, D., Yi, S., Yu, G., Yuan, J., Zhang, M., Zhang, Z., Goddard, A., Wood, W.I., Godowski, P. and Gray, A. *Genome Res*. **2003**. 13 (10), 2265-2270.

15. Ahn, S.-M. and Simpson, R.J. *Proteomics Clin. Appl.* **2007**. (1), 1004-1015.
16. Carmeliet, P. and Jain, R.K. *Nature*. **2000**. 407 (6801), 249-257.
17. Simpson, R.J., Lim, J.W., Moritz, R.L., Mathivanan, S. *Expert review of proteomics*. **2009**. 6 (3), 267-283.
18. Celis, J.E., Gromov, P., Cabezon, T., Moreira, J.M., Ambartsumian, N., Sandelin, K., Rank, F., Gromova, I. *Mol. Cell. Proteomics*. **2004**. 3 (4), 327-344.
19. Haslene-Hox, H., Oveland, E., Berg, K.C., Kolmannskog, O., Woie, K., Salvesen, H.B., Tenstad, O., Wiig, H. *PLoS ONE*. **2011**. 6 (4), e19217.
20. Stone, M.D., Odland, R.M., McGowan, T., Onsongo, G., Tang, C., Rhodus, N.L., Jagtap, P., Bandhakavi, S., Griffin, T.J. *Clin Proteomics*. **2010**. 6 (3), 75-82.
21. Hathout, Y. *Expert review of proteomics*. **2007**. 4 (2), 239-248.
22. Whitehead, R.H., Macrae, F.A., St John, D.J., Ma, J. *J. Natl. Cancer Inst.* **1985**. 74 (4), 759-765.
23. Bernhard, O.K., Kapp, E.A., Simpson, R.J. *Journal of proteome research*. **2007**. 6 (3), 987-995.
24. Zhang, H., Li, X.J., Martin, D.B., Aebersold, R. *Nature biotechnology*. **2003**. 21 (6), 660-666.
25. Zhou, Y., Aebersold, R., Zhang, H. *Anal Chem*. **2007**. 79 (15), 5826-5837.
26. Gessner, R. and Tauber, R. *Annals of the New York Academy of Sciences*. **2000**. 915 136-143.
27. Wachsberger, P.R., Burd, R., Marero, N., Daskalakis, C., Ryan, A., McCue, P., Dicker, A.P. *Clinical Cancer Res.* **2005**. 11 (2 Pt 1), 835-842.
28. Celis, J.E., Moreira, J.M., Cabezon, T., Gromov, P., Friis, E., Rank, F., Gromova, I. *Mol. Cell. Proteomics*. **2005**. 4 (4), 492-522.
29. Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J., Klenk, D.C. *Anal Biochem*. **1985**. 150 (1), 76-85.
30. Kapp, E.A., Schutz, F., Reid, G.E., Eddes, J.S., Moritz, R.L., O'Hair, R.A., Speed, T.P., Simpson, R.J. *Anal Chem*. **2003**. 75 (22), 6251-6264.
31. Tauro, B.J., Greening, D.W., Mathias, R.A., Ji, H., Mathivanan, S., Scott, A.M., Simpson, R.J. *Methods*. **2012**. 56 (2), 293-304.
32. Ji, H., Greening, D.W., Kapp, E.A., Moritz, R., Simpson, R.J. *Proteomics Clin. Appl.* **2009**. 3 433–451.
33. Wiig, H., Aukland, K., Tenstad, O. *American Journal of Physiology - Heart and Circulatory Physiology*. **2003**. 284 (1), H416-424.
34. Ito, R., Oue, N., Yoshida, K., Kunimitsu, K., Nakayama, H., Nakachi, K., Yasui, W. *Virchows Arch*. **2005**. 447 (4), 717-722.
35. Takamura, M., Ichida, T., Matsuda, Y., Kobayashi, M., Yamagiwa, S., Genda, T., Shioji, K., Hashimoto, S., Nomoto, M., Hatakeyama, K., Ajioka, Y., Sakamoto, M., Hirohashi, S. and Aoyagi, Y. *Cancer letters*. **2004**. 212 (2), 253-259.
36. Bech-Serra, J.J., Santiago-Josefat, B., Esselens, C., Saftig, P., Baselga, J., Arribas, J., Canals, F. *Molecular and cellular biology*. **2006**. 26 (13), 5086-5095.
37. Mathivanan, S., Lim, J.W., Tauro, B.J., Ji, H., Moritz, R.L., Simpson, R.J. *Mol. Cell. Proteomics*. **2010**. 9 (2), 197-208.
38. van Niel, G., Raposo, G., Candalh, C., Boussac, M., Hershberg, R., Cerf-Bensussan, N., Heyman, M. *Gastroenterology*. **2001**. 121 (2), 337-349.

39. Caby, M.P., Lankar, D., Vincendeau-Scherrer, C., Raposo, G., Bonnerot, C. *Int Immunol*. **2005**. 17 (7), 879-887.
40. Simpson, R.J., Jensen, S.S., Lim, J.W. *Proteomics*. **2008**. 8 (19), 4083-4099.
41. Kulasingam, V. and Diamandis, E.P. *Mol. Cell. Proteomics*. **2007**. 6 (11), 1997-2011.
42. Roth, J. *Chemical reviews*. **2002**. 102 (2), 285-303.
43. Ohtsubo, K. and Marth, J.D. *Cell*. **2006**. 126 (5), 855-867.
44. Radisky, D.C. and Bissell, M.J. *Science (New York, N.Y.* **2004**. 303 (5659), 775-777.
45. Lopez-Otin, C. and Matrisian, L.M. *Nat Rev Cancer*. **2007**. 7 (10), 800-808.
46. Liotta, L.A. and Kohn, E.C. *Nature*. **2001**. 411 (6835), 375-379.
47. Sardana, G., Marshall, J., Diamandis, E.P. *Clinical chemistry*. **2007**. 53 (3), 429-437.
48. Lescuyer, P., Hochstrasser, D., Rabilloud, T. *Journal of proteome research*. **2007**. 6 (9), 3371-3376.
49. Ulmer, T.A., Keeler, V., Loh, L., Chibbar, R., Torlakovic, E., Andre, S., Gabius, H.J., Laferte, S. *Journal of cellular biochemistry*. **2006**. 98 (5), 1351-1366.
50. Ishida, H., Murata, N., Hayashi, Y., Tada, M., Hashimoto, D. *Surg Today*. **2003**. 33 (12), 885-892.
51. Berndorff, D., Gessner, R., Kreft, B., Schnoy, N., Lajous-Petter, A.M., Loch, N., Reutter, W., Hortsch, M., Tauber, R. *The Journal of cell biology*. **1994**. 125 (6), 1353-1369.
52. Wendeler, M.W., Drenckhahn, D., Gessner, R., Baumgartner, W. *Journal of molecular biology*. **2007**. 370 (2), 220-230.
53. Kreft, B., Berndorff, D., Bottinger, A., Finnemann, S., Wedlich, D., Hortsch, M., Tauber, R., Gessner, R. *The Journal of cell biology*. **1997**. 136 (5), 1109-1121.
54. Hinoi, T., Lucas, P.C., Kuick, R., Hanash, S., Cho, K.R., Fearon, E.R. *Gastroenterology*. **2002**. 123 (5), 1565-1577.
55. Takamura, M., Sakamoto, M., Ino, Y., Shimamura, T., Ichida, T., Asakura, H., Hirohashi, S. *Cancer science*. **2003**. 94 (5), 425-430.
56. Wang, X.Q., Luk, J.M., Leung, P.P., Wong, B.W., Stanbridge, E.J., Fan, S.T. *Clinical Cancer Res.* **2005**. 11 (2 Pt 1), 483-489.
57. Motoshita, J., Nakayama, H., Taniyama, K., Matsusaki, K., Yasui, W. *Pathology international*. **2006**. 56 (4), 200-205.
58. Ge, J., Chen, Z., Wu, S., Yuan, W., Hu, B. *Clin Oncol (R Coll Radiol)*. **2008**. 20 (4), 275-283.
59. Tian, M.M., Zhao, A.L., Li, Z.W., Li, J.Y. *World journal of gastroenterology : WJG*. **2007**. 13 (23), 3189-3198.
60. Park, S.S., Kang, S.H., Park, J.M., Kim, J.H., Oh, S.C., Lee, J.H., Chae, Y.S., Kim, S.J., Kim, C.S., Mok, Y.J. *Annals of surgical oncology*. **2007**. 14 (1), 94-99.
61. Panarelli, N.C., Yantiss, R.K., Yeh, M.M., Liu, Y., Chen, Y.T. *American journal of clinical pathology*. **2012**. 138 (2), 211-222.
62. Ai, J., Tang, Q., Wu, Y., Xu, Y., Feng, T., Zhou, R., Chen, Y., Gao, X., Zhu, Q., Yue, X., Pan, Q., Xu, S., Li, J., Huang, M., Daugherty-Holtrop, J., He, Y., Xu, H.E., Fan, J., Ding, J. and Geng, M. *J Natl Cancer Inst*. **2011**. 103 (22), 1696-1712.

63. Dong, W., Yu, Q., Xu, Y. *Dig Dis Sci*. **2007**. 52 (2), 536-542.
64. Mathivanan, S., Ji, H., Simpson, R.J. *Journal of proteomics*. **2010**. 73 (10), 1907-1920.
65. Stoorvogel, W., Kleijmeer, M.J., Geuze, H.J., Raposo, G. *Traffic*. **2002**. 3 (5), 321-330.
66. Skog, J., Wurdinger, T., van Rijn, S., Meijer, D.H., Gainche, L., Sena-Esteves, M., Curry, W.T., Jr., Carter, B.S., Krichevsky, A.M., Breakefield, X.O. *Nature Cell. Biol.* **2008**. 10 (12), 1470-1476.
67. Tauro, B.J., Greening, D.W., Mathias, R.A., Mathivanan, S., Ji, H., Simpson, R.J. *Mol. Cell. Proteomics*. **2012**.
68. Bernhard, O.K., Mathias, R.A., Barnes, T.W., Simpson, R.J. *Methods in molecular biology*. **2011**. 728 195-206.
69. Liotta, L.A., Ferrari, M., Petricoin, E. *Nature*. **2003**. 425 (6961), 905.

Table 1 - Glycoproteins identified in LIM1215 CCM and Tif

Accession Number	Gene Name	Protein Description	Subcellular Localisation	Function	LIM1215 CCM		LIM1215 preTif	
					Identified peptides[a]	Sequence coverage (%)[b]	Identified peptides[a]	Sequence coverage (%)[b]
Q13443	ADAM9	Disintegrin and metalloproteinase domain-containing protein 9	Plasma membrane (Type I)	Protease	1	2.4		
Q09328	MGAT5	Alpha-1,6-mannosylglycoprotein 6-beta-N-acetylglucosaminyltransferase A	Vesicular membrane (Type II)	Non-protease enzyme	1	1.3		
P15514	AREG	Amphiregulin	Secreted	Growth/Signaling	1	4.4		
Q06481	APLP2	Amyloid-like protein 2	Plasma membrane (Type I)	Regulation of hemostasis/Coagulation	1	1.3		
P02647	APOA1	Apolipoprotein A-1	Secreted	Cholestrol Transport			1	8.6
Q75882	ATRN	Attractin	Secreted	Growth/Signaling	1	0.6		
P02749	APOH	Beta-2-glycoprotein 1	Secreted	Binds negatively charged substances	1	3.8		
P61769	B2M	Beta-2-microglobulin	Secreted	Defense	1	11.8		
P07686	HEXB	Beta-hexosaminidase subunit beta	Vesicular lumen	Non-protease enzyme	1	4.2		
P43251	BTD	Biotinidase	Secreted	Non-protease enzyme	1	4		
Q12864	CDH17	Cadherin-17	Plasma membrane (Type I)	Adhesion	2	4.4		
P07339	CTSD	Cathepsin D	Vesicular lumen	Protease	4	6.8		
P09668	CTSH	Cathepsin H	Vesicular lumen	Protease	1	3.6		
P16070	CD44	CD44 antigen	Plasma membrane (Type I)	Adhesion	3	4.2		
Q9BY67	CADM1	Cell adhesion molecule 1	Plasma membrane (Type I)	Adhesion	1	7.5		
Q14126	DSG2	Desmoglein-2	Plasma membrane (Type I)	Adhesion	1	0.9		
Q14118	DAG1	Dystroglycan	Plasma membrane (Type I)	Adhesion	1	2.5		
P29323	EPHB2	Ephrin type-B receptor 2	Plasma membrane (Type I)	Growth/Signaling	1	1.2		
P20827	EFNA1	Ephrin-A1	Plasma membrane (GPI anchor)	Growth/Signaling	1	6.3		
P52797	EFNA3	Ephrin-A3	Plasma membrane (GPI anchor)	Growth/Signaling	2	9.7		
P52798	EFNA4	Ephrin-A4	Plasma membrane (GPI anchor)	Defense	1	5.5		
Q08380	LGALS3BP	Galectin-3-binding protein	Secreted	Adhesion	4	8.9	1	2.2
Q92820	GGH	Gamma-glutamyl hydrolase	Secreted	Non-protease enzyme	1	4.1		
Q8NBJ4	GOLM1	Golgi membrane protein 1	Vesicular membrane (Type II)	Defense	3	9		
Q99988	GDF15	Growth/differentiation factor 15	Secreted	Growth/Signaling	1	6.8		
P08581	MET	Hepatocyte growth factor receptor	Plasma membrane (Type I)	Protease	1	0.9		
P30447	HLA-A	HLA class I histocompatibility antigen, A-23 alpha chain	Plasma membrane (Type I)	Defense	2	6.8		
P10321	HLA-C	HLA class I histocompatibility antigen, Cw-7 alpha chain	Plasma membrane (Type I)	Defense	1	4.1		
P30508	HLA-C	HLA class I histocompatibility antigen, Cw-12 alpha chain	Plasma membrane (Type I)	Defense	1	4.1		
P01033	TIMP1	Metalloproteinase inhibitor 1	Secreted	Protease inhibitor	1	7.2		
P41271	NBL1	Neuroblastoma suppressor of tumorigenicity 1	Secreted	Growth/Signaling	1	7.8		
Q92823	NRCAM	Neuronal cell adhesion molecule	Plasma membrane (Type I)	Adhesion	1	0.8		
P80188	LCN2	Neutrophil gelatinase-associated lipocalin	Secreted	Transport	1	6.6		
P50897	PPT1	Palmitoyl-protein thioesterase 1	Vesicular lumen	Non-protease enzyme	1	6.2		
P15151	PVR	Poliovirus receptor	Plasma membrane (Type I)	Defense	2	7.2		
P07602	PSAP	Proactivator polypeptide	Vesicular lumen	Degradation of sphingolipids	1	3.2		
P01833	PIGR	Polymeric immunoglobulin receptor	Plasma membrane (Type I)	Transport			1	2.9
O95084	PRSS23	Serine protease 23	Secreted	Protease	1	3.7		
Q9Y5Y6	ST14	Suppressor of tumorigenicity protein 14	Plasma membrane (Type II)	Protease	2	5		
Q9GZX9	TWSG1	Twisted gastrulation protein homolog 1	Secreted	Growth/Signaling	1	4.5		
Q13308	PTK7	Tyrosine-protein kinase-like 7	Plasma membrane (Type I)	Adhesion	3	4.4		

a Number of identified peptides, based on - (i) All peptide matches with a rank >1 or an associated E-value of >1 were removed (refer to Supplemental Table S1 and S2 for corresponding E-values for each identified peptide), (ii) Peptides with a Protein Score < Identity score that have no deamidation in a glycosylation motif were removed, (iii) Peptide matches to queries which had already been matched were removed, (iv) All peptides were BLAST searched (http://au.expasy.org/cgi-bin/blast.pl) against the mammalian entries of the Swiss-Prot database (v04/2009) to determine all proteins to which the peptide maps taking into account the rules of parsimony, and (v) Proteins were considered identified by either 2 peptides or one fully tryptic glycopeptide. MS/MS of each individual spectra was analysed.

b Sequence coverage, expressed as the number of amino acids spanned by the assigned peptides from the respective experiments divided by the length of the unmodified sequence as listed in the Swiss-Prot Database (v04/2009)

Table 2 - Identified human prototypic glycopeptides from LIM1215 xenograft-derived tumour interstitial fluid (Tif)

Accession Number	Gene Name	Protein Description	Peptide Sequence[a]	Mascot Peptide Ions Score[b]	E-value[c]	Coverage (%)	Charge	Mr (Calc)[d]	Technical replicate[e]
P02647	APOA1	Apolipoprotein A-I	R.pyQGLLPVLESFKVSFLSALEEYTK.K	95.33	0.0000094	8.6	2	2580.38	1,2,3
P02647	APOA1	Apolipoprotein A-I	R.QGLLPVLESFKVSFLSALEEYTK.K	85.32	0.0000094	8.6	2	2597.4	1,2
Q12864	CDH17	Cadherin-17	K.APKPVEM*VEN^STDPHPIK.I	31	0.19	4.5	3	2004.98	3
Q12864	CDH17	Cadherin-17	K.KQDTPQYN^LTIEVSDKDFK.T	52	0.11	4.5	3	2269.11	1,3
Q08380	LGALS3BP	Galectin-3 binding protein	R.ALGFEN^ATQALGR.A	67.4	0.0032	2.2	2	1347.68	1,3
P01833	PIGR	Polymeric-immunoglobulin receptor	R.LSLLEEPGN^GTFTVILNQLTSR.D	113	0.00000018	2.9	2	2402.27	1,2,3
Q13308	PTK7	Tyrosine-protein kinase-like 7	R.M*HIFQN^GSLVIHDVAPEDSGR.Y	49.86	0.32	2	3	2338.1	1,2,3

[a] Amino acid abbreviations used are: N^, deamidated asparagine; M*, oxidized methionine; pyQ, N-terminal pyroglutamic acid. Amino acids flanking the peptide on both sides are shown separated from the peptide sequence by a full stop.

[b] Mascot™ peptide ions scores which is -10*Log10(P), where P is the absolute probability that the match is a random event (http://www.matrixscience.com/help/scoring_help.html).

[c] Mascot peptide expectation value (a value of 1 indicates that 1 other peptide could correspond to the same score based on random matching).

[d] Calculated monoisotopic masses of uncharged peptides

[e] Indicates in which of the three LC-MS/MS analyses that were performed on the sample the peptide was identified.

Shaded proteins indicate protein/peptide sequence identified in the LIM1215 glyco-secretome (Supplemental Table T1)

Figure legends

Figure 1 - Characterization of Cadherin-17 secretion from LIM1215 cells. Panel A, detection of cadherin-17 by Western blotting in LIM1215 CCM, cell lysate, Tif and plasma samples. The lanes contain the following samples: Tif (20 µg protein); CCM (20 µg protein); detergent cell lysate (20 µg protein) and plasma (20 µg protein) from a mouse bearing LIM1215 xenograft. Cadherin-17 was detected with a polyclonal antibody against cadherin-17 ectodomain in all but the plasma samples predominantly as a 97 kDa form. Panel B, human xenograft-derived Cadherin-17 can be detected in mouse blood: Cadherin-17 was precipitated from 75 µL plasma from mice bearing a LIM1215 xenograft or control mice bearing no xenograft with a monoclonal antibody against ectodomain (n=3). Western detection with a polyclonal antibody (to the cadherin-17 ectodomain) shows the 97 K of cadherin-17 in samples from xenograft bearing animals but not in the control. Panel C, cadherin-17 is secreted as a C-terminally truncated form. Secreted or cellular samples from LIM1215 cells were probed with polyclonal antibodies against the ectodomain (top, 75 kDa) and C-terminal (bottom, 95 kDa) domain of cadherin-17. The lanes contain the following samples: concentrated CM (20 µg protein); Tif (20 µg protein); detergent lysate (10^5 cells); cadherin-17 IP from 1.0 mL detergent lysate using a monoclonal antibody; control IP from 1.0 mL detergent lysate using IgG1. The C-terminal domain is only detected in the cell-associated samples demonstrating that the molecule is C-terminally truncated when secreted. Panel D, plasma cadherin-17 also lacks its C- terminus. Cadherin-17 was immunoprecipitated with a monoclonal antibody or isotype control (IgG1) from LIM1215 detergent lysate or 75 µL plasma from mice carrying either a LIM1215 or no xenograft. Western detection with a polyclonal antibody against the cadherin-17 ectodomain or the C-terminus shows that plasma cadherin-17 is C-terminally truncated. The 75Kform of cadherin-17 is detectable in the IP

from detergent lysate but not in the IP from plasma in the upper panel below the 97K form. Bands from immunoglobulins and non-specific bands are visible below the cadherin-17 bands in both panels.

Figure 2 - Cadherin-17 is non-classically secreted via exosomes. OptiPrep™ density gradient fractions (F6-10) containing LIM1215-derived exosomes were analyzed by SDS-PAGE, Western blot and detected with antibodies to the cadherin-17 ectodomain (top), the cadherin-17 C-terminus (middle) and the exosomal marker CD9 (bottom). Cadherin-17 is present in LIM1215 exosomal fractions.

Figure 1

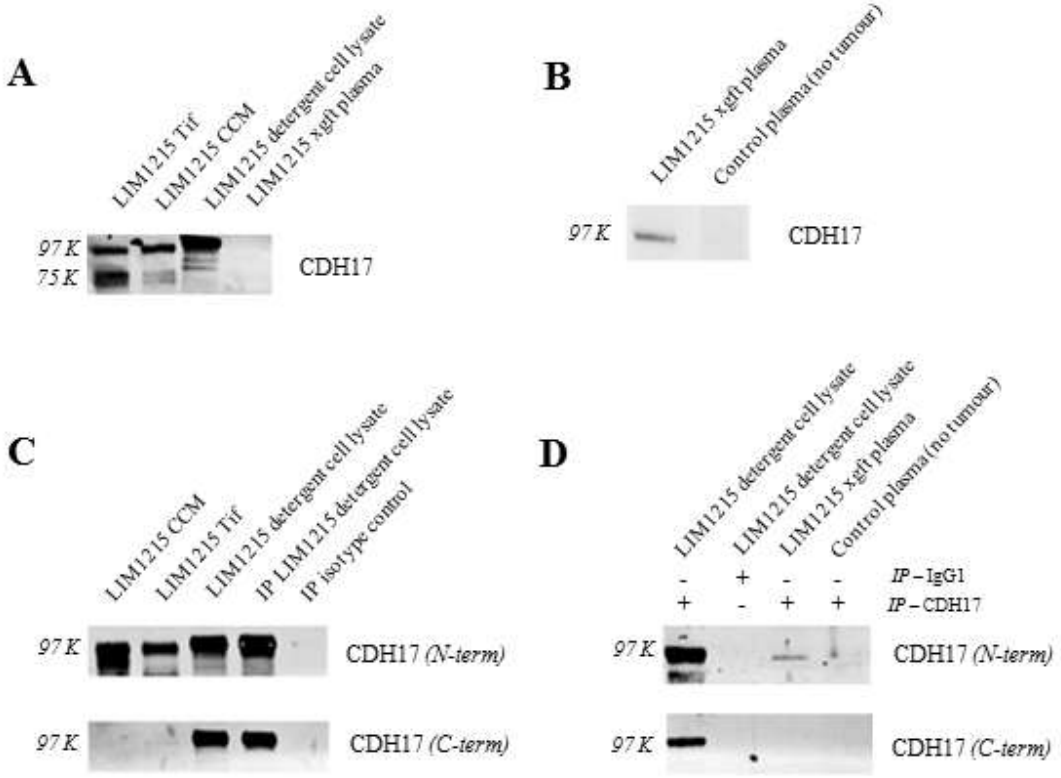
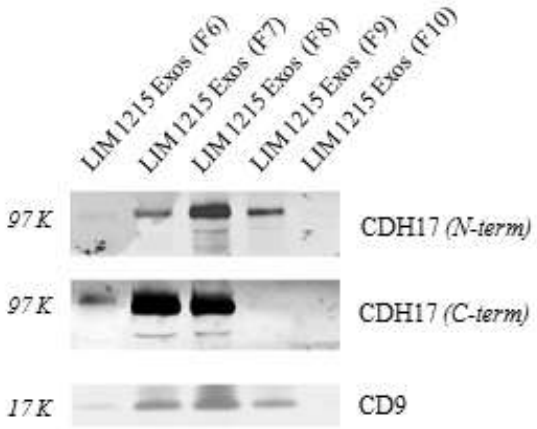

A
LIM1215 Tif
LIM1215 CCM
LIM1215 detergent cell lysate
LIM1215 xgft plasma
97 K
75 K
CDH17
B
LIM1215 xgft plasma
Control plasma (no tumour)
97 K
CDH17
C
LIM1215 CCM
LIM1215 Tif
LIM1215 detergent cell lysate
IP LIM1215 detergent cell lysate
IP isotype control
97 K
CDH17 (N-term)
97 K
CDH17 (C-term)
D
LIM1215 detergent cell lysate
LIM1215 detergent cell lysate
LIM1215 xgft plasma
Control plasma (no tumour)
- + - - IP – IgG1
+ - + + IP – CDH17
97 K
CDH17 (N-term)
97 K
CDH17 (C-term)

Figure 2


LIM1215 Exos (F6)
LIM1215 Exos (F7)
LIM1215 Exos (F8)
LIM1215 Exos (F9)
LIM1215 Exos (F10)
97 K
CDH17 (N-term)
97 K
CDH17 (C-term)
17 K
CD9

Supporting information

Supplementary Table S1 - Identified glycopeptides from LIM1215 CCM

Supplementary Table S2 - Identified peptides containing N-glycosylation sites in secretomes (LIM1215 CCM and LIM1215 Tif)

Supplementary Table S3 - Proteins identified in the glyco-enriched LIM1215 Tif