PROTEOMICS IDENTIFICATION OF INTRACELLULAR AND SECRETED PROTEINS INVOLVED IN METASTASIS FROM A PAIR OF ISOGENIC COLORECTAL CANCER CELL LINES

LIN QIFENG

B. Sc. (Hons.), NUS

**A THESIS SUBMITTED
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY**

DEPARTMENT OF BIOCHEMISTRY

NATIONAL UNIVERSITY OF SINGAPORE

2014

Declaration

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Lin Qifeng

1 December 2014

Acknowledgements

First and most important of all, I express my greatest appreciation to my supervisor A/Prof Maxey Chung, for the opportunity to embark on this remarkable journey, and for his unwavering guidance and support throughout these five years. I have been and always will be deeply grateful that he was willing to accept a student with no experience in any protein work and had nothing but a keen interest to learn about proteomics. It has been a life-changing and rewarding experience to realise that proteomics is truly my lifelong interest, and for that I am profoundly thankful.

I also gratefully thank Dr Tan Hwee Tong, for all the countless advice and assistance, and the many occasions when I simply needed a second opinion. Many thanks go to Lim Teck Kwang, who has generously shared with me so much of his technical knowledge on LC and MS, as well as for being more a best friend than just a colleague. I am deeply obliged to Hendrick Loei for his helpful suggestions in the initial development of the hollow fibre culture system, and also to my three Honours students, Ryan Lim, Hannah Lim and Lin Hui Ling, for their assistance in the secretome projects. Additionally, I appreciatively acknowledge all past and present members of the Oncoproteomics group: Cynthia, Gek San, Aida, Vincent, Hou Qian, Wu Wei and Seow Chong, for the many helpful tips and the friendly companionship.

I am deeply indebted to Avery Khoo, Xin Xiu Sam, and Dr Tony Lim from the Department of Pathology, SGH, for their assistance on the immunohistochemistry work. I also thank Dr Cheong Wai Kit, Dr Rajeev Singh and Dr Eng Chon Boon (NUH Tissue Repository), as well as Nicole Kong, Prof Pierce Chow, A/Prof Tang Choong Leong and A/Prof Cheah Peh Yean (SingHealth Tissue Repository and SGH), for their support in acquiring the patient serum samples for ELISA work. Sincere thanks go to Prof Hooi Shing Chuan from Department of Physiology, NUS, for his generous gift of the E1 cell line.

In addition, I am thankful for the support from the staff of the Protein and Proteomics Centre, Say Tin, Xian Hui and Michelle, where our laboratory is based. I gratefully acknowledge Jason Neo and Dr Justin Lim from AB SCIEX as well, for their technical advice and support. Furthermore, I would like to take this opportunity to express my sincere appreciations to Dr Laszlo Orban, of the Reproductive Genomics Group at Temasek Life Sciences Laboratory, for the wonderful undergraduate research experience in his laboratory. I also gratefully appreciate the companionship of my friends, Shun Qiang, Zijie, Eddy, Jia Mei and Adeline, who begun this journey together with me. Last but certainly not the least of all: to my parents who have been my support and source of strength, I dedicate this work to them.

Table of Contents

Summary

Colorectal cancer is the fourth most common cause of cancer deaths and the high mortality rate is mostly ascribed to liver metastasis. In order to improve the clinical management of colorectal cancer metastasis and patient outcome, there is currently an urgent need to further the understanding of colorectal cancer metastasis, as well as to develop biomarkers with high sensitivity and specificity for prognosis and disease monitoring. To address these clinical needs, we applied proteomics approaches to identify intracellular and secreted proteins differentially expressed in the colon adenocarcinoma cell line HCT-116 and its metastatic derivative, E1.

In the first part of the study, we compared the intracellular protein profiles of HCT-116 and E1 cells using the iTRAQ technology, in order to identify potential novel metastatic-related proteins. Drebrin (DBN1), an actin-binding protein involved in cytoskeletal remodelling, was found to be overexpressed in E1. In comparison to primary colon adenocarcinoma tissue sections, DBN1 levels were significantly higher in matched lymph node and liver metastases, suggesting that DBN1 could be clinically relevant in colorectal cancer metastasis. Our findings represent the first time that DBN1 has been shown to be involved in colorectal cancer metastasis.

In the second part, we performed an iTRAQ-based comparison between the HCT-116 and E1 secretomes to search for novel biomarkers for prognosis and disease monitoring. We utilised the hollow fibre culture (HFC) system to

collect conditioned media (CM) samples from HCT-116 and E1 cells, and also showed that the HFC system is a highly attractive and viable alternative for CM preparation. Among the differentially secreted proteins identified from the E1 secretome, several well-known secreted proteins in colorectal cancer, such as GDF15, SPARC and SERPINE1, as well as potentially novel players such as PLOD3 and MAN1A1, were identified.

Finally, to delve deeper into the HCT-116 and E1 secretomes in search for more low-abundance secreted proteins, we used the multi-lectin affinity chromatography (MLAC) approach to enrich for secreted glycoproteins. We compared the MLAC-enriched "glycosecretomes" of HCT-116 and E1 using the label-free quantitative SWATH-MS technology, and observed that a potentially novel secreted protein in colorectal cancer, Laminin subunit β-1 (LAMB1), was oversecreted in E1 cells. LAMB1 levels were also significantly higher in patient serum samples as compared to healthy controls when measured using ELISA. ROC analyses indicated that LAMB1 performed better at discriminating between colorectal cancer patients from controls with an AUC of 0.86 (64% sensitivity, 96% specificity), as compared to CEA with an AUC of 0.74 (53% sensitivity, 89% specificity). More importantly, when LAMB1 is used in combination with CEA, the diagnostic performance was further improved, showing an AUC of 0.91 (80% sensitivity, 92% specificity).

List of Figures

List of Tables

List of standard abbreviations

2-DE	Two-dimensional electrophoresis
AUC	Area under the curve
ACN	Acetonitrile
CID	Collision-induced dissociation
CM	Conditioned media
Con A	Concanavalin A
DAB	3,3'-diaminobenzidine
DDA	Data-dependent acquisition
DIA	Data-independent acquisition
ECM	Extracellular matrix
ECS	Extra-cappillary space
ELISA	Enzyme-linked immunosorbent assay
ER	Endoplasmic reticulum
ESI	Electrospray ionisation
FBS	Fetal bovine serum
FDR	False discovery rate
HFC	Hollow fibre culture
HRP	Horseradish peroxidase
IHC	Immunohistochemistry
IPI	International Protein Index
iTRAQ	Isobaric tags for relative and absolute quantitation
JAC	Jacalin
LC	Liquid chromatography

MALDI	Matrix-assisted laser desorption ionisation
MLAC	Multi-lectin affinity chromatography
MMTS	Methyl methane-thiosulfonate
MRM	Multi-reaction monitoring
MS	Mass spectrometry
MSMS	Tandem mass spectrometry
MW	Molecular weight
MWCO	Molecular weight cut-off
NK cells	Natural Killer cells
PCR	Polymerase chain reaction
PTM	Post-translational modification
PPM	Parts per million
ROC	Receiver operator characteristics
RP	Reversed phase
RT	Retention time
S.D.	Standard deviation
SCX	Strong cation exchange
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate - polyacrylamide gel electrophoresis
SRM	Selected reaction monitoring
TCEP	Tris(2-carboxyethyl)phosphine
TEAB	Triethyl ammonium bicarbonate
TOF	Time of flight
WGA	Wheat germ agglutinin
XIC	Extracted-ion chromatogram

1. Introduction

1.1. Colorectal cancer epidemiology

Colorectal cancer is currently ranked as the third most widespread cancer in the world, with over 1.2 million new cases estimated in 2008. The highest incidence rates of colorectal cancer occur in developed countries, such as Australia, New Zealand, Europe and North America. With more than 608 000 deaths estimated, colorectal cancer is the fourth most common cause of cancer deaths (Jemal *et al.*, 2011). Similar trends are observed in Singapore, where colorectal cancer has the highest incidence rate and second most frequent cause of death among all cancers in males, and the second most common and third in cancer-related deaths among females (Lee *et al.*, 2014).

1.2. Colorectal cancer aetiology

Colorectal cancer develops from the epithelial cells of the interior lining of the large intestines and the rectum. The transformation of normal epithelial cells into a benign precursor lesion (adenoma) and subsequently to an adenocarcinoma requires a complex interplay of changes at the molecular level. Constitutive activation of the Wnt signalling pathway is one of the most common genetic aberrations and can be found in about 93% of all colorectal cancer cases (de Wit *et al.*, 2013). This is most often caused by mutations in the adenomatous polyposis coli (*APC*) tumour suppressor gene (Cancer Genome Atlas, 2012; Kinzler and Vogelstein, 1996). There are also a number of other genetic alterations which occur at high frequency in colorectal cancer, including activating mutations in KRAS and BRAF, inactivation of TP53, alterations of PI3K/Akt, and the TGFβ signalling pathways (Jones *et al.*, 2008).

Many have also recognised that the introduction of genomic instability is one of the major factors that contributes to colorectal cancer malignancy (de Wit *et al.*, 2013). Approximately 70% of colorectal carcinomas would display chromosomal instability (CIN), where large portions of chromosomes could be gained or lost (Pino and Chung, 2010). These changes are usually associated with accumulation of mutations in oncogenes and tumour suppressor genes which are crucial for colorectal cancer progression (de Wit *et al.*, 2013). Another less frequently observed form of genomic instability is microsatellite instability (MSI), which can be found in about 15% of colorectal cancer patients (Boland and Goel, 2010). In these cases, a defect in the DNA mismatch repair mechanism results in accumulation of mutations in microsatellites. This leads to frameshift insertions or deletions, which affect the transcription of genes involved in DNA repair, cell cycle regulation, proliferation and apoptosis (Boland and Goel, 2010; de Wit *et al.*, 2013).

1.3. Colorectal cancer staging systems

Cancer staging is one of the most crucial elements during diagnosis to define the tumour and determine the most appropriate treatment for the patient. In colorectal cancer, the two most commonly employed staging systems are the Dukes' classification (Dukes, 1932), and the tumour node metastasis (TNM) system (maintained collaboratively by the American Joint Committee on Cancer and the Union for International Cancer Control). The Dukes' classification (Table 1) was originally described in 1932 by a pathologist

Cuthbert E. Dukes for the classification for rectal cancers, where three classes, A, B and C, were used to describe the extent of rectal tumour spread (Dukes, 1932). Class C was later divided into C_1 and C_2, to further differentiate between the extent of lymph node metastases (Gabriel *et al.*, 1932). Subsequently, the Dukes' classification was extended to include colon tumours as well (Simpson and Mayo, 1939). Although not described in the original classification, an additional Class D has also become commonly used to denote tumours with distant metastasis (Deans *et al.*, 1992). However, the Dukes' classification has been complicated by other further modifications, resulting in frequent confusion whenever the classification is used (Deans *et al.*, 1992).

Table 1. The modified Dukes' classification for colorectal cancer staging

Modified Dukes' classification	Description
A	**Tumour growth is limited to the mucous membrane and submucosa**
B	**Tumour invades through intestinal walls, but no metastasis in regional lymph nodes**
C	**Involvement of lymph nodes**
C_1	Only regional lymph nodes involved
C_2	Metastasis spread to lymph nodes at point of ligature of blood vessels
D	**Distant metastasis**

The TNM system for colorectal cancer is largely compatible with the Dukes' classification, but is more precise in identifying sub-groups which have different prognosis. Colorectal cancer staging using the TNM system categorises the tumours based on the depth of tumour invasion into or beyond the intestinal walls (T), number of regional lymph nodes involved (N), and the presence of distant metastases (M). Combinations of the T, N and M characteristics are then used to group patients into different prognostic stages (Table 2).

Table 2. The TNM system for colorectal cancer staging
The tumour node metastasis (TNM) system for staging colorectal cancers, based on the AJCC Cancer Staging Handbook, Seventh Edition

Table 2a. Definitions of TNM

Primary tumour (T)

TX	Primary tumour cannot be assessed
T0	No evidence of primary tumour
Tis	Carcinoma in situ: intraepithelial or invasion of lamina propria
T1	Tumour invades submucosa
T2	Tumour invades muscularis propria
T3	Tumour invades through the muscularis propria into pericolorectal tissues
T4a	Tumour penetrates into surface of the visceral peritoneum
T4b	Tumour directly invades or is adherent to other organs or structures

Regional lymph nodes (N)

NX	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	Metastasis in one to three regional lymph nodes
N1a	Metastasis in one regional lymph node
N1b	Metastasis in two to three regional lymph nodes
N1c	Tumour satellite deposits in subsierose or in non peritonealised tissues
N2	Metastases in more than four regional lymph nodes
N2b	Metastases in four to six regional lymph nodes
N2c	Metastases in seven or more regional lymph nodes

Distant metastasis (M)

M0	No distant metastasis
M1	Distant metastasis
M1a	Metastases confined to one organ or site (e.g. liver, lung, ovary or non-regional node)
M1b	Metastases in more than one organ/site or the peritoneum

Table 2b. Anatomic stage/Prognostic groups

AJCC/UICC staging	T	N	M	Dukes
0	Tis	N0	M0	-
I	T1	N0	M0	A
	T2	N0	M0	A
IIa	T3	N0	M0	B
IIb	T4a	N0	M0	B
IIc	T4b	N0	M0	B
IIIa	T1-T2	N1/N1c	M0	C
	T1	N2a	M0	C
IIIb	T3-T4a	N1/N1c	M0	C
	T2-T3	N2a	M0	C
	T1-T2	N2b	M0	C
IIIc	T4a	N2a	M0	C
	T3-T4a	N2b	M0	C
	T4b	N1-N2	M0	C
IVa	Any T	Any N	M1a	D
IVb	Any T	Any N	M1b	D

As mentioned earlier, staging plays an important role in colorectal cancer disease management as it is the main factor that influences prognosis and treatment decisions. The five-year survival rate for patients who present with early stage disease (stage I) is more than 90%, while stage II patients would have about 70% to 80% survival rates. However, when metastasis to the regional lymph nodes have occurred for stage III patients, the five-year survival is reduced to 40% to 80%, depending on the number of lymph nodes involved. For stage IV patients who have colorectal cancer with distant metastasis, the five-year survival rate for patients is only 8% (O'Connell *et al.*, 2004). From this, it is clear that the high mortality rate in colorectal cancer is

due to distant metastasis. Unfortunately, due to the asymptomatic nature of colorectal cancer in the early stages, many patients would already present with metastatic disease during diagnosis (Paschos and Bird, 2008).

1.4. Clinical needs in colorectal cancer detection and management

The development from a benign adenoma to a metastatic adenocarcinoma progresses through several well-defined stages, and the disease outcome is heavily dependent on the stage at which the disease is detected (de Wit *et al.*, 2013). At every stage of colorectal cancer progression, there are specific clinical needs to improve detection and management of the disease (Figure 1).

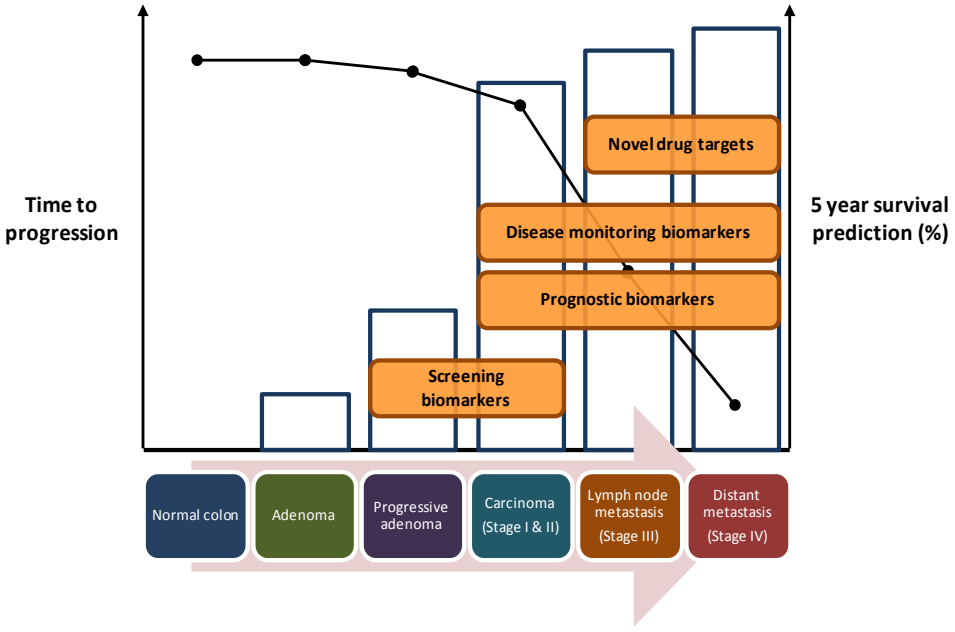


Figure 1. Clinical needs in colorectal cancer management
The stage-wise progression of colorectal cancer is shown on the x-axis. On the primary y-axis, the time to progression is shown using bars with blue outline, while on the secondary y-axis, the five year survival rate is shown using black lines with dots. The different clinical needs are shown in orange boxes at the stages of the disease where they are most pertinent. Screening biomarkers are required at early stages of the disease where clinical intervention is most effective. Prognostic biomarkers are needed from Stage I onwards to predict outcome, and similarly so for disease monitoring biomarkers, where they are used for monitoring disease recurrence. Novel drug targets are particularly important for patients who present with late stage colorectal cancer, in order to reduce the high mortality rate and improve patient outcome (de Wit *et al.*, 2013).

1.4.1. Screening biomarkers: early detection of colorectal cancer

First of all, since it takes about 17 years for an adenoma to progress to a carcinoma (Jones *et al.*, 2008), developing methods for detection of the disease at early stages would be one of the most logical approaches to reduce mortality in colorectal cancer. Moreover, surgical resection is highly curative at early stages of the disease (O'Connell *et al.*, 2004). Currently, the most widely employed procedures for colorectal cancer detection are colonoscopy and the faecal occult blood test (FOBT), where stools are tested for the presence of haemoglobin as an indication of tumour bleeding. However, colonoscopy has poor patient compliance as it is expensive and invasive. Moreover, even though colonoscopy is generally regarded as the gold standard in detection of colorectal cancers and precancerous polyps, it has a miss rate of 25% for detection of both polyps and adenomas (Leufkens *et al.*, 2012). On the other hand, the FOBT is non-invasive, but has poor specificity as the presence of blood in the stools could be due to non-tumour lesions. As such, there is a need for biomarkers which are directly related to the molecular characteristics of the tumour tissue, in order to improve the sensitivity and specificity of the detection methods currently available (de Wit *et al.*, 2013).

1.4.2. Biomarkers for prognosis and disease monitoring

Next, as mentioned earlier, prognosis and treatment decisions for colorectal cancer patients are mainly guided by the clinical and/or pathological staging of the tumour. Patients who are diagnosed with stage I or II tumours would undergo surgical resection only, while stage III patients would be

recommended for adjuvant chemotherapy (Labianca *et al.*, 2013). However, 25-30% of patients would develop metachronous hepatic metastases (Paschos and Bird, 2008). Currently, the only biomarker that is clinically used for monitoring of colorectal cancer recurrence and metastasis is the carcinoembryonic antigen (CEA). Levels of CEA are routinely monitored in colorectal cancer patients following surgical resection as high CEA levels have been shown to be associated with poor patient outcome (Wang *et al.*, 1994). However, elevation of CEA can also occur in response to inflammatory conditions such as hepatitis and inflammatory bowel disease (van der Schouw *et al.*, 1992). Furthermore, there is conflicting data in the literature regarding the prognostic utility of CEA (Carpelan-Holmstrom *et al.*, 1996; Carriquiry and Pineyro, 1999; Harrison *et al.*, 1997; Moertel *et al.*, 1986). Thus, there is also an urgent need for more biomarkers with higher sensitivity and specificity to improve prognosis and decision making for colorectal cancer patients.

1.4.3. Novel drug targets for intervention of colorectal cancer metastasis

Even though metastasis is the major contributing factor in colorectal cancer mortality, the mechanisms that drive the metastatic transformation remain poorly understood. Moreover, as compared to the progression from adenoma to carcinoma, the duration required for a carcinoma to further acquire metastatic traits appears to be much shorter (approximately 2 years) (Jones *et al.*, 2008). Currently, the treatment strategy for patients who present with stage IV disease or develop metachronous metastasis is mainly palliative

chemotherapy, especially when resection is no longer a suitable option. This is typically a "one-size-fits-all" strategy (de Wit *et al.*, 2013), where the treatment for most patients consists of a fluoropyrimidine (fluorouracil) in various combinations and schedules (Van Cutsem *et al.*, 2010). However, these combinatorial fluorouracil therapies are not targeted against metastasis, but simply induce cell death through inhibition of RNA and DNA synthesis (Longley *et al.*, 2003).

Recently, targeted drug therapies such as monoclonal antibodies against vascular endothelial growth factor (VEGF) and epidermal growth factor receptor (EGFR) have been recommended to be used in conjunction with chemotherapy as they seemed to improve outcome in some colorectal cancer patients. These anti-VEGF and anti-EGFR treatments are targeted against angiogenesis (Hurwitz *et al.*, 2004) and the MAPK-signalling pathway (Jonker *et al.*, 2007) respectively. Unfortunately, there have been some complications and restrictions reportedly associated with these therapies. For instance, the effectiveness of the anti-EGFR drugs are restricted to KRAS wild-type tumours only (Van Cutsem *et al.*, 2010). These further underscore the urgency to improve the understanding of the molecular mechanisms involved in colorectal cancer metastasis, which could then lead to more effective patient treatment.

1.5. Colorectal cancer liver metastasis

Mortality in colorectal cancer can be mostly ascribed to metastasis. As mentioned earlier, the five-year survival prediction of stage IV colorectal cancer patients is 8.1%, a drastic reduction from the 44.3% for stage IIIc patients (O'Connell *et al.*, 2004). The most prominent form of distant metastasis that is observed in colorectal cancer patients is liver metastasis. At the point of diagnosis, about 25% of patients would present with hepatic metastasis (synchronous metastasis), while a further 25-30% would eventually develop secondary tumours in the liver (Paschos and Bird, 2008). Even for those patients who are suitable and undergo surgical resection, about 1 in 2 cases would develop recurrence, and more than 60% of these cases would be hepatic-related (de Jong *et al.*, 2009).

The progression from a local adenocarcinoma to metastasis involves a complex series of events known as the invasion-metastasis cascade (Valastyan and Weinberg, 2011). To generate a clinically detectable neoplasm at secondary sites (Figure 2), the epithelial cells in the primary tumour must first acquire the ability to invade into the surrounding extracellular matrix (ECM) and stromal cells, and then intravasate into the blood vessels supplying the primary tumour. Once in the blood circulation, the tumour cells must subsequently endure the rigors of transport through the vasculature and evade the host immune system, before arresting at a distant organ site. Finally, the tumour cells would need to extravasate into the distant tissue, and then survive in the hostile microenvironment to proliferate and colonise the metastatic site.

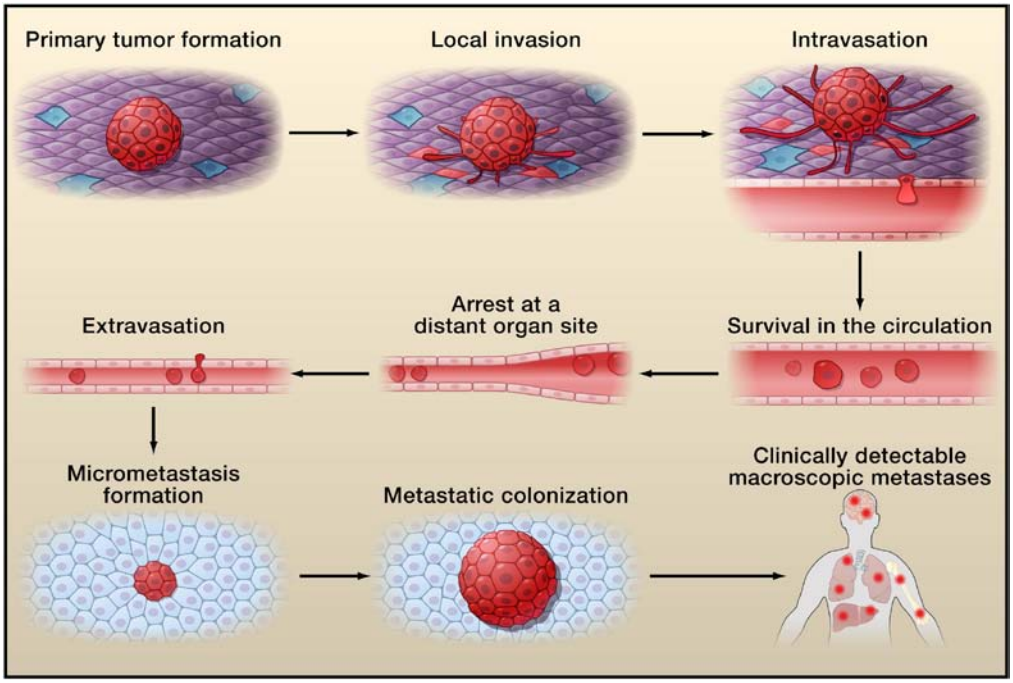


Figure 2. The invasion-metastasis cascade
The invasion-metastasis cascade is a progression of the events that results in the formation of clinically detectable metastases from a primary tumour. Metastasising cells must first acquire the ability to leave the primary tumour site through local invasion and intravasation. They must then be able to survive in the circulation before arresting and extravasating at a distant organ site. Finally, the tumour cells must adapt to and colonise the foreign environment of the secondary site. Carcinoma cells are shown in red. Figure reproduced from Valastyan and Weinberg, 2011, with permission from Cell, Elsevier Inc.

However, although the general metastatic process is understood as such, the complex cellular pathways and mechanisms that drive colorectal cancer metastasis still remain largely unknown. Therefore, in order to address the clinical needs in colorectal cancer metastasis intervention, a more comprehensive understanding of the molecular mechanisms that underlie the metastatic process is required. However, the players (*i.e.* proteins) involved in these molecular mechanisms hardly act as isolated units, but are interconnected in complex and intricate pathways with other proteins and

cellular components. As such, it would be logical to apply a systems-based approach (as opposed to traditional single gene/protein analyses) to more effectively elucidate and understand the proteins involved in metastasis. In this respect, the comparative analysis of the global protein expression profiles using proteomics between metastatic and primary colorectal cancer has been shown to be an extremely valuable and appealing approach (Bitarte *et al.*, 2007).

1.6. Quantitative proteomics technologies

1.6.1. Gel-based and liquid chromatography (LC)-based quantitative proteomics

Comparative quantitative proteomics involves systemic identification and quantitation of the proteins expressed between related biological samples, and is most often applied to evaluate global changes in protein abundances that occur in response to any physiological events. The traditional workhorse of quantitative proteomics is the two-dimensional gel electrophoresis (2-DE) method, where all the proteins in a complex protein sample are first fractionated according to their respective isoelectric points, and then separated according to molecular weight (O'Farrell, 1975). However, there are limitations to the 2-DE method, such as inter-gel variation (Righetti *et al.*, 2004) and poor resolution for highly basic proteins and those smaller than 10 kDa (Bitarte *et al.*, 2007).

An alternative to overcome these problems is to switch to liquid chromatography (LC)-based platforms to separate the complex protein samples. LC-based platforms are advantageous over gel-based methods in that the separation of the complex protein samples can be easily automated and directly coupled to the mass spectrometer for detection and identification of proteins. Moreover, LC separations can be enhanced by performing multi-dimensional separations through digestion of the proteins into peptides, and then exploiting the different chemical interactions between the peptides and the stationary phase (Wu *et al.*, 2006). This vastly improves the separation of complex biological samples and enables the identification of higher numbers of proteins as compared to gel-based methods (Wang and Hanash, 2003).

