Colon tumour secretopeptidome: Insights into endogenous proteolytic cleavage events in the colon tumour microenvironment

David W. Greening[1+], Eugene A. Kapp[2+], Hong Ji[1], Terry P. Speed[2] and Richard J. Simpson[1*]

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

[2] *Bioinformatics Division, Walter & Eliza Hall Institute of Medical Research, Victoria, 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

[+] Note these authors contributed equally to this work

Running title: Cancer secretopeptidome

Keywords: peptidome, secretome, low-molecular weight, peptides, proteolytic cleavage, ectodomain, intramembrane proteolysis (RIP), MSPro, Percolator, MEROPS, intracytoplasmic domain, cancer, tumour

Abbreviations

ADAM, a disintegrin and metalloproteinase; APP, amyloid precursor protein; BSG, basigin; CALR, calreticulin; CDH1, epithelial cadherin; CEA, carcinoembryonic antigen; CM, culture medium; CRC, colorectal cancer; CTF, carboxy-terminal fragment; DG, dystroglycan; DSG2, desmoglein-2; ECM, extracellular matrix; FDR, false discovery rate; GDF15, growth/differentiation factor 15; GPA33, cell surface A33 antigen; GPC1, glypican-1; GPI, glycosylphosphatidylinositol; HNPCC, hereditary nonpolyposis colorectal cancer; ICD, intracellular domain; intramembrane-cleaving proteases, I-CLiPs; IDE, insulin degrading enzyme, insulysin; MMP, matrix metalloproteinase; MT-MMP, membrane-type MMP; NMWL, nominal molecular weight limit; NTF, N-terminal fragment; PEP, posterior error probability; PM, plasma membrane; PSM, peptide spectral matches; RIP, regulated intramembrane proteolysis; sTIC, summed total ion current; TGFBI, transforming growth factor-beta-induced protein ig-h3; TIC, total ion current; TMHMM, transmembrane hidden markov model; TP53I11, tumor protein p53-inducible protein 11; TNFRSF9, tumor necrosis factor receptor superfamily member 9; XCorr, cross-correlation score.

ABSTRACT

The secretopeptidome comprises endogenous peptides derived from proteins secreted into the tumour microenvironment through classical and non-classical secretion. This study characterised the low-M_r (<3kDa) component of the human colon tumour (LIM1215, LIM1863) secretopeptidome, as a first step towards gaining insights into extracellular proteolytic cleavage events in the tumour microenvironment. Based on two biological replicates, this secretopeptidome isolation strategy utilised differential centrifugal ultrafiltration in combination with analytical RP-HPLC and nanoLC-MS/MS. Secreted peptides were identified using a combination of Mascot and post-processing analyses including MSPro re-scoring, extended feature sets and Percolator, resulting in 474 protein identifications from 1228 peptides (≤1% *q*-value, ≤5% PEP) - a 36% increase in peptide identifications when compared with conventional Mascot (homology ionscore thresholding). In both colon tumour models, 122 identified peptides were derived from 41 cell surface protein ectodomains, 23 peptides (12 proteins) from regulated intramembrane proteolysis (RIP), and 12 peptides (9 proteins) generated from intracellular domain proteolysis. Further analyses using the protease/substrate database MEROPS (http://merops.sanger.ac.uk/), revealed 335 (71%) proteins classified as originating from classical/non-classical secretion, or the cell membrane. Of these, peptides were identified from 42 substrates in MEROPS with defined protease cleavage sites, while peptides generated from a further 205 substrates were fragmented by hitherto unknown proteases. A salient finding was the identification of peptides from 88 classical/ non-classical secreted substrates in MEROPS, implicated in tumour progression and angiogenesis (FGFBP1, PLXDC2), cell-cell recognition and signalling (DDR1, GPA33), and tumour invasiveness and metastasis (MACC1, SMAGP); the nature of the proteases responsible for these proteolytic events is unknown. To confirm reproducibility of peptide fragment abundance in this study, we report the identification of a specific cleaved peptide fragment in the secretopeptidome from the colon-specific GPA33 antigen in 4/14 human CRC models. This improved secretopeptidome isolation and characterisation strategy has extended our understanding of endogenous peptides generated through proteolysis of classical/ non-classical secreted proteins, extracellular proteolytic processing of cell surface membrane proteins, and peptides generated through

RIP. The novel peptide cleavage site information in this study provides a useful first step in detailing proteolytic cleavage associated with tumourigenesis and the extracellular environment.

INTRODUCTION

Extracellular proteolysis represents a dynamic role in cell regulation, signalling, and tissue homeostasis. In cancer, proteolytic activity is an important component regulating intercellular communication throughout the surrounding microenvironment, with altered proteolysis promoting deregulated tumour growth, tissue remodelling, inflammation, tissue invasion, and metastasis [1]. Tumour cells continually modulate the expression of cell surface adhesion components in response to the requirements of the microenvironment. Recent evidence indicates that different proteolytic mechanisms control the expression and activity of adhesion components on tumour cells and bioactive peptide fragments throughout the extracellular milieu (reviewed [2]). Various proteases have been demonstrated to promote controlled release of cell surface membrane protein fragments. These peptide fragments can function as reporters or effectors of signalling pathways, modulating cell adhesion, proteolytic activity, and transcriptional activation [2, 3]. Typically, these substrates include members of the cadherin and catenin families, integrins and ephrins, in addition to the immunoglobulin superfamily. During tumour progression, it has been reported that different mechanisms exist whereby cleavage of soluble membrane proteins influence cell-cell adhesion and/or activating signals that promote cell proliferation and/or migration [4]. Cleaved fragments have been proposed to associate either with integrins and/or components of the extracellular matrix (ECM) [5, 6], act in an autocrine or paracrine manner to activate distinct signalling events [6, 7], or cytoplasmic fragments may activate aberrant signalling or potentiate normal signalling when not anchored to the plasma membrane (PM) [8]. Ectodomain shedding and intramembrane proteolysis are becoming critical elements of many diverse intra- and intercellular signalling events mediated by small peptide fragments.

Ectodomain shedding is a proteolytic mechanism involved in the coordinated cleavage and release of extracellular domains of cell membrane proteins [9]. This process, and the sheddases involved, control the biological activity and function of various cell surface membrane proteins [10]. Ectodomain shedding is typically mediated by proteases of the metalloproteinase (MMP) and a

disintegrin and metalloproteinase (ADAM) families. The diversity of proteins which undergo ectodomain shedding indicates that the process is essential not only during normal mammalian development but also in pathological conditions including inflammation and cancer [11, 12]. These sheddases modulate proteolysis of the ECM, proteolytic activation, processing and release of cytokines, growth factors and receptors, and regulate signal transduction [2]. Ectodomain shedding has been reported in various substrates, including basigin (BSG), CD44, CD166 (ALCAM), desmoglein-2 (DSG2), E-cadherin (CDH1), EpCAM, syndecans (SDC1/4), integrins, and ephrin receptors [2, 12]. Cleavage of CDH1 has been demonstrated to yield a soluble CDH1 fragment (CTF1 chain) which subsequently prevents cell-cell aggregation [13, 14], induces cell invasiveness [13, 15], and promotes epithelial-to-mesenchymal transition in mammary epithelial cells [16]. Further, soluble CDH1 fragments derived through ectodomain shedding have recently been identified as diagnostic markers for late-stage human colorectal carcinoma and familial adenomatous polyposis [17].

Proteolysis of integral membrane proteins can also occur within transmembrane domains [18]. This type of regulated intramembrane proteolysis (RIP) occurs following processing of the ectodomain (to less than 30 amino acids), allowing further cleavage of products within the cellular membrane, releasing the intracellular domain (ICD) into the cytoplasm [19]. Such intramembrane proteolysis is an important mechanism for regulating cellular responsiveness, cell adhesion, and protein trafficking [3, 18]. Intramembrane proteolysis is mediated by multi-spanning intramembrane-cleaving proteases (I-CLiPs) [20], which include the zinc metalloproteases (site-2 proteases; S2P), aspartyl proteases (presenilins and signal peptide peptidases) and serine proteases (rhomboids) [3]. Substrates include amyloid precursor protein (APP), CD44, EpCAM, pro-TNF-α, TNF receptors, IL-1R2, IL-2-α receptor, IL-6 receptor, Notch ligands Jagged and Delta, Notch receptor, and ErbB4 [19]. ICD domains from several different substrates including EpCAM [8] and Notch [21], have been shown to target the nucleus, modulating gene transcription. In addition to PM-tethered proteins, I-CLiPs can also liberate intracellular membrane-tethered proteins to the lumen of the endoplasmic reticulum and Golgi [3], with C-terminal fragments released into the extracellular space by hitherto unknown

mechanisms. Detailed knowledge about the function of I-CLiPs, the spectrum of substrates, and regulation of intra- or extracellular processes remains to be elucidated.

Recent efforts have been directed towards developing high throughout proteomic screens to identify protease cleavage events and protease substrates in complex biological samples [22-27]. However, due to their molecular weight (<3kDa) and solubility, issues with identifying specific cleavage sites and coverage of peptide fragments remain challenging. Peptidomics is currently focused towards comprehensively studying peptides cleaved from precursor proteins by endogenous proteases [28-30]. Peptidomics has the potential to define processing sites of precursor proteins in their native, intact form without the requirement for enzymatic digestion in typical proteomic workflows. Using a peptide profiling approach, we have characterized the secretopeptidome of human CRC cells LIM1215 [31] and LIM1863 [32] based on two biological replicates, and identified various extracellular endogenous proteolytic cleavage events associated with cell surface membrane proteins. We demonstrate that these cleavage products are not degradation products of the proteasome, based on comparison with several studies investigating human endogenous protein ubiquitination sites with mass spectrometry [33, 34]. We also demonstrate that the combined use of the Mascot search algorithm, MSPro cross-correlation (XCorr) re-scoring, extended feature sets, and the support vector machine-learning Percolator algorithm significantly increases the number of peptides identified compared with the conventional Mascot homology ionscore thresholding search strategy. We have extended these studies to compare sequence and substrate cleavage information in the protease and substrate database MEROPS (http://merops.sanger.ac.uk/), identifying various substrates involved in cell-cell adhesion (CADM1), tumour progression and angiogenesis (EFNA1, FGFBP1, PLXDC2), and tumour invasiveness and metastasis (CD9, MACC1, SMAGP) processed by hitherto unidentified proteases. To confirm reproducibility of peptide fragment abundance in this study, we report the identification of a specific cleaved peptide fragment in the secretopeptidome from the colon-specific GPA33 antigen in 4/14 human CRC models. This improved secretopeptidome isolation and characterisation strategy described here has extended our understanding of endogenous peptides generated through proteolysis of classical/non-classical secreted proteins from colon tumour cells, the

extracellular proteolytic processing of cell surface membrane proteins, and peptides generated through RIP. The derived peptide cleavage site information in this study provides a useful first step in detailing proteolytic cleavage associated with tumourigenesis and the extracellular environment.

EXPERIMENTAL PROCEDURES

Cell culture and reagents – Human colorectal cancer (CRC) cell lines LIM1215 [31] and LIM1863 [32] were from the Ludwig Institute for Cancer Research, Melbourne. LIM1215 cells were grown as adherent monolayers (3×10^8 cells/10 dishes for each analysis) in RPMI-1640 medium (Invitrogen, Carlsbad, CA) supplemented with 5% FCS (CSL, Melbourne), 100 U penicillin and 100 mg/mL streptomycin (Sigma-Aldrich), and incubated at 37 °C in a humidified atmosphere containing 10% CO_2 until sub-confluent. At ~80% confluency, cells were washed (×4) with 10 mL of RPMI-1640 medium, and cultured (15 mL) for a further 24 h in serum-free RPMI media supplemented with 0.8% insulin transferrin-selenium (ITS) solution (Invitrogen). LIM1863 cells were cultured as floating organoids in RPMI-1640 medium (Invitrogen) supplemented with 5% FCS (CSL), α-thioglycerol (10 μM), insulin (25 units/L), hydrocortisone (1 mg/L), with 10% CO_2 at 37 °C. LIM1863 cells (1.5×10^8 cells/3 flasks for each analysis) were washed (×4) with 30 mL of RPMI-1640 medium, and cultured for 24 h in 150 mL of serum-free RPMI media supplemented with 0.8% ITS solution (Invitrogen). All remaining cell lines (SW1222, LS174T, LIM1899, LIM2537, HT29, LIM2405, LIM2550, LOVO, LIM2099, LIM2408, HCA7, and CACO2) were grown as above. Culture medium (CM) from two biological replicates was harvested (LIM1215, 150 mL; LIM1863, 450 mL), and protease inhibitors (Complete Protease inhibitor cocktail tablets, Roche). CM was clarified by centrifugation for 5 min at 480*g* at 4 °C, and for 10 min at 2000*g* at 4 °C.

Differential centrifugal ultrafiltration – CM supernatants from each cell line were fractionated through a series of Amicon® Ultra-15 (Millipore, MA, USA) and Macrosep® Omega (Pall Life Sciences) centrifugal filter devices. Membranes were pre-rinsed with deionized water (A10-

Synthesis™ water-polishing system, Millipore, MA, USA), and subsequently RPMI-1640 medium. CM supernatants were filtered initially through Amicon® Ultra-15 100,000 nominal molecular weight limit (NMWL) centrifugal filter devices (3,000*g*, 4 °C), with each filtrate (flow-through, <100K) subsequently filtered through an Amicon® Ultra-15 3,000 NMWL filter (3,000*g*, 4 °C). Filtrates (flow-through, <3kDa) were concentrated to ~2 mL using a 1,000 NMWL Macrosep® Omega centrifugal device (2,500*g*, 4 °C), to obtain the secretopeptidome (1-3K). Centrifugation conditions were optimized as previously described [35] to ensure 95% (v/v) filtrate recovery. Protein concentration was determined by absorbance at A280 nm using ND-2000 spectrophotometer (NanoDrop, Wilmington, DE, USA).

Analytical RP-HPLC fractionation - Analytical RP-HPLC was performed using an 1100 HPLC system (Agilent Technologies), with column eluent monitored using a diode-array detector fitted with a standard 13-µL flow-cell (Agilent Model G1315B) and a multi-wavelength fluorescent detector fitted with a standard 8-µL flow-cell (Agilent Model G1321A) coupled in-line. Peptidomic fractions (150 µg, 0.5 mL) were manually injected (2 mL sample loop) onto a Brownlee Aquapore RP-300 cartridge (100 × 2.1 mm id, octylsilica 300Å pore size, 7 µm *d*p (Perkin-Elmer)). Sample injections were made up with 1:1 solvent A, 0.1% (v/v/) aqueous TFA. The column was developed at 100 µL/min over 75 min at room temperature using a linear 60-min gradient from 0–100% solvent B; 0.1% aqueous TFA, 60% CH_3CN. Column eluent was monitored for UV absorbance at 215 nm. 100 µL fractions were collected (t = 14-58 min) into 96-well low-protein binding plates (Agilent Technologies), after correcting for the post-column dead volume (50 µL). Fractions were concentrated to ~20 µL by centrifugal lyophilisation (SpeedVac AES 1010, Savant, U.S.A.) prior to sample injection for nanoLC-MS/MS analysis.

nanoLC-MS/MS - NanoLC-MSMS experiments were performed with a 1200 series nanoLC system (Agilent) equipped with a UPLC nano-Acquity® C18 150 × 0.15 mm i.d. column (Waters, Milford, MA). The system was developed with a linear 60-min gradient with a flow rate of 0.8 µL/min at 45 °C from 0-100% solvent B, where solvent A was 0.1% aqueous formic acid and solvent

B was 0.1% aqueous formic acid/60% CH_3CN. The nanoHPLC was coupled on-line to an LTQ-Orbitrap mass spectrometer equipped with a nanoelectrospray ion source (Thermo Fisher Scientific) for automated MS/MS, as described [35, 36].

Database searching and bioinformatic analyses - Peak lists (MGF files) were generated using Mascot Distiller (v2.3) using a default parameter template for Orbitrap LTQ instruments (high-res MS1 and low-res MS2). MGF files were searched using the Mascot v2.3.01 search algorithm (Matrix Science) against the Q112 LudwigNR protein sequence database with a taxonomy filter for human (Homo sapiens) comprising 140,018 entries. The search parameters were as follows: search unrestricted (no-enzyme) with N-terminal protein acetylation (+42 Da) as a variable modification. A peptide mass tolerance of ±30 ppm, and fragment ion mass tolerance of ±0.7 Da was used. The automatic decoy (random) database sequence option was enabled to allow false-discovery rate estimation. MSPro, previously described [37], was used for collating Mascot search result files, extracting and re-scoring all peptide identifications using a fast cross-correlation scoring scheme, instantiating the Percolator program and summarizing the results (Supplemental Data). Peptides were identified using two alternative analysis strategies; Mascot with homology ionscore thresholding (*conventional*), and a combination of Mascot, MSPro re-scoring and Percolator (*Multi-scoring*) with a 1% *q*-value (FDR) and peptide significance threshold of ≤1% or ≤5% PEP. Raw mass spectrometry data is deposited in the PeptideAtlas and can be accessed at http://www.peptideatlas.org/PASS/PASS00227 [38-40].

SignalP 4.0 [41] and SecretomeP 2.0 [42] algorithms (Center for Biological Sequence Analysis) were used to predict classical and non-classical secretion modes, respectively. A SecretomeP score >0.5 indicates a high probability of secretion via a non-classical pathway. Transmembrane proteins were predicted based on Kyte-Doolittle hydropathy scores [43]. The MEROPS peptidase database (http://merops.sanger.ac.uk/) was used as resources for annotating proteolytic events and determining substrate specificity [44].

Label-free spectral counting – The relative abundance of a protein within a sample was estimated using semi-quantitative normalised spectral count ratios (N_{SC}) [36]. For each individual protein for each search strategy, the number of significant scoring peptide spectra matches (PSMs) were summed and normalised by the total number of significant scoring PSMs in the sample (Eqn.1). Note that a PSM is deemed significant if PEP score based on Percolator is ≤0.05 (≤5%).

$$N_{SC} = (n+f) / (t-n+f) \qquad \textit{Eqn. 1}$$

where *n* is the number of significant PSMs for each protein in the sample, *t* is the total number of significant PSMs in the sample and *f* is the correction factor set to 1.25.

RESULTS and DISCUSSION

The experimental strategy for isolating the secretopeptidome developed in this study is shown in Figure 1. *Ex vivo* proteolysis of polypeptides in the secretome was restricted by using a protease inhibitor cocktail set as described in Experimental Procedures. Due to the fact that proteolysis differs from degradation processes, we examined our data for proteins shown previously to be associated with human endogenous proteasomal degradation and protein ubiquitination [33, 34] (Supplemental Table 1). In addition, to discriminate between peptide 'laddering' (*i.e.,* sequential truncated peptide sequences) through in-source fragmentation and endogenous proteolytic processing [45-47], peptide elution (retention) time is detailed in Supplemental Table 2.

Multi-scoring strategy increases peptide identifications in the secretopeptidome

To overcome issues with top-down peptidomics data analysis (large search space) two alternative post-processing analysis strategies were examined, - (i) Mascot with homology ionscore thresholding (*conventional*), and - (ii) a combination of Mascot, MSPro re-scoring and Percolator (*Multi-scoring*) with a 1% *q*-value (FDR) and 1% or 5% PEP (Figure 2) (Table 1). Based on ≤5% PEP, 474 proteins with 1228 unique peptides were identified, including 267 single peptide identifications (MS/MS for

single peptide identifications: Supplemental Figure 2, Supplemental Table 3), with 368 proteins (1103 peptides) reported with ≤1% PEP. These results demonstrate that the Multi-scoring approach significantly increased the number of peptides identified compared with the conventional approach (36% increase based on ≤5% PEP and 29% based on ≤1% PEP). While Brosch *et al.*, [48] recently reported a similar improvement (39%) when using Mascot+Percolator over a conventional Mascot search, their analysis was limited to tryptic peptides which constitute a much smaller database search space compared with the current study.

Analysis of secretopeptidome based on predicted protein secretion algorithms

Using the combined Multi-scoring approach, 474 proteins (based on 1228 unique peptides) were identified from biological replicates of both colon tumour cell models (LIM1215 and LIM1863) at a 1% FDR, of which 267 were single peptide identifications (Supplemental Tables 1-3). To demonstrate that the cleavage products are of reproducible abundance, we performed biological replicates for each cell line, identifying on average 50% of proteins in common for LIM1215 (Figure 3A), and 55% in common for LIM183 (Figure 3B) (Supplemental Table 4). Between datasets, 117 proteins were identified in common for LIM1215 and LIM1863, while 149 and 208 proteins were identified unique to each dataset, respectively (Figure 3C). Mutational differences between both of these cell lines may be attributed to differences in identified proteins for each dataset (Figure 3D). Recently, secreted proteins encoded by mutated genes (mutant proteins) have been identified using mass spectrometry from a panel of 18 human CRC cell lines [49]. Interestingly, in this study we report 13 secreted proteins with detected mutations, including CD166 antigen (ALCAM), Actin, cytoplasmic 2 (ACTG1), Low-density lipoprotein receptor-related protein 4 (LRP4), of which 7 proteins were unique to the LIM1215 secretopeptidome, 2 proteins unique to the LIM1863 secretopeptidome, and 4 proteins identified in both datasets.

Based on UniProt annotation of the 474 proteins, 98 (21%) were classified as secreted, extracellular, or cell membrane-derived. Using Kyte-Doolittle hydropathy scores [43], 48 proteins were predicted to contain at least one transmembrane-spanning domain. Using predictive algorithms, a total of 335

proteins (70%) from the secretopeptidome were identified as being secreted by classical and non-classical means (Table 2). These comprise 88 proteins predicted by SignalP to be classically secreted, 238 proteins predicted to be non-classically secreted based on SignalP and SecretomeP (<0.5). A further 9 proteins were predicted to be integral membrane proteins based on Kyte-Doolittle hydropathy scores. Endogenous proteasomal degradation and protein ubiquitination was monitored in the secretopeptidome [33, 34], with 113 proteins identified attributed with degradation (Table 2, Supplemental Table 1). The remaining 139 proteins were not classified as classically secreted, non-classically secreted, or integral membrane proteins.

Analysis of secretopeptidome using the protease database, MEROPS, identifies novel proteolytic substrates

We next examined the 335 predicted secreted proteins using the protease and substrate database MEROPS (http://merops.sanger.ac.uk/) [44] (Supplemental Table 1). MEROPS analysis revealed 42 substrates (classified as M) having peptide cleavage sequences resulting from proteases of known specificity (*e.g.*, APP, AREG, DSG3, and SDC1). A further 205 proteins (classified as MU) were shown to contain substrates listed in MEROPS, but which contained no peptide cleavage sequence information resulting from unknown proteases (*e.g.*, ADAM10, BSG, CD99, DPEP1, MEP1A, and MARCKSL1). A salient finding was the identification of 88 substrates in MEROPS with no matched-peptide sequence or cleavage information (classified as U) (Supplemental Table 5). Interestingly, only Transmembrane and coiled-coil domain-containing protein 1 (TMCO1), and Bax inhibitor 1 (TMBIM6) were identified as ubiquitinated proteins, highlighting that a significant proportion of these substrates were derived through extracellular cleavage. These substrates included fibroblast growth factor-binding protein 1 (FGFBP1), involved in angiogenesis and tumour growth [50]; inactive tyrosine-protein kinase 7 (PTK7), implicated in epithelial tissue organization, angiogenesis, cell polarity, actin cytoskeleton reorganization and apoptosis [51]; and the colon-specific cell surface A33 antigen (GPA33) [52]. To investigate the protease specificity for substrates not reported in MEROPS, we identified cleavage site positions of selected substrates including Cadherin-17 (CDH17), CD166, Ephrin type A receptor 1 (EPHA1) and 2 (EPHA2), Cell adhesion molecule 1 (CADM1), Cell surface

A33 antigen (GPA33), Mucin-13 (MUC13), and Plexin domain-containing protein 2 (PLXDC2) (Figure 4). Intriguingly, in many cases when N- and C-terminal amino acids of peptides juxtapositioned to each cleavage site (*i.e.*, amino acids P3, P2, P1 and P'3, P'2, P'1) were searched in MEROPS, no proteases specific for these cleavage sites were identified. This finding highlights the importance of the secretopeptidome as a first step in the identification of active proteases involved in tumourigenesis and the extracellular environment.

Cleaved peptide substrate derived from cell surface A33 antigen in different human colon models

To further investigate the expression of substrates derived for the A33 antigen we noted that a specific cleaved peptide (^{22}ISVETPQDVLR^{32}) was generated only in the LIM1215 secretopeptidome, in comparison to LIM1863. We expanded this secretopeptidome analysis to include an additional 12 CRC cell lines, including SW1222, LS174T, LIM1899, LIM2537, HT29, LIM2405, LIM2550, LOVO, LIM2099, LIM2408, HCA7, and CACO2. All cell line conditions were maintained as described in Experimental Procedures. In addition to LIM1215, only SW1222, LS174T, and LIM1899 were shown to express the N-terminal cleaved peptide fragment (^{22}ISVETPQDVLRASQGK^{37}) substrate in the secretome (Figure 5, Supplemental Figure 3). Cell line characteristics of these lines are shown in Supplemental Figure 4. The A33 antigen is a 43 kDa glycoprotein belonging to the immunoglobulin superfamily, containing a single transmembrane domain [53]. In humans, the expression of A33 antigen is highly restricted to the epithelial cells of the small and large intestine and is associated with >95% of human CRC and >50% of gastric cancers [54, 55]. Although a definitive role for the A33 antigen remains to be defined, a variety of indirect evidence suggests roles in cell adhesion, cell trafficking, and in the intestinal immune response [56-58]. Although there are no known proteases to cleave the A33 antigen, it is reported that an unknown peptidase cleaves the hydrophobic signal peptide between alanine (A^{21}) and isoleucine (I^{22}), which is required to produce the mature protein with the N-terminal sequence of the native molecule [53]. What protease(s) results in C-terminal cleavage at R^{32} or K^{37}, and the role of this protease(s) in colon tumour progression remains to be determined. For all lines used in this study, similarities associated with growth properties,

tumour stage and origin, microsatellite stability (MS) status, and mutational differences (e.g., *APC*, *p53*, *K-Ras*, *B-raf*, *PI3-K*) refer Supplemental Figure 4.

Post-translation modifications and GPI-anchored proteins in the secretopeptidome

Several post-translational sequence modifications in the secretopeptidome were identified, including 195 peptides (135 proteins) containing an N-terminal acetylation residue (*q*-value 1%, PEP 5%) (Supplemental Table 2). Protein acetylation has been attributed with non-classical secretion, with secretion of HMGB1 in response to inflammatory stimuli controlled by protein acetylation, acetylation shifting equilibrium from the nucleus towards cytoplasm and cell surface [59]. Likewise, several glycosylphosphatidylinositol (GPI)-anchored proteins [60] were identified, including ephrin-A1 (EFNA1), glypican-1 (GPC1), Xaa-Pro aminopeptidase 2 (XPNPEP2), and the membrane anchored protease, dipeptidase 1 (DPEP1). Identified peptide sequences of EFNA1 (6 peptides) and GPC1 (1 peptide) were derived predominantly from the C-terminus, while DPEP1 (2 peptides) were derived from the N-terminus.

Secreted peptides resulting from ectodomain shedding

Proteolytic processing of proteins which transverse or peripherally-associated with the PM play an important role in tumour development and progression [61]. More than 40 cell surface proteins have been shown to undergo proteolytic cleavage [62]. In this study, 122 ectodomain peptide fragments from 41 cell surface membrane ectodomains were observed (Figure 6, Table 3, Supplemental Tables 1-2). Several adhesion proteins known to be proteolytically cleaved include CD166 (ALCAM), BSG (CD147), EpCAM, integrins (α3, α6), and cadherin family proteins, calsyntenin-1 (CLSTN1), DSG2, and E-cadherin (CDH1). In this study, the ectodomain peptide ^{698}AQPVEAGLQIPA^{709} from CDH1 was identified, the sequence cleavage of which is not reported in the MEROPS database. CDH1 shedding has been to modulate epithelial cell-cell adhesion and migration, and subcellular localization and downstream signalling of β-catenin [14]. CDH1 ectodomain fragments have been shown to prevent cell-cell aggregation [13], induce cell invasion [13, 63], and promotes epithelial-to-mesenchymal transition [64]. Further, we report a unique peptide ^{514}ISDENREKVNDQAK^{527}

according to MEROPS, derived from the ectodomain of ALCAM [65]. Ectodomain shedding of ALCAM, which can be cleaved from the cell membrane by MMP-2 and ADAM-17 [66], has been shown *in vivo* localized at the invasive front of primary melanomas [67], capable of inducing tumour growth and progression and facilitate metastatic dissemination [68].

In addition to the ectodomain proteolysis of cell surface adhesion proteins, we observe proteolysis of specific ectodomains of cell surface proteins. These include Amyloid beta A4 protein (APP), where 8 ectodomain peptides were identified, of which the peptide sequence cleavage of 3 different peptides [^{688}LVFFAEDVGSNK^{689}, ^{672}DAEFRHDSGYEVHHQK^{687}, ^{673}AEFRHDSGYEVHHQK^{687}] agrees with proteolysis by the ADAM10 peptidase [69]. The elution (retention) times of these peptides differ indicating that these peptides are derived through endogenous proteolysis and were not artefacts of in-source fragmentation. Further, APP was not identified associated with human endogenous protein ubiquitination sites [33, 34], indicating that identification of cleaved APP was derived through extracellular proteolysis, and not an artifact of proteasomal degradation.

In the secretopeptidome we observe proteolytic ectodomain cleavage of the proteases themselves, including MMP-15 (MT2-MMP), shown to mediate direct-acting proteases capable of basement membrane remodelling during cell migration [70]; matripase (ST14), associated with increased cell motility, cancer invasion and metastasis through activation of an ECM-degrading protease system at the cell surface [71]; and the ECM protease, ADAM10 in the secretopeptidome.

Regulated intramembrane and intracytoplasmic proteolysis

Regulated intramembrane proteolysis describes the proteolytic processing of type I or type II transmembrane proteins, where a membrane-bound ectodomain fragment is subsequently cleaved by I-CliPs into an intracellular domain (ICD) and a soluble, ectodomain [3, 18, 19]. In the secretopeptidome, 23 intramembrane domain fragments from 12 substrates were identified, including APP [72], APLP2 [36], DSG2 [73], and SDC1 [73] (Figure 6, Table 3).

A proteomic study recently identified various type I transmembrane protein substrates from RIP, including APP, APLP2, SDC1, and DSG2 (from a total of 13 different substrates) [73]. In comparison, several peptide sequences from the intramembrane, derived through proteolysis were identified in the secretopeptidome, including APP, where one peptide [^{688}LVFFAEDVGSNK^{699}], and 7 peptides which overlapped with this ectodomain sequence were identified, and DSG2, where the peptide ^{605}DSYVGLGPAAIA^{616} was identified.

Further, an intact peptide fragment from the substrate Interferon alpha-inducible protein 6 (IFI6) was identified which transverses the transmembrane domain region [^{84}GGGVPAGGLVATL^{96}] (Table 3). This is the first report of IFI6 cleavage in MEROPS. IFI6, a membrane receptor involved in type I interferon-mediated signalling, is a component of JAK/STAT signalling cascade [62]. It is suggested intramembrane cleavage might be part of a general mechanism for receptor protein turnover, or generating a soluble receptor, which could bind and neutralize extracellular type I IFNs and attenuate activity [62].

In addition to the proteolytic cleavage and generation of C-terminal tails of PM proteins, I-CliPs hydrolyse C-terminal peptides from different subcellular organelles, including the endoplasmic reticulum, Golgi and endosome compartments. We report the identification of 24 ICD fragments (12 cytoplasmic, 11 lumenal, 1 vesicle membrane) in the secretopeptidome, including Integral membrane protein 2B (ITM2B), Vesicular integral-membrane protein VIP36 (LMAN2), Cation-independent mannose-6-phosphate receptor (IGF2R), and synaptotagmin-7 (SYT7) (Table 3, Supplemental Tables 1, 3). IGF2R, a multifunctional glycoprotein and receptor for various ligands including the Leukemia inhibitory factor (LIF) [74], was identified in the secretopeptidome based on ICD and lumenal peptides. IGF2R promotes aggregation and internalization into clathrin-coated vesicles, M6P-mediated transport to late endosomes [75] and is recycled to the secretory pathway or PM [76, 77]. Recently, tumour necrosis factor convertase (TACE, ADAM-17) has been shown to mediate ectodomain release of membrane-associated IGF2R from human endothelial cells [78], and a β-site APP cleaving enzyme 1 (BACE1) substrate [79]. Further, IGF2R was not identified associated with

human endogenous protein ubiquitination sites [33, 34], indicating that identification of cleaved IGF2R was derived through extracellular proteolysis, and not an artifact of proteasomal degradation.

In addition, SYT7, a component of the exocytic machinery involved in cytoplasmic vesicle release, is involved in secretory vesicle trafficking. In this study, the peptide [^{4}DPEAASPGAPSR^{15}] was identified associated with the vesicular domain. Proteolytic processing of SYT7 has been suggested to regulate the rate of exocytosis in rat cerebral cortex [69]. SYT7 was not identified associated with human endogenous protein ubiquitination sites, again indicating SYT7 is derived through extracellular proteolysis, and not an artifact of proteasomal degradation. This is the first report of SYT7 proteolysis in MEROPS, and of a soluble peptide derived from cytoplasmic vesicles in the tumour-derived secretome. Recently, Sidhu *et al*., [80] demonstrated the release of EMMPRIN from the surface of NCI-H460 cells via microvesicle shedding. Following secretion, these microvesicles are rapidly degraded to release bioactive EMMPRIN peptide fragments. Vesicle-mediated release of EMMPRIN has been further confirmed in several tumour cell lines, including colorectal adenocarcinoma [81]. In our study peptides derived from the ectodomain were identified, providing an extensive insight into the contribution of endogenous proteolysis in the extracellular milieu. It is interesting to note that shedding of tumour surface antigens from membrane vesicles has been implicated as an important feature of malignant transformation [82-84].

Conclusion

The improved, reproducible secretopeptidome isolation and characterisation strategy described in this study promises to extend our understanding of endogenous (non-tryptic) peptides generated through proteolysis of classical/non-classical secreted proteins from colon tumour cells, the extracellular proteolytic processing of cell surface membrane proteins, and peptides generated through RIP. Further studies need to be undertaken to establish whether these <3 kDa endogeneous peptides have functional activity throughout the extracellular region and tumour microenvironment and the specific proteases responsible for such proteolysis. The expression of GPA33 was monitored in 14 different

human colon tumour models, with further studies directed towards analysis of *in vivo* secretome models of the tumour microenvironment (including xenograft analyses) in order to identify if specific cleavage products are of biological relevance. Given the underrepresentation of this class of low-M_r peptides, the secretopeptidome strategy outlined here promises to contribute an important element to the fledging Human Proteome Project. The discovery of several novel proteolytic cleavage sites using this peptidome approach is an important first step towards identifying new proteases that may provide therapeutic targets in cancer biology.

Competing interests. The authors have declared that no competing interests exist.

Acknowledgements. Funding was provided, in part, by the Australian National Health and Medical Research Council under Program Grant 487922 (RJS, JH and DWG). We acknowledge the NHMRC-funded Australian Proteomics Computational Facility (APCF) under Enabling Grant 381413. This work was supported by funds from the Operational Infrastructure Support Program provided by the Victorian Government, Australia.

References

1. Lopez-Otin, C. and Matrisian, L.M. (2007) Emerging roles of proteases in tumour suppression. Nat Rev Cancer. 7(10). 800-808.
2. van Kilsdonk, J.W., van Kempen, L.C., van Muijen, G.N., Ruiter, D.J., Swart, G.W. (2010) Soluble adhesion molecules in human cancers: sources and fates. Eur J Cell Biol. 89(6). 415-427.
3. Weihofen, A. and Martoglio, B. (2003) Intramembrane-cleaving proteases: controlled liberation of proteins and bioactive peptides. Trends Cell Biol. 13(2). 71-78.
4. Craig, S.E. and Brady-Kalnay, S.M. (2011) Cancer Cells Cut Homophilic Cell Adhesion Molecules and Run. Cancer Res. 71(2). 303-309.
5. Mechtersheimer, S., Gutwein, P., Agmon-Levin, N., Stoeck, A., Oleszewski, M., Riedle, S., Postina, R., Fahrenholz, F., Fogel, M., Lemmon, V. and Altevogt, P. (2001) Ectodomain shedding of L1 adhesion molecule promotes cell migration by autocrine binding to integrins. J Cell Biol. 155(4). 661-673.
6. Conacci-Sorrell, M., Kaplan, A., Raveh, S., Gavert, N., Sakurai, T., Ben-Ze'ev, A. (2005) The shed ectodomain of Nr-CAM stimulates cell proliferation and motility, and confers cell transformation. Cancer Res. 65(24). 11605-11612.
7. Najy, A.J., Day, K.C., Day, M.L. (2008) The ectodomain shedding of E-cadherin by ADAM15 supports ErbB receptor activation. J Biol Chem. 283(26). 18393-18401.
8. Maetzel, D., Denzel, S., Mack, B., Canis, M., Went, P., Benk, M., Kieu, C., Papior, P., Baeuerle, P.A., Munz, M. and Gires, O. (2009) Nuclear signalling by tumour-associated antigen EpCAM. Nat Cell Biol. 11(2). 162-171.
9. Arribas, J. and Borroto, A. (2002) Protein ectodomain shedding. Chem Rev. 102(12). 4627-4638.
10. Kheradmand, F. and Werb, Z. (2002) Shedding light on sheddases: role in growth and development. Bioessays. 24(1). 8-12.
11. Garton, K.J., Gough, P.J., Raines, E.W. (2006) Emerging roles for ectodomain shedding in the regulation of inflammatory responses. J Leukoc Biol. 79(6). 1105-1116.
12. Murphy, G. (2008) The ADAMs: signalling scissors in the tumour microenvironment. Nat Rev Cancer. 8(12). 929-941.
13. Noe, V., Fingleton, B., Jacobs, K., Crawford, H.C., Vermeulen, S., Steelant, W., Bruyneel, E., Matrisian, L.M., Mareel, M. (2001) Release of an invasion promoter E-cadherin fragment by matrilysin and stromelysin-1. J Cell Sci. 114(Pt 1). 111-118.
14. Maretzky, T., Reiss, K., Ludwig, A., Buchholz, J., Scholz, F., Proksch, E., de Strooper, B., Hartmann, D., Saftig, P. (2005) ADAM10 mediates E-cadherin shedding and regulates epithelial cell-cell adhesion, migration, and beta-catenin translocation. Proc Natl Acad Sci U S A. 102(26). 9182-9187.
15. Ryniers, F., Stove, C., Goethals, M., Brackenier, L., Noe, V., Bracke, M., Vandekerckhove, J., Mareel, M., Bruyneel, E. (2002) Plasmin produces an E-cadherin fragment that stimulates cancer cell invasion. Biol Chem. 383(1). 159-165.
16. Lochter, A., Galosy, S., Muschler, J., Freedman, N., Werb, Z., Bissell, M.J. (1997) Matrix metalloproteinase stromelysin-1 triggers a cascade of molecular alterations that leads to stable epithelial-to-mesenchymal conversion and a premalignant phenotype in mammary epithelial cells. J Cell Biol. 139(7). 1861-1872.
17. Weiss, J.V., Klein-Scory, S., Kubler, S., Reinacher-Schick, A., Stricker, I., Schmiegel, W., Schwarte-Waldhoff, I. (2011) Soluble E-cadherin as a serum biomarker candidate: Elevated levels in patients with late-stage colorectal carcinoma and FAP. Int J Cancer. 128(6). 1384-1392.
18. Medina, M. and Dotti, C.G. (2003) RIPped out by presenilin-dependent gamma-secretase. Cell Signal. 15(9). 829-841.
19. Wolfe, M.S. (2009) Intramembrane proteolysis. Chem Rev. 109(4). 1599-1612.
20. Erez, E., Fass, D., Bibi, E. (2009) How intramembrane proteases bury hydrolytic reactions in the membrane. Nature. 459(7245). 371-378.

21. LaVoie, M.J. and Selkoe, D.J. (2003) The Notch ligands, Jagged and Delta, are sequentially processed by alpha-secretase and presenilin/gamma-secretase and release signaling fragments. J Biol Chem. 278(36). 34427-34437.
22. Dean, R.A. and Overall, C.M. (2007) Proteomics discovery of metalloproteinase substrates in the cellular context by iTRAQ labeling reveals a diverse MMP-2 substrate degradome. Mol Cell Proteomics. 6(4). 611-623.
23. Dix, M.M., Simon, G.M., Cravatt, B.F. (2008) Global mapping of the topography and magnitude of proteolytic events in apoptosis. Cell. 134(4). 679-691.
24. Gevaert, K., Goethals, M., Martens, L., Van Damme, J., Staes, A., Thomas, G.R., Vandekerckhove, J. (2003) Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. Nat Biotechnol. 21(5). 566-569.
25. Tam, E.M., Morrison, C.J., Wu, Y.I., Stack, M.S., Overall, C.M. (2004) Membrane protease proteomics: Isotope-coded affinity tag MS identification of undescribed MT1-matrix metalloproteinase substrates. Proc Natl Acad Sci U S A. 101(18). 6917-6922.
26. Timmer, J.C., Enoksson, M., Wildfang, E., Zhu, W., Igarashi, Y., Denault, J.B., Ma, Y., Dummitt, B., Chang, Y.H., Mast, A.E., Eroshkin, A., Smith, J.W., Tao, W.A. and Salvesen, G.S. (2007) Profiling constitutive proteolytic events in vivo. Biochem J. 407(1). 41-48.
27. Doucet, A. and Overall, C.M. (2011) Broad coverage identification of multiple proteolytic cleavage site sequences in complex high molecular weight proteins using quantitative proteomics as a complement to edman sequencing. Mol Cell Proteomics. 10(5). M110.003533.
28. Sasaki, K., Sato, K., Akiyama, Y., Yanagihara, K., Oka, M., Yamaguchi, K. (2002) Peptidomics-based approach reveals the secretion of the 29-residue COOH-terminal fragment of the putative tumor suppressor protein DMBT1 from pancreatic adenocarcinoma cell lines. Cancer Res. 62(17). 4894-4898.
29. Ahram, M., Adkins, J.N., Auberry, D.L., Wunschel, D.S., Springer, D.L. (2005) A proteomic approach to characterize protein shedding. Proteomics. 5(1). 123-131.
30. Clynen, E., Baggerman, G., Veelaert, D., Cerstiaens, A., Van der Horst, D., Harthoorn, L., Derua, R., Waelkens, E., De Loof, A., Schoofs, L. (2001) Peptidomics of the pars intercerebralis-corpus cardiacum complex of the migratory locust, Locusta migratoria. Eur J Biochem. 268(7). 1929-1939.
31. Whitehead, R.H., Macrae, F.A., St John, D.J., Ma, J. (1985) A colon cancer cell line (LIM1215) derived from a patient with inherited nonpolyposis colorectal cancer. J. Natl. Cancer Inst. 74(4). 759-765.
32. Whitehead, R.H., Jones, J.K., Gabriel, A., Lukies, R.E. (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.
33. Danielsen, J.M., Sylvestersen, K.B., Bekker-Jensen, S., Szklarczyk, D., Poulsen, J.W., Horn, H., Jensen, L.J., Mailand, N., Nielsen, M.L. (2011) Mass spectrometric analysis of lysine ubiquitylation reveals promiscuity at site level. Molecular & cellular proteomics : MCP. 10(3). M110 003590.
34. Shi, Y., Chan, D.W., Jung, S.Y., Malovannaya, A., Wang, Y., Qin, J. (2011) A data set of human endogenous protein ubiquitination sites. Molecular & cellular proteomics : MCP. 10(5). M110 002089.
35. Greening, D.W. and Simpson, R.J. (2010) A centrifugal ultrafiltration strategy for isolating the low-molecular weight (</=25K) component of human plasma proteome. J Proteomics 73(3). 637-648
36. Tauro, B.J., Greening, D.W., Mathias, R.A., Ji, H., Mathivanan, S., Scott, A.M., Simpson, R.J. (2012) Comparison of ultracentrifugation, density gradient separation, and immunoaffinity capture methods for isolating human colon cancer cell line LIM1863-derived exosomes. Methods. 56(2). 293-304.
37. Greening, D.W., Glenister, K.M., Kapp, E.A., Moritz, R.L., Sparrow, R.L., Lynch, G.W., Simpson, R.J. (2008) Comparison of human platelet-membrane cytoskeletal proteins with the plasma proteome: Towards understanding the platelet-plasma nexus. Proteomics Clin. Appl. 2(63-77.

38. Deutsch, E.W., Lam, H., Aebersold, R. (2008) PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. EMBO Rep. 9(5). 429-434.
39. Desiere, F., Deutsch, E.W., King, N.L., Nesvizhskii, A.I., Mallick, P., Eng, J., Chen, S., Eddes, J., Loevenich, S.N., Aebersold, R. (2006) The PeptideAtlas project. Nucleic Acids Res. 34(Database issue). D655-658.
40. Desiere, F., Deutsch, E.W., Nesvizhskii, A.I., Mallick, P., King, N.L., Eng, J.K., Aderem, A., Boyle, R., Brunner, E., Donohoe, S., Fausto, N., Hafen, E., Hood, L., Katze, M.G., Kennedy, K.A., Kregenow, F., Lee, H., Lin, B., Martin, D., Ranish, J.A., Rawlings, D.J., Samelson, L.E., Shiio, Y., Watts, J.D., Wollscheid, B., Wright, M.E., Yan, W., Yang, L., Yi, E.C., Zhang, H. and Aebersold, R. (2005) Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. Genome Biol. 6(1). R9.
41. Bendtsen, J.D., Nielsen, H., von Heijne, G., Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol. 340(4). 783-795.
42. Bendtsen, J.D., Jensen, L.J., Blom, N., Von Heijne, G., Brunak, S. (2004) Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng Des Sel. 17(4). 349-356.
43. Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol. 157(1). 105-132.
44. Rawlings, N.D., Barrett, A.J., Bateman, A. (2010) MEROPS: the peptidase database. Nucleic Acids Res. 38(Database issue). D227-233.
45. Baczek, T., Wiczling, P., Marszall, M., Heyden, Y.V., Kaliszan, R. (2005) Prediction of peptide retention at different HPLC conditions from multiple linear regression models. J Proteome Res. 4(2). 555-563.
46. Krokhin, O.V., Ens, W., Standing, K.G. (2003) Characterizing degradation products of peptides containing N-terminal Cys residues by (off-line high-performance liquid chromatography)/matrix-assisted laser desorption/ionization quadrupole time-of-flight measurements. Rapid Comm. Mass Spec. 17(22). 2528-2534.
47. Krokhin, O.V., Craig, R., Spicer, V., Ens, W., Standing, K.G., Beavis, R.C., Wilkins, J.A. (2004) An improved model for prediction of retention times of tryptic peptides in ion pair reversed-phase HPLC: its application to protein peptide mapping by off-line HPLC-MALDI MS. Mol. Cell. Proteomics. 3(9). 908-919.
48. Brosch, M., Yu, L., Hubbard, T., Choudhary, J. (2009) Accurate and sensitive peptide identification with Mascot Percolator. J Proteome Res. 8(6). 3176-3181.
49. Mathivanan, S., Ji, H., Tauro, B.J., Chen, Y.S., Simpson, R.J. (2012) Identifying mutated proteins secreted by colon cancer cell lines using mass spectrometry. J Proteomics. 76 Spec No.(141-149.
50. Czubayko, F., Liaudet-Coopman, E.D., Aigner, A., Tuveson, A.T., Berchem, G.J., Wellstein, A. (1997) A secreted FGF-binding protein can serve as the angiogenic switch in human cancer. Nat Med. 3(10). 1137-1140.
51. Peradziryi, H., Tolwinski, N.S., Borchers, A. (2012) The many roles of PTK7: a versatile regulator of cell-cell communication. Archives of biochemistry and biophysics. 524(1). 71-76.
52. Ackerman, M.E., Chalouni, C., Schmidt, M.M., Raman, V.V., Ritter, G., Old, L.J., Mellman, I., Wittrup, K.D. (2008) A33 antigen displays persistent surface expression. Cancer immunology, immunotherapy : CII. 57(7). 1017-1027.
53. Heath, J.K., White, S.J., Johnstone, C.N., Catimel, B., Simpson, R.J., Moritz, R.L., Tu, G.F., Ji, H., Whitehead, R.H., Groenen, L.C., Scott, A.M., Ritter, G., Cohen, L., Welt, S., Old, L.J., Nice, E.C. and Burgess, A.W. (1997) The human A33 antigen is a transmembrane glycoprotein and a novel member of the immunoglobulin superfamily. Proc Natl Acad Sci U S A. 94(2). 469-474.
54. Welt, S., Scott, A.M., Divgi, C.R., Kemeny, N.E., Finn, R.D., Daghighian, F., Germain, J.S., Richards, E.C., Larson, S.M., Old, L.J. (1996) Phase I/II study of iodine 125-labeled monoclonal antibody A33 in patients with advanced colon cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 14(6). 1787-1797.
55. Sakamoto, J., Kojima, H., Kato, J., Hamashima, H., Suzuki, H. (2000) Organ-specific expression of the intestinal epithelium-related antigen A33, a cell surface target for antibody-

based imaging and treatment in gastrointestinal cancer. Cancer chemotherapy and pharmacology. 46 Suppl(S27-32.
56. Johnstone, C.N., Tebbutt, N.C., Abud, H.E., White, S.J., Stenvers, K.L., Hall, N.E., Cody, S.H., Whitehead, R.H., Catimel, B., Nice, E.C., Burgess, A.W. and Heath, J.K. (2000) Characterization of mouse A33 antigen, a definitive marker for basolateral surfaces of intestinal epithelial cells. American journal of physiology. Gastrointestinal and liver physiology. 279(3). G500-510.
57. Abud, H.E., Johnstone, C.N., Tebbutt, N.C., Heath, J.K. (2000) The murine A33 antigen is expressed at two distinct sites during development, the ICM of the blastocyst and the intestinal epithelium. Mech Dev. 98(1-2). 111-114.
58. Kagnoff, M.F. (1998) Current concepts in mucosal immunity. III. Ontogeny and function of gamma delta T cells in the intestine. The American journal of physiology. 274(3 Pt 1). G455-458.
59. Bonaldi, T., Talamo, F., Scaffidi, P., Ferrera, D., Porto, A., Bachi, A., Rubartelli, A., Agresti, A., Bianchi, M.E. (2003) Monocytic cells hyperacetylate chromatin protein HMGB1 to redirect it towards secretion. Embo J. 22(20). 5551-5560.
60. Chatterjee, S. and Mayor, S. (2001) The GPI-anchor and protein sorting. Cell Mol Life Sci. 58(14). 1969-1987.
61. Black, P.H. (1980) Shedding from normal and cancer-cell surfaces. N Engl J Med. 303(24). 1415-1416.
62. Ito, E., Yana, I., Fujita, C., Irifune, A., Takeda, M., Madachi, A., Mori, S., Hamada, Y., Kawaguchi, N., Matsuura, N. (2010) The role of MT2-MMP in cancer progression. Biochem Biophys Res Commun. 393(2). 222-227.
63. Klucky, B., Mueller, R., Vogt, I., Teurich, S., Hartenstein, B., Breuhahn, K., Flechtenmacher, C., Angel, P., Hess, J. (2007) Kallikrein 6 induces E-cadherin shedding and promotes cell proliferation, migration, and invasion. Cancer Res. 67(17). 8198-8206.
64. Zheng, G., Lyons, J.G., Tan, T.K., Wang, Y., Hsu, T.T., Min, D., Succar, L., Rangan, G.K., Hu, M., Henderson, B.R., Alexander, S.I. and Harris, D.C. (2009) Disruption of E-cadherin by matrix metalloproteinase directly mediates epithelial-mesenchymal transition downstream of transforming growth factor-beta1 in renal tubular epithelial cells. The American journal of pathology. 175(2). 580-591.
65. Steinberger, D. (2000) Medical billing outsourcing: is it worth the risk? Journal of ophthalmic nursing & technology. 19(5). 218-219.
66. Bech-Serra, J.J., Santiago-Josefat, B., Esselens, C., Saftig, P., Baselga, J., Arribas, J., Canals, F. (2006) Proteomic identification of desmoglein 2 and activated leukocyte cell adhesion molecule as substrates of ADAM17 and ADAM10 by difference gel electrophoresis. Mol Cell Biol. 26(13). 5086-5095.
67. Hofmann, U.B., Westphal, J.R., Zendman, A.J., Becker, J.C., Ruiter, D.J., van Muijen, G.N. (2000) Expression and activation of matrix metalloproteinase-2 (MMP-2) and its co-localization with membrane-type 1 matrix metalloproteinase (MT1-MMP) correlate with melanoma progression. The Journal of pathology. 191(3). 245-256.
68. van Kempen, L.C., Meier, F., Egeblad, M., Kersten-Niessen, M.J., Garbe, C., Weidle, U.H., Van Muijen, G.N., Herlyn, M., Bloemers, H.P., Swart, G.W. (2004) Truncation of activated leukocyte cell adhesion molecule: a gateway to melanoma metastasis. J Invest Dermatol. 122(5). 1293-1301.
69. Amour, A., Knight, C.G., English, W.R., Webster, A., Slocombe, P.M., Knauper, V., Docherty, A.J., Becherer, J.D., Blobel, C.P., Murphy, G. (2002) The enzymatic activity of ADAM8 and ADAM9 is not regulated by TIMPs. FEBS Lett. 524(1-3). 154-158.
70. Hotary, K., Li, X.Y., Allen, E., Stevens, S.L., Weiss, S.J. (2006) A cancer cell metalloprotease triad regulates the basement membrane transmigration program. Genes & development. 20(19). 2673-2686.
71. Netzel-Arnett, S., Hooper, J.D., Szabo, R., Madison, E.L., Quigley, J.P., Bugge, T.H., Antalis, T.M. (2003) Membrane anchored serine proteases: a rapidly expanding group of cell surface proteolytic enzymes with potential roles in cancer. Cancer Metastasis Rev. 22(2-3). 237-258.

72. Sastre, M., Steiner, H., Fuchs, K., Capell, A., Multhaup, G., Condron, M.M., Teplow, D.B., Haass, C. (2001) Presenilin-dependent gamma-secretase processing of beta-amyloid precursor protein at a site corresponding to the S3 cleavage of Notch. EMBO Rep. 2(9). 835-841.
73. Hemming, M.L., Elias, J.E., Gygi, S.P., Selkoe, D.J. (2008) Proteomic profiling of gamma-secretase substrates and mapping of substrate requirements. PLoS Biol. 6(10). e257.
74. Hebert, E. (2006) Mannose-6-phosphate/insulin-like growth factor II receptor expression and tumor development. Bioscience reports. 26(1). 7-17.
75. Ghosh, P., Dahms, N.M., Kornfeld, S. (2003) Mannose 6-phosphate receptors: new twists in the tale. Nat. Rev. Mol. Cell Biol. 4(3). 202-212.
76. Hawkes, C. and Kar, S. (2004) The insulin-like growth factor-II/mannose-6-phosphate receptor: structure, distribution and function in the central nervous system. Brain Res. Reviews. 44(2-3). 117-140.
77. Mulholland, D.J., Dedhar, S., Coetzee, G.A., Nelson, C.C. (2005) Interaction of nuclear receptors with the Wnt/beta-catenin/Tcf signaling axis: Wnt you like to know? Endocr Rev. 26(7). 898-915.
78. Leksa, V., Loewe, R., Binder, B., Schiller, H.B., Eckerstorfer, P., Forster, F., Soler-Cardona, A., Ondrovicova, G., Kutejova, E., Steinhuber, E., Breuss, J., Drach, J., Petzelbauer, P., Binder, B.R. and Stockinger, H. (2011) Soluble M6P/IGF2R released by TACE controls angiogenesis via blocking plasminogen activation. Circul. Res. 108(6). 676-685.
79. Hemming, M.L., Elias, J.E., Gygi, S.P., Selkoe, D.J. (2009) Identification of beta-secretase (BACE1) substrates using quantitative proteomics. PLoS ONE. 4(12). e8477.
80. Schechter, I. (2012) Reprint of "on the size of the active site in proteases. I. Papain". Biochem Biophys Res Commun. 425(3). 497-502.
81. Smits, A.H., Jansen, P.W., Poser, I., Hyman, A.A., Vermeulen, M. (2012) Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. Nucleic Acids Res.
82. Black, P.H. (1980) Shedding from the cell surface of normal and cancer cells. Advances in cancer research. 32(75-199.
83. Baj-Krzyworzeka, M., Szatanek, R., Weglarczyk, K., Baran, J., Urbanowicz, B., Branski, P., Ratajczak, M.Z., Zembala, M. (2006) Tumour-derived microvesicles carry several surface determinants and mRNA of tumour cells and transfer some of these determinants to monocytes. Cancer immunology, immunotherapy : CII. 55(7). 808-818.
84. Stoeck, A., Keller, S., Riedle, S., Sanderson, M.P., Runz, S., Le Naour, F., Gutwein, P., Ludwig, A., Rubinstein, E., Altevogt, P. (2006) A role for exosomes in the constitutive and stimulus-induced ectodomain cleavage of L1 and CD44. Biochem J. 393(Pt 3). 609-618.
85. Kall, L., Canterbury, J.D., Weston, J., Noble, W.S., MacCoss, M.J. (2007) Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat Methods. 4(11). 923-925.
86. Yang, X., Dondeti, V., Dezube, R., Maynard, D.M., Geer, L.Y., Epstein, J., Chen, X., Markey, S.P., Kowalak, J.A. (2004) DBParser: web-based software for shotgun proteomic data analyses. J Proteome Res. 3(5). 1002-1008.
87. Rowan, A.J., Lamlum, H., Ilyas, M., Wheeler, J., Straub, J., Papadopoulou, A., Bicknell, D., Bodmer, W.F., Tomlinson, I.P. (2000) APC mutations in sporadic colorectal tumors: A mutational "hotspot" and interdependence of the "two hits". Proc Natl Acad Sci U S A. 97(7). 3352-3357.
88. Luchtenborg, M., Weijenberg, M.P., Roemen, G.M., de Bruine, A.P., van den Brandt, P.A., Lentjes, M.H., Brink, M., van Engeland, M., Goldbohm, R.A., de Goeij, A.F. (2004) APC mutations in sporadic colorectal carcinomas from The Netherlands Cohort Study. Carcinogenesis. 25(7). 1219-1226.
89. Zhang, H.H., Walker, F., Kiflemariam, S., Whitehead, R.H., Williams, D., Phillips, W.A., Mikeska, T., Dobrovic, A., Burgess, A.W. (2009) Selective inhibition of proliferation in colorectal carcinoma cell lines expressing mutant APC or activated B-Raf. International journal of cancer. Journal international du cancer. 125(2). 297-307.
90. Liu, Y. and Bodmer, W.F. (2006) Analysis of P53 mutations and their expression in 56 colorectal cancer cell lines. Proc Natl Acad Sci U S A. 103(4). 976-981.

91. Seth, R., Crook, S., Ibrahem, S., Fadhil, W., Jackson, D., Ilyas, M. (2009) Concomitant mutations and splice variants in KRAS and BRAF demonstrate complex perturbation of the Ras/Raf signalling pathway in advanced colorectal cancer. Gut. 58(9). 1234-1241.
92. Philp, A.J., Campbell, I.G., Leet, C., Vincan, E., Rockman, S.P., Whitehead, R.H., Thomas, R.J., Phillips, W.A. (2001) The phosphatidylinositol 3'-kinase p85alpha gene is an oncogene in human ovarian and colon tumors. Cancer Res. 61(20). 7426-7429.
93. Schechter, I. and Berger, A. (1968) On the active site of proteases. 3. Mapping the active site of papain; specific peptide inhibitors of papain. Biochem Biophys Res Commun. 32(5). 898-902.

Figure and Table Legends

Figure 1 - Experimental overview of the secretopeptidome. LIM1215 and LIM1863 **c**ell culture media (CM) was collected for both biological replicates following 24 h serum-free media (SFM) incubation and the secretome prepared by differential centrifugation. A protease inhibitor cocktail set was included to limit exogenous proteolysis, with samples immediately placed on ice. The peptidome (<3K) was enriched through a series of centrifugal ultrafiltration stages of decreased nominal molecular weight limit (NMWL) membrane pore size (100K, 3K, 1K) (*i.e.*, the filtrate, flow through, from the 100K membrane was filtered through a subsequent 3K membrane, before concentrating the 1-3K secretome using a 1K filter). Centrifugation conditions for each filtration stage were optimised to ensure 95% (v/v) recovery of the filtrate. Each secretopeptidome (150 µg, 0.5 mL) was fractionated using RP-HPLC (Supplemental Figure 1) over a linear 60-min gradient of 60% CH_3CN, fractions collected, concentrated by lyophilisation, and subsequently analysed using nanoLC-MS/MS. No tryptic digestion was performed for these studies, analysing the intact proteome. Peptides were identified using two alternative search strategies detailed in Figure 2. Peptide retention times were obtained to ensure peptides were not artefacts of in-source fragmentation. Peptide sequence data was subsequently analysed for cleavage/proteolytic sites reported in MEROPS. For this study raw mass spectrometry data is deposited in the PeptideAtlas and can be accessed at http://www.peptideatlas.org/PASS/PASS00227.

Figure 2 – Data analysis search strategies of the secretopeptidome. Peptides were identified using two alternative search strategies, using (*conventional*) Mascot with homology ionscore thresholding, and a combination of Mascot, MSPro re-scoring and Percolator (*Multi-scoring*) at a 1% *q*-value (FDR) and 1% or 5% Posterior Error Probability (PEP). For these analyses, the automatic decoy (random) database sequence option was enabled to allow FDR estimation. MSPro was used to collate all Mascot results, extract peptide identifications, re-score and instantiate the Percolator program and summarise results (Supplemental Data) [37]. For the conventional Mascot approach peptides were deemed significant if the Mascot ionscore was ≥ the homology score threshold. For the Multi-scoring

approach, all peptides with ionscores >5 were included in the Percolator analysis (v1.2) [48, 85]. A Mascot delta score (Δ) of 1 was used for each MS/MS spectrum essentially allowing more than 1 peptide hit per MS/MS spectrum to be analysed. Significant scoring peptides were assigned to protein groups based on the principle of parsimonious analysis [86] and all peptides labelled as unique or duplicate along with their status (*i.e.*, razor, unique or degenerate).

Figure 3 - Protein identifications for each secretopeptidome biological replicate, including LIM1215 (A) and LIM1863 (B). A total of 474 proteins were identified in combined biological replicates for the comparison of LIM1215 and LIM1863, with 117 proteins common to both human CRC cell lines (C). These 474 proteins (based on 1228 unique peptides) were identified at a 1% FDR ($q \leq 0.01$ and PEP ≤ 0.05), of which 267 were single peptide identifications. A comparison with secreted proteins encoded by mutated genes (mutant proteins) from a panel of 18 human CRC cell lines [49] was performed, of which 13 secreted proteins were identified in the secretopeptidome; 7 proteins unique to LIM1215 secretopeptidome, 2 proteins unique to LIM1863 secretopeptidome, and 4 proteins identified in both datasets. Mutational differences between LIM1215 and LIM1863 are shown (D), based on APC mutation status [87, 88], P53 mutation status [89, 90], K-Ras mutation status [89], B-Raf mutation status [89, 91], and PI3Ks mutation status [92].

Figure 4 – MEROPS database analysis of selected substrates and their cleaved peptides in the secretopeptidome. For selected proteins not reported in MEROPS (Supplemental Table 5) we investigated the protease specificity for the identified peptide sequence, both the N- and C-terminus. The general nomenclature of cleavage site positions of the substrate designate the cleavage site between P1-P1', incrementing the numbering in the N-terminal direction of the cleaved peptide bond (P2, P3) [80, 93]. On the carboxyl side of the cleavage site numbering is incremented (P1', P2', P3'). MEROPS was used to identify no known proteases specific for these cleaved peptide sequence regions. Reported function and expression of these proteins is indicated.

Figure 5 – Secretopeptidome analysis of cleaved peptide substrate derived from cell surface GPA33 antigen. In addition to the human CRC cell lines LIM1215 and LIM1863, we expanded our analysis of the secretopeptidome for additional 12 human CRC cell lines. We identified the specific N-terminal-derived peptide, and laddered C-terminal sequence in LIM1215, SW1222, LS174T, and LIM1899. No peptides for GPA33 were identified in the secretopeptidome for all remaining cell lines. A schematic structure of GPA33 is shown, indicating the ecotodomain region where identified peptides were derived. Refer Supplemental Figure 4 for detail about associated cell growth properties, tumour stage and origin, microsatellite stability (MS) status, and mutational differences for cell line utilised in this study.

Figure 6 – Proteolytic cleavage fragments identified in the secretopeptidome from colorectal cancer cells. Identification of peptides spanning extracellular domains of transmembrane proteins detected in the secretome is consistent with a shedding process. Red traces indicate extracellular domains, blue traces indicate intracellular domains, green traces represent luminal domains and orange traces represent vesicle membrane domains. Transmembrane regions are indicated in yellow. Black regions indicate identified peptides while grey regions indicate identified peptides spanning a transmembrane domain. The number of significant peptides identified ($q \leq 0.01$ and PEP ≤ 0.05) and the type of proteolytic cleavage is indicated: extracellular region (*Ecto*), intramembrane region (*Intramembrane proteolysis*), intracytoplasmic region (*ICD*), lumenal region (*Lumen*), vesicle membrane region (*vesicle membrane proteolysis*).

Figure 1

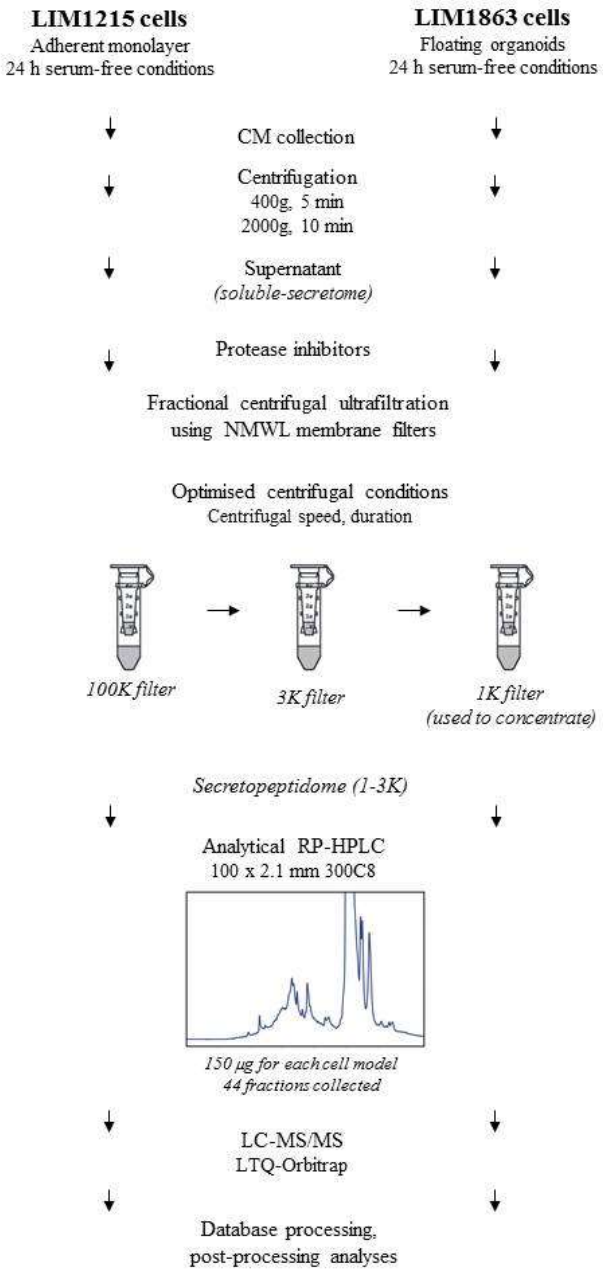

LIM1215 cells
Adherent monolayer
24 h serum-free conditions
LIM1863 cells
Floating organoids
24 h serum-free conditions
CM collection
Centrifugation
400g, 5 min
2000g, 10 min
Supernatant
(soluble-secretome)
Protease inhibitors
Fractional centrifugal ultrafiltration
using NMWL membrane filters
Optimised centrifugal conditions
Centrifugal speed, duration
100K filter
3K filter
1K filter
(used to concentrate)
Secretopeptidome (1-3K)
Analytical RP-HPLC
100 x 2.1 mm 300C8
150 µg for each cell model
44 fractions collected
LC-MS/MS
LTQ-Orbitrap
Database processing,
post-processing analyses

Figure 2

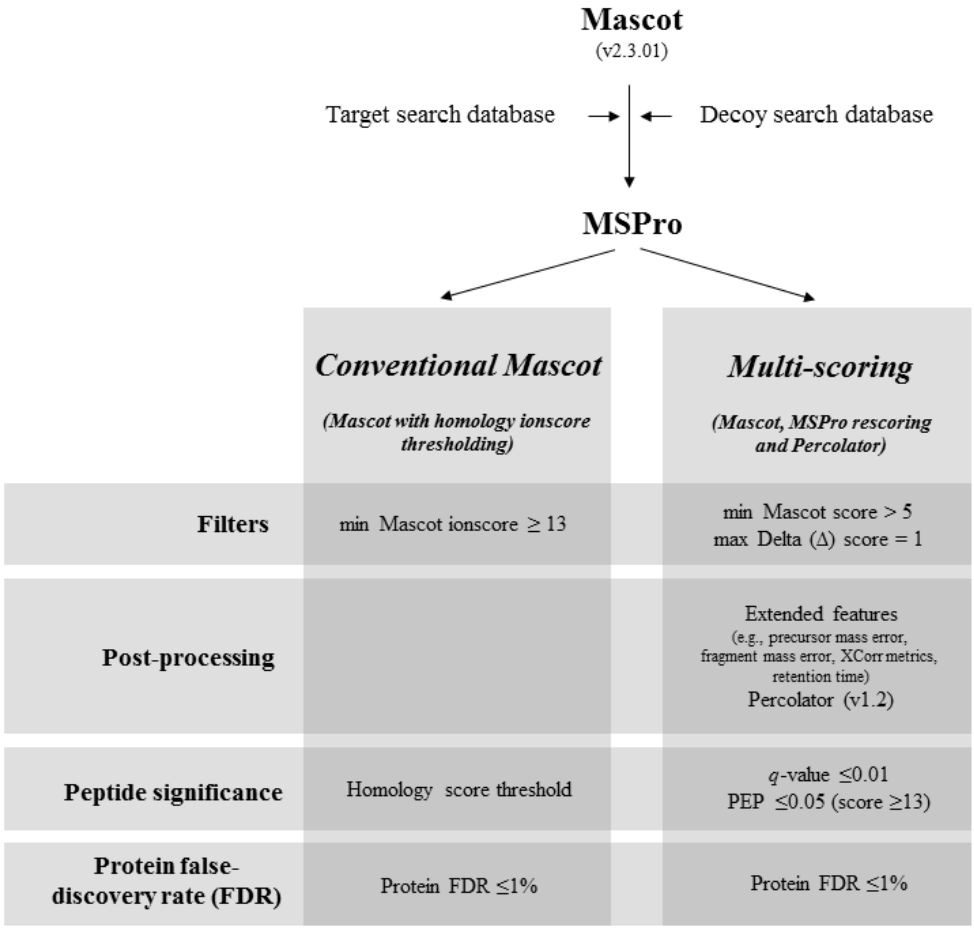
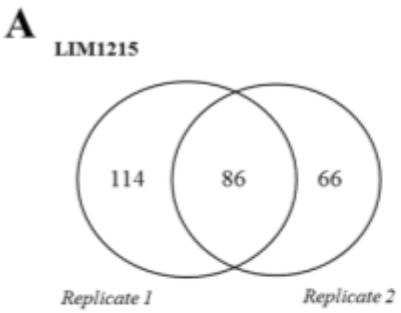
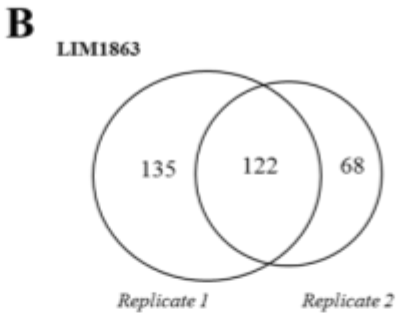
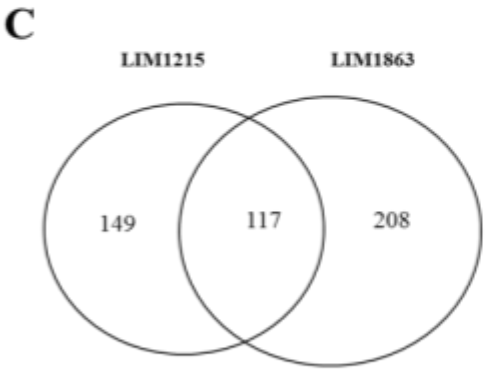

Mascot
(v2.3.01)
Target search database
Decoy search database
MSPro
Conventional Mascot
(Mascot with homology ionscore thresholding)
Multi-scoring
(Mascot, MSPro rescoring and Percolator)
Filters
min Mascot ionscore ≥ 13
min Mascot score > 5
max Delta (Δ) score = 1
Post-processing
Extended features
(e.g., precursor mass error, fragment mass error, XCorr metrics, retention time)
Percolator (v1.2)
Peptide significance
Homology score threshold
q-value ≤0.01
PEP ≤0.05 (score ≥13)
Protein false-discovery rate (FDR)
Protein FDR ≤1%
Protein FDR ≤1%

Figure 3







D

Accession Number	Gene Name	Protein Description	LIM1215 Secretopep Six Peptides	LIM1863 Secretopep Six Peptides
P05386	RPLP1	60S acidic ribosomal protein P1		4
Q8CBM7	DHCR7	7-dehydrocholesterol reductase	1	2
P60281	ACTG1	Actin, cytoplasmic 2	24	22
Q13740	ALCAM	CD166 antigen	1	
P27695	APEX1	DNA-(apurinic or apyrimidinic site) lyase		1
Q99988	GDF15	Growth differentiation factor 15	4	
O75096	LRP4	Low-density lipoprotein receptor-related protein 4	2	
Q14515	MEP1A	Meprin A subunit alpha		1
P13747	HLA-B	MHC class I antigen	10	
P30462	HLA-A	MHC class I antigen (Fragment)	11	
P35579	MYH9	Myosin-9	2	1
P02545	LMNA	Prelamin-A/C	1	
Q8NBP7	PCSK9	Proprotein convertase subtilisin/kexin type 9	1	3

	LIM1215	LIM1863
Growth Properties	adherent	floating organoid
APC Mutation[a]	wt	stop aa 1506
p53 Mutation[b]	wt	aa 234 Y to H
K-Ras Mutation[c]	wt	wt
B-Raf Mutation[d]	wt	wt
PI3 Kinase Mutation[e]	wt	wt

Figure 4

						N-terminal			C-terminal		
Accession Number	**Gene Name**	**Protein Description**	**Function/Expression**	**Peptide Sequence**		**P3,P2,P1**	**P1',P2',P3'**	**Cleavage**	**P1',P2',P3'**	**P3,P2,P1**	**Cleavage**
Q12864	CDH17	Cadherin 17	cell adhesion membrane-associated gastrointestinal glycoprotein	(N)DGGRPPLEGIVS(L)	748 - 759	RIN	DGG	Unknown protease	LPV	IVS	Unknown protease
				(R)HTDFEER(E)	732 - 738	STR	HTD	Unknown protease	EYV	EER	Unknown protease
				(I)NDGGRPPLEGIVS(L)	747 - 759	IRI	NDG	Unknown protease	LPV	IVS	Unknown protease
Q13740	ALCAM	CD166 antigen	cell adhesion transmembrane glycoprotein	(E)ISDENREKVNDQAK(L)	514 - 527	ADE	ISD	Unknown protease	LIV	QAK	Unknown protease
P29317	EPHA2	Ephrin type-A receptor 2	receptor tyrosine kinase, contact-dependent bidirectional signaling	(A)LTQEGQGAGSK(V)	510 - 520	VQA	LTQ	Unknown protease	VHE	GSK	Unknown protease
Q9BY67	CADM1	Cell adhesion molecule 1	cell-cell adhesion transmembrane glycoprotein	(R)AGEEGSIRA(V)	363 - 371	DSR	AGE	Unknown protease	VDH	IRA	Unknown protease
P21709	EPHA1	Ephrin type-A receptor 1	receptor tyrosine kinase, contact-dependent bidirectional signaling, induces cell attachment to extracellular matrix	(T)SPPVSRGLTGGEIVA(V)	536 - 550	FRT	SPP	Unknown protease	VIF	IVA	Unknown protease
Q9H3R2	MUC13	Mucin-13	cell signaling transmembrane protein, expressed in epithelial tissues	(S)DNEKTVTEKINK(A)	296 - 307	TTS	DNE	Unknown protease	AIR	INK	Unknown protease
Q99795	GPA33	Cell surface A33 antigen	transmembrane protein associated with intestinal epithelium expression and human colon cancer. Roles in cell-cell recognition and signaling	(K)NRVSISNNAEQS(D)	87 - 98	LYK	NRV	Unknown protease	DAS	EQS	Unknown protease
Q6UX71	PLXDC2	Plexin domain-containing protein 2	expressed in endothelial cells of the tumor stroma, roles in tumor angiogenesis	(H)LKDNGASTDDSAAEKKGGT(L)	433 - 451	ALH	LKD	Unknown protease	LHA	GGT	Unknown protease

Figure 5

	Cell Line	Peptide Sequence	Page on Supp Figure 3
1	*LIM1215*	ISVETPQDVLR	2
	LIM1215	ISVETPQDVLR	3
	LIM1215	ISVETPQDVLRA	4
	LIM1215	ISVETPQDVLRA	9
2	*SW1222*	ISVETPQDVLRASQGK	1
3	*LS174T*	ISVETPQDVLRASQGK	5
	LS174T	ISVETPQDVLRASQGK	6
4	*LIM1899*	ISVETPQDVLRASQGK	7
	LIM1899	ISVETPQDVLRASQGK	8

5	LIM2537	*N-terminal peptide N/Id*
6	HT29	*N-terminal peptide N/Id*
7	LIM2405	*N-terminal peptide N/Id*
8	LIM2550	*N-terminal peptide N/Id*
9	LOVO	*N-terminal peptide N/Id*
10	LIM2099	*N-terminal peptide N/Id*
11	LIM2408	*N-terminal peptide N/Id*
12	HCA7	*N-terminal peptide N/Id*
13	CACO2	*N-terminal peptide N/Id*
14	LIM1863	*N-terminal peptide N/Id*

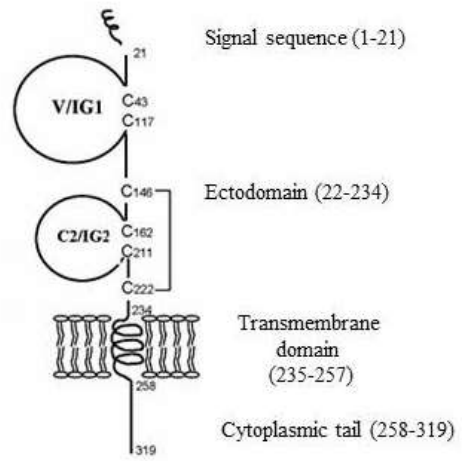


Figure 6

Gene Name	Cleaved peptide fragments	Sequence annotation
ADAM10	Ecto (3)	
ALCAM	Ecto (1)	
APLP2	Ecto (7) / Intramembrane proteolysis (1)	
APP	Ecto (8) / Intramembrane proteolysis (11)	
BTN3	Ecto (3)	
BSG	Ecto (3)	
CA9	Ecto (18)	
CADM1	Ecto (1)	
CD81	Ecto (1)	
CD9	Ecto (2)	
CD99	Ecto (5) / Intramembrane proteolysis (1) / ICD cleavage (2)	
CDH1	Ecto (1)	
CDH17	Ecto (3)	
CLPTM1	Ecto (1)	
CLSTN1	Ecto (3)	
DDR1	Ecto (1) / Intramembrane proteolysis (2)	
DSG2	Ecto (1) / Intramembrane proteolysis (1) / ICD cleavage (3)	
EFNB1	Ecto (2)	
EPCAM	Ecto (3)	
EPHA1	Intramembrane proteolysis (1)	
EPHA2	Ecto (1)	
EPHB4	Ecto (1)	
F11R	Ecto (1)	
FXYD3	Ecto (4)	
GPA33	Ecto (1)	
HLA-E	Ecto (5)	
ICOSLG	Ecto (5)	
IFI6	Intramembrane proteolysis (TM) (1)	
IGF1R	ICD cleavage (1) / Lumen (1)	
ITGA3	Ecto (1)	
ITGA6	Ecto (2)	
ITM2B	Lumen cleavage (2)	
LMAN2	Lumen cleavage (7)	
LRP4	Ecto (2)	
MEP1A	Ecto (1)	
MMP15	Ecto (2)	
PLXDC2	Ecto (1)	
PTK7	Ecto (2)	
SDC1	Ecto (9) / Intramembrane proteolysis (1)	
SDC4	Ecto (11)	
SEZ6L2	Ecto (1)	
SLC11A2	ICD cleavage (1)	
SLC16A1	ICD cleavage (1)	
SMAGP	Intramembrane proteolysis (1)	
SPINT2	Ecto (1) / Intramembrane proteolysis (1)	
ST14	Ecto (1)	
SVT7	Vesicle membrane proteolysis (1)	
TMEM200A	Ecto (1)	
TNFRSF1A	Intramembrane proteolysis (1)	
TPSH11	ICD cleavage (1)	

Table 1 - Mascot, calculated retention time, MSPro and Percolator (*Multi-scoring*) increases peptide identification rate in comparison to (*conventional*) Mascot with homology ionscore thresholding

Data Analysis Strategy[#]		**Comparison of Conventional and Multi-scoring**
Conventional Mascot[≠]	***Multi-scoring***[≠] (≤ 1% *q*-value)	
FDR ≤ 1% 238 proteins (788 peptides) 67 single peptide identifications	***≤ 5% PEP*** 474 proteins (1228 peptides) 267 single peptide identifications[*]	788/1228 = **36%** increase PSM identifications
	≤ 1% PEP 368 proteins (1103 peptides) 173 single peptide identifications	788/1103 = **29%** increase PSM identifications

[#] Refer to Figure 2 for overview of alternative data analysis strategies utilized in this study

[≠] Protein false-discovery rate (FDR) is ≤1% for each data analysis strategy

* MS/MS Spectra for single peptide identifications in the secretopeptidome refer Supplemental Figure 2

Table 2 - Predicted protein secretion pathways identified in the secretopeptidome

Classical secretion (SignalP)[a]	**Non-classical secretion** (SignalP and SecretomeP)[b]	**Integral membrane proteins** (KD>1.6)[c]	**Protein Ubiquitinilation**[d]	**Other**[e]	***Predicted secreted** (%)*
88	238	9	113	139	*70.5*

a Proteins predicted by SignalP to be classically secreted (SignalP v5.0)

b Proteins predicted to be non-classically secreted using SignalP and SecretomeP (<0.5) (v2.0)

c Integral membrane proteins predicted to contain transmembrane spanning domains by Kyte-Doolittle (>1.6) and not predicted by either SignalP or SecretomeP (<0.5)

d Ubiquitination proteins identified with mass spectrometry based on [33, 34]

e Proteins not classified as either classically secreted, non-classically secreted, or integral membrane proteins

Refer Supplemental Table 1 for protein information relating to these calculations

Table 3 - Extracellular and intramembrane cleavage of membrane proteins from secretopeptidome of colorectal cancer cells

Proteolytic cleavage[a]	Substrates (Gene name)	Protein substrates (#)[b]	Cleaved peptide fragments (#)[c]
Extracellular cleavage	ADAM10, ALCAM, B7H6, BSG, CA9, , CD9, CD81, CDH1, CDH17, CLPTM1, CLSTN1, EFNB1, EPCAM, EPHA2, EPHB4, F11R, FXYD3, GPA33, HLA-E, ICOSLG, ITGA3, ITGA6, LRP4, MEP1A, MMP15, MUC13, PLXDC2, PTK7, SDC4, SEZ6L2, ST14, TMEM200A, TNFRSF19	33	89 Ecto
Extracellular cleavage/ Intramembrane proteolysis	APLP2, APP, CADM1, DDR1, SDC1, SPINT2	6	27 Ecto / 17 Intra
Extracellular cleavage/ Intramembrane proteolysis/ Cytoplasmic cleavage	CD99, DSG2	2	6 Ecto / 2 Intra/ 5 Cyto
Intramembrane proteolysis	EPHA1, SMAGP, TNFRSF1A	3	3 Intra
Intramembrane proteolysis (TM)	IFI6	1	1 Intra (TM)
Cytoplasmic cleavage	ACSL1, ATP1A1, SLC11A2, SLC16A1, TP53I11, VAPB	6	6 Cyto
Lumen cleavage	ITM2B, LMAN2, XYLT1	3	10 Lumen
Cytoplasmic/ Lumen cleavage	IGF2R	1	1 Cyto/ 1 Lumen
Vesicle membrane proteolysis	SYT7	1	1 Vesicle membrane

Extracellular cleavage (Ecto)		**41**	**122**
Intramembrane proteolysis (RIP)		**12**	**23**
ICD proteolysis	**Cytoplasmic cleavage**	**9**	**12**
	Lumenal cleavage	**4**	**11**
	Vesicle membrane proteolysis	**1**	**1**

a Type of sequence cleavage based on peptide sequence (refer Supplemental Table 2 for peptide sequence for each substrate)

b Number of protein substrates undergoing specified cleavage (refer Supplemental Table 2 for peptide sequence for each substrate)

c Number of identified peptides (based on *q*-value ≤0.01, PEP ≤0.05), with the proteolytic cleavage type specified (refer Supplemental Table 2 for peptide sequence for each substrate). Substrate cleavage was determined based on identified peptide sequence. Extracellular cleavage refers to substrate cleavage of ectodomain, Intramembrane proteolysis (RIP) refers to substrate cleavage within transmembrane domain, where the peptide sequence is identified derived from within the transmembrane domain entirely referred to as Intramembrane proteolysis (TM). For Intracellular domain (ICD) proteolysis, substrate where derived from the cytoplasmic, lumen or vesicle membrane domains. Substrate where indicated were observed to undergo various types of proteolysis based on identified peptides from these domains.