1.6.2. Stable isotopic labelling strategies for LC-based quantitative proteomics

The introduction of stable isotopic labelling strategies together with high-throughput multi-dimensional LC platforms was an important factor in driving the popularity of comparative quantitative proteomics analyses between different biological samples. These strategies involve the incorporation of non-radioactive stable isotopic tags onto specific amino acids through metabolic or chemical means. Metabolic labelling entails the use of culture media containing modified essential amino acids (with deuterium, ^{13}C, or ^{15}N) that is incorporated *in vivo* into proteins synthesised by the cells during culture (Ong *et al.*, 2002). However, this procedure, known as stable isotope labelling by

amino acids in culture (SILAC), can only be performed on *in vivo* biological samples (*i.e.* cell line and animal models).

On the other hand, chemical labelling techniques can be performed on any biological samples, as the labelling is usually performed post-protein isolation using isotopic labelling reagents. One such method is the isobaric tags for relative and absolute quantitation (iTRAQ) technology. The iTRAQ approach was originally based on four isobaric (same mass) reagents that are covalently attached to the N-termini and lysine side chains of post-tryptic digest peptides (Ross *et al.*, 2004). Each of the iTRAQ reagents consists of a reporter group, a balance group and the amine-reactive group. The reporter groups of the four isobaric reagents have masses ranging from 114 to 117 *m/z*, and the balance groups have varying masses to maintain the same overall mass for all four reagents.

The typical iTRAQ labelling workflow involves the simultaneous tryptic digestion of up to four different samples followed by labelling of the four samples with each of the four isobaric iTRAQ reagents. The resultant iTRAQ-derivatised peptides can then be pooled and analysed in a single LC-MS experiment. During MS, because of the isobaric nature of the iTRAQ reagents, identical peptides from different samples (labelled with different iTRAQ reagents) are indistinguishable from each other because they will have the same precursor *m/z*. Subsequently, during collision induced dissociation (CID) of the precursor ions in MSMS mode, the reporter groups are released from the peptides and generate strong peak signals which can then be analysed to

infer the relative abundance of the labelled peptides. Interestingly, the iTRAQ reagents have been shown to enhance the fragmentation efficiency of the labelled peptides, leading to improved signals from the daughter ions and overall better sequence coverage (Hardt *et al.*, 2005). Moreover, the iTRAQ technology has been shown to be complementary to gel-based techniques in terms of protein identification, but is much more sensitive (Wu *et al.*, 2006). Currently, the iTRAQ technology has been improved for use of up to eight isobaric reagents (ranging from 113 to 121 *m/z*, excluding 120 *m/z*). Therefore, this allows for concurrent analysis of up to eight different biological samples in a single LC-MS/MS experiment, enabling iTRAQ to remain as one of the most powerful tools for LC-based quantitative proteomics.

1.6.3. SWATH-MS - a label-free alternative for quantitative proteomics

Although labelling-based strategies are known to be more accurate in protein quantitation, they have several restrictions, which include limited number of samples analysable in a single run, requiring expensive isotope labels and complex sample preparations (Neilson *et al.*, 2011; Patel *et al.*, 2009). Therefore, there is increasing interest recently in exploring quantitative proteomics strategies which do not require any stable isotopic labelling (*i.e.* label-free quantitative proteomics). Currently, there are two different approaches for label-free quantitation: area under the curve (AUC) measurement and spectral counting (Neilson *et al.*, 2011). AUC measurement, also known as ion counts, uses the integrative measurement of the precursor

ion abundance at a specific retention time as a reflection of the respective peptide abundance (Podwojski *et al.,* 2010). In contrast, spectral counting involves quantitation by counting the number of MSMS spectra assigned to a particular peptide, and is based on the premise that more abundant peptides will have higher tendency to be selected for fragmentation, and therefore should produce more MSMS spectra (Liu *et al.*, 2004).

SWATH-MS is an AUC measurement-based label-free quantitation method that was recently developed. Conceptually, SWATH-MS is closer to selected reaction monitoring (SRM, also known as multiple reaction monitoring or MRM) in targeted proteomics. SRM involves the monitoring of a pre-determined pairs of peptide precursors and fragment ions (known as transitions). Quantitation is performed by measuring the intensities of the fragment ions, and is therefore essentially AUC measurement-based (Neilson *et al.,* 2011). Although highly accurate and reproducible, SRM is limited by the number of proteins that can be analysed in a single LC-MS experiment. Therefore, SRM is most often used as a validation tool for both label-free and labelled quantitative proteomics experiments.

SWATH-MS overcomes the limitations of SRM by merging the high throughput protein identification capabilities of discovery proteomics with the reproducible and accurate quantitation of targeted proteomics (Gillet *et al.*, 2012). This is achieved by using an unbiased data independent acquisition (DIA) method, where all precursor ions within a specific mass window are fragmented for MSMS analysis (Figure 3). This is unlike the traditional data

dependent acquisition (DDA) method used in discovery proteomics, where the most abundant ions are selected for fragmentation by the mass spectrometer.

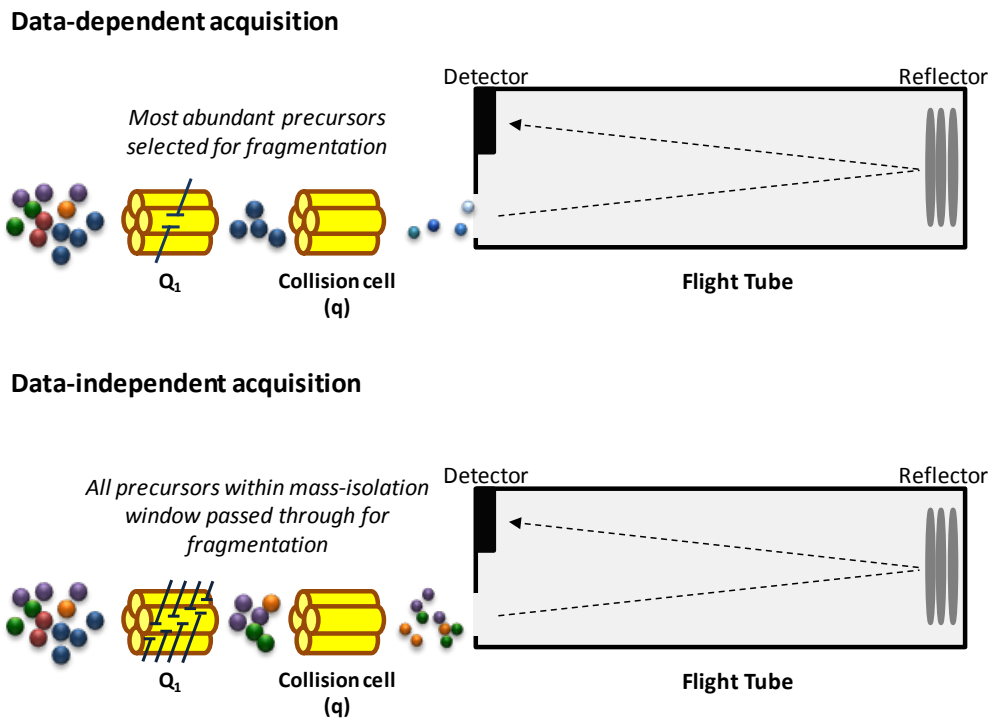


Figure 3. Data-dependent acquisition (DDA) *vs.* data-independent acquisition (DIA)
In DDA, the most abundant precursors are selected by the mass spectrometer for fragmentation, and the resultant spectra is composed of the fragment ions of the selected precursor peptide only. However, in DIA, a pre-determined mass isolation window is specified, and all precursors within this mass range are allowed to pass through for fragmentation. Thus, the resultant spectra is a recording of all fragment ions that are derived from all the precursors that were within the isolation window.

Specifically, the SWATH-MS DIA workflow involves the repeated and continuous acquisition of fragment ion spectra from all precursor ions within sequential series of 25 Da isolation windows throughout a specified mass range of 350 to 1250 m/z (termed as "swaths", see Figure 4). The resultant output is therefore composite spectra of all the fragment ions that are derived from all the precursor ions within each swath. In order to identify peptides that are present within these composite spectra, a spectra ion library that contains peptide identities with their corresponding fragment ion information would be required, and is usually generated by traditional DDA and database searching. Quantitation is then performed by AUC measurement of the highest intensity fragment ion signals. Due to the high resolution of the quadrupole time-of-flight (QqTOF) instrument, it has been shown that the accuracy and reproducibility of SWATH-MS quantitative analysis is comparable to that of SRM, which is considered as the gold standard in quantitative mass spectrometry analysis (Liu *et al.*, 2013). Considering these, SWATH-MS could become a powerful label-free alternative for performing unbiased quantitative proteomics comparisons between biological samples.

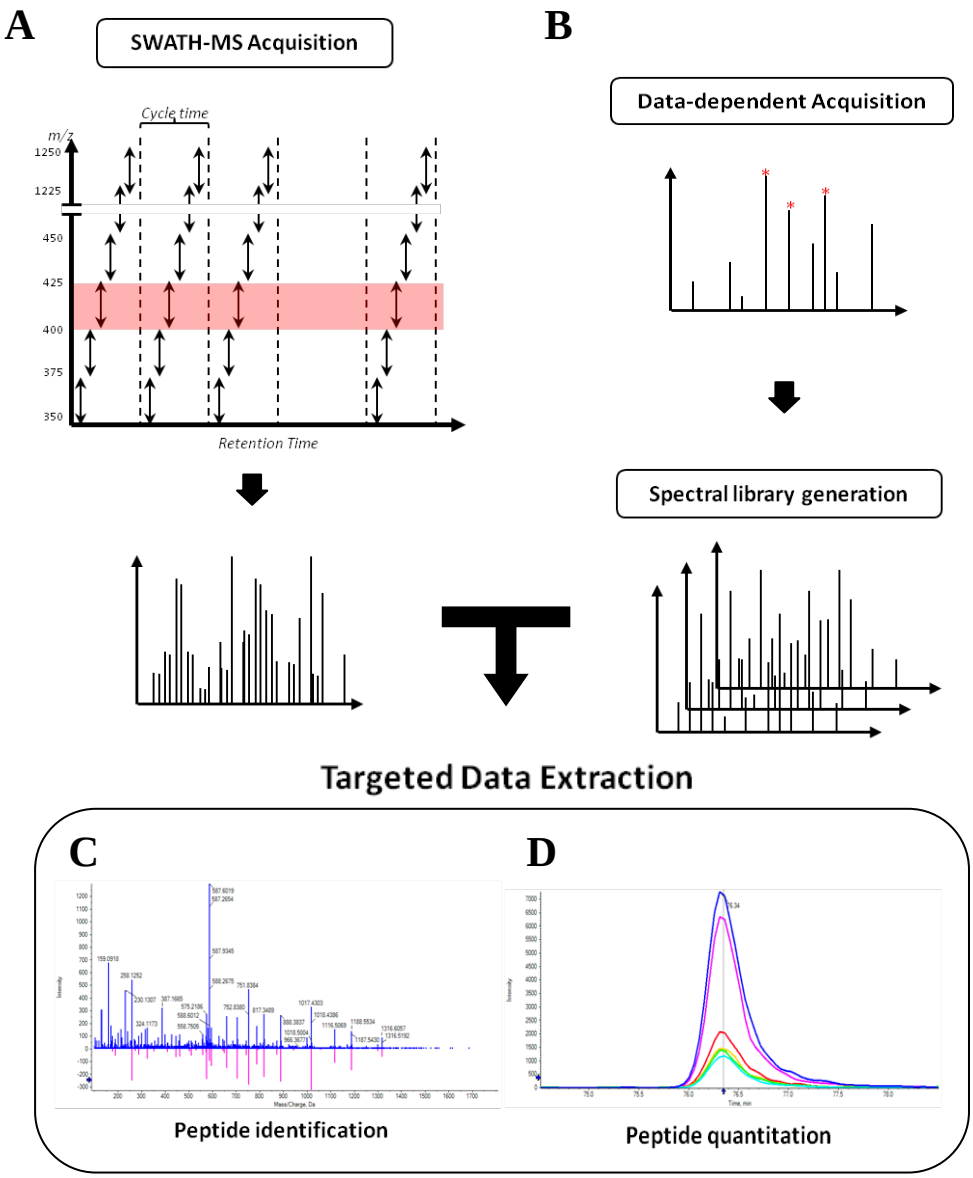


Figure 4. The SWATH-MS workflow
(A) SWATH-MS data-independent acquisition (DIA) consists of the continuous acquisition of high resolution fragment ion spectra across the entire chromatographic elution time range (retention time) by repeatedly stepping through distinct precursor isolation windows of 25 Da width (black double arrow) throughout the specified mass range of 350 – 1250 *m/z*. The series of isolation windows acquired for a specific mass range (e.g. 400 – 425 *m/z*) across the retention time range is denoted as a "swath" (highlighted in red) (Gillet *et al.*, 2012). The result of the SWATH-MS acquisition are spectra that are composite of all fragment ions derived from the precursors within each swath. **(B)** In order to identify the peptides within the spectra, traditional data-dependent acquisition (DDA) is used to generate a spectra ion library containing fragment ion spectra with corresponding peptide identity information. **(C)** Peptide identification is performed by matching the fragment ion spectra from the SWATH-MS analysis (top, blue) against that of the spectra ion library (bottom, purple), while **(D)** quantitation is performed by extracting the highest intensity fragment ions and overlaying them in a SRM-like extracted ion chromatogram (XIC). Peptide abundance is then estimated by AUC measurement of the fragment ion signals in the XIC.

1.7. Addressing clinical needs in management of colorectal cancer metastasis using proteomics

1.7.1. Identification of novel intracellular proteins involved in colorectal cancer metastasis using comparative quantitative proteomics

Proteomics approaches are ideal for discovering new and unexpected protein relationships because of its unbiased nature. As such, comparative quantitative proteomics analyses between primary and metastatic colorectal cancer will be highly useful for elucidating novel changes in protein expression levels that could be related to metastasis. Although comparing tissue extracts from synchronous or metachronous liver metastasis with the primary tumour would be the ideal case study, such samples are often difficult to obtain, and yet the intrinsic biological heterogeneity among individual patients would mean that larger samples numbers are required (Chen and Yates, 2007; Chen *et al.*, 2006). Cell lines, on the other hand, are renewable resources and highly suitable for the large amount of samples required for proteomic analyses (Chen and Yates, 2007). Furthermore, they are mostly homogenous and easily manipulated for downstream functional studies.

Various groups around the world have reported utilising comparative proteomics approaches on either cell lines or patient tissues to study colorectal cancer liver metastasis. Using comparative 2-DE analyses, a number of interesting targets have been identified, such as the mitochondrial F_0F_1-ATP synthase (Chang *et al.*, 2007), hnRNPA1 (Ma *et al.*, 2009), RhoGDI (Zhao *et*

al., 2008), HMGB1 (Liu *et al.*, 2007; Zhao *et al.*, 2007) and HSP27 (Liu *et al.*, 2007; Pei *et al.*, 2007; Zhao *et al.*, 2007).

Considering the advantages of LC-based quantitative proteomics approaches over gel-based methods mentioned earlier, iTRAQ would be a highly appealing approach to discover more novel proteins involved in colorectal cancer metastasis. Currently, there have been several reports which have utilised the iTRAQ technology to discover diagnostic biomarkers in colorectal cancer tissues (Besson *et al.*, 2011; Fan *et al.*, 2012) and plasma samples (Zhang *et al.*, 2012a), as well as investigating the effects of butyrate treatment (Tan *et al.*, 2008) and a novel phosphoinositide-3 kinase inhibitor drug (Mallawaaratchy *et al.*, 2012) on colorectal cancer cells. However, only Ghosh and colleagues (2011) performed an iTRAQ-based comparison between the primary colon cancer cell line SW480 and its lymph node metastatic variant SW620 in search of proteins involved in colorectal cancer metastasis. It should be additionally noted that the abovementioned studies in the previous paragraph only utilised the SW480 and SW620 pair of cell lines as well, and thus may potentially be introducing a biased perception of the colorectal cancer metastasis mechanism (Schaaij-Visser *et al.*, 2013). As such, it will be useful to examine other isogenic colorectal cancer cell lines with different metastatic potentials to gain a different perspective on other novel proteins involved in metastasis.

1.7.2. Identification of serological biomarkers by analysing the colorectal cancer secretome

In clinical screening, blood-based assays are most commonly employed because of their simplicity and non-invasiveness. Unfortunately, biomarker discovery using patient blood samples can be extremely challenging due to the broad dynamic range of serum and plasma protein concentrations. The presence of the high abundance proteins, such as albumin, haptoglobin, transferrins and immunoglobins, hinders the detection of tumour-specific biomarkers, which are usually at very low concentration ranges of nanograms per millilitre (Zhang and Chan, 2007). Consequently, there is increasing interest in proteomics analyses of other proximal biological fluids and in particular, the “secretome”: proteins secreted from cancer tissue specimens and cell lines.

The term "secretome" was originally used by Tjalsma and colleagues (2000) to describe the total proteins that are released by a cell, tissue or organism. These secreted proteins constitute approximately 10-15% of the total proteins encoded by the human genome and are known to be involved in important physiological processes including immune defense, blood coagulation, matrix remodelling and cell signalling (Karagiannis *et al.*, 2010). Proteins can be released into the extracellular space through two mechanisms: the classical secretory pathway and the non-classical secretory pathways. In the classical secretory pathway, proteins targeted for extracellular release are synthesised as protein precursors which usually contain signal peptides located at the N-terminus. These signal peptides direct the proteins to the rough endoplasmic

reticulum (ER) and subsequently to the Golgi apparatus, following which the proteins would then be released into the extracellular environment in Golgi-derived secretory vesicles. Alternatively, proteins can be exported through ER/Golgi-independent mechanisms which are also known as the non-classical secretory pathways. In these non-classical secretory pathways, proteins may be exported by targeting endosomes recycling back to the plasma membrane, directly translocating across the plasma membrane, or through exosomal secretion (Nickel, 2003). Exosomes are intralumenal vesicles (ILVs) which are formed by inward budding from the limiting membrane of multivesicular bodies (MVBs). During the formation process, some cytosolic proteins may be incorporated into the invaginating membrane and become engulfed in the ILVs. As a result, when a MVB eventually fuses with the plasma membrane, the ILVs within are released extracellularly together with their cargo of cytosolic proteins (Simpson *et al.*, 2008).

The neoplasm is by no means a stand-alone entity. Tumour cells constantly interact with their extracellular environment to create favourable conditions for tumour progression, such as inducing angiogenesis or degrading the extracellular matrix to facilitate metastasis. These interactions are mediated by a variety of proteins secreted by the tumour cells, including growth factors, chemokines, cytokines, adhesion molecules, proteases and shed receptors. In the same way, surrounding stromal cells are also recruited by tumour cells to actively release proteins which further the progression of the neoplasia (Karagiannis, 2010). Thus, the cancer secretome can be described as constituting of proteins released from cancer-associated stromal cells, as well

as all proteins that are secreted by cancer cells through classical or non-classical secretory pathways, or shed from the cell surface (Makridakis and Vlahou, 2010). Since these secreted proteins are released into the extracellular environment, they have the highest likelihood of entering the blood circulation and thus present a highly promising source for serological biomarker discovery.

By definition, the cancer secretome would include all the secreted proteins that are found in the tissue interstitial fluids and other proximal biological fluids. However, it is now more commonly associated with the conditioned media (CM) samples collected from cancer cell lines (Karagiannis *et al.*, 2010), as many cancer secretome discovery studies would choose to analyse CM samples due to the challenges associated with using biological fluids. A couple of previous studies have successfully identified potential serological biomarkers from CM samples of colorectal cancer cell lines. Wu and colleagues (2008) analysed the CM from 21 different cell lines across 12 cancer types and found that secretion of CRMP2 (collapsin response mediator protein-2) was specific to colorectal cancer. They further showed that levels of plasma CRMP2 were significantly higher in colorectal cancer patients and thus proposed that CRMP2 could be a potential colorectal cancer-specific diagnostic marker. The Schwarte-Waldhoff group discovered that high levels of sE-cadherin (soluble E-cadherin) was secreted from SW620, HT29 and SW948 cells (Diehl *et al.*, 2007) and subsequently showed that serum sE-cadherin was significantly higher in patients with Stage III and IV carcinomas (Weiss *et al.*, 2011). In a similar study, Xue and colleagues (2010) compared

the CM from SW480 and SW620 cells and found that serum TFF3 (trefoil factor 3) and GDF15 (growth/differentiation factor 15) levels could be used for diagnostic discrimination for patients with colorectal cancer lymph node metastasis. These studies illustrate the potential of using colorectal cancer cell line CM for discovering potential serological biomarkers for colorectal cancer.

However, one challenge of using CM samples for biomarker discovery is that the concentration of the secreted proteins is usually very low, due to the high dilution from the culture media. As such, in order to obtain sufficient starting material for proteomics analyses, typical CM preparation procedures would involve the culture of the cells in multiple culture flasks, followed by collecting of the CM samples and concentrating the secreted proteins into a smaller volume. Precipitation and ultrafiltration are two classical concentration methods routinely used in secretome studies. Precipitation using trichloroacetic acid (TCA) is an effective procedure, but often results in poor yield, especially during resolubilisation. Ultrafiltration using molecular weight cut-off columns is a simpler method, but is time-consuming and loss of low-molecular weight proteins may occur (Mbeunkui *et al.*, 2006). Another challenge associated with using CM samples is that detection of secreted proteins may be confounded by the presence of intracellular proteins, which are released during cell lysis. This is exacerbated by the usage of serum-free media during CM sample preparation. However, this is almost unavoidable as serum proteins would interfere with detection of secreted proteins, because of their high abundances and sequence homologies with secreted proteins (Karagiannis *et al.*, 2010).

In a novel breakthrough to circumvent the above issues, the Liao group proposed the use of the commercially available hollow fibre culture (HFC) system to simultaneously collect and concentrate secreted proteins from cultured cells (Chiu *et al.*, 2013; Wu *et al.*, 2009). The HFC system consists of a hollow fibre cartridge coupled to an oxygenator coil, a culture media reservoir and a perfusion pump (Figure 5). The cells are cultured within the cartridge containing a total of 2800 hollow fibres, providing a surface area sufficient to support the adherence and growth of up to 10^9 cells. The culture media is supplied into the extra-capillary space (ECS) of the cartridge through the intralumen space in the hollow fibres and continuously recirculated throughout the system. The hollow fibres have a molecular weight cut-off of 5 kDa, which allows for exchange of nutrients and waste material from the ECS, but retains cells and secreted proteins larger than 5 kDa within the ECS. Since the volume of the ECS is only 15 ml, the HFC system enriches for secreted proteins in CM samples by facilitating high-density cell culture while retaining secreted proteins in a small media volume. Moreover, due to the large surface area for cell growth and dynamic removal of waste, cell lysis rates could be reduced significantly, thus lessening the extent of intracellular protein contamination in the CM samples. Although the HFC system appears to be a highly appealing method for preparation of CM, there are only a few studies which have utilised it thus far, such as in the secretome analyses of nasopharyngeal carcinoma (Chang *et al.*, 2009; Wu *et al.*, 2009), hepatocellular carcinoma (Wen *et al.*, 2011), non-small cell lung cancer (Chang *et al.*, 2012) and cholangiocarcinoma (Weeraphan *et al.*, 2012).

Therefore, given the practical usefulness of the HFC system, it could be worth applying the HFC system for collection of CM samples from colorectal cancer cell lines to identify secreted proteins that could be used as potential serological biomarkers.

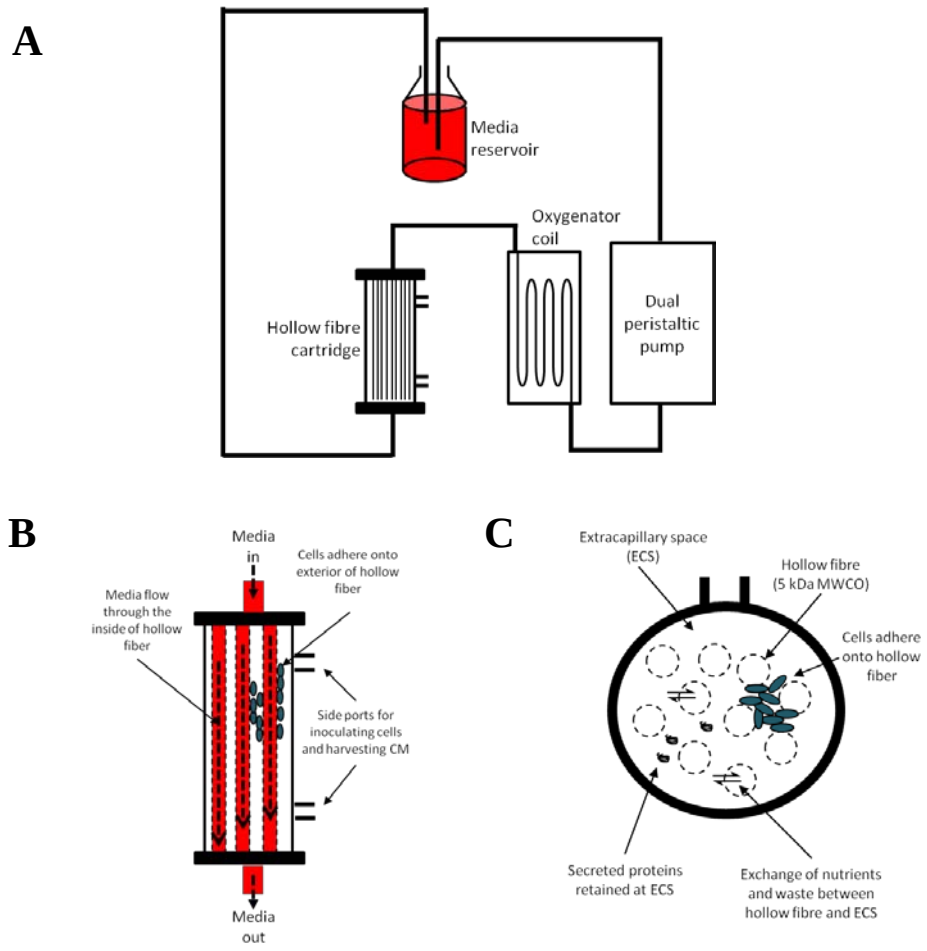


Figure 5. The hollow fibre culture (HFC) system
(A) Schematic diagram of the HFC system setup. Culture media is circulated from the media reservoir throughout the entire setup in a continuous loop by the action of the dual peristaltic pump. The dual peristaltic pump allows for the simultaneous culture of two hollow fibre cartridges. Before entering the hollow fibre cartridge, the culture media is equilibrated to the air composition and temperature of the incubator through the oxygenator coil. **(B)** Longitudinal section of the hollow fibre cartridge. The hollow fibres are sealed within the cartridge and the culture media would flow through the intralumen space of the hollow fibres. Cells would be inoculated into the system through one of the side ports and adhere onto the exterior of the hollow fibres. **(C)** Transverse section of the hollow fibre cartridge. Due to the 5 kDa MWCO of the hollow fibres, the culture media can freely exchange between the hollow fibres and the extracapillary space (ECS). However, secreted proteins larger than 5 kDa in size will be trapped in the ECS.

1.7.3. Delving deeper into the "glycosecretome" - enriching for differentially secreted glycoproteins between HCT-116 and E1

There is also an increasing interest in the enrichment of secreted proteins from the CM in order to enhance the detection of more low abundance tumour specific biomarkers. Various strategies for enrichment of secreted proteins have been introduced, including the use of resin-bound hexapeptide libraries (Colzani *et al.,* 2009) and metabolic labelling coupled with click chemistry (Eichelbaum *et al.,* 2012). Among these, one of the most commonly employed approaches in secretome studies is glycoprotein enrichment.

Glycosylation is one of the most prevalent and biologically important post-translational modifications in proteins. Protein glycosylation most commonly occurs on either asparagine residues (N-linked glycosylation), or serine or threonine resisdues (O-linked glycosylation). As the majority of serum proteins, as well as secreted proteins, are glycoyslated, targeting of the carbohydrate moieties could be used as a method to enrich for secreted glycoproteins, as well as eliminate intracellular contamination (Ahn *et al.*, 2010; Zhang *et al.*, 2012b). Many different techniques for glycoprotein enrichment have been developed, but the hydrazide chemistry and lectin capture methods are most widely employed (Pan *et al.*, 2011).

The use of hydrazide chemistry to enrich for glycoproteins was first described by Zhang and colleagues (2003) as a means to selectively capture N-linked glycoproteins in human serum. The carbohydrate moieties on glycoproteins are first oxidised to form aldehydes, which are subsequently covalently bound

to hydrazide groups immobilised on resins. In theory, this approach should be able to capture most of the glycoproteins present in any biological sample. However, the caveat lies in the limitation of methods available to release the bound glycoproteins or glycopeptides (Pan *et al.*, 2011). For N-linked glycoproteins, this is easily accomplished by using the peptide-N-glycosidase F (PNGase F) enzyme, which specifically cleaves between the oligosaccharide group and the asparagine residue. Unfortunately, for O-linked glycoproteins, there are no straightforward approaches for releasing the proteins from the resin. The oligosaccharide chain must be sequentially cleaved using a series of exoglycosidases until the core structure can be removed using either O-glycosidase or chemical methods, such as β-elimination (Wei and Li, 2009).

On the other hand, the lectin capture strategy, or lectin affinity chromatography, is more straightforward, flexible and cost-effective (Wei and Li, 2009). Lectins are proteins that have high affinity for carbohydrate moieties, and different lectins can selectively bind to different sets of oligosaccharide epitopes (some may have overlapping affinities). Lectin affinity enrichment experiments are usually designed with the desired lectin immobilised on solid supports such as agarose or silica in varying chromatographic formats, and bound glyoproteins are easily eluted using either low pH or competitive displacement with inhibitory sugars (Fanayan *et al.*, 2012; Wei and Li, 2009). However, a major drawback of lectin affinity chromatography is that no single lectin can comprehensively cover the entire glycoproteome (Fanayan *et al.*, 2012).

A simple solution to this problem was proposed by Yang and Hancock (2004). Instead of using a single lectin, the authors showed that a mixture of different lectins in a single column (termed multi-lectin affinity chromatography, MLAC) could provide a more complete enrichment of glycoproteins from human serum. This was achieved by combining the different affinities of three lectins: Concanavalin A (Con A), wheat germ agglutinin (WGA) and Jacalin (JAC). Con A has high affinity for α-mannose structures, WGA binds to *N*-acetylglucosamine (GlcNAc) and sialic acid residues, while JAC recognises *N*-acetylgalactosamine (GalNAc) moieties (West and Goldring, 2004). Since most N-glycans have a core structure of two GlcNAc residues followed by three mannosyl residues, these structures will be targeted by Con A and WGA (Figure 6). On the other hand, JAC has a high affinity for GalNAc, which is a common feature in O-glycan core structures. Additionally, many O-glycans also terminate with sialic acid (Pahlsson *et al.*, 1994). Therefore, the combination of these three lectins would, in theory, be able to capture a large majority of the glycoproteins present in a biological sample.

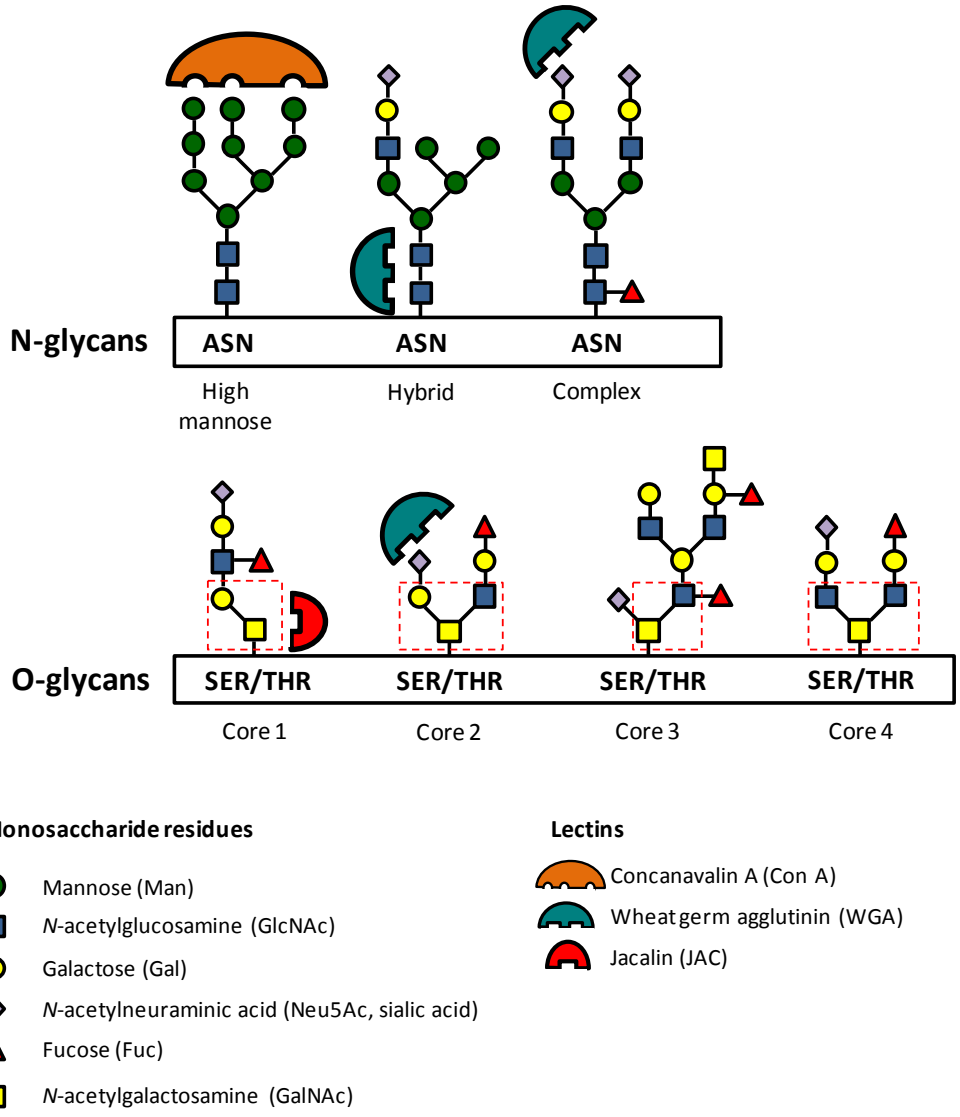


Figure 6. Overview of glycan structures targeted by MLAC
All N-glycans have a common core consisting of $Man_3GlcNAc_2$ attached to asparagine (in a consensus sequence of NX[ST]), and are classified according to three types: high mannose, where only mannose residues are attached to the core; complex, where a variety of diverse monosaccharides are attached to the core; and hybrid, where only mannose residues are present on one arm of the core, and other types of monosaccharides can be found on the other arm (Stanley *et al.*, 2009). O-glycans are linked through GalNAc residues to the –OH groups on serines and threonines. There are four main types of core structures in O-glycans which are more commonly observed, which are delineated by the red dotted boxes (Brockhausen *et al.*, 2009). Among the three different lectins used in MLAC, Con A targets high mannose type structures, WGA has high affinity for the GlcNAc residues in the core structures in N-glycans, while JAC binds to the GalNAc residues common in the cores of O-glycans. Moreover, WGA can also recognise sialic acid residues, which can be found in complex and hybrid N-glycans, as well as in the extended chains of O-glycans.

Moreover, when multiple lectins are mixed together, their binding affinities are enhanced considerably as compared to individual lectins. This is due to a phenomenon known as the glycoside cluster effect, where the clustering of both lectins and carbohydrates have been shown to lead to increased binding affinities (Lundquist and Toone, 2002). Because of this effect, it is possible to use MLAC to enrich for even mid- to low-abundance glycoproteins from biological samples (Taylor *et al.*, 2009).

There have been a number of studies which have reported the use of MLAC for glycoprotein enrichment for discovery of cancer biomarkers, and demonstrated that MLAC could be effectively used on various biological samples. For instance, MLAC can be used directly with serum samples to identify serum glycoproteins associated with breast (Yang *et al.*, 2006) and lung cancer (Heo *et al.*, 2007), or the serum samples could be immuno-depleted first before MLAC enrichment to enhance the identification of low abundance proteins (Zeng *et al.*, 2011). MLAC can also be applied on tissue samples, as shown by Na and colleagues (2009), who identified carboxylesterase 1 as a potential serological biomarker for hepatocellular carcinoma. Qi and colleagues (2014) similarly used MLAC enrichment on HCC tissues and found that pro-Cathepsin D was elevated in HCC patients. Applying MLAC on CM cultured from tumour tissue explants, Yao and colleagues (2012) discovered that EFEMP2 could be used as a biomarker for early detection for colorectal cancer. Although MLAC has been successfully utilised on these different kinds of samples, there are as yet no studies which have applied it on cell line CM. As such, it might be useful to explore the

applicability of MLAC on enriching for secreted glycoproteins from CM collected from colorectal cancer cell lines.

1.8. Aim of our study

1.8.1. Modelling colorectal cancer liver metastasis with the isogenic HCT-116 and E1 cell lines

Hepatic metastasis is one of the major factors that contributes to the high mortality rate in colorectal cancer. To improve patient outcome, there is an urgent need to identify novel proteins involved colorectal cancer metastasis, as well as to discover biomarkers with better sensitivities and specificities for prognosis and monitoring. Therefore, in our study, we utilised an isogenic pair of cell lines consisting of the poorly metastatic colon adenocarcinoma cell line HCT-116 and its liver metastatic derivative, E1, to model for colorectal cancer liver metastasis. The E1 cell line was generated by *in vivo* passaging of HCT-116 cells and generation of liver metastasis in athymic nude mice (Tay *et al.*, 2010). E1 cells were shown to display features characteristic of epithelial-mesenchymal transition, and possess increased motility and invasiveness as compared to the parental HCT-116 cells. Moreover, karyotype analyses on the E1 cells showed similar non-reciprocal translocations as those observed in the parental HCT-116 cell line, indicating that their genetic backgrounds were largely similar (Tay *et al.*, 2010). Therefore, we reasoned that any differences that are observed in a comparison between the protein expression profiles of HCT-116 and E1 cells might potentially contribute to metastasis.

1.8.2. A tripartite approach to address clinical needs in colorectal cancer management using proteomics

The aim of our study was to apply proteomics approaches in addressing the clinical needs in colorectal cancer metastasis management in three ways (summarised in Figure 7). First, to uncover novel proteins involved in colorectal cancer liver metastasis, we used the iTRAQ technology to compare between the HCT-116 and E1 intracellular proteomes, in order to identify proteins that were differentially expressed in E1. Second, to identify more sensitive and specific biomarkers for prognosis and monitoring, we analysed the secretomes of HCT-116 and E1 cells in search of metastasis-related secreted proteins which could be translated into serological biomarkers. Using the HFC system for preparation of CM samples from HCT-116 and E1 cells, we subsequently performed an iTRAQ-based comparison between the HCT-116 and E1 secretomes to identify differentially secreted proteins in E1 cells. We also explored here the applicability of the SWATH-MS technology as a MS-based tool for large-scale verification of the iTRAQ analysis of the HCT-116 and E1 secretomes. Finally, in order to fractionate the CM to enhance the detection of more low abundance secreted proteins, we enriched for secreted glycoproteins (glycosecretome) in the HCT-116 and E1 cell lines using the MLAC approach. After the MLAC enrichment, we then compared between the HCT-116 and E1 glycosecretomes using SWATH-MS as an unbiased label-free quantitative proteomics technology, in search for differentially secreted glycoproteins in the E1 CM samples.

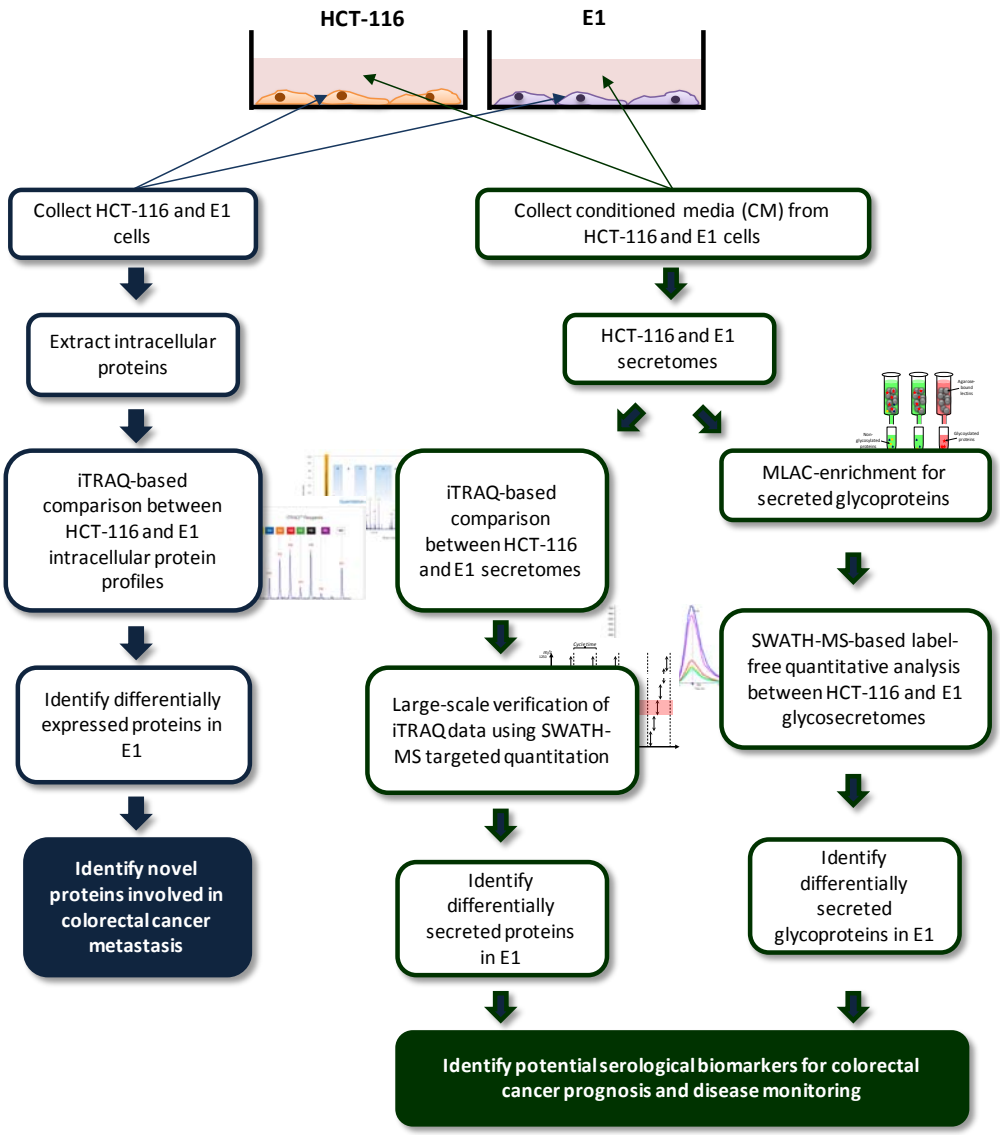


Figure 7. A tripartite proteomics approach to address clinical needs in colorectal cancer management

The aims of our study were threefold. First, intracellular proteins were extracted from HCT-116 and E1 cells and were compared using the iTRAQ technology to identify differentially expressed proteins in E1. Some of these could be potentially novel metastasis-related proteins. Second, secreted proteins from HCT-116 and E1 cells were collected from their respective CM samples and compared using the iTRAQ technology. Large-scale verification of the iTRAQ data was performed using the SWATH-MS technology as a targeted quantitative proteomics approach. Finally, we enriched for secreted glycoproteins from HCT-116 and E1 cells using the MLAC approach, in search for more low-abundance secreted proteins. An unbiased label-free SWATH-MS-based quantitative comparison between the HCT-116 and E1 glycosecretomes was performed. The second and third aims of our study were targeted towards the identification of potential serological biomarkers for colorectal cancer prognosis and disease monitoring.

2. Materials and Methods

2.1. Cell lines

The HCT-116 human colorectal adenocarcinoma cell line was obtained from the American Type Culture Collection (ATCC, Rockville, MD). The metastatic derivative E1 cell line was a kind gift from Professor SC Hooi (Department of Physiology, Yong Loo Lin School of Medicine, National University of Singapore). The two cell lines were cultured in McCoy's 5A media (Sigma-Aldrich, St. Louis, MO) supplemented with 10% fetal bovine serum (Life Technologies Gibco, Carlsbad, CA), in a humidified incubator (37^oC, 5% CO_2).

2.2. HFC system cell culture and collection of CM

HCT-116 and E1 cells were initially cultured in McCoy's 5A media supplemented with 10% FBS in a humidified incubator (37^oC, 5% CO_2). Cells were harvested using trypsin, and the cell number and viability was determined using the trypan blue exclusion test. Harvested cells were resuspended in McCoy's 5A media and subsequently inoculated into a HFC system cartridge (C2008, FiberCell Systems Inc, Frederick, MD). Before inoculation, the HFC cartridge was pre-cultured twice with sterile PBS, and then twice with McCoy's 5A media supplemented with 10% FBS. The FiberCell Systems Duet Pump (FiberCell Systems Inc) was used for simultaneous control of the circulation of two HFC cartridges in a humidified incubator (37^oC, 5% CO_2).

HCT-116 and E1 cells were inoculated into separate HFC cartridges. After allowing the cells to adhere, the reservoir media was substituted with Hyclone serum-free McCoy's 5A media without phenol red (Thermo Scientific HyClone, South Logan UT) supplemented with CDM-HD serum replacement (FiberCell Systems Inc). In addition, the extra capillary space (ECS) of the cartridges was flushed with 20 ml of HyClone serum replacement media multiple times until all traces of serum proteins were removed as determined using SDS-PAGE.

The cells were further incubated in the HyClone serum replacement media for 24 h before commencing daily collection of the CM samples. Aliquots of media were aspirated from the reservoir daily and glucose concentrations in the aliquots were measured using Accu-Check Advantage Meter Kit (Roche Diagnostics, Basel, Switzerland). The glucose consumption rates of the cells were used as an indicator of cell growth. The media reservoir was replaced with a fresh bottle once the glucose concentration was reduced by 50%.

Each of the CM sample collected from the ECS was centrifuged at 350 x *g* for 5 min at 15°C, and the supernatant was filtered using a 0.2 μm filter (Sartorius Stedim Biotech S.A., Aubagne Cedex, France). The filtrate was then concentrated using the Amicon Ultra-15 Centrifugal Filter Device (5 kDa MWCO; Millipore, Bedford/Billerica, MA). The concentrated CM samples were further subjected to buffer exchange on the same centrifugal filter device by washing thrice with 0.5 M triethylammonium bicarbonate (TEAB; Sigma Aldrich, St Louis, MO). Six CM samples per cell line were pooled. Whole cell

lysate samples were obtained from the respective cells cultured in the HFC cartridges by boiling in 0.5 M TEAB and 1% (w/v) SDS buffer at 100°C for 10 min.

2.3. Multi-lectin affinity chromatography (MLAC) enrichment

CM samples were prepared in a similar manner as described in the above section. The multi-lectin columns were prepared by mixing 1.0 ml each of agarose-bound ConA, WGA, and Jacalin (Vector Laboratories, Burlingame, CA) in empty PD-10 disposable columns (Amersham Biosciences, Piscataway, NJ). The columns were first equilibrated twice with 10 ml of loading buffer (20 mmol/L Tris, 0.15 mol/L NaCl, 1 mmol/L Mn^{2+}, and 1 mmol/L Ca^{2+}, pH 7.4). HCT-116 and E1 CM samples, each containing 1 mg of protein, were diluted with loading buffer to 2 ml before loading into the MLAC column. The columns were then incubated overnight at 4^{o}C with rocking to maximise glycoprotein binding. The unbound proteins were collected using 10 ml of loading buffer, followed by three washes using 10 ml of loading buffer each. Subsequently, the bound glycoproteins were released using 10 ml of elution buffer (20 mmol/L Tris, 0.5 mol/L NaCl, 0.17 mol/L methyl-α-*D*-mannopyranoside, 0.17mol/L *N*-acetylglucosamine and 0.27 mol/L galactose, pH 7.4). The eluted glycoprotein fractions were then concentrated using Amicon Ultra-15 Centrifugal Filter Devices (10 kDa MWCO).

2.4. iTRAQ labelling

2.4.1. iTRAQ labelling of whole cell lysate samples

Four biological replicates each of HCT-116 and E1 cells were harvested. The cells were lysed in a buffer consisting of 0.5 M TEAB and 1% SDS with boiling at 100^{o}C for 10 min. The resulting cell lysates were then centrifuged at 17,000 x *g* for 1 hr at 15^{o}C to remove cell debris. Protein quantitation was performed using the Coomassie Plus Protein Assay Reagent kit (Pierce Biotechnology) following the manufacturer's instructions. The samples were labelled with iTRAQ tags according to the manufacturer's recommendations (AB SCIEX, Foster City, CA). Briefly, 50 µg of protein was subjected to reduction using 5 mM tris-(2-carboxyethyl)-phosphine (TCEP) at 60^{o}C for 1 hr. This was followed by alkylation of cysteine residues with 10 mM methyl methane-thiosulfonate (MMTS) at room temperature for 10 min. Each of the samples was subsequently diluted to 0.05% SDS before digestion with trypsin (Trypsin with $CaCl_2$, AB SCIEX) at 37^{o}C for 16 hr. The resulting tryptic digests were labelled for 2 hr with the iTRAQ 8-plex reagents in the following order: Tag_{113} - Tag_{116} to four biological replicates of HCT-116, and Tag_{117} - Tag_{121} to four biological replicates of E1. These eight iTRAQ-derivatised samples were then pooled. Interfering substances were removed using a cation-exchange cartridge system (AB SCIEX) according to the manufacturer's instructions. The ion-exchange eluate was desalted using a Sep-Pak C18 cartridge (Waters, Milford, MA), and subsequently lyophilised and reconstituted in 5 mM KH_2PO_4 in 5% acetonitrile (pH 3) for two-dimensional liquid chromatography (2-D LC) separation.

2.4.2. iTRAQ labelling of whole secretome samples

iTRAQ labelling was performed similar to the previous section. The tryptic digests were labelled with the iTRAQ 8-plex reagents in the following order: Tag_{113} and Tag_{117}: biological replicates of HCT-116 CM samples; Tag_{114} and Tag_{118}: biological replicates of E1 CM samples; Tag_{115} and Tag_{119}: biological replicates of HCT-116 lysate samples; and, Tag_{116} and Tag_{121}: biological replicates of E1 lysate samples. The lyophilized iTRAQ-labelled peptide mixture was reconstituted in 5% acetonitrile (ACN) with 0.05% formic acid (FA) for 2-D LC.

2.5. LC-MS analysis

2.5.1. 2-D LC - MALDI-TOF/TOF analysis for cell lysate samples

The iTRAQ-labelled peptide mixture was separated by 2-D LC consisting of strong cation exchange (SCX) followed by reversed phase (RP) separation. The 2-D LC separation was performed using the Ultimate 3000 LC system (Thermo Scientific Dionex, Sunnyvale, CA) coupled to a Probot MALDI spotting device (Thermo Scientific Dionex) using parameters as described previously (Loei *et al.*, 2012). Briefly, the reconstituted peptides were separated using a SCX 300Å NanoEase Trap column (Waters). A total of nine SCX fractions were collected and further separated using a RP column (Symmetry C18 300 µm x 150 mm NanoEase Column; Waters). Eluted RP fractions were then mixed with MALDI matrix (7 mg/ml α-cyano-4-

hydroxycinnamic acid with 130 µg/ml ammonium citrate in 75% ACN) and spotted onto 1232-well stainless steel MALDI target plates (AB SCIEX).

MS and MS/MS analyses were performed using a 4800 MALDI TOF/TOF Analyser (AB SCIEX). The analyses were carried out in positive ion mode and laser output was set to 4500 for MS acquisition and 4800 for MS/MS. 1000 laser shots were accumulated from each sample well and precursor ions were selected from MS spectra ranging from 920 to 3900 Da with a minimum S/N ratio of 40. MS/MS analyses were performed for the seven most abundant precursor ions per well. Air was used for CID, with collision energy set at 1 kV and collision gas pressure of $1x10^{-6}$ Torr. A total of 5000 shots were accumulated for each MS/MS spectrum.

All the MS/MS spectra generated were processed using the ProteinPilot Software 2.0.1 (AB SCIEX) for relative quantification and protein identification. The processing method was set to the Paragon algorithm and all the spectra were searched against the International Protein Index (IPI) Human database (version 3.66, 86 839 total protein sequences). Search parameters were adjusted to indicate tryptic digestion and allow for cysteine alkylation by MMTS and biological modifications. The detected protein threshold score was set at 1.3 and the auto bias correction was applied for global normalisation. The spectra were also searched against a decoy database to determine the false discovery rate (FDR). This decoy database consisted of a randomised version of the IPI Human v3.66 database. The threshold for determining fold change was set at 1.3 based on a previous study (Tan *et al.*, 2008). This fold change

threshold was derived from the standard deviation (S.D.) calculated from equal amounts of six-protein mixtures labelled with two iTRAQ reagents and analysed by LC-MS/MS. The S.D. of all the iTRAQ ratios from the identified peptides was 0.15 and therefore, 1.3 (1 + 2 S.D.) was determined to be the significant cut-off threshold. Furthermore, to ensure that the differential expressions observed were robust across biological variations, we applied an additional criterion: in order for any protein to be considered as differentially expressed, the difference must be observed in all the comparisons between all the HCT-116 and E1 biological replicates. Finally, the differentially expressed proteins were classified and annotated according to Gene Ontology Biological Processes using the AmiGO browser (Carbon *et al.*, 2009).

2.5.2. 2-D LC - ESI-qTOF analysis of whole secretome samples

The first dimension peptide separation was performed using the Ultimate LC system (Dionex-LC-Packings, Sunnyvale, CA) connected to a Zorbax Bio-SCX II column (Agilent, SantaClara, CA). The SCX solvent A was 5% ACN (with 0.05% FA) and solvent B was 5% ACN, 500 mM NaCl (with 0.05% FA). A total of 105 fractions were eluted (20µl each) over a gradient of 0 - 20% solvent B over 90 min at a flow rate of 10 µl/min. The eluted fractions were subsequently combined to 18 fractions (approximately 4µg of proteins per fraction) and then desalted with the Sep-Pak tC18 µElution Plate (Waters) using a vacuum manifold (Millipore). The desalted fractions were then subject to a second-dimension reversed-phase (RP) chromatography on the NanoLC-

Ultra system (Eksigent, Dublin, CA) coupled with the cHiPLC-Nanoflex system (Eksigent).

For each SCX fraction, 2 µg of peptides were trapped on a precolumn (200 µm x 0.5 mm) and then eluted on an analytical column (75µm x 150 mm) for separation. Both columns were packed with Reprosil-Pur C18-AQ 3 µm 120Å beads (Eksigent, Dublin, CA). The RP solvent A was 2% ACN (with 0.1% FA) while solvent B was 98% ACN (with 0.1% FA). The peptides were eluted from the analytical column in a linear gradient of 12 - 30% solvent B over 90 min at a flow rate of 300 nL/min and then directly injected into the mass spectrometer for analysis.

The MS analysis was performed using a 5600 TripleTOF analyzer (QqTOF; AB SCIEX) in the Information Dependent Acquisition mode, or DDA mode. Precursor ions were selected across the mass range of 350-1250 *m/z* using 250 ms accumulation time per spectrum. A maximum of 20 precursors per cycle from each MS spectra were selected for MS/MS analyses with 100 ms minimum accumulation time for each precursor across a mass range of 100-1800 *m/z* and dynamic exclusion for 15s. MS/MS analysis was recorded in high sensitivity mode with rolling collision energy on and iTRAQ reagent collision energy adjustment on.

Protein identification and relative iTRAQ quantification were performed with the ProteinPilot Software 4.5 (AB SCIEX) using the Paragon algorithm for peptide identification and the ProGroup algorithm for grouping similar protein

IDs to reduce redundancy. The MS/MS spectra were searched against the International Protein Index (IPI) human database (version 3.87, 91 444 entries) with search parameters adjusted for cysteine alkylation by MMTS and biological modifications specified within the algorithm. The detected protein threshold was set at 0.05. The MS/MS spectra were also searched against a decoy database to estimate the false discovery rate (FDR) for peptide identification. The decoy database consisted of reversed protein sequences from the IPI human v3.87 database.

We adopted a similar approach that we have previously published (Loei *et al.*, 2012) to distinguish between genuine secreted proteins from the intracellular proteins released from cell lysis. The relative abundance of all proteins in the CM of each cell line was compared with that of its respective whole cell lysate, and a cut-off ratio of > 1.5 was used to identify truly secreted proteins. In addition, these proteins should fulfil the cut-off requirement in both biological replicates. The proteins which fulfilled the above-mentioned filtering criteria were analysed using SecretomeP 2.0 (Bendtsen *et al.*, 2004) to predict the possibility of classical and non-classical secretion. A query against the ExoCarta exosomal protein database (Mathivanan *et al.*, 2012) was also employed to identify potential proteins released in exosomes. A 1.3 fold change cut-off was selected to determine differentially regulated proteins. Thus, proteins with HCT-116 CM / E1 CM ratios of > 1.3 would be considered as oversecreted, while those with ratios < 0.77 would be undersecreted. Gene Ontology Biological Process annotation of the differentially secreted proteins was performed using the AmiGO browser.

2.6. SWATH-MS sample preparation and analysis

CM samples from HCT-116 and E1 cells (30 µg total protein each) were first reduced with 5 mM TCEP and alkylated with 10 mM MMTS. Subsequently, the CM proteins were digested using trypsin (Promega, Madison, WI) at a trypsin-to-protein ratio of 1:10 (w/w) at 37°C for 16 h. The peptides were then desalted using the Sep-Pak tC18 µElution Plate (WATERS) and a vacuum manifold (Millipore), lyophilised and finally reconstituted in 5% ACN (with 0.1% FA) for MS analysis. Equal proportions from each of the processed conditioned media samples were combined to create a pooled sample, which were used for the generation of the reference spectral ion library as part of the SWATH-MS analysis.

The MS analyses were performed on a TripleTOF 5600 Analyzer (QqTOF; AB SCIEX). The pooled sample was analysed in the Information Dependent Mode (Data Dependent Acquisition, DDA) as described in section 2.5.2, except that the iTRAQ adjustment for the rolling collision energy was removed. The individual conditioned media samples were analysed in the SWATH-MS Mode (Data Independent Acquisition, DIA), which specified for a quadrupole resolution of 25 Da/mass selection (Gillet *et al.*, 2012). The precursor mass range was scanned across 350-1250 *m/z*, generating a set of 36 overlapping windows. The collision energy for each window was set at 35 eV with a spread of 15 eV. Fragment ion spectra for each window were recorded with an accumulation time of 80 ms across the range of 100-1800 *m/z* each at

high sensitivity mode, resulting in a total cycle time of 2.98 s. Three technical replicates of the SWATH-MS acquisition were performed for each sample.

To create a spectral library of all the detectable peptides in the samples, the MS/MS spectra generated from the pooled sample in DDA were searched using the ProteinPilot 4.5 software. Spectra identification was performed by searching against either the IPI (version 3.87) or the SwissProt database with search parameters specified for *Homo sapiens* sequences (20 254 entries, released on March 06, 2013), cysteine alkylation by MMTS and biological modifications specified within the algorithm. The detected protein threshold was set at 0.05 and a false discovery rate (FDR) analysis was also performed by searching against a decoy database consisting of reversed protein sequences from either of the IPI or SwissProt databases.

2.6.1. SWATH-MS analysis for large-scale verification of iTRAQ data from HCT-116 and E1 secretome samples

SWATH-MS data analysis was performed using the MultiQuant workflow for SRM-like quantitation. Up to three peptides with three transitions each from each protein were manually selected using the SWATH Acquisition MicroApp 1.0.0.653 in the PeakView 1.2.03 software (AB SCIEX) with the following parameters: 25 ppm ion library tolerance, 4 min XIC extraction window, 0.01 Da XIC width, and considering only peptides with at least 99% confidence and excluding those which were shared or contained modifications. Peak area extraction was performed using the MultiQuant 2.1 software (AB SCIEX). All

transition peaks were detected using a retention-time (RT) window of 45 s. A 1-point Gaussian smooth was applied to all transition peaks, and peak areas were integrated using the following parameters: noise percentage at 40%, baseline subtraction window of 2 min and peak splitting at 2 points. Global normalisation was performed by adjusting the peak areas for each transition against the total area sums of all transitions from all proteins detectable in the SWATH-MS analysis. The fold change for each protein was calculated from the quotient of the normalised total area sum of all transition peaks of the protein in E1 and that in HCT-116.

2.6.2. SWATH-MS analysis for label-free quantitative comparison between MLAC-enriched HCT-116 and E1 CM samples

SWATH-MS quantitation was performed using the MarkerView workflow. Up to five peptides with three transitions per protein were automatically highlighted by the SWATH Acquisition MicroApp 1.0.0.653 in the PeakView 1.2.03 software with the same parameters as described in the previous section. However, to ensure reliable quantitation, only proteins which had 3 or more peptides available for quantitation were selected for XIC peak area extraction and exported for analysis in the MarkerView 1.2.1 software (AB SCEIX). Global normalisation was performed according to the Total Area Sums of all detected proteins in the samples. The Student's T-test was used to perform two-sample comparisons between the averaged area sums of all the transitions derived for each protein across the three E1 and HCT-116 replicate runs, in order to identify proteins that were significantly differentially secreted in E1.

2.7. Western blot

2.7.1. Validation of differentially expressed proteins in E1 cells

Equal aliquots of whole cell lysates from HCT-116 and E1 were resolved by 1-D SDS-PAGE. The gels were subsequently equilibrated in Towbin transfer buffer (25 mM Tris, 192 mM glycine, 0.1% SDS, 20% methanol) before electroblotting onto PVDF membranes. The blots were then blocked overnight with 5% non-fat dry milk in TBS-T (20 mM Tris-HCl, 150 mM NaCl, 0.1% Tween-20, pH 7.5). The mouse anti-TCTP (Translationally controlled tumour protein; sc-133131) antibody was purchased from Santa Cruz Biotechnology, Inc. Mouse anti-AKAP12 (A-kinase anchor protein; H00009590-M01) and anti-DBN1 (Drebrin; ab12350) antibodies were purchased from Abnova and Abcam respectively. HRP-conjugated anti-mouse IgG (NXA931, GE-Healthcare) was used as the secondary antibody. The immunoblotting conditions for each antibody are shown in Supplementary Table 1. Actin (A2066, Sigma-Aldrich) was used as the loading control. Visualisation was performed using the Enhanced Chemiluminescence system (ECL, GE-Healthcare) or Super Signal West Dura (Pierce).

2.7.2. Validation of differentially secreted proteins in the whole E1 secretome

The following proteins were selected for validation by western blot: GDF15 (growth/differentiation factor 15), PAI1 (plasminogen activator inhibitor 1; SERPIN E1), SPARC (secreted protein, acidic, cysteine rich; osteonectin), MAN1A1 (mannosyl-oligosaccharide 1,2-alphamannosidase IA), and PLOD3

(procollagen-lysine, 2-oxoglutarate 5-dioxygenase 3). Equal aliquots of CM and whole cell lysate samples from both HCT-116 and E1 cells were resolved by SDS-PAGE and blotted as described in the previous section. Rabbit anti-SPARC (15274-1-AP) was purchased from Proteintech Group (Chicago, IL). Rabbit anti-GDF15 (ab14586) and mouse anti-PLOD3 (ab89263) antibodies were purchased from Abcam (Cambridge, UK). Mouse anti-PAI1 (sc-5297) and rabbit anti-MAN1A1 (M3694) were purchased from Santa Cruz Biotechnology (Santa Cruz, CA) and Sigma-Aldrich respectively. The immunoblotting conditions for the antibodies used are detailed in Supplementary Table 1. The secondary antibodies used were HRP-conjugated anti-mouse IgG (NXA931, GE-Healthcare) and HRP-conjugated anti-rabbit IgG (1858415, Pierce Biotechnology). Visualisation was performed using either the Enhanced Chemiluminescence (ECL) substrate (GE-Healthcare) or Super Signal West Dura (Pierce Biotechnology).

2.7.3. Validation of differentially secreted proteins in E1 from the MLAC-enriched glyco-secretomes

Equal aliquots of CM and whole cell lysate samples from HCT-116 and E1 cells were resolved by SDS-PAGE and blotted as described in the previous section. Rabbit anti-COL6A2 (14853-1-AP) was purchased from Proteintech group (Chicago), while mouse anti-LAMB1 (sc-17810) was purchased from Santa Cruz Biotechnology (Santa Cruz). The immunoblotting conditions for the antibodies used are detailed in Supplementary Table 1. The secondary antibodies used were HRP-conjugated anti-mouse IgG (NXA931, GE-

Healthcare) and HRP-conjugated anti-rabbit IgG (1858415, Pierce Biotechnology). Visualisation was performed using Super Signal West Dura (Pierce Biotechnology) together with the ChemiDoc MP System (Bio-Rad, Hercules, CA) and the ImageLab v4.1 software (Bio-Rad).

2.8. Immunohistochemical analysis in clinical patient tissues

Seven colorectal cancer patient adenocarcinoma tissue sections with matched lymph node and liver metastases were obtained from the Department of Pathology, Singapore General Hospital, with the appropriate ethics approval. The clinical information of the seven patient tissue samples is detailed in Supplementary Table 2. The tissue slides were stained using the REAL EnVision Detection Systems Peroxidase/DAB (Dako, Denmark) according to the manufacturer's recommendations with some modifications. Antigen retrieval was performed using Tris-EDTA (pH 9) at 100^{o}C for 5 min. Endogenous peroxidase activity was inhibited by treating the slides with REAL Peroxidase-Blocking solution (Dako) for 10 min. The slides were then incubated with the anti-DBN1 antibody (ab12350, Abcam) diluted in REAL Antibody diluent (Dako) at 1:200 for 2 hrs. Secondary antibody (Dako REAL EnVision/HRP, Rabbit/Mouse) incubation was performed for 30 min, followed by the DAB chromogen buffer (Dako REAL DAB+ Chromogen) for 15 min. The slides were then counterstained with haematoxylin. Serial staining of the colon adenocarcinoma tissue sections against neurofilaments (NF) was performed on the Leica Bond-III automated IHC stainer (Leica Microsystems,

Germany) using the Bond Polymer Refine Detection method with an anti-neurofilament antibody (M0762, Dako).

Scoring was performed by Dr. KH Lim (Department of Pathology, Singapore General Hospital). The diagnoses for the patient samples were histologically confirmed in the stained tissue sections and the scorer was blind to the clinicopathological information. The immunostaining was assessed as follows: staining intensity (I) was graded according to negative (0), mild staining (1), moderate staining (2) and strong staining (3). Percentage of cells stained (P) were classified as: <10% (0), 10–29% (1), 30–49% (2), 50–79% (3) and >80% (4). The overall staining index for each slide was therefore evaluated as I x P. Statistical analyses were performed using the OriginPro 8.5 software (OriginLab, Northampton, MA). The paired Wilcoxon signed rank test was used to compare the staining indices of the primary tumour tissues to their matched lymph node and liver metastases.

2.9. ELISA analysis on clinical patient serum samples

A total of 19 serum samples from colorectal cancer patients, as well as 47 serum samples from healthy individuals free of any cancer, were obtained from the NUH-NUS (National University Hospital - National University of Singapore) Tissue Repository. An additional 26 serum samples from colorectal cancer patients were further obtained from the SingHealth Tissue Repository. The clinical information of the 45 colorectal cancer patient samples that could be accessed is detailed in Supplementary Table 3. This

study was conducted with approval from the relevant Institutional Review Boards.

The concentrations of LAMB1 in the serum samples were measured using commercial ELISA kits from USCN Life Sciences Inc. (SEA184Hu), according to the manufacturer's protocol. All plate washing steps were carried out using the ELx50 Microplate strip washer (BioTek, Winooski, VT). The absorbance readings were measured using the Tecan Infinite M200 spectrophotometer (Tecan, Männedorf, Switzerland). The ELISA data was analyzed with the statistical software OriginPro 8.6 (OriginLabs) and XLSTAT software v2.01 (Addinsoft SARL, Paris, France).

3. Results

3.1. Part One. iTRAQ-based comparison between HCT-116 and E1 intracellular proteomes

3.1.1. Identification of differentially expressed intracellular proteins in E1 cells as compared to HCT-116

In order to identify differentially expressed proteins in E1 as compared to HCT-116 cells, we compared the intracellular proteomes of HCT-116 and E1 cells using the iTRAQ technology. We identified a total of 547 proteins in both cell lines with at least 95% confidence, after considering only those which were identified with two peptides and above, as well as discarding any hits that were false positives or without iTRAQ ratios.

Due to problems in protein quantitation of the HCT-116 and E1 replicates that were labelled with the iTRAQ Tag_{113} and Tag_{121} respectively, these were excluded from downstream analyses. Consequently, only iTRAQ ratios from three replicates each of HCT-116 and E1 were considered for determining differences in expression levels of the proteins. Two criteria were applied to determine the list of differentially expressed proteins in E1: (i) only proteins with iTRAQ ratios of above 1.3 or below 0.77 would be considered as differentially expressed; (ii) the protein must also be differentially expressed in all three E1 biological replicates across all nine comparisons. With these two criteria, a total of 31 proteins were determined to be differentially expressed in E1 (Table 3). 20 proteins were found to be overexpressed in E1, while the list of underexpressed proteins consisted of 11 members. Gene Ontology Biological Processes classification of these 31 proteins indicated that they are involved in metastasis-related processes such as gene expression, signal transduction, cell proliferation and cytoskeletal organisation (Table 3).

3.1.2. Western blot validation of iTRAQ data

In order to confirm the altered expression of differentially expressed proteins in E1, a selection of three proteins was chosen for validation. The differential expressions of two overexpressed proteins, DBN1 and TCTP, and one underexpressed protein, AKAP12, were confirmed with western blot using commercially available antibodies (see Figure 8A). Parallel immunoblotting of Actin was used as loading control. The visualised antibody signals were quantified by scanning the respective blots on a densitometer (summarised in Figure 8B). The expression of TCTP was significantly increased in all three E1 replicates (p = 0.0125). Similarly, DBN1 was also significantly overexpressed in E1 (p = 0.0094). Conversely, the expression levels of AKAP12 in E1 were significantly reduced as compared to HCT-116 (p = 0.0095).

Table 3. Differentially expressed proteins in E1 cell line identified from iTRAQ analysis
Table reproduced from Lin *et al.*, 2014, with permission from Proteomics, John Wiley and Sons.

IPI Accession[a]	Uniprot Accession[b]	Name	Unused[c]	Total[d]	% Cov[e]	Peptides (95%)[f]	Average E1 / HCT-116 ratio[g]	s.d.[h]	Gene Ontology Biological Process[i]
IPI00464952.2	Q05519	SFRS11 Splicing factor, arginine/serine-rich 11	4.82	4.82	24.7	2	3.13	1.60	
IPI00856098.1	Q9P2E9	RRBP1 p180/ribosome receptor	7.72	7.75	30.5	3	2.40	0.72	
IPI00328715.4	Q86UE4	MTDH Protein LYRIC	10.18	10.19	20.1	5	1.58	0.20	Gene expression, transcription and translation
IPI00025329.1	P84098	RPL19 60S ribosomal protein L19	8.76	9.68	38.8	5	0.64	0.05	
IPI00395998.5	P62910	RPL32 60S Ribosomal protein L32	3.69	3.69	45.8	2	0.50	0.13	
IPI00221089.5	P62277	RPS13 40S ribosomal protein S13	8	8	42.4	4	0.47	0.09	
IPI00021831.1	P10644	PRKAR1A cAMP-dependent protein kinase type I-alpha regulatory subunit	6.26	6.26	23.4	3	1.85	0.30	
IPI00021405.3	P02545	LMNA Lamin-A/C	40.48	40.54	55	25	1.78	0.24	
IPI00025252.1	P30101	PDIA3 Protein disulfide-isomerase A3	37.56	39.97	53.5	24	1.68	0.23	Signal transduction, cellular response and regulation of apoptosis
IPI00550900.1	P13693	TPT1 Translationally-controlled tumor protein	9.2	9.2	40.1	4	1.61	0.24	
IPI00329801.12	P08758	ANXA5 Annexin A5	10.02	10.02	36.6	5	1.60	0.14	
IPI00218733.6	P00441	SOD1 Superoxide dismutase [Cu-Zn]	11.56	11.56	89.6	7	1.58	0.18	
IPI00867627.1	Q02952	AKAP12 A-kinase anchor protein 12	46.24	46.24	25.2	24	0.65	0.05	
IPI00479997.4	P16949	STMN1 Stathmin	21.53	21.53	73.8	10	2.04	0.26	
IPI00013808.1	O43707	ACTN4 Alpha-actinin-4	42.99	42.99	43.7	23	1.54	0.12	Cytoskeletal organisation
IPI00295624.5	Q16643	DBN1 Drebrin	13.29	13.29	30.6	6	1.45	0.11	
IPI00215611.5	P50238	CRIP1 Cysteine-rich protein 1	4.09	4.09	55.8	2	3.12	1.37	
IPI00217465.5	P16403	HIST1H1C Histone H1.2	37.17	37.2	74.6	22	1.78	0.27	
IPI00419585.9	P62937	PPIA Peptidyl-prolyl cis-trans isomerase A	19	19	72.7	11	1.61	0.18	DNA replication, cell proliferation and cell cycle regulation
IPI00412771.1	Q9Y5K6	CD2AP CD2-associated protein	6.45	6.45	10.5	3	1.54	0.13	
IPI00472442.4	P25786	PSMA1 Proteasome subunit alpha type-1	5.41	5.41	34.2	3	0.65	0.05	
IPI00004233.2	P46013	MKI67 Antigen KI-67	8.18	8.19	12.7	4	0.59	0.10	

Table 3. Continued

IPI Accession[a]	Uniprot Accession[b]	Name	Unused[c]	Total[d]	% Cov[e]	Peptides (95%)[f]	Average E1 / HCT-116 ratio[g]	s.d.[h]	Gene Ontology Biological Process[i]
IPI00299149.1	P61956	SUMO2 Small ubiquitin-related modifier 2	3.23	3.23	62.1	4	1.56	0.10	Protein modification and protein folding
IPI00939174.1	Q96FW1	OTUB1 Ubiquitin thioesterase protein OTUB1	4.02	4.02	16.9	2	0.60	0.10	
IPI00784090.2	P50990	CCT8 T-complex protein 1 subunit theta	8.02	8.02	14.1	4	0.60	0.09	
IPI00893358.1	P50991	CCT4 T-complex protein 1, delta subunit	5.47	5.47	12.5	2	0.50	0.05	
IPI00796366.2	P60660	MYL6 Myosin light polypeptide 6	9.81	9.81	42.9	8	1.82	0.16	Cell differentiation and cell development
IPI00220327.4	P04264	KRT1 Keratin, type II cytoskeletal 1	2	10.1	20.2	6	0.30	0.18	
IPI00169383.3	P00558	PGK1 Phosphoglycerate kinase 1	27.92	27.92	43.9	15	1.43	0.10	Metabolic processes
IPI00442073.5	P21291	CSRP1 Cysteine and glycine-rich protein 1	4	4	25.9	2	1.78	0.31	Unknown
IPI00020075.4	Q9NUJ1	ABHD10 Abhydrolase domain-containing protein 10, mitochondrial	4	4	19.6	2	0.48	0.05	

[a] Accession number of the identified proteins in the IPI Human v3.66 database

[b] Corresponding UniProt Accession of the identified proteins

[c] ProteinPilot Unused Score. Total score of the unique peptides detected for a protein. Used as a measure of protein confidence for the identified protein

[d] ProteinPilot Total Score. Total score of all the peptides detected for a protein

[e] Percentage sequence coverage of identified peptides

[f] Number of distinct peptides identified with at least 95% confidence

[g] Average fold change ratio of each protein in E1 against HCT-116 cells, calculated by the average of nine comparisons between three HCT-116 and E1 replicates

[h] Standard deviation for the average fold change ratio

[i] Gene Ontology Biological Process annotation from AMIGO browser

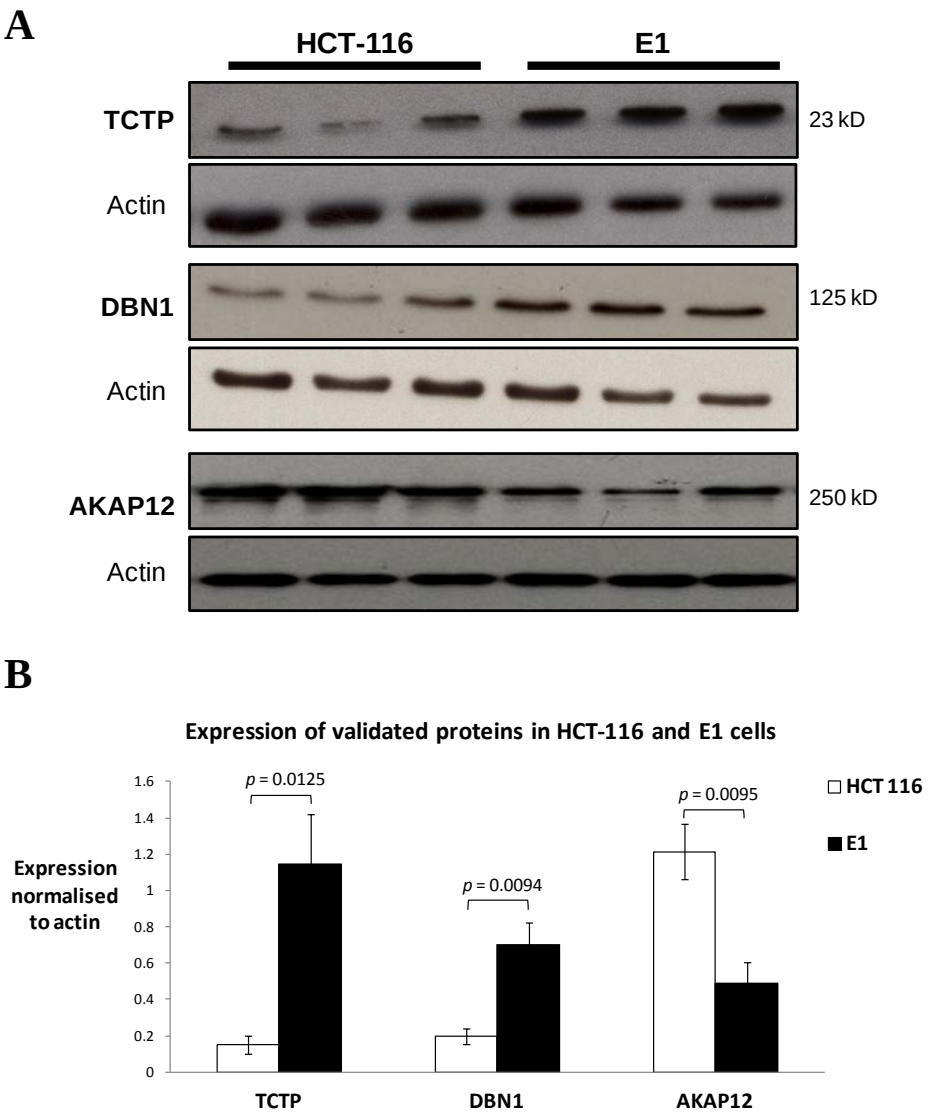


Figure 8. Western blot validation of selected differentially expressed proteins in E1
(A) Differential expression of TCTP, DBN1 and AKAP12 suggested from the iTRAQ data, validated using western blot in three biological replicates each from HCT-116 and E1 cells. The immunoreactive bands are shown here together with their respective actin loading controls. **(B)** Quantification of the immunoreactive bands detected from TCTP, DBN1 and AKAP12 using a densitometer. The averaged readings from the three biological replicates normalised against actin are shown with the error bar denoting standard deviation. Statistical analyses were performed using the Student's T-test. Figure reproduced from Lin *et al.*, 2014, with permission from Proteomics, John Wiley and Sons.

3.1.3. Immunohistochemistry analyses of DBN1

The relevance of DBN1 overexpression in clinical colorectal cancer metastasis was assessed using immunohistochemical (IHC) analyses of colorectal cancer patient tissues (Figure 9). DBN1 expression was observed to be negative in most of the primary tumour sections (Figure 9A), with the exception of one case which was stained weakly positive (Figure 9B). Since DBN1 is a neuronal protein, we compared the localisation of DBN1 in the primary tumour sections against serial neurofilaments (NF) immunostain, and observed that DBN1 localisation indeed corresponded with that of NF (Figure 10). Therefore, we used the presence of DBN1 expression in the neuronal cells in the primary tumour sections as an internal positive control. In contrast, in the matched lymph node and liver metastasis sections, DBN1 stained positive for all seven cases (Figures 9C-D). Localisation of DBN1 was observed to be mainly cytoplasmic in both lymph node and liver metastases, with indications of juxtamembrane enrichment observable at higher magnifications (see Figures 9E-F). The staining indices for the sections are summarised in Figure 9G. Paired Wilcoxon signed-rank tests confirmed that DBN1 was indeed significantly overexpressed in the matched lymph node ($p = 0.0078$, n = 7) and liver metastases ($p = 0.0078$, n = 7).

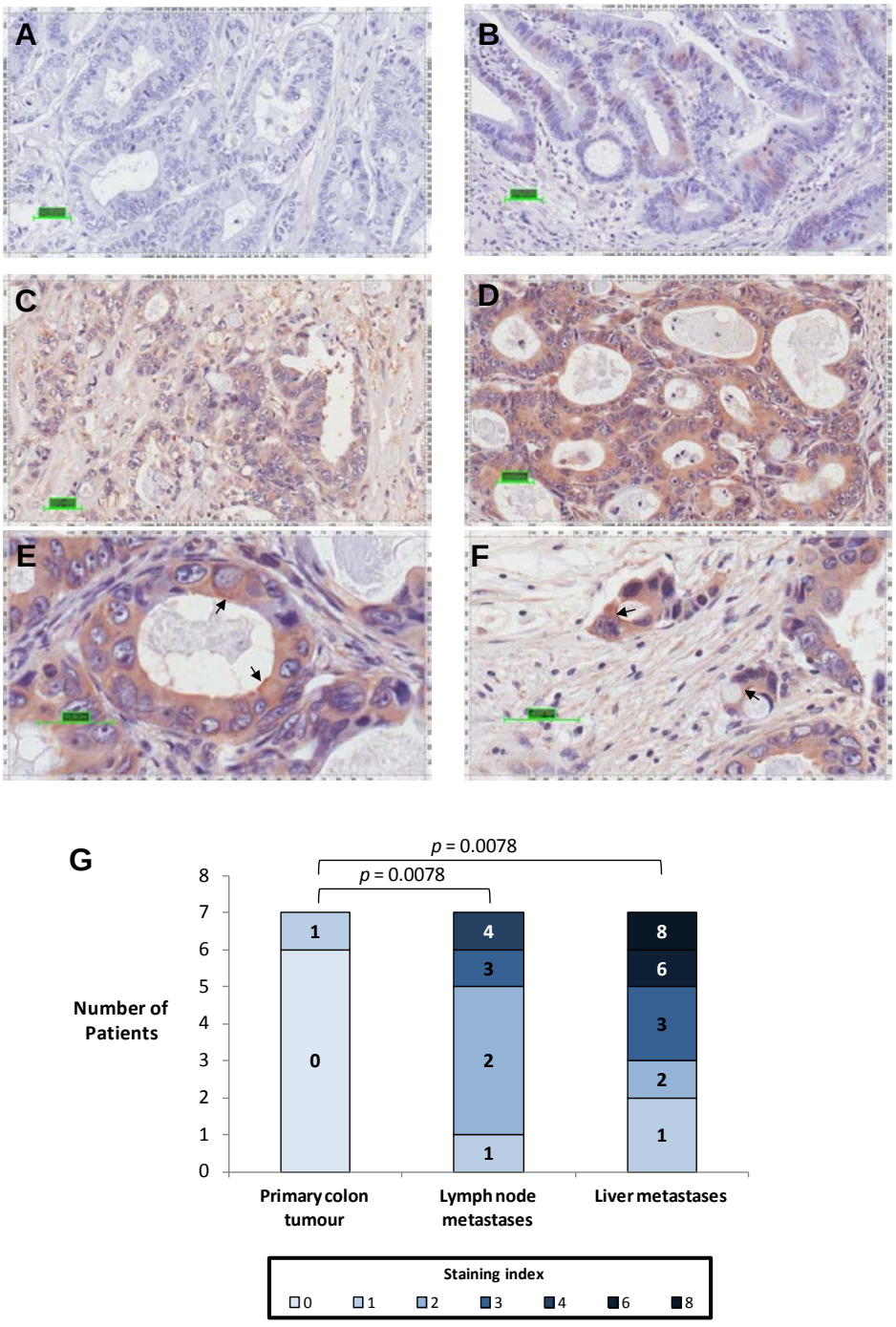


Figure 9. Immunohistochemical analyses of DBN1 expression in clinical colorectal cancer patient samples

DBN1 immunostaining in primary tumour tissue sections **(A-B)**, with matched lymph node **(C)** and liver metastases **(D)** sections (20x magnification). **(E-F)** Higher magnification (40x) reveal indications of DBN1 juxtamembrane enrichment (indicated by arrows). Scale bars indicate 50.18 µm. **(G)** Distribution of the number of patient tissue sections in the primary adenocarcinoma sections with matched lymph node and liver metastases, plotted against staining indices scored for DBN1 expression. The colours of the bars denote the staining indices scored with darker shades of blue indicating higher staining index, while the number within each bar indicate its respective staining index. Statistical analyses were performed using the paired Wilcoxon signed-rank tests. Figure reproduced from Lin *et al.*, 2014, with permission from Proteomics, John Wiley and Sons.

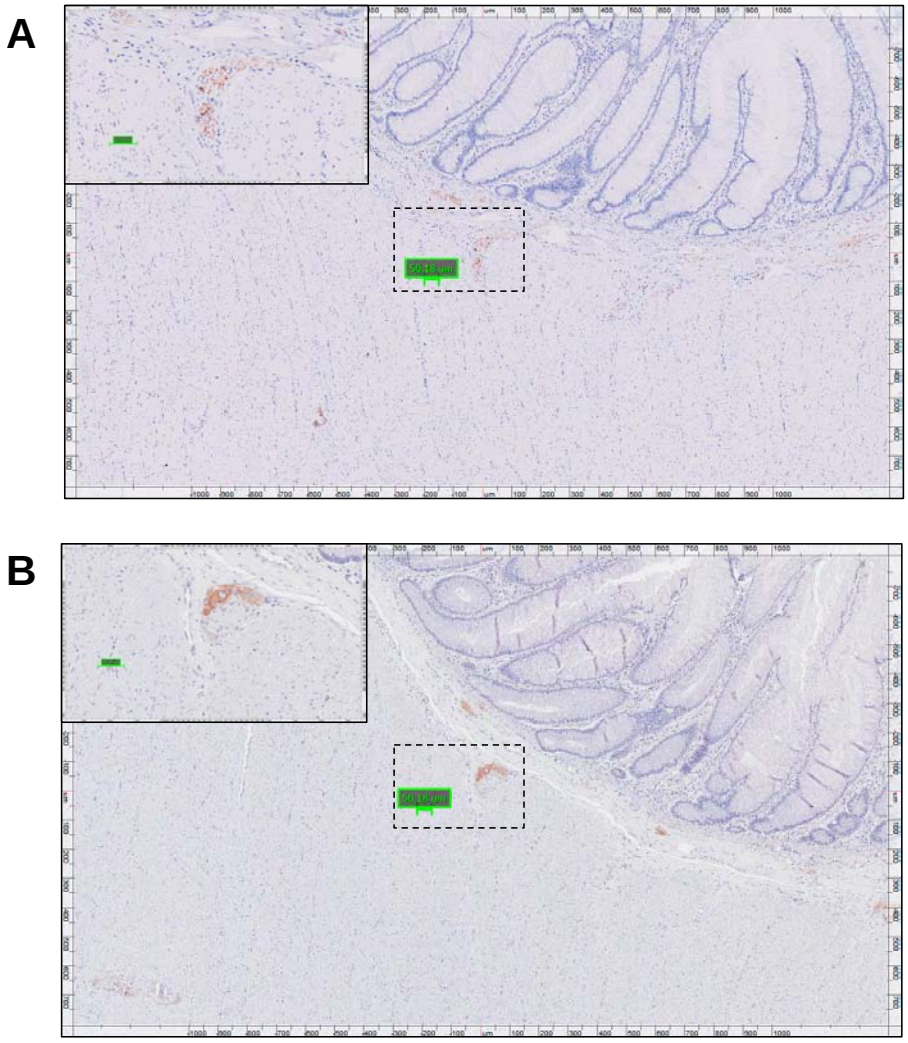


Figure 10. Serial staining of colon adenocarcinoma tissue sections against DBN1 and neurofilament (NF)
Comparisons between serial staining of colon adenocarcinoma tissue sections against NF **(A)** and DBN1 **(B)** showed that localisation of the DBN1 signal corresponded with that of NF. Images at 4x magnification. The dotted box indicates the area that is represented in the magnified inset image (at 20x magnification). Scale bars indicate 50.18 μm. Figure reproduced from Lin *et al.*, 2014, with permission from Proteomics, John Wiley and Sons.

3.2. Part Two. Comparative analysis of the HCT-116 and E1 secretomes using the hollow fibre culture (HFC) system

3.2.1. The HFC system is a viable alternative for CM preparation

When Wu and colleagues (2009) proposed the use of the HFC system as an alternative for CM preparation, they also studied the levels of intracellular contaminants in the HFC system. By analysing the levels of housekeeping intracellular proteins (such as G3PDH, Hsp60, tubulin and actin) in the CM prepared from the HFC system as opposed to dish culture, they discovered that the cell lysis rate in the HFC system was 0.001-0.022%, as compared to the 0.32-1.84% in dish culture (Wu *et al.*, 2009).

To further investigate the culture conditions in the HFC system as compared to traditional culture methods, we performed an iTRAQ-based comparison between the whole cell lysate profiles of HCT-116 and E1 cells cultured in flasks and the HFC system. A total of 990 proteins were identified within 1% local FDR and matched by at least two peptides. Using a 1.3 fold-change cut-off threshold, 57 and 70 proteins were observed to be overexpressed in HCT-116 and E1 cells cultured in the HFC system as compared to flask culture, with an overlap of 26 proteins in both cell lines. Similarly, 144 and 89 proteins were underexpressed in HCT-116 and E1 cells cultured in the HFC system, of which 76 proteins were common to both cell lines (Figure 11A, see Supplementary Table 4-5 for list of differentially expressed proteins). Gene Ontology Biological Process annotation enrichment analysis indicated that metabolic processes, gene expression and biosynthetic processes were enriched in the proteins which were overexpressed in both cell lines when cultured in the HFC system (Figure 11B). Conversely, terms related to stress

response and regulation of cell death were enriched in underexpressed proteins in both cell lines which were cultured in the HFC system (Figure 11C).

Our data extends the observations reported by Wu and colleagues (2009) that the HFC system provides better growth conditions for CM collection. More interestingly, we found that even though in serum-deprived conditions, the cells cultured in the HFC system had higher levels of proteins associated with proliferation and growth, while lower levels of those involved in stress response and cell death. The implications of this are particularly important, since serum-deprivation has often been criticised of inducing reductions in cell proliferation (Cooper, 2003; Shin *et al.*, 2008), increasing cell death (Hasan *et al.*, 1999), and alterations in protein synthesis and secretion patterns (Zander and Bemark, 2008). Therefore, the large surface area for growth and the dynamic removal of waste provided by the HFC system seems to mitigate to some extent the detrimental effects of serum-deprivation. In addition, coupled with the capability for concentrating secreted proteins in a small culture volume, the HFC system appears to be a highly attractive and viable alternative for preparation of CM samples.

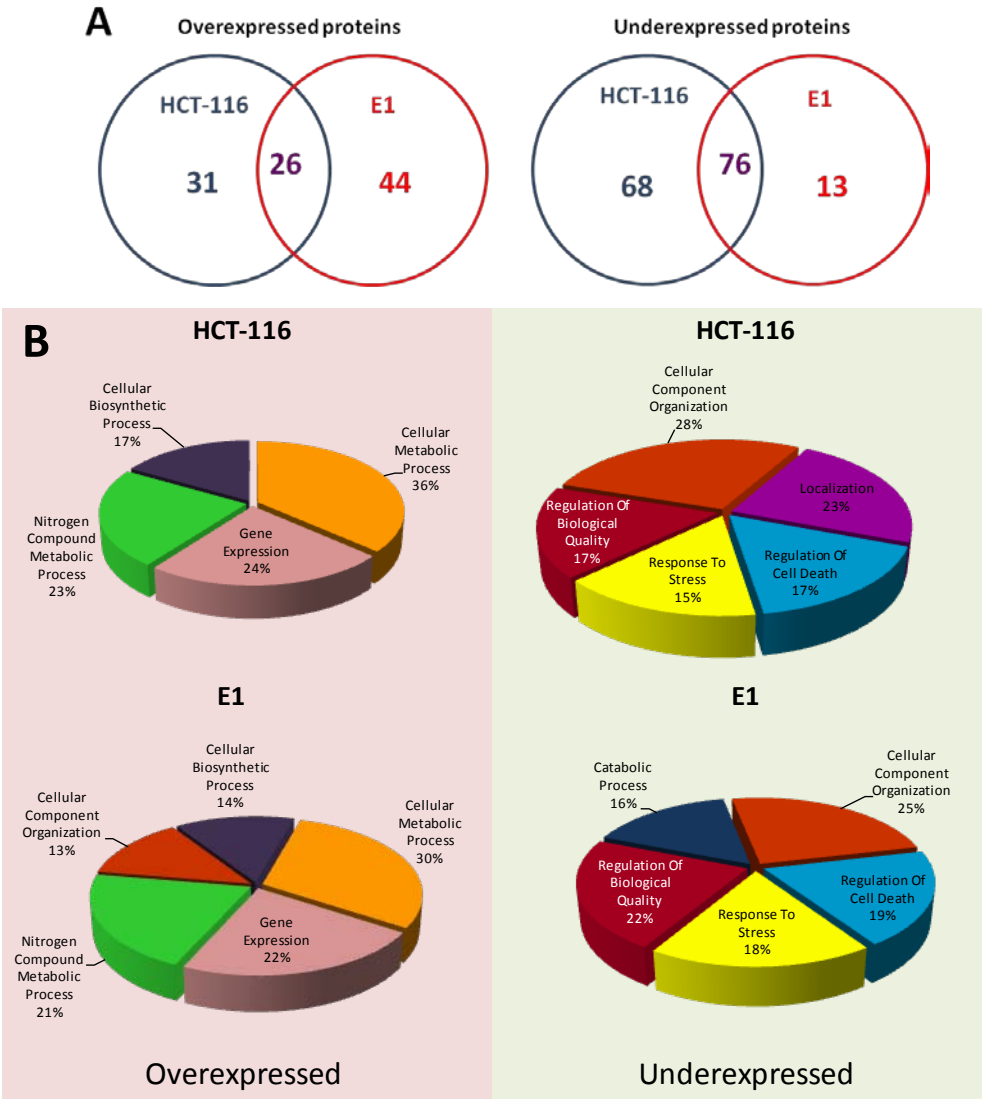


Figure 11. iTRAQ analysis of proteins differentially expressed in HCT-116 and E1 cells cultured in the HFC system as compared to culture flasks

(A) Number of proteins that were differentially expressed in HCT-116 and E1 cells cultured in the HFC system as compared to culture flasks **(B)** Gene Ontology Biological Process annotation of the differentially expressed proteins

3.2.2. iTRAQ analysis of HCT-116 and E1 secretomes

To identify differentially secreted proteins between HCT-116 and E1 cells which could be potential biomarkers for colorectal cancer metastasis, we performed an iTRAQ-based comparison between HCT-116 and E1 CM samples prepared using the HFC system. A total of 1201 proteins were identified using a 1% local FDR and with at least two peptides. To distinguish between genuine secreted proteins from intracellular contamination, we utilised an approach that we have previously published whereby the relative abundance of the proteins in the CM of each cell line were compared with that of its respective whole cell lysate (Loei *et al.*, 2012). This was based on the premise that secreted proteins would most likely be in higher abundance in the CM as compared to the cell lysate. Therefore, by applying a CM / Lysate cut-off ratio of more than 1.5, we would be able to differentiate between genuine secreted proteins from intracellular proteins released from cell lysis. Using this approach, we identified a total of 127 putative secreted proteins, of which 110 were common to both cell lines, nine were secreted only in HCT-116 cells, and eight were secreted only in E1 cells (Figure 12A).

When analysed using the SecretomeP secretion prediction server, 91 (72%) of the 127 putative secreted proteins were predicted to contain a N-terminal signal peptide, while 17 (13%) were predicted to be secreted using non-classical secretory mechanisms (Figure 12B). The remaining 19 unconfirmed proteins were further queried against the ExoCarta exosomal protein database and among these, 12 were reportedly secreted through exosomal release. In

total, 120 (94%) could be confirmed to be secreted through various mechanisms.

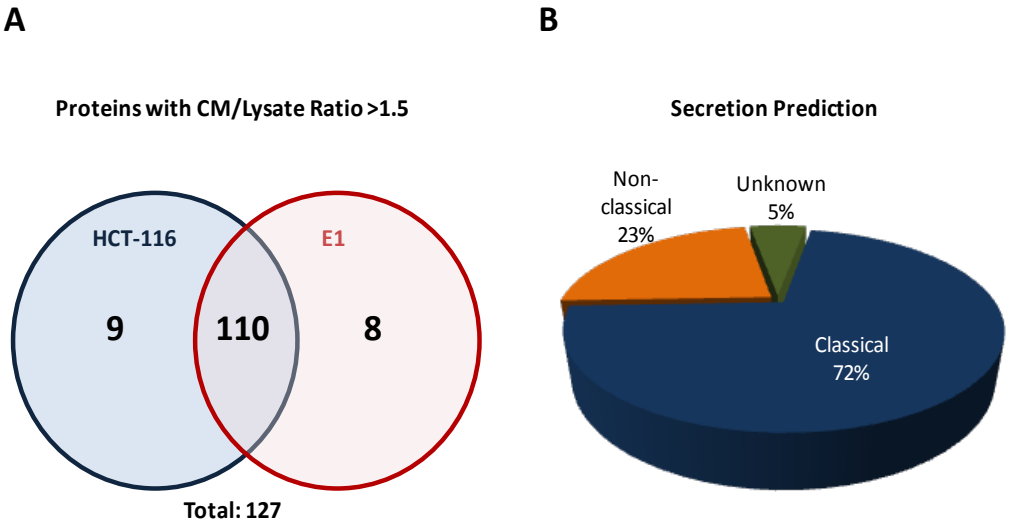


Figure 12. Overview of HCT-116 and E1 secretomes
(A) Number of proteins with CM/Lysate ratios of more than 1.5, indicating possible secretion. A total of 127 putative secreted proteins were identified, with 110 proteins common to both cell lines, nine secreted in HCT-116 only and eight secreted in E1 only. (**B)** Secretion prediction analysis of the 127 putative secreted proteins using the SecretomeP software and ExoCarta database. 91 proteins contained a signal peptide sequence, which would target the protein for secretion through the classical secretory pathway, and 29 proteins were predicted to be secreted non-classically. The remaining seven could not be confirmed.

Subsequently, to identify differentially secreted proteins in colorectal cancer metastasis, we compared the levels of the 127 secreted proteins in HCT-116 and E1 cells. Using a fold-change cut-off of 1.3, a total of 38 proteins were observed to be differentially secreted in the E1 CM (Table 4), of which 15 proteins were oversecreted in E1 cells, while 23 were undersecreted in E1 cells.

Table 4. Differentially secreted proteins identified from iTRAQ analysis of HCT-116 and E1 secretomes

IPI Accession[a]	Uniprot Accession[b]	Name	Unused[c]	Total[d]	% Cov[e]	Peptides (95%)[f]	E1 CM / HCT-116 CM Ratio (Replicate 1)[g]	E1 CM / HCT-116 CM Ratio (Replicate 2)[g]	SecretomeP[h]	ExoCarta[i]
IPI00014572.2	P09486	SPARC Secreted protein, acidic, cysteine-rich (Osteonectin)	21.67	22.2	52.7	22	6.30	7.35	0.889/Signal P	Present
IPI00291866.5	P05155	SERPING1 Plasma protease C1 inhibitor	11.12	11.12	51.8	9	4.58	5.22	0.676/Signal P	Present
IPI00847635.1	P01011	SERPINA3 Alpha-1-antichymotrypsin	3.1	5.25	27.2	3	4.32	3.43	0.839/Signal P	Present
IPI00783987.2	P01024	C3 Complement C3	5.73	6.16	19.7	5	3.85	9.60	0.618/Signal P	Present
IPI00007118.1	P05121	SERPINE1 Plasminogen activator inhibitor 1	26.27	26.6	53.7	38	3.21	3.57	0.644/Signal P	Present
IPI00016150.1	Q99574	SERPINI1 Neuroserpin	20.75	20.83	55.1	22	3.14	2.46	0.806/Signal P	Absent
IPI00009720.1	P15018	LIF Leukemia inhibitory factor	6.15	6.3	42.6	5	2.73	5.78	0.501/Signal P	Present
IPI00297252.6	Q8IWU5	SULF2 Extracellular sulfatase Sulf-2	7.74	7.95	30.8	4	2.45	2.47	0.607/Signal P	Absent
IPI00844511.1	P33908	MAN1A1 Mannosyl-oligosaccharide 1,2-alpha-mannosidase IA	51.01	51.48	58.2	63	2.31	1.67	0.437/Signal P	Present
IPI00023673.1	Q08380	LGALS3BP Galectin-3-binding protein	80.04	80.25	61.4	111	2.16	1.77	0.738/Signal P	Present
IPI00030255.1	O60568	PLOD3 Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3	21.31	21.39	41.6	19	1.87	2.00	0.561/Signal P	Present
IPI00306543.5	Q99988	GDF15 Growth/differentiation factor 15	10.02	10.05	55.2	7	1.85	3.43	0.883/Signal P	Absent
IPI00218201.2	P09603	CSF1 Macrophage colony-stimulating factor 1	4	4.01	20.6	5	1.78	1.45	0.416/Signal P	Present
IPI00982225.1	P05067	APP Amyloid beta A4 protein	38.02	38.02	54.3	41	1.44	1.64	0.215	Present
IPI00023728.1	Q92820	GGH Gamma-glutamyl hydrolase	25.96	26.28	52.2	24	1.37	1.49	0.628/Signal P	Present
IPI00103175.1	Q8WVQ1	CANT1 Soluble calcium-activated nucleotidase 1	30.7	30.83	71.8	35	0.77	0.72	0.635	Absent
IPI00177543.6	P19021	PAM Peptidyl-glycine alpha-amidating monooxygenase	21.43	21.54	36.8	17	0.74	0.63	0.358/Signal P	Present
IPI00215628.1	P13611	VCAN Versican core protein	12.06	12.26	16.4	12	0.72	0.73	0.470/Signal P	Present
IPI00807460.1	Q15642	TRIP10 Cdc42-interacting protein 4	4.06	4.06	34	4	0.71	0.76	0.47	Present
IPI00979124.1	Q15262	PTPRK receptor-type tyrosine-protein phosphatase kappa	18.93	19.16	18.9	20	0.69	0.56	0.357/Signal P	Present
IPI00009276.2	Q9UNN8	PROCR Endothelial protein C receptor precursor	11.97	12.07	29.5	9	0.64	0.72	0.862/Signal P	Absent
IPI00007778.1	Q01459	CTBS Di-N-acetylchitobiase	7.97	7.97	40.8	9	0.60	0.76	0.819/Signal P	Absent
IPI00604624.1	Q9GZT8	NIF3L1 NIF3-like protein 1	2.03	2.05	48.8	2	0.58	0.75	0.653	Absent
IPI00024129.1	P45877	PPIC Peptidyl-prolyl cis-trans isomerase C	8.02	10.03	25.9	6	0.58	0.56	0.878/Signal P	Absent
IPI00018387.1	P09958	FURIN Furin	5.62	5.67	12.9	3	0.58	0.77	0.388/Signal P	Present

Table 4. Continued

IPI Accession[a]	Uniprot Accession[b]	Name	Unused[c]	Total[d]	% Cov[e]	Peptides (95%)[f]	E1 CM / HCT-116 CM Ratio (Replicate 1)[g]	E1 CM / HCT-116 CM Ratio (Replicate 2)[g]	SecretomeP[h]	ExoCarta[i]
IPI00003590.2	O00391	QSOX1 Sulfhydryl oxidase 1	57.37	57.51	70.8	75	0.57	0.68	0.611/Signal P	Present
IPI00419724.2	Q9NPR2	SEMA4B semaphorin-4B precursor	5.43	5.5	20.3	4	0.57	0.62	0.446/Signal P	Absent
IPI00294004.1	P07225	PROS1 Vitamin K-dependent protein S	24.28	24.28	34.8	26	0.56	0.52	0.5/Signal P	Present
IPI00025846.3	Q02487	DSC2 Desmocollin-2	15.26	15.42	32.4	20	0.55	0.55	0.313/Signal P	Present
IPI00807403.1	Q13740	ALCAM CD166 antigen	22.25	22.3	44.6	22	0.50	0.55	0.461/Signal P	Present
IPI00025861.3	P12830	CDH1 Cadherin-1	26.94	27.44	38.4	20	0.49	0.59	0.179/Signal P	Present
IPI00015953.3	Q9NR30	DDX21 Nucleolar RNA helicase 2	6.03	6.21	44.7	3	0.47	0.61	0.508	Present
IPI00290039.7	Q9H5V8	CDCP1 CUB domain-containing protein 1	12.63	12.63	33.5	14	0.46	0.57	0.181/Signal P	Absent
IPI00012023.1	P15514	AREG Amphiregulin	8.15	8.18	39.7	10	0.43	0.56	0.645/Signal P	Absent
IPI00293088.7	P10253	GAA Lysosomal alpha-glucosidase	16.82	17.07	30.5	16	0.40	0.67	0.796/Signal P	Present
IPI00028911.2	Q14118	DAG1 Dystroglycan	22.67	22.79	30.1	18	0.40	0.47	0.110/Signal P	Present
IPI00465315.6	P99999	CYCS Cytochrome c	6.33	6.33	58.4	4	0.32	0.59	0.465	Absent
IPI00017968.1	P35318	ADM ADM	3.07	3.08	53	4	0.31	0.41	0.608/Signal P	Absent

[a] Accession number of the identified proteins in the IPI Human v3.87 database.

[b] Corresponding UniProt Accession of the identified proteins.

[c] ProteinPilot Unused Score. Total score of the unique peptides detected for a protein. Used as a measure of protein confidence for the identified protein.

[d] ProteinPilot Total Score. Total score of all the peptides detected for a protein.

[e] Percentage sequence coverage of identified peptides.

[f] Number of distinct peptides identified with at least 95% confidence.

[g] Fold change ratio of each secreted protein in E1 against HCT-116 CM samples, in biological replicates 1 and 2.

[h] Secretion prediction using SecretomeP software. Results are presented in the form of "NN-score" followed by the "Signal P" annotation, which would denote secretion using the classical secretory pathway. A NN-score of > 0.5 would indicate non-classical secretion.

[i] Secretion prediction using the ExoCarta database. Proteins which are denoted as "present" have been found to be present in exosomes.

3.2.3. Large-scale verification of iTRAQ data using SWATH-MS for SRM-like targeted quantitation

To increase the confidence in the list of differentially secreted proteins that were identified by the iTRAQ analysis, we decided to perform a MS-based large-scale verification of the iTRAQ data by using SWATH-MS acquisition to perform SRM-like targeted quantitation. A total of three technical replicates from each of the two HCT-116 and E1 CM biological replicates were analysed using SWATH-MS acquisition, while the spectral library was generated using DDA on a pooled sample from all the CM samples. For targeted quantitative analysis of the above-mentioned 38 differentially secreted proteins in E1, peptides and transitions suitable for quantitation were manually selected using the SWATH Acquisition MicroApp in the PeakView software. Subsequently, extraction of the peak areas from the extracted ion chromatograms (XICs) of the selected peptides and transitions was performed using the MultiQuant software. For each protein, the peak areas from its respective peptides and transitions extracted from the three technical replicates were summed (denoted as "total area sum"), and this was used as a representation of the relative abundance of the protein in the sample. Finally, the fold change ratio of the protein was then calculated by the quotient of the total area sum of the protein in the E1 CM sample divided by that in the HCT-116 CM sample (Table 5). Using this approach for targeted quantitative analysis, we observed that among the 38 differentially secreted proteins in E1, 25 proteins (66%) were corroborated by the SWATH-MS analysis (Table 5).

Table 5. Differentially secreted proteins which showed similar trends in both iTRAQ and SWATH-MS analysis

Protein	iTRAQ		SWATH-MS						
	E1 CM / HCT-116 CM Ratio (Replicate 1)[a]	E1 CM / HCT-116 CM Ratio (Replicate 2)[a]	E1 CM Replicate 1 Total Area Sum[b]	HCT CM Replicate 1 Total Area Sum[b]	E1 / HCT-116 Ratio (Replicate 1)[c]	E1 CM Replicate 2 Total Area Sum	HCT CM Replicate 2 Total Area Sum	E1 / HCT-116 Ratio (Replicate 2)	Peptides used[d]
SPARC	**6.30**	**7.35**	1.73E+06	9.45E+03	**182.72**	2.10E+06	9.13E+03	**229.83**	3
SERPING1	**4.58**	**5.22**	4.06E+05	2.56E+04	**15.84**	2.60E+05	2.42E+04	**10.73**	3
SERPINA3	**4.32**	**3.43**	2.53E+04	1.67E+03	**15.14**	2.58E+04	2.37E+03	**10.87**	1
C3	**3.85**	**9.60**	1.81E+05	7.72E+04	**2.35**	1.61E+05	5.54E+04	**2.91**	3
SERPINE1	**3.21**	**3.57**	1.05E+06	2.01E+05	**5.22**	1.07E+06	2.22E+05	**4.84**	3
SERPINI1	**3.14**	**2.46**	8.69E+05	2.08E+05	**4.19**	6.11E+05	1.78E+05	**3.43**	3
SULF2	**2.45**	**2.47**	8.98E+04	1.57E+04	**5.70**	1.04E+05	2.50E+04	**4.17**	3
MAN1A1	**2.31**	**1.67**	1.58E+06	5.24E+05	**3.01**	2.01E+06	1.15E+06	**1.75**	3
LGALS3BP	**2.16**	**1.77**	3.07E+06	1.10E+06	**2.79**	3.36E+06	1.57E+06	**2.13**	3
PLOD3	**1.87**	**2.00**	1.98E+05	6.51E+04	**3.04**	1.04E+05	4.34E+04	**2.40**	3
GDF15	**1.85**	**3.43**	2.53E+05	7.31E+04	**3.47**	2.24E+05	8.71E+04	**2.57**	3
CSF1	**1.78**	**1.45**	5.57E+04	2.27E+04	**2.45**	7.58E+04	2.46E+04	**3.09**	2
APP	**1.44**	**1.64**	2.31E+06	1.05E+06	**2.19**	1.96E+06	1.31E+06	**1.50**	3
CTBS	**0.60**	**0.76**	4.89E+04	9.37E+04	**0.52**	4.74E+04	6.86E+04	**0.69**	2
NIF3L1	**0.58**	**0.75**	3.73E+04	6.78E+04	**0.55**	3.35E+04	4.49E+04	**0.75**	2
PPIC	**0.58**	**0.56**	1.43E+05	3.06E+05	**0.47**	1.44E+05	2.86E+05	**0.50**	3
SEMA4B	**0.57**	**0.62**	6.73E+04	9.17E+04	**0.73**	7.48E+04	1.40E+05	**0.53**	1
PROS1	**0.56**	**0.52**	5.14E+05	7.55E+05	**0.68**	3.85E+05	1.01E+06	**0.38**	3
DSC2	**0.55**	**0.55**	2.06E+05	3.82E+05	**0.54**	1.97E+05	5.34E+05	**0.37**	2
ALCAM	**0.50**	**0.55**	1.21E+05	1.79E+05	**0.68**	9.62E+04	1.76E+05	**0.55**	3
CDH1	**0.49**	**0.59**	6.16E+05	1.03E+06	**0.60**	5.77E+05	1.20E+06	**0.48**	3
AREG	**0.43**	**0.56**	1.96E+05	2.69E+05	**0.73**	1.08E+05	2.33E+05	**0.46**	1
GAA	**0.40**	**0.67**	5.79E+04	1.35E+05	**0.43**	4.08E+04	6.73E+04	**0.61**	3
DAG1	**0.40**	**0.47**	5.76E+04	1.10E+05	**0.52**	4.52E+04	1.00E+05	**0.45**	2
CYCS	**0.32**	**0.59**	3.59E+05	1.58E+06	**0.23**	3.54E+05	8.76E+05	**0.40**	3

[a] Fold change ratio of each secreted protein in E1 against HCT-116 CM samples reported in the iTRAQ analysis, in biological replicates 1 and 2.

[b] Total sum of the areas under the peak signals obtained for each protein across three technical replicates, from up to three peptides each with three transitions.

[c] SWATH-MS quantitation ratio for each protein, obtained by the quotient of the total area sums of all the peptides for a particular protein in E1 against that of HCT-116.

[d] Number of peptides used for SWATH-MS quantitation

Eight proteins (LIF, GGH, TRIP10, PTPRk, PROCR, FURIN, DDX21 and ADM) were not detected in the SWATH-MS analysis, while the differential secretion of five (CANT1, PAM, VCAN, QSOX1 and CDCP1) reported by the SWATH-MS analysis differed from that in the iTRAQ analysis. Therefore, downstream analyses were only performed on the 25 differentially secreted proteins which were corroborated by both iTRAQ and SWATH-MS quantitative analyses. Gene Ontology Biological Process annotation of these 25 differentially secreted proteins in E1 indicated that they could be involved in processes such as signal transduction, adhesion and migration, as well as protein processing and post-translational modifications (Figure 13).

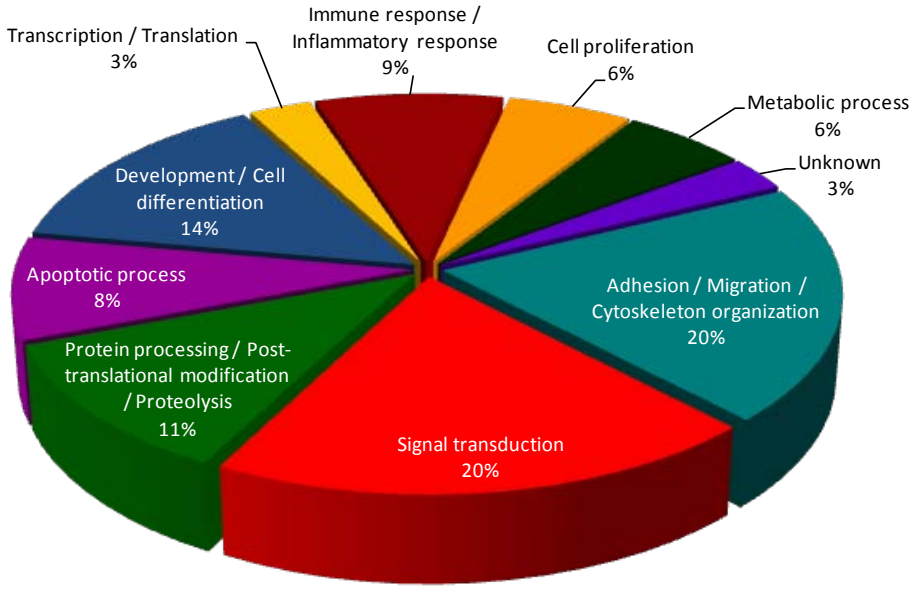


Figure 13. Gene Ontology Biological Process annotation of the 25 differentially secreted proteins in the E1 secretome
Gene Ontology Biological Process annotation was performed on the 25 secreted proteins which showed similar differential trends in both iTRAQ and SWATH-MS analyses.

3.2.4. Western blot validation of selected candidate proteins differentially secreted in E1 cells

To verify the differential secretion of the proteins in E1 as suggested by the iTRAQ data, we selected five proteins for western blot validation. These five proteins were oversecreted in E1 cells and included well-known dysregulated secreted proteins in colorectal cancer such as GDF15, PAI1, SPARC, as well as potentially novel players such as MAN1A1 and PLOD3.

The levels of the abovementioned five proteins were assessed in both the CM and lysates of HCT-116 and E1 cells using commercially available antibodies. The oversecretion of the five selected proteins in the HCT-116 and E1 CM were consistent with both the iTRAQ and SWATH-MS data (Figure 14). The molecular weights (MW) of the immunoreactive bands from MAN1A1 in the CM samples were lower than its theoretical MW and presumably represented a truncated form of the protein with its signal peptide cleaved. The immunoreactive bands from MAN1A1 in the whole cell lysates samples corresponded to its theoretical MW, while the levels of the other 4 proteins were too low to be detected in the whole cell lysates of HCT-116 and E1 under the immunoblotting conditions. In addition, we also probed for the presence of Lamin B, a component of the nuclear lamina, as a control to assess the level of intracellular contamination in the CM samples. Typical housekeeping proteins, such as GAPDH and actin, were not used because they have been detected to be present in exosomes previously (Mathivanan *et al.*, 2012). We were unable to detect any Lamin B in HCT-116 and E1 CM samples from both biological replicates, although strong immunoreactive

bands were detected from the whole cell lysate samples. Therefore, this indicated that our CM samples should have minimal intracellular contamination.

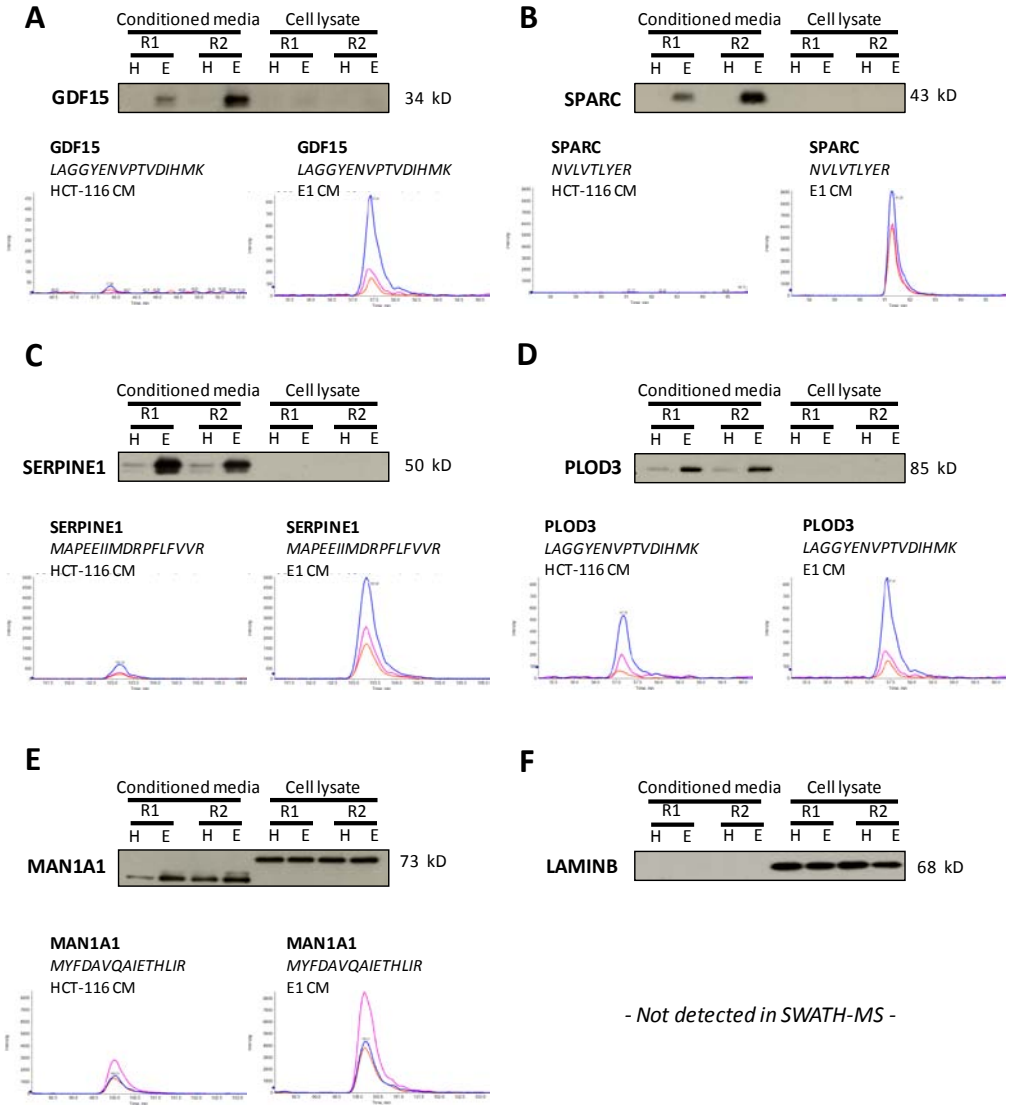


Figure 14. SWATH-MS and western blot analysis of selected differentially secreted proteins in E1

Differential secretion of **(A)** GDF15, **(B)** SPARC, **(C)** SERPINE1, **(D)** PLOD3 and **(E)** MAN1A1 suggested by the iTRAQ data, validated using both SWATH-MS and western blot. SWATH-MS analyses were performed on CM samples only, and representative extracted ion chromatograms (XICs) from one peptide of each protein are shown here. Western blot analyses were performed on both CM and whole cell lysate samples. Samples from HCT-116 *[H]* and E1 *[E]* from both biological replicates *[R1 and R2]* were analysed. **(F)** LaminB, a component of the nuclear lamina, was used as a control to confirm the absence of intracellular contamination in the CM samples.

3.3. Part Three. SWATH-MS quantitative analysis of HCT-116 and E1 MLAC-enriched secretomes

In the spectral library generated using DDA from the pooled sample, a total of 568 proteins were identified using a 1% local FDR and with at least two peptides. Using the criteria of selecting only those proteins which had 3 or more peptides suitable for quantitation, 421 proteins were selected for further quantitative analyses. Using SecretomeP in the same manner as described earlier, 287 of the 421 proteins (68%) were predicted to contain an N-terminal signal peptide, while 53 (13%) were predicted to be secreted using non-classical secretion pathways (Figure 15). The remaining 81 unconfirmed proteins were then searched against the ExoCarta database, and 66 of these have been previously discovered in exosomes. In total, 406 of the 421 proteins (96%) could be confirmed to be secreted through various mechanisms. In addition, when queried against the UniProt database, 294 of the 421 proteins (70%) have been previously shown to contain at least a single glycosylation modification.

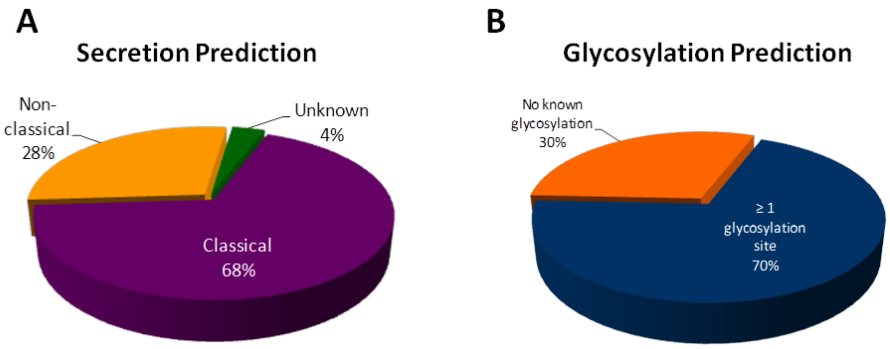


Figure 15. Overview of the HCT-116 and E1 glycosecretomes
(A) Secretion prediction on the 421 proteins selected for quantitative analyses using the SecretomeP software and ExoCarta database. A total of 287 of these proteins were predicted to be secreted by the classical secretory pathway, while 119 proteins were predicted to be secreted non-classically. The remaining 15 proteins could not be confirmed. **(B)** A total of 294 proteins were found to contain at least a single glycosylation site when queried against the UniProt database, while there was no evidence of any known or potential glycosylations for the remaining 127 proteins.

A total of three technical replicates from each HCT-116 and E1 MLAC-enriched CM sample were analysed using the SWATH-MS acquisition. For quantitative analysis of the above-mentioned 421 proteins, peptides and transitions suitable for quantitation were automatically selected using the SWATH Acquisition MicroApp in the PeakView software, and their respective XIC peak areas were extracted and exported for analysis in the MarkerView software. To control for uneven sample loss across the different samples during the sample preparation process, we performed a global normalisation based on the total sum of all the peak areas extracted from all the peptides and transitions across the three technical replicates of each CM sample. The basis for this was that given the similar genetic backgrounds of the HCT-116 and E1 cell lines, majority of the proteins should have relatively similar abundances across the two cell lines.

After the normalisation process, we used the Student's T-test to perform a statistical comparison of the average extracted peak area for each protein in the E1 MLAC-enriched CM, with respect to that in HCT-116. An overview of the analysis is shown in Figure 16, where the logarithm of the fold change for each of the 421 selected proteins, in E1 as compared to HCT-116, is plotted with its respective *p*-value from the Student's T-test analysis. To select for statistically significant differentially secreted proteins in the E1 MLAC-enriched CM, we only considered secreted glycoproteins which showed more than the 1.3 fold-change difference with *p*-values of less than 0.05. Using these criteria, a total of 149 glycoproteins were observed to be differentially secreted in E1 cells (Table 6).

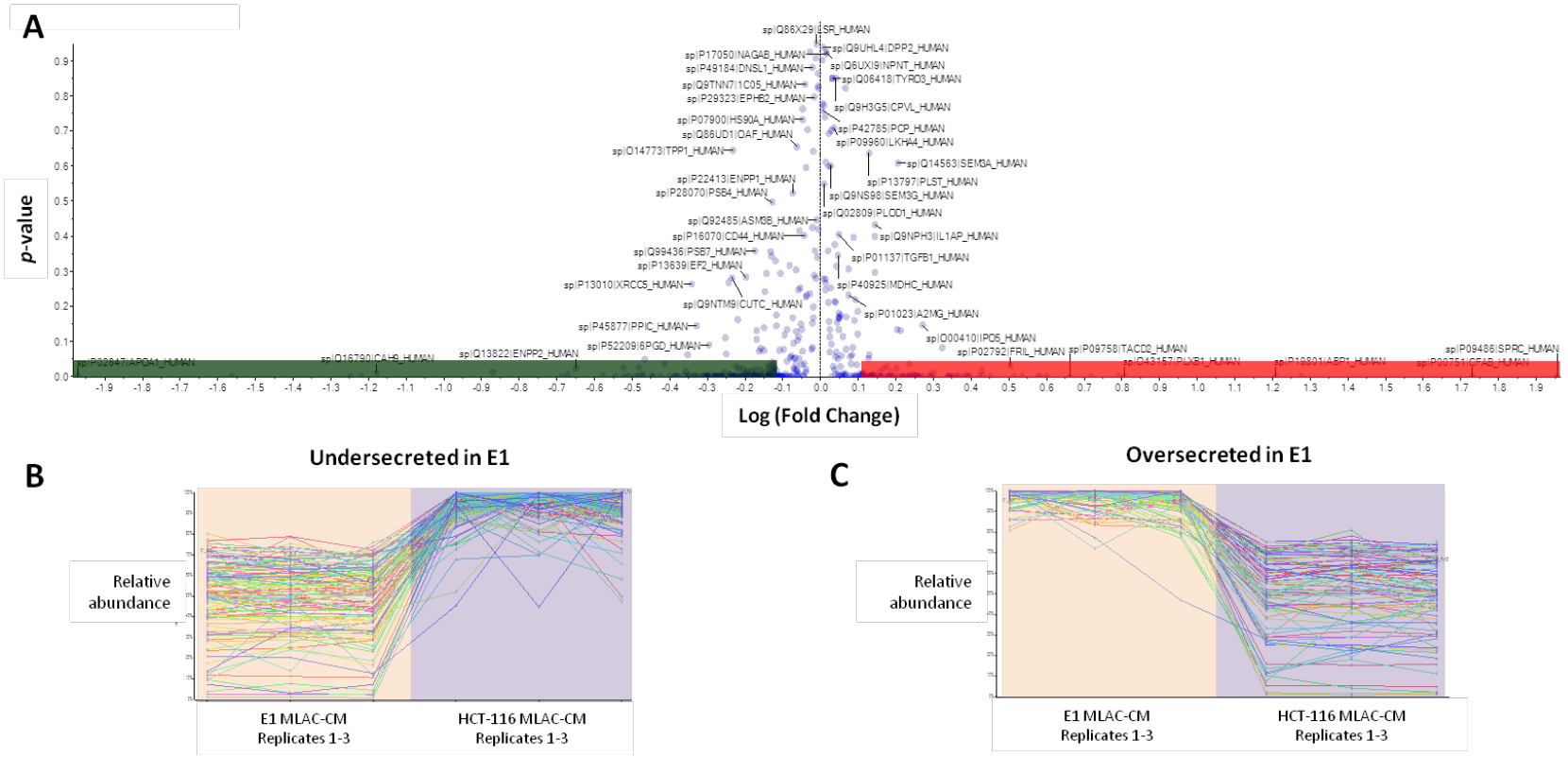


Figure 16. SWATH-MS quantitative analysis of HCT-116 and E1 MLAC-enriched CM

(A) Quantitative analysis of the 421 proteins identified in both HCT-116 and E1 MLAC-enriched CM samples, depicted in the form of a volcano plot. The logarithm of the fold change ratio for each protein in the E1 CM as compared to HCT-116 was plotted with respect to its respective *p*-value. The secreted glycoproteins which show statistically significant undersecretion in the E1 CM are highlighted in green, with their relative abundances in each sample plotted in **(B)**, and those with significant oversecretion in the E1 CM are highlighted in red, with their relative abundances plotted in **(C)**.

Table 6. Differentially secreted glycoproteins identified from SWATH-MS analysis

Uniprot Accession	Name	Unused[a]	Total[b]	% Cov[c]	Peptides(95%)[d]	Glycoprotein[e]	SecretomeP[f]	ExoCarta[g]	E1 Mean[h]	HCT-116 Mean	E1 Sigma[i]	HCT-116 Sigma	E1/HCT-116 ratio[j]	*p* -value[k]
P09486	SPARC SPARC	16.43	16.43	46.5	10	Yes	0.942/SignalP	Present	3.3E+05	3.6E+03	9.6E+03	8.8E+02	90.20	5.2E-07
P00751	CFB Complement factor B	14.21	14.21	23.7	8	Yes	0.457/SignalP	Present	1.5E+05	2.8E+03	8.1E+03	4.8E+02	53.69	6.0E-06
P05155	SERPING1 Plasma protease C1 inhibitor	26.55	26.55	34.2	20	Yes	0.676/SignalP	Present	1.3E+06	6.7E+04	1.5E+04	4.5E+03	18.92	1.7E-08
P19801	ABP1 Amiloride-sensitive amine oxidase [copper-containing]	9.92	9.92	21	5	Yes	0.656/SignalP	Present	9.7E+04	6.0E+03	1.2E+04	4.6E+03	16.15	3.0E-04
O43157	PLXNB1 Plexin-B1	8.9	8.9	7.1	5	Yes	0.552/SignalP	Absent	4.4E+04	6.8E+03	5.4E+03	1.7E+03	6.41	3.4E-04
P36955	SERPINF1 Pigment epithelium-derived factor	26.19	26.19	53.6	15	Yes	0.825/SignalP	Present	2.3E+05	3.8E+04	6.3E+03	6.8E+02	6.19	7.1E-07
P20061	TCN1 Transcobalamin-1	9.4	9.4	17.6	9	Yes	0.793/SignalP	Present	5.7E+04	1.1E+04	1.1E+03	3.6E+03	5.22	2.8E-05
P09758	TACSTD2 Tumor-associated calcium signal transducer 2	6	6	15.2	3	Yes	0.316/SignalP	Present	3.3E+04	7.1E+03	1.5E+03	3.3E+03	4.60	2.6E-04
P09603	CSF1 Macrophage colony-stimulating factor 1	15.24	15.24	16.4	14	Yes	0.488/SignalP	Present	5.2E+05	1.3E+05	2.0E+04	8.5E+03	3.98	6.0E-06
Q8IWU5	SULF2 Extracellular sulfatase Sulf-2	6	6	9.4	3	Yes	0.607/SignalP	Absent	3.8E+04	9.9E+03	5.8E+02	1.2E+03	3.84	3.4E-06
P05121	SERPINE1 Plasminogen activator inhibitor 1	151.06	151.06	83.8	174	Yes	0.644/SignalP	Present	7.9E+06	2.1E+06	5.4E+05	7.5E+04	3.76	5.1E-05
Q99574	SERPINI1 Neuroserpin	44.87	44.87	62.2	40	Yes	0.806/SignalP	Absent	1.0E+06	2.9E+05	4.9E+03	4.2E+03	3.43	4.5E-09
P08195	SLC3A2 4F2 cell-surface antigen heavy chain	22.37	22.37	37.6	12	Yes	0.644	Present	1.1E+05	3.2E+04	7.4E+02	1.6E+03	3.26	2.1E-07
P01011	SERPINA3 Alpha-1-antichymotrypsin	6.31	6.31	22.9	4	Yes	0.839/SignalP	Present	5.4E+04	1.9E+04	5.2E+03	1.7E+03	2.81	3.7E-04
Q08431	MFGE8 Lactadherin	7.68	7.68	39	6	Yes	0.732/SignalP	Absent	9.5E+04	3.5E+04	3.2E+03	1.4E+04	2.70	2.1E-03
P34059	GALNS N-acetylgalactosamine-6-sulfatase	8.1	8.1	15.1	5	Yes	0.779/SignalP	Absent	8.5E+03	3.2E+03	6.4E+02	1.1E+03	2.64	2.1E-03
P05154	SERPINA5 Plasma serine protease inhibitor	10.37	10.37	23.4	6	Yes	0.838/SignalP	Present	4.1E+04	1.6E+04	1.7E+03	5.6E+03	2.49	1.9E-03
P53634	CTSC Dipeptidyl peptidase 1	15.23	15.36	33.1	13	Yes	0.767/SignalP	Present	1.9E+05	7.9E+04	7.7E+03	9.3E+03	2.46	8.2E-05
P15018	LIF Leukemia inhibitory factor	7.23	7.23	37.1	4	Yes	0.501/SignalP	Present	2.0E+04	8.4E+03	2.1E+03	1.7E+03	2.44	1.6E-03
P30530	AXL Tyrosine-protein kinase receptor UFO	10	10	14	6	Yes	0.363/SignalP	Present	2.4E+05	1.0E+05	1.3E+04	9.7E+03	2.38	1.1E-04
P00533	EGFR Epidermal growth factor receptor	5.8	5.8	7.2	3	Yes	0.512/SignalP	Present	8.2E+03	3.6E+03	7.7E+02	5.4E+02	2.30	1.0E-03
Q14766	LTBP1 Latent-transforming growth factor beta-binding protein 1	18.09	18.09	15.1	10	Yes	0.387/SignalP	Present	2.7E+04	1.2E+04	2.6E+03	1.7E+03	2.25	1.1E-03
Q6PCB0	VWA1 von Willebrand factor A domain-containing protein 1	17.45	17.45	50.3	10	Yes	0.517/SignalP	Present	1.6E+05	7.0E+04	8.1E+03	3.5E+02	2.22	5.3E-05
P01024	C3 Complement C3	43.25	43.25	36.7	32	Yes	0.618/SignalP	Present	1.3E+05	6.0E+04	7.0E+03	4.4E+03	2.21	1.1E-04
P32004	L1CAM Neural cell adhesion molecule L1	18.6	18.62	16.9	13	Yes	0.554/SignalP	Present	1.4E+05	6.5E+04	3.2E+03	1.6E+03	2.18	3.4E-06
Q08380	LGALS3BP Galectin-3-binding protein	75.64	75.64	50.4	112	Yes	0.738/SignalP	Present	3.2E+06	1.5E+06	2.4E+04	4.0E+04	2.18	3.6E-07
P02751	FN1 Fibronectin	18.61	18.61	14.5	11	Yes	0.369/SignalP	Present	2.1E+04	1.1E+04	1.9E+03	3.6E+03	2.00	1.0E-02
O60568	PLOD3 Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3	85.86	85.86	63	73	Yes	0.561/SignalP	Present	2.3E+06	1.2E+06	4.9E+04	2.5E+04	1.98	3.5E-06
Q92820	GGH Gamma-glutamyl hydrolase	28.31	28.31	45.3	25	Yes	0.628/SignalP	Present	2.3E+06	1.2E+06	3.0E+04	3.2E+04	1.96	1.7E-06
Q99988	GDF15 Growth/differentiation factor 15	6.21	6.21	28.6	7	Yes	0.883/SignalP	Present	3.8E+04	2.0E+04	3.3E+03	2.5E+03	1.94	1.5E-03
Q7Z304	MAMDC2 MAM domain-containing protein 2	14.73	14.73	30.6	8	Yes	0.701/SignalP	Present	5.5E+04	2.9E+04	1.7E+03	4.0E+03	1.87	5.1E-04
O14672	ADAM10 Disintegrin and metalloproteinase domain-containing protein 10	9.57	9.57	14.4	5	Yes	0.158/SignalP	Present	7.6E+04	4.1E+04	3.7E+03	3.1E+03	1.83	2.5E-04
Q92876	KLK6 Kallikrein-6	12.99	12.99	49.6	9	Yes	0.771/SignalP	Absent	4.7E+05	2.6E+05	4.0E+04	4.9E+03	1.83	7.9E-04
P00749	PLAU Urokinase-type plasminogen activator	9.71	9.71	30.4	6	Yes	0.65/SignalP	Present	1.3E+05	7.0E+04	2.3E+03	4.3E+03	1.82	3.6E-05
Q93063	EXT2 Exostosin-2	7.13	7.13	19.5	6	Yes	0.705	Present	5.2E+04	2.9E+04	1.1E+03	4.1E+03	1.79	7.5E-04
P27797	CALR Calreticulin	13.69	13.69	41.3	8	Yes	0.366/SignalP	Present	3.6E+04	2.0E+04	4.0E+03	1.8E+03	1.79	3.4E-03
P33908	MAN1A1 Mannosyl-oligosaccharide 1,2-alpha-mannosidase IA	58.43	58.43	47.3	59	Yes	0.437/SignalP	Present	8.7E+05	4.9E+05	7.6E+04	1.5E+04	1.78	1.1E-03
Q9BZR6	RTN4R Reticulon-4 receptor	21.84	21.84	46.1	20	Yes	0.562/SignalP	Absent	3.4E+05	1.9E+05	3.8E+04	2.7E+04	1.77	5.2E-03
Q7Z7M9	GALNT5 Polypeptide N-acetylgalactosaminyltransferase 5	18.75	18.76	11.5	10	Yes	0.286/SignalP	Present	1.3E+05	7.7E+04	1.4E+04	2.8E+04	1.73	3.6E-02

Table 6. Continued

Uniprot Accession	Name	Unused[a]	Total[b]	% Cov[c]	Peptides(95%)[d]	Glycoprotein[e]	SecretomeP[f]	ExoCarta[g]	E1 Mean[h]	HCT-116 Mean	E1 Sigma[i]	HCT-116 Sigma	E1/HCT-116 ratio[j]	*p* -value[k]
P02750	LRG1 Leucine-rich alpha-2-glycoprotein	12.15	12.15	35.7	8	Yes	0.72/SignalP	Present	5.6E+04	3.3E+04	4.6E+03	4.3E+03	1.73	2.9E-03
P12821	ACE Angiotensin-converting enzyme	20.93	20.93	18.8	15	Yes	0.668/SignalP	Present	1.2E+05	7.2E+04	2.8E+03	7.7E+03	1.72	3.8E-04
Q9HAT2	SIAE Sialate O-acetylesterase	42	42	51.4	41	Yes	0.786/SignalP	Present	3.9E+05	2.3E+05	1.6E+04	2.3E+04	1.71	5.6E-04
Q14697	GANAB Neutral alpha-glucosidase AB	28.08	28.08	33.1	16	Yes	0.64/SignalP	Present	7.5E+04	4.5E+04	2.9E+03	1.3E+03	1.68	7.5E-05
P23142	FBLN1 Fibulin-1	66.53	66.53	52.2	51	Yes	0.589/SignalP	Present	6.5E+05	4.0E+05	6.1E+03	1.6E+04	1.65	1.4E-05
Q9GZM7	TINAGL1 Tubulointerstitial nephritis antigen-like	20.02	20.02	41.8	16	Yes	0.935/SignalP	Present	3.5E+05	2.1E+05	3.4E+03	1.6E+04	1.64	1.2E-04
Q8NBI3	DRAXIN Draxin	4	4	9.5	2	Yes	0.592/SignalP	Absent	4.6E+04	2.8E+04	1.9E+03	1.4E+03	1.62	2.0E-04
P55268	LAMB2 Laminin subunit beta-2	16.26	16.31	12.4	8	Yes	0.261/SignalP	Present	2.0E+04	1.3E+04	1.6E+03	3.3E+03	1.59	2.4E-02
Q6YHK3	CD109 CD109 antigen	192.8	192.8	58.9	177	Yes	0.6/SignalP	Present	3.0E+06	1.9E+06	2.8E+04	5.5E+04	1.58	6.2E-06
P18827	SDC1 Syndecan-1	2.89	2.89	5.5	6	Yes	0.539/SignalP	Present	1.5E+06	9.5E+05	1.8E+04	3.5E+04	1.55	2.2E-05
P12110	COL6A2 Collagen alpha-2(VI) chain	38.02	38.02	44.8	29	Yes	0.214/SignalP	Present	5.5E+05	3.6E+05	2.3E+04	1.9E+04	1.54	3.8E-04
Q9NZ53	PODXL2 Podocalyxin-like protein 2	8.67	8.67	18	9	Yes	0.284/SignalP	Absent	2.6E+05	1.7E+05	2.9E+04	2.1E+04	1.54	1.2E-02
Q9NY97	B3GNT2 UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 2	10.32	10.32	32.5	6	Yes	0.535/SignalP	Absent	5.3E+04	3.5E+04	6.6E+02	2.3E+03	1.52	2.0E-04
Q8NBP7	PCSK9 Proprotein convertase subtilisin/kexin type 9	101.9	101.9	53.6	93	Yes	0.761/SignalP	Present	2.6E+06	1.7E+06	1.9E+04	2.6E+04	1.51	1.1E-06
P07339	CTSD Cathepsin D	22.63	22.63	47.8	15	Yes	0.758/SignalP	Present	1.4E+05	9.0E+04	2.5E+03	2.7E+03	1.51	2.7E-05
Q9NT99	LRRC4B Leucine-rich repeat-containing protein 4B	6.5	6.5	15.2	5	Yes	0.098/SignalP	Absent	1.7E+04	1.1E+04	2.9E+03	6.5E+02	1.50	2.7E-02
P98160	HSPG2 Basement membrane-specific heparan sulfate proteoglycan core protein	243.69	243.69	48.4	175	Yes	0.51/SignalP	Present	2.4E+06	1.6E+06	1.1E+04	5.9E+04	1.48	2.5E-05
P09668	CTSH Pro-cathepsin H	10.21	10.21	55.2	9	Yes	0.574/SignalP	Absent	1.1E+05	7.2E+04	1.3E+03	1.6E+04	1.46	2.4E-02
Q9Y4L1	HYOU1 Hypoxia up-regulated protein 1	15.29	15.29	19.4	13	Yes	0.32/SignalP	Present	3.4E+05	2.3E+05	1.0E+04	8.4E+03	1.45	1.7E-04
P19438	TNFRSF1A Tumor necrosis factor receptor superfamily member 1A	11.66	11.66	15.4	13	Yes	0.603/SignalP	Present	2.1E+05	1.4E+05	5.3E+03	5.0E+03	1.44	1.2E-04
P11717	IGF2R Cation-independent mannose-6-phosphate receptor	37.71	37.71	17.8	22	Yes	0.29/SignalP	Present	9.9E+04	6.9E+04	8.7E+03	8.5E+03	1.44	1.3E-02
Q08345	DDR1 Epithelial discoidin domain-containing receptor 1	39.79	39.79	29.1	34	Yes	0.094/SignalP	Present	1.8E+06	1.3E+06	4.8E+04	2.4E+04	1.43	5.6E-05
Q9UNW1	MINPP1 Multiple inositol polyphosphate phosphatase 1	38.1	38.1	65.9	29	Yes	0.684/SignalP	Absent	4.0E+05	2.8E+05	1.9E+04	2.0E+04	1.40	2.0E-03
Q9NZ08	ERAP1 Endoplasmic reticulum aminopeptidase 1	72.12	72.12	54.7	61	Yes	0.609/SignalP	Present	6.6E+05	4.7E+05	9.5E+03	4.6E+03	1.39	7.1E-06
O00468	AGRN Agrin	179.66	179.66	54.8	161	Yes	0.293/SignalP	Present	2.8E+06	2.0E+06	6.1E+04	4.3E+04	1.38	5.8E-05
P28799	GRN Granulins	9.46	9.46	24.3	6	Yes	0.59/SignalP	Present	4.6E+04	3.4E+04	3.5E+03	2.5E+03	1.37	7.1E-03
P07942	LAMB1 Laminin subunit beta-1	40.98	40.98	22.4	35	Yes	0.352/SignalP	Present	3.5E+05	2.6E+05	9.7E+03	2.2E+04	1.36	2.4E-03
P10619	CTSA Lysosomal protective protein	10.53	10.53	18.3	7	Yes	0.909/SignalP	Present	5.6E+04	4.1E+04	1.8E+03	2.2E+03	1.36	7.8E-04
P01033	TIMP1 Metalloproteinase inhibitor 1	38.66	38.66	60.9	37	Yes	0.765/SignalP	Present	1.4E+06	1.0E+06	3.3E+04	8.4E+03	1.35	5.5E-05
Q86SQ4	GPR126 G-protein coupled receptor 126	14.38	14.38	12.5	10	Yes	0.404/SignalP	Absent	4.3E+05	3.2E+05	4.7E+03	7.5E+03	1.35	2.5E-05
P48551	IFNAR2 Interferon alpha/beta receptor 2	6.66	6.66	13.6	4	Yes	0.864/SignalP	Absent	8.6E+04	6.4E+04	2.6E+03	2.4E+03	1.34	4.4E-04
Q9Y5Y6	ST14 Suppressor of tumorigenicity 14 protein	83.2	83.2	53	61	Yes	0.678	Absent	1.4E+06	1.1E+06	4.8E+04	2.9E+04	1.32	4.5E-04
P12109	COL6A1 Collagen alpha-1(VI) chain	178.54	178.54	59.9	269	Yes	0.234/SignalP	Present	1.2E+07	8.9E+06	1.4E+05	1.5E+05	1.32	1.8E-05
Q9UBQ6	EXTL2 Exostosin-like 2	17.98	17.98	42.4	20	Yes	0.702/SignalP	Absent	5.2E+05	4.0E+05	5.6E+03	1.9E+04	1.30	4.5E-04
Q16610	ECM1 Extracellular matrix protein 1	30.77	30.77	41.9	20	Yes	0.711/SignalP	Present	4.2E+05	3.2E+05	2.8E+04	2.6E+04	1.30	1.1E-02
P35052	GPC1 Glypican-1	54.49	54.49	52.3	90	Yes	0.33/SignalP	Present	4.7E+06	6.2E+06	1.1E+05	6.0E+04	0.76	4.2E-05
Q9H9K5	ERVMER34-1 HERV-MER_4q12 provirus ancestral Env polyprotein	7.55	7.55	9.4	4	Yes	0.364/SignalP	Absent	4.1E+04	5.4E+04	2.1E+03	1.5E+03	0.76	1.0E-03
P09958	FURIN Furin	8.02	8.02	14.9	4	Yes	0.388/SignalP	Present	9.6E+03	1.3E+04	1.5E+03	4.7E+02	0.76	2.6E-02
Q16787	LAMA3 Laminin subunit alpha-3	17.5	21.67	13.2	15	Yes	0.328/SignalP	Present	3.0E+04	4.0E+04	2.2E+03	3.6E+03	0.75	1.6E-02
P10253	GAA Lysosomal alpha-glucosidase	47.06	47.06	45.6	35	Yes	0.796/SignalP	Present	2.5E+05	3.4E+05	3.4E+03	1.5E+04	0.75	6.9E-04
Q9UBR2	CTSZ Cathepsin Z	26.73	26.73	57.1	20	Yes	0.861/SignalP	Present	9.0E+04	1.2E+05	1.2E+04	4.9E+03	0.75	1.4E-02
Q9UBX7	KLK11 Kallikrein-11	6.31	6.31	23.8	4	Yes	0.665	Absent	1.4E+05	1.9E+05	6.4E+03	1.0E+04	0.74	1.9E-03
P25786	PSMA1 Proteasome subunit alpha type-1	5.87	5.87	25.9	5	Yes	0.371	Present	5.0E+04	6.7E+04	3.7E+03	2.6E+03	0.74	2.5E-03
Q5VU97	CACHD1 VWFA and cache domain-containing protein 1	45.06	45.06	35.5	28	Yes	0.247/SignalP	Absent	1.4E+05	1.8E+05	4.8E+03	3.0E+03	0.73	1.1E-04

Table 6. Continued

Uniprot Accession	Name	Unused[a]	Total[b]	% Cov[c]	Peptides(95%)[d]	Glycoprotein[e]	SecretomeP[f]	ExoCarta[g]	E1 Mean[h]	HCT-116 Mean	E1 Sigma[i]	HCT-116 Sigma	E1/HCT-116 ratio[j]	*p* -value[k]
Q92729	PTPRU Receptor-type tyrosine-protein phosphatase U	17.52	17.52	17.1	9	Yes	0.365/SignalP	Absent	7.5E+04	1.0E+05	1.2E+04	5.5E+03	0.73	2.2E-02
Q96HE7	ERO1L ERO1-like protein alpha	11.37	11.37	22	6	Yes	0.746/SignalP	Present	5.1E+04	7.0E+04	3.8E+03	4.4E+03	0.73	4.7E-03
Q8NFZ4	NLGN2 Neuroligin-2	10.97	10.97	17.4	6	Yes	0.138/SignalP	Present	5.1E+04	7.1E+04	3.7E+03	5.9E+03	0.73	9.1E-03
Q13740	ALCAM CD166 antigen	61.15	61.15	55.2	52	Yes	0.463/SignalP	Present	4.2E+05	5.8E+05	1.2E+04	2.9E+04	0.73	1.1E-03
P18065	IGFBP2 Insulin-like growth factor-binding protein 2	8.49	8.49	34.5	4	Yes	0.886/SignalP	Absent	7.6E+04	1.0E+05	2.9E+03	1.5E+04	0.73	3.2E-02
Q99650	OSMR Oncostatin-M-specific receptor subunit beta	18.76	18.76	23.7	12	Yes	0.354/SignalP	Present	8.1E+04	1.1E+05	3.8E+03	2.5E+03	0.72	2.8E-04
O43405	COCH Cochlin	48.13	48.13	51.5	30	Yes	0.655/SignalP	Absent	3.5E+05	5.0E+05	1.3E+03	2.8E+04	0.70	7.7E-04
P50895	BCAM Basal cell adhesion molecule	7.08	7.08	16.6	5	Yes	0.095/SignalP	Present	1.4E+04	2.0E+04	1.3E+03	8.2E+02	0.70	2.6E-03
P61916	NPC2 Epididymal secretory protein E1	8.03	8.03	46.4	6	Yes	0.931/SignalP	Absent	4.6E+05	6.7E+05	2.0E+04	2.0E+04	0.69	1.9E-04
O00462	MANBA Beta-mannosidase	38.49	38.49	29.8	23	Yes	0.776/SignalP	Absent	1.8E+05	2.6E+05	1.5E+04	1.3E+03	0.68	6.9E-04
Q92859	NEO1 Neogenin	47.68	47.68	33.4	30	Yes	0.246/SignalP	Absent	2.5E+05	3.7E+05	2.9E+04	3.2E+03	0.68	1.9E-03
Q7LGC8	CHST3 Carbohydrate sulfotransferase 3	8	8	27.6	4	Yes	0.811	Absent	5.2E+04	7.8E+04	4.1E+03	1.1E+04	0.67	2.1E-02
O75326	SEMA7A Semaphorin-7A	52.42	52.42	47.8	45	Yes	0.484/SignalP	Absent	6.9E+05	1.0E+06	2.0E+04	2.6E+04	0.67	5.6E-05
O00391	QSOX1 Sulfhydryl oxidase 1	146.67	146.67	60.1	156	Yes	0.611/SignalP	Present	2.6E+06	3.9E+06	3.1E+05	1.7E+05	0.67	3.1E-03
O75144	ICOSLG ICOS ligand	6	6	20.2	3	Yes	0.416/SignalP	Absent	1.2E+05	1.8E+05	4.7E+03	4.8E+02	0.67	2.4E-05
P15514	AREG Amphiregulin	32.88	32.88	31	35	Yes	0.645/SignalP	Absent	8.3E+05	1.2E+06	1.3E+04	1.2E+05	0.67	4.5E-03
Q92854	SEMA4D Semaphorin-4D	10.2	10.2	10.9	7	Yes	0.349/SignalP	Absent	1.0E+05	1.5E+05	7.7E+03	1.2E+04	0.66	3.7E-03
Q14118	DAG1 Dystroglycan	35.83	35.83	30.3	26	Yes	0.11/SignalP	Present	8.1E+05	1.2E+06	1.2E+04	3.8E+04	0.66	4.9E-05
Q07954	LRP1 Prolow-density lipoprotein receptor-related protein 1	12.01	12.01	5.2	6	Yes	0.34/SignalP	Present	4.3E+04	6.5E+04	1.6E+03	9.2E+03	0.66	1.5E-02
O95274	LYPD3 Ly6/PLAUR domain-containing protein 3	4.6	6.71	20.2	6	Yes	0.635/SignalP	Absent	3.6E+04	5.6E+04	1.7E+03	5.7E+03	0.65	4.5E-03
Q5R3F8	ELFN2 Protein phosphatase 1 regulatory subunit 29	7.49	7.49	11.5	6	Yes	0.243/SignalP	Absent	2.4E+04	3.7E+04	2.3E+03	5.0E+03	0.64	1.4E-02
P19823	ITIH2 Inter-alpha-trypsin inhibitor heavy chain H2	10.64	10.64	13.2	6	Yes	0.509/SignalP	Present	3.7E+04	5.8E+04	6.9E+03	2.0E+03	0.64	7.0E-03
O43909	EXTL3 Exostosin-like 3	19.71	19.71	28	13	Yes	0.649	Absent	5.3E+04	8.3E+04	6.0E+03	5.0E+03	0.64	2.6E-03
P23284	PPIB Peptidyl-prolyl cis-trans isomerase B	27.58	27.58	44	30	Yes	0.853/SignalP	Present	3.2E+05	5.0E+05	2.6E+03	2.8E+04	0.63	3.7E-04
O94985	CLSTN1 Calsyntenin-1	128.78	128.78	54.9	137	Yes	0.436/SignalP	Present	5.4E+06	8.6E+06	1.5E+05	7.7E+05	0.63	2.1E-03
Q96MU8	KREMEN1 Kremen protein 1	6.58	6.58	24.5	4	Yes	0.38/SignalP	Absent	8.7E+03	1.4E+04	2.3E+03	9.2E+02	0.63	2.1E-02
P11047	LAMC1 Laminin subunit gamma-1	42.98	42.98	25.1	24	Yes	0.284/SignalP	Present	7.3E+04	1.2E+05	4.2E+03	9.3E+03	0.62	1.7E-03
Q9GZX9	TWSG1 Twisted gastrulation protein homolog 1	14.9	14.9	39.9	15	Yes	0.751/SignalP	Present	2.0E+06	3.2E+06	2.9E+04	1.1E+05	0.62	5.0E-05
Q15262	PTPRK Receptor-type tyrosine-protein phosphatase kappa	42.45	42.45	25	28	Yes	0.36/SignalP	Present	4.0E+05	6.6E+05	4.1E+03	1.9E+04	0.61	2.2E-05
P05156	CFI Complement factor I	5.28	5.28	19.6	3	Yes	0.62/SignalP	Present	1.4E+04	2.3E+04	1.6E+03	4.0E+03	0.61	2.4E-02
Q9H3T3	SEMA6B Semaphorin-6B	25.71	25.72	26.8	22	Yes	0.161/SignalP	Absent	3.2E+05	5.3E+05	1.9E+04	1.8E+04	0.60	1.5E-04
P07093	SERPINE2 Glia-derived nexin	69.82	69.82	60.8	68	Yes	0.692/SignalP	Present	7.3E+05	1.2E+06	1.5E+04	2.9E+04	0.59	1.2E-05
O76061	STC2 Stanniocalcin-2	18.02	18.02	37.1	15	Yes	0.601/SignalP	Present	1.2E+05	2.1E+05	5.2E+03	1.5E+04	0.59	7.8E-04
P10909	CLU Clusterin	8.43	8.43	27.6	5	Yes	0.826/SignalP	Present	4.4E+04	7.6E+04	1.2E+03	2.7E+03	0.58	5.2E-05
P43251	BTD Biotinidase	53.06	53.06	50.8	56	Yes	0.72/SignalP	Absent	1.1E+06	1.8E+06	1.8E+04	2.4E+04	0.58	1.6E-06
P28066	PSMA5 Proteasome subunit alpha type-5	11.21	11.21	50.2	6	Yes	0.485	Present	2.1E+04	3.8E+04	1.9E+03	4.2E+03	0.55	2.9E-03
Q9UNN8	PROCR Endothelial protein C receptor	26.64	26.64	32.8	23	Yes	0.916/SignalP	Absent	8.4E+05	1.5E+06	1.8E+04	8.4E+04	0.55	1.6E-04
P24592	IGFBP6 Insulin-like growth factor-binding protein 6	10.39	10.39	62.1	10	Yes	0.675/SignalP	Present	2.5E+05	4.7E+05	2.0E+04	2.6E+04	0.54	3.2E-04
Q12913	PTPRJ Receptor-type tyrosine-protein phosphatase eta	22	22	9.5	11	Yes	0.402/SignalP	Present	1.5E+05	2.8E+05	3.2E+03	2.9E+03	0.53	7.2E-07
Q92673	SORL1 Sortilin-related receptor	27.47	27.47	20	15	Yes	0.534/SignalP	Present	5.2E+04	1.0E+05	1.7E+03	1.0E+03	0.53	2.3E-06
Q02487	DSC2 Desmocollin-2	41.7	41.7	40.8	25	Yes	0.313/SignalP	Present	3.3E+05	6.2E+05	7.5E+03	2.9E+03	0.52	3.6E-07
P07711	CTSL1 Cathepsin L1	36.77	36.77	51.4	30	Yes	0.514/SignalP	Absent	5.4E+05	1.0E+06	1.0E+04	1.6E+04	0.52	1.6E-06
Q9NZN1	IL1RAPL1 Interleukin-1 receptor accessory protein-like 1	10.56	10.56	18.1	6	Yes	0.117/SignalP	Present	2.3E+04	4.5E+04	4.7E+03	2.3E+03	0.50	1.8E-03
Q14517	FAT1 Protocadherin Fat 1	108.79	108.79	28.5	74	Yes	0.462/SignalP	Present	1.4E+05	2.7E+05	3.8E+03	1.0E+04	0.50	2.9E-05

Table 6. Continued

Uniprot Accession	Name	Unused[a]	Total[b]	% Cov[c]	Peptides(95%)[d]	Glycoprotein[e]	SecretomeP[f]	ExoCarta[g]	E1 Mean[h]	HCT-116 Mean	E1 Sigma[i]	HCT-116 Sigma	E1/HCT-116 ratio[j]	*p* -value[k]
P07225	PROS1 Vitamin K-dependent protein S	40.13	40.13	54.4	46	Yes	0.5/SignalP	Present	5.9E+05	1.2E+06	3.7E+04	6.5E+04	0.50	1.6E-04
Q86UN3	RTN4RL2 Reticulon-4 receptor-like 2	6.54	6.54	19.5	7	Yes	0.787/SignalP	Absent	6.5E+04	1.3E+05	9.2E+03	3.1E+03	0.50	3.2E-04
P14314	PRKCSH Glucosidase 2 subunit beta	22.69	22.69	29	15	Yes	0.308/SignalP	Present	2.8E+04	6.0E+04	1.3E+03	2.6E+03	0.47	4.7E-05
O94766	B3GAT3 Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 3	9.24	9.24	28.7	5	Yes	0.947/SignalP	Present	3.5E+04	7.5E+04	9.4E+02	4.6E+03	0.46	1.2E-04
Q9NPR2	SEMA4B Semaphorin-4B	27.9	27.9	35.5	22	Yes	0.447/SignalP	Present	1.2E+05	2.7E+05	6.0E+03	7.8E+03	0.45	1.3E-05
Q9UMF0	ICAM5 Intercellular adhesion molecule 5	17.52	17.52	22.9	11	Yes	0.227/SignalP	Present	9.5E+04	2.3E+05	3.5E+03	1.2E+04	0.42	5.7E-05
P31431	SDC4 Syndecan-4	9.44	9.44	19.2	7	Yes	0.557/SignalP	Present	2.2E+05	5.4E+05	1.2E+03	1.9E+04	0.41	8.7E-06
P49862	KLK7 Kallikrein-7	14.45	14.45	70	11	Yes	0.702/SignalP	Absent	7.4E+04	2.1E+05	4.3E+03	1.0E+04	0.36	3.3E-05
Q15828	CST6 Cystatin-M	4.55	4.55	34.9	3	Yes	0.944/SignalP	Absent	4.0E+04	1.2E+05	1.6E+03	5.9E+03	0.34	2.4E-05
Q96L58	B3GALT6 Beta-1,3-galactosyltransferase 6	6	6	17.9	3	Yes	0.593/SignalP	Absent	8.3E+03	2.5E+04	5.1E+03	1.4E+03	0.34	5.7E-03
O94907	DKK1 Dickkopf-related protein 1	10.38	10.38	25.2	11	Yes	0.952/SignalP	Present	1.9E+04	5.9E+04	5.7E+02	6.0E+02	0.33	1.3E-07
P16870	CPE Carboxypeptidase E	35.72	35.72	53.2	29	Yes	0.464/SignalP	Absent	4.5E+05	1.4E+06	5.2E+03	1.8E+04	0.32	9.9E-08
Q99715	COL12A1 Collagen alpha-1(XII) chain	20.15	20.15	10.9	10	Yes	0.332/SignalP	Present	2.3E+04	7.5E+04	7.7E+02	3.2E+03	0.31	1.2E-05
Q9Y4K0	LOXL2 Lysyl oxidase homolog 2	14.72	14.72	18.9	10	Yes	0.711/SignalP	Present	3.4E+04	1.1E+05	6.5E+02	4.9E+03	0.30	1.1E-05
Q13822	ENPP2 Ectonucleotide pyrophosphatase/phosphodiesterase family member 2	3.83	3.83	11.4	3	Yes	0.488/SignalP	Absent	3.6E+03	1.6E+04	9.3E+02	6.1E+03	0.22	2.4E-02
Q9BY76	ANGPTL4 Angiopoietin-related protein 4	6	6	20.7	4	Yes	0.829/SignalP	Present	3.2E+03	1.5E+04	9.4E+02	3.3E+03	0.21	3.9E-03
O14786	NRP1 Neuropilin-1	21.34	21.34	25.1	13	Yes	0.594/SignalP	Present	2.4E+04	1.8E+05	1.2E+03	6.3E+04	0.14	1.3E-02
Q16790	CA9 Carbonic anhydrase 9	8.03	8.03	26.1	5	Yes	0.804/SignalP	Absent	2.3E+03	3.5E+04	9.1E+02	1.2E+04	0.07	1.1E-02
Q9BQ16	SPOCK3 Testican-3	2.98	2.98	17	4	Yes	0.392/SignalP	Absent	5.2E+03	8.5E+04	2.2E+03	1.8E+03	0.06	1.0E-06
Q9UI42	CPA4 Carboxypeptidase A4	7.47	7.47	32.1	4	Yes	0.805/SignalP	Absent	4.4E+03	7.8E+04	3.6E+03	1.0E+04	0.06	3.3E-04
P02768	ALB Serum albumin	44.43	44.43	36.8	31	Yes	0.467/SignalP	Present	7.0E+04	2.6E+06	3.2E+03	1.1E+05	0.03	2.6E-06
P02647	APOA1 Apolipoprotein A-I	5.94	5.94	31.8	4	Yes	0.848/SignalP	Present	1.6E+03	1.5E+05	4.1E+02	1.3E+04	0.01	4.4E-05

[a] ProteinPilot Unused Score. Total score of the unique peptides detected for a protein. Used as a measure of protein confidence for the identified protein.

[b] ProteinPilot Total Score. Total score of all the peptides detected for a protein.

[c] Percentage sequence coverage of identified peptides.

[d] Number of distinct peptides identified with at least 95% confidence.

[e] Presence of glycosylation annotation in the UniProt database

[h] Secretion prediction using SecretomeP software. Results are presented in the form of "NN-score" followed by the "Signal P" annotation, which would denote secretion using the classical secretory pathway. A NN-score of > 0.5 would indicate non-classical secretion.

[g] Secretion prediction using the ExoCarta database. Proteins which are denoted as "present" have been found to be present in exosomes.

[h] Average of the total area sums from the peak signals derived from all the quantitated peptides in a protein, across three technical replicates

[i] Standard deviation for the mean calculated in Column h

[j] Ratio of average area sum from E1 samples over that from HCT-116 samples

[k] p-value for the Student's T-test comparison for significant difference between E1 average and HCT-116 average. Indication of the possibility that the differential levels of a protein observed between E1 and HCT-116 is due to chance only.

Of the 149 differentially secreted glycoproteins identified, 74 glycoproteins were oversecreted in the E1 MLAC-enriched CM as compared to that of HCT-116, while 75 were undersecreted. According to the Gene Ontology Biological Process annotation, a large portion of these proteins could be mapped to processes important in metastasis, such as adhesion and ECM organisation, signalling, as well as migration and cytoskeletal organisation (Figure 17).

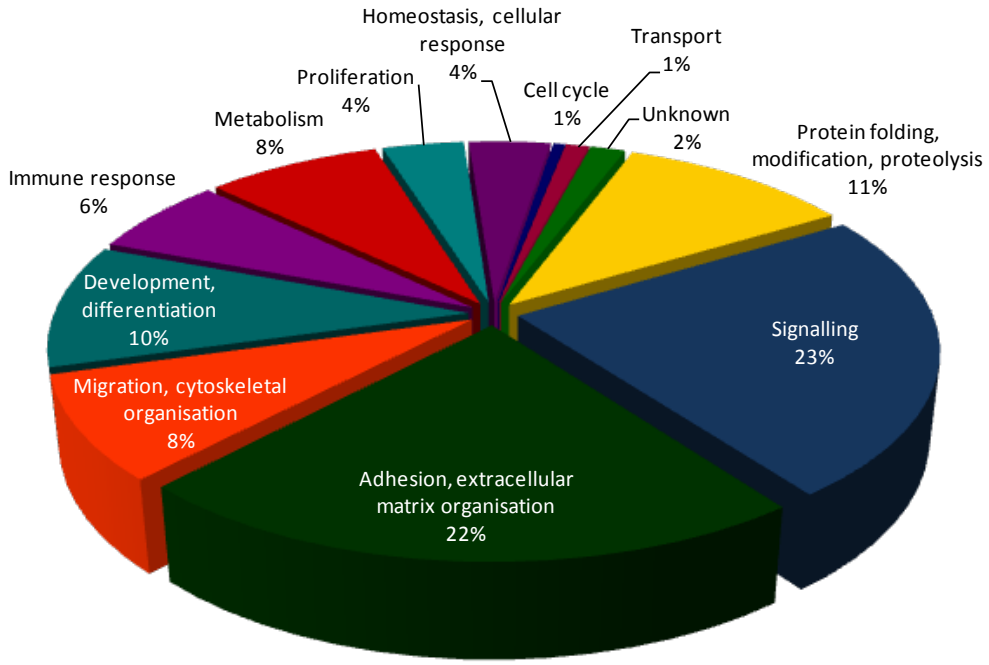


Figure 17. Gene Ontology Biological Process annotation of the 149 differentially secreted proteins in the E1 glycosecretome
Gene Ontology Biological Process annotation was performed on the 149 secreted glycoproteins which showed differential secretion levels in the SWATH-MS analysis of HCT-116 and E1 glyosecretomes.

3.3.1. Western blot validation of selected candidate glycoproteins differentially secreted in E1 cells

All of the five secreted proteins that were selected for western blot validation in the previous chapter (GDF15, SPARC, SERPINE1, PLOD3 and MAN1A1; see section 3.2.4) were also identified to be oversecreted in the E1 MLAC-enriched glycosecretome. Therefore, to further verify the differential secretion of the glycoproteins in the E1 CM as suggested by the SWATH-MS quantitative analysis, we selected two other ECM proteins, collagen type VI, α-2 (COL6A2) and laminin, β-1 (LAMB1) for validation using western blot.

The levels of COL6A2 and LAMB1 were assessed in both CM and lysate samples from HCT-116 and E1 cells using commercially available antibodies. The oversecretions of COL6A2 and LAMB1 in the E1 CM were shown to be as consistent as indicated by the SWATH-MS data (Figure 18A-B). Similar to the other differentially secreted proteins assessed by western blot in the earlier section, levels of COL6A2 and LAMB1 were too low to be detected in the whole cell lysates from both HCT-116 and E1 cells under the immunoblotting conditions. In addition, the absence of Lamin B indicated that the CM samples were relatively free of intracellular contamination (Figure 18C).

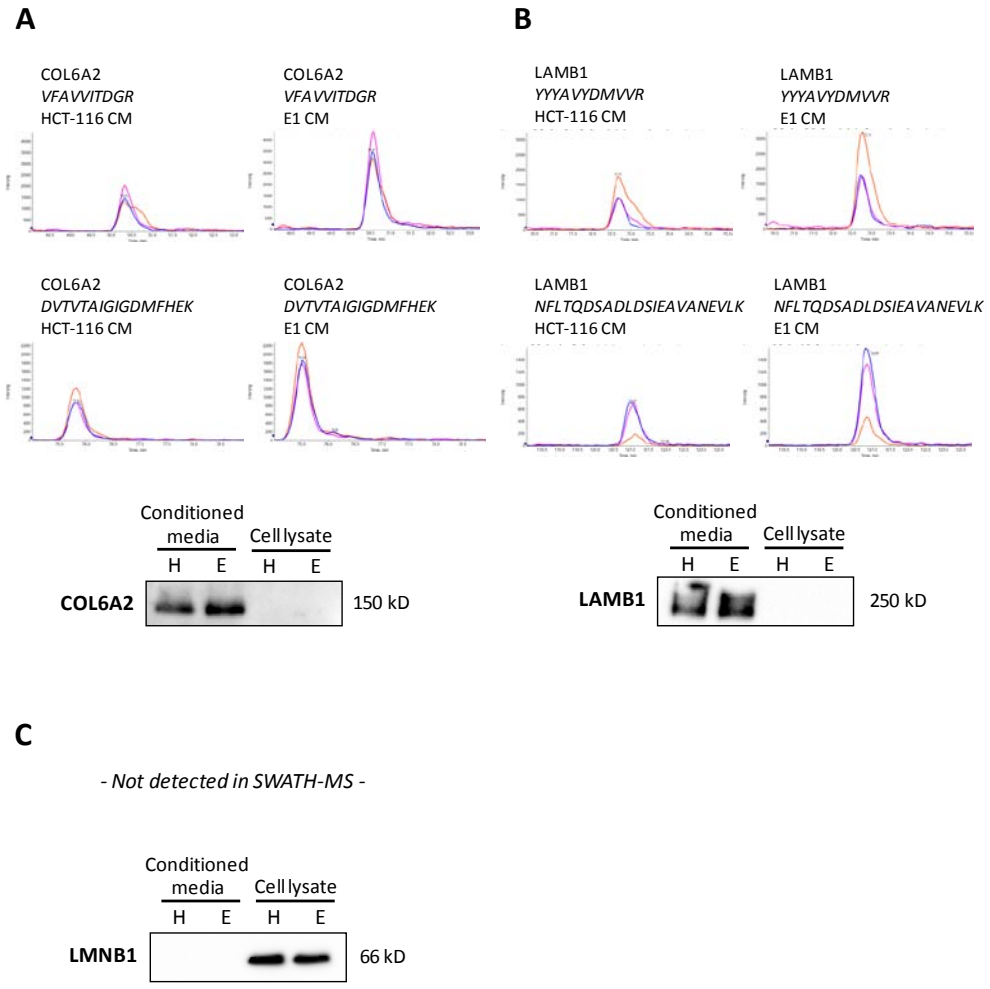


Figure 18. Representative SWATH-MS spectra and western blot analysis of COL6A2 and LAMB1 levels in HCT-116 and E1 CM samples

Differential secretion of **(A)** COL6A2 and **(B)** LAMB1, suggested by the iTRAQ data, were validated using western blot. Representative extracted ion chromatograms (XICs) from two peptides of each protein are shown here. Western blot analyses were performed on both CM and whole cell lysate samples. Samples from HCT-116 *[H]* and E1 *[E]* conditioned media and whole cell lysate samples were analysed. **(C)** Lamin B (LMNB1), a component of the nuclear lamina, was again used here as a control to confirm the absence of intracellular contamination in the CM samples.

3.3.2. LAMB1 levels were significantly higher in serum of colorectal cancer patients as compared to healthy controls

The levels of LAMB1 in the 45 colorectal cancer patient serum samples were measured to be 282.0 ± 405.1 ng/ml (mean ± SD), which were significantly higher than in the 47 healthy control samples, 28.6 ± 17.8 ng/ml (p = 1.39 x 10^{-9}, Mann-Whitney test; Figure 19A). The 45 colorectal cancer patients were also further classified according to the modified Dukes' classification (Supplementary Table 3), but there appeared to be no significant difference between LAMB1 serum levels in patients with different Dukes' staging (p = 0.27, Kruskal-Wallis ANOVA, Figure 19C). In addition, we performed a receiver operating characteristic (ROC) analysis to assess the diagnostic performance of LAMB1 for colorectal cancer. In discriminating between colorectal cancer patients and healthy controls, the area under the ROC curve (AUC) was 0.86 (95% confidence interval, 0.786 - 0.934; Figure 20A). When a cut-off value of 58.0 ng/ml was applied, LAMB1 could distinguish between colorectal cancer patients from healthy controls with a sensitivity of 64% and a specificity of 96%.

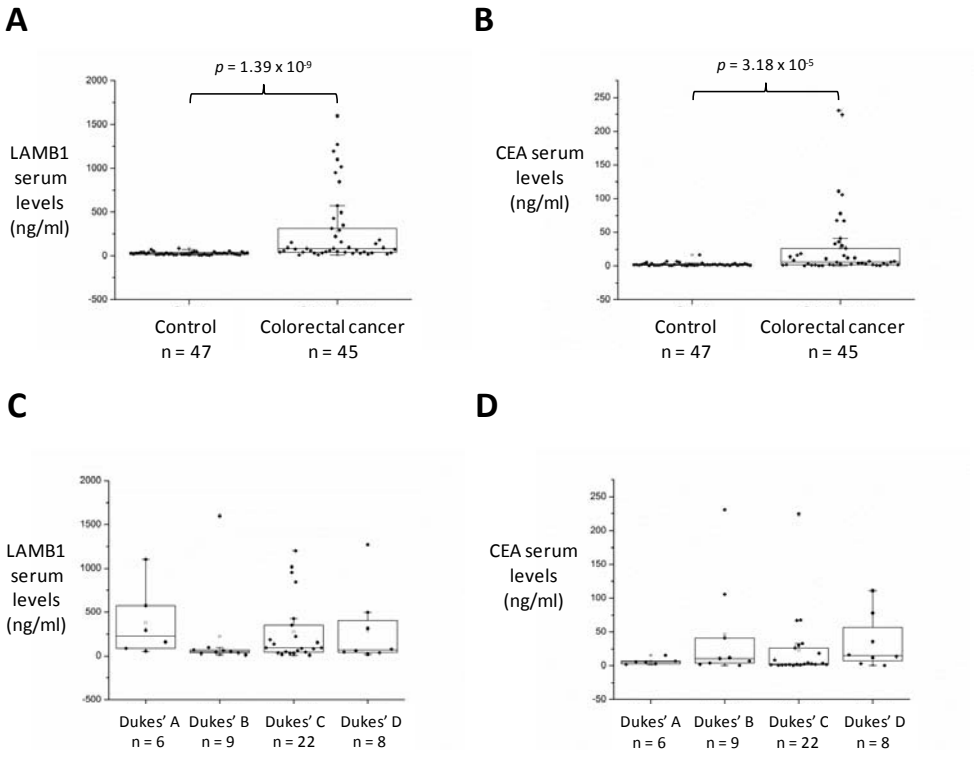
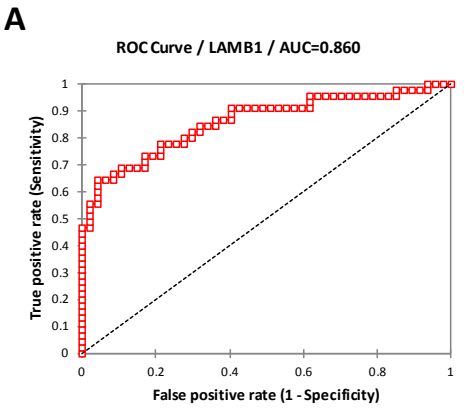


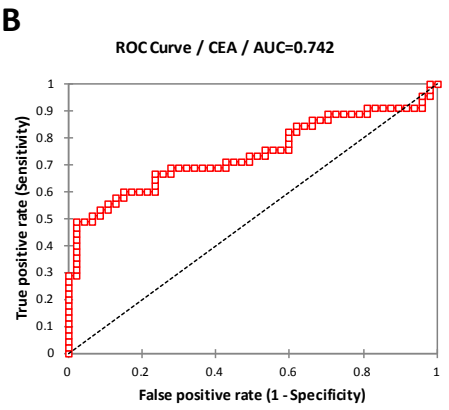
Figure 19. LAMB1 and CEA serum levels in colorectal cancer patients and healthy controls

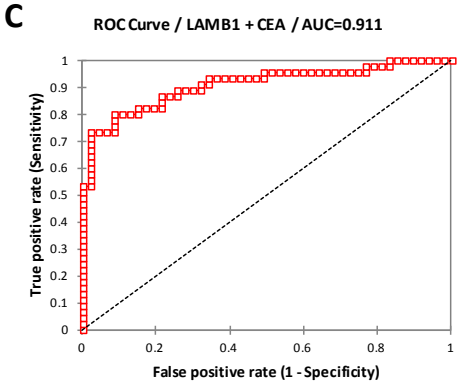
The levels of **(A)** LAMB1 and **(B)** CEA were measured in serum samples from 45 colorectal cancer patients and 47 healthy individuals using ELISA. Serum LAMB1 and CEA levels were both significantly higher in colorectal cancer patients as compared to controls. Statistical analyses were performed using the Mann-Whitney Test. **(C-D)** The colorectal cancer patients were further categorised according to the modified Dukes' classification. However, there was no significant difference in both serum LAMB1 and CEA levels between patients of different Dukes' staging. Data are represented as the upper and lower quartiles and interquartile range (box), the median (horizontal line), the median (dot in box), middle 90% distribution (dashes), and minimum and maximum values (crosses).

To assess the performance of LAMB1 against the current gold standard in the clinic, we also measured the levels of CEA in the same serum samples. Similarly, CEA levels in the serum of colorectal cancer patients were significantly higher than in the healthy controls (27.2 ± 51.5 ng/ml *vs.* 2.5 ± 2.6 ng/ml, $p = 3.18 \times 10^{-5}$; Figure 19B). In addition, there was no significant difference between CEA serum levels in patients with different Dukes' staging as well ($p = 0.30$, Figure 19D). The AUC for CEA in distinguishing between colorectal cancer patients from healthy controls was 0.74 (95% confidence interval, 0.64 - 0.85; Figure 20B), and with the typical cut-off value of 5.0 ng/ml used in the clinic (Jeon *et al.*, 2013), the sensitivity and specificity were 53% and 89% respectively.

Our results showed that LAMB1 performed better than CEA in discriminating colorectal cancer patients from healthy controls, with higher sensitivity and specificity. More importantly, when LAMB1 was used in combination with CEA, the AUC was further improved to 0.91 (95% confidence interval, 0.85 - 0.97; Figure 20C), and the sensitivity was increased to 80%, while the specificity was slightly reduced to 92% (Figure 20D).







D

	Sensitivity	Specificity
LAMB1	64%	96%
CEA	53%	89%
LAMB1 + CEA	80%	92%

Figure 20. ROC analysis of the diagnostic performance of LAMB1 in comparison to and combination with CEA.
ROC curves of **(A)** LAMB1, **(B)** CEA, and **(C)** LAMB1 in combination with CEA in discriminating colorectal cancer patients from healthy controls. **(D)** Diagnostic performance of LAMB1 (at cut-off levels of 58.0 ng/ml), as compared to that of CEA (at cut-off levels of 5.0 ng/ml).

4. Discussion

4.1. Part One. iTRAQ analysis of HCT-116 and E1 cells suggests overexpression of DBN1 during colorectal cancer liver metastasis

In order to further the understanding of the molecular mechanisms that drive colorectal cancer metastasis, we sought to identify potentially novel metastasis-related proteins by comparing the intracellular proteomes of the isogenic pair of colon cancer cell lines, HCT-116 and E1. The key advantage of using the HCT-116 and E1 cell lines as a model for studying colorectal cancer liver metastasis lies in the fact that E1 was generated from HCT-116. Karyotype analyses of the parental HCT-116 cells and metastatic-derivative E1 cells confirmed that their genetic backgrounds are largely similar (Tay *et al.*, 2010). Therefore, we postulated that any differences between E1 and HCT-116 could potentially be involved in conferring the aggressive metastatic phenotype on E1. Using an iTRAQ-based comparative analysis of the HCT-116 and E1 intracellular proteomes, a total of 31 proteins were determined to be differentially expressed in E1. Furthermore, these proteins appeared to be involved in metastasis-related processes such as gene expression, signal transduction, cell proliferation and cytoskeletal organisation (Table 3).

4.1.1. Differential expression of proteins involved in gene expression, transcription and translation

Several proteins involved in gene expression, such as protein LYRIC (MTDH) and p180/ribosome receptor (RRBP1), were observed to be overexpressed in E1 cells. MTDH is involved in diverse aspects of tumour progression and its overexpression has been observed to correlate to poor clinical outcomes in many different tumours (Hu *et al.*, 2009). Overexpression of RRBP1 has been previously observed in breast (Telikicherla *et al.*, 2012) and lung (Tsai *et al.*,

2013) cancer patient tissues, and was suggested to be involved in promoting tumour cell survival by protecting against ER stress (Tsai *et al.,* 2013).

4.1.2. Perturbation of proteins involved in signalling and regulation of apoptosis

During tumourigenesis and metastatic transformation, the regulation of apoptosis is inevitably perturbed in order to enhance survival. Therefore, we were hardly surprised to observe the overexpression of proteins such as annexin A5 (ANXA5), Lamin-A/C (LMNA) and the translationally controlled tumour protein (TCTP) in E1 cells. ANXA5 has been reported to be correlated to colorectal cancer staging and increased expression of ANXA5 has been shown to be prognostic for liver metastasis (Xue *et al.*, 2009). Expression of LMNA has been similarly found to be related to mortality in colorectal cancer and is thought to influence metastasis through up-regulation of T-plastin which consequentially down-regulates E-cadherin (Willis *et al.,* 2008).

Overexpression of TCTP is a common observation in diverse types of tumours, since TCTP is a highly conserved, ubiquitous and multi-functional survival factor. It has been shown to regulate cell growth in diverse cell types across different organisms and involved in various processes such as cytoskeletal organisation and protein synthesis (Telerman and Amson, 2009). TCTP can also influence reprogramming of somatic cell nuclei to express proteins characteristic of embryo cells by transcriptional activation of *oct4* and *nanog* genes, which are well-known in establishing pluripotency (Koziol *et al.*, 2007).

Furthermore, TCTP has also been suggested to be anti-apoptotic by interfering with BAX-induced apoptosis through its interactions with BCL-XL and MCL1 (Susini *et al.*, 2008). More interestingly, TCTP has been identified to be most strongly down-regulated in tumour reversion models generated from human leukaemia and breast cancer cells (Tuynder *et al.*, 2004).

Using western blot, we further validated that TCTP levels were indeed significantly increased in all three E1 replicates (p = 0.0125; Figure 8). Our findings corroborated those reported by Chung and colleagues (2000), who found that expression of TCTP are up-regulated at both mRNA and protein levels in colorectal cancer tissues. A recent study further revealed that knockdown of TCTP in colon cancer cell lines resulted in reduction of proliferation, migration and invasion (Ma *et al.*, 2010).

Another regulator of apoptosis identified in this study was the A-kinase anchor protein 12 (AKAP12; Gravin, AKAP250). AKAP12 was originally discovered as a minor autoantigen that correlated to poor prognosis in myasthenia gravis patients (Gordon *et al.*, 1992; Sasaki *et al.*, 2001). It is currently understood to function as a scaffolding protein and coordinates signalling and cytoskeletal reorganisation through binding of key players such as protein kinases A and C (PKA and PKC) (Gelman, 2010). There are many lines of evidence to indicate that AKAP12 is likely to be a tumour suppressor. The AKAP12 gene has been mapped to chromosome 6q24-25.2, which has been shown to be a deletion hotspot in prostate, breast and ovarian cancers (Gelman, 2002). Furthermore,

expression of AKAP12 has been reported to be down-regulated in a large variety of cancers (reviewed in Gelman, 2010).

Our iTRAQ data indicated a similar reduction in the levels of AKAP12 in E1 cells, and this was also confirmed to be significant using western blot (p = 0.0095; Figure 8). This suggests that AKAP12 could likely function likewise as a tumour suppressor in colorectal cancer metastasis. Indeed, the AKAP12 promoter region has been found to be hypermethylated in several types of colorectal cancer cell lines and primary colorectal cancer tissue samples, and the extent of methylation could be correlated with advanced Dukes' stages (Liu *et al.*, 2010a; Mori *et al.*, 2006). Moreover, when AKAP12 was re-expressed in the colorectal cancer lymph node metastatic-derived LoVo cells, apoptosis was induced and reductions in migration, invasion and anchorage-independent growth were observed (Liu *et al.*, 2011). Although the authors only showed that LoVo-cells overexpressing AKAP12 were less effective in metastasis to the lung in nude mice, our data indicates that AKAP12 might play a similar tumour suppressive role in colorectal cancer liver metastasis. As such, AKAP12 may be a promising target for novel therapeutic interventions against colorectal cancer metastasis.

4.1.3. Dysregulation of proteins involved in reorganisation of the cytoskeleton

One of the most important steps in cancer metastasis is the acquisition of enhanced motility in tumour cells through dynamic remodelling of the

cytoskeleton. Therefore, proteins that are involved in cytoskeletal reorganisation are inevitably targets of dysregulation during tumourigenesis and metastasis. Three such proteins were identified to be overexpressed in E1 cells: actinin-4 (ACTN4), stathmin-1 (STMN1) and drebrin (DBN1). ACTN4 has been previously reported to be involved in colorectal cancer lymph node metastasis (Honda *et al.,* 2005), while we (Tan *et al.,* 2012) and others (Abal *et al.,* 2007; Zheng *et al.,* 2010) have shown that overexpression of STMN1 in colorectal cancer resulted in increased metastatic capabilities in tumour cells, and can also be correlated to poorer patient prognosis.

Drebrin (developmentally regulated brain protein; DBN1) is an actin-binding protein that was originally discovered from chicken embryo brains (Shirao *et al.,* 1988). Alternative splicing results in four isoforms of DBN1: E1 and E2 (embryonic), A (adult) and s-A (truncated form of A), although only the E2 and s-A isoforms have been identified in mammals (Jin *et al.,* 2002; Kojima *et al.*, 1993; reviewed in Dun and Chilton, 2010). DBN1 was initially believed to be a neuronal-specific protein which regulates migration and differentiation of axons and dendrites during brain development. However, there have been several studies reporting the presence of DBN1 in various non-neuronal tissues, such as stomach, kidney, bladder and eccrine sweat glands (Keon *et al.*, 2000; Peitsch *et al.,* 2005; Peitsch *et al.,* 2003). There is some evidence to indicate that DBN1 may be similarly involved in regulating actin plasticity in non-neuronal cells. For example, it has been found to accumulate at the extended apical membrane of parietal cells and also able to induce a fibroblast-like phenotype in cultured epithelial cells (Keon *et al.,* 2000).

DBN1 has been reported to be up-regulated in epithelial skin tumours (Peitsch *et al.*, 2005), recurrent non-small cell lung cancer (Mitra *et al.*, 2011), and in lymphoblastic leukaemia (Vaskova *et al.*, 2011). Our results from both iTRAQ analysis and subsequent western blot validation indicated that levels of DBN1 were significantly increased in E1 cells ($p = 0.0094$, Figure 8). To the best of our knowledge, there is currently no evidence in the literature reporting the involvement of DBN1 in colorectal cancer, or in any other gastrointestinal cancers. As such, we selected DBN1 as a target for further validation in clinical colorectal cancer patient tissue samples.

4.1.4. Clinical relevance of DBN1 overexpression in colorectal cancer metastasis

We assessed the levels of DBN1 in primary colorectal cancer tissue sections against matched lymph node and liver metastases, and found that the staining intensity of DBN1 in the metastatic tissues were significantly higher than in the primary tumour (Figure 9). These results suggest that DBN1 may be potentially involved in the metastatic progression of colorectal cancer. However, the exact mechanism by which DBN1 influences metastasis remains unknown.

Interestingly, we observed some indications that DBN1 may be enriched along intercellular membranes, which was similar to reports from Peitsch and colleagues (2005), who found that DBN1 was enriched along adherens

junctions at cell-cell boundaries in epithelial tumour cells. Since DBN1 is an actin-binding protein, these results suggest that DBN1 could serve as a physical linker between the adherens junctions and the actin cytoskeleton network (Peitsch *et al.*, 2005). Cell migration is believed to involve the coupling of both cell adhesion and membrane protrusion machineries, and one of the ways this could occur is direct coupling through proteins such as vinculin (DeMali and Burridge, 2003). Therefore, DBN1 may similarly function as a mediator between cell adhesion and protrusion in order to promote migration during cancer metastasis. In support of this, there has been evidence to show that DBN1 could stabilise cellular adhesions and was co-expressed with vinculin in neuronal cells (Ikeda *et al.*, 1996). On the other hand, DBN1 has also been found to be a binding partner of the gap junction protein connexin-43 as well, indicating that DBN1 could be alternatively involved in reorganisation of the actin cytoskeleton in response to external stimuli (Butkevich *et al.*, 2004). However, the exact mechanism by which DBN1 contributes to colorectal cancer metastasis will have to be explored in future studies.

4.1.5. Summary

Using an iTRAQ-based comparison between the intracellular protein profiles of HCT-116 and E1 cells, we identified DBN1, a neuronal actin-binding protein, which was overexpressed in the metastatic-derived E1 cells. We further showed that DBN1 levels were significantly increased in lymph node and liver metastasis tissues, as compared to primary colon adenocarcinoma tissues, indicating the clinical relevance of DBN1 overexpression in colorectal cancer metastasis. Our findings represent the first time that DBN1 has been shown to be involved in colorectal cancer metastasis. Taking all the evidence in the literature in hand, DBN1 is highly likely to be an important player in the metastatic progression. Whether DBN1 may have a direct or indirect effect on cell motility in colorectal cancer remains unknown and definitely warrants further investigation.

4.2. Part Two. Comparative analysis of the HCT-116 and E1 secretomes using the hollow fibre culture (HFC) system

The cancer secretome is a highly valuable resource for identification of serological biomarkers, since secreted proteins are highly likely of entering the blood circulation. Therefore, by comparing between the CM samples collected from HCT-116 and E1 cells, we hoped to identify proteins that were differentially secreted in E1 cells, which could be translated into potential serological biomarkers for colorectal cancer prognosis and disease monitoring. By combining an iTRAQ analysis of HCT-116 and E1 CM samples, with a large-scale targeted quantitative verification using SWATH-MS acquisition, we identified a total of 25 differentially secreted proteins in the E1 secretome. Majority of these differentially secreted proteins appeared to be involved in processes such as signal transduction, adhesion and migration, as well as protein processing and post-translational modifications

4.2.1. Dysregulation of secreted proteins involved in signalling

The interaction between the tumour cells and their extracellular environment is critical in order to create favourable conditions for tumour progression and metastasis. This interaction is mediated by a variety of secreted proteins, which serve as signalling molecules to transmit information in both autocrine and paracrine fashions (Karagiannis *et al.*, 2010). Therefore, it is expected that a large majority of secreted proteins that were dysregulated in the E1 secretome would be involved in signalling. One of these proteins was a well-known oversecreted protein in colorectal cancer, known as growth/differentiation factor 15 (GDF15).

GDF15 has a plethora of roles in many physiological processes, and is also known by many other names, such as macrophage inhibitory cytokine 1 (MIC1) and non-steroidal anti-inflammatory drug (NSAID) activated gene 1 (NAG1). GDF15 belongs to the transforming growth factor-β (TGF-β) superfamily, and is involved in mediating cellular responses to stress, inflammation, tissue repair, and may also be an anti-apoptotic and potent survival factor (Mimeault and Batra, 2010). Considering these, it is hardly a surprise to find that elevation of GDF15 expression has been observed in a wide variety of cancers, including brain, lung, thyroid, breast and prostate among many others (reviewed in Mimeault and Batra, 2010).

In our iTRAQ analysis of the HCT-116 and E1 secretomes, we observed that levels of secreted GDF15 were higher in the CM of E1 cells, and verified this using both SWATH-MS and western blot (Figure 14A). Our observations were consistent with the study performed by Xue and colleagues (2010), who also observed the oversecretion of GDF15 in the CM of the lymph node metastatic SW620 cell line. Furthermore, increased levels of GDF15 in both tissue and serum of colorectal cancer patients have been observed in multiple studies, and can be correlated with disease stage, recurrence and metastasis (Barderas *et al.*, 2013; Brown *et al.*, 2003; Wallin *et al.*, 2011; Xue *et al.*, 2010).

However, current knowledge on GDF15 appears to suggest a paradoxical functional role in colorectal cancer. In contrary to the elevation of GDF15 in both patient tissue and serum samples, *in vitro* studies in colorectal cancer

cells seem to indicate that GDF15 may be tumour suppressive instead. In HCT-116 cells, elevation of GDF15 expression induced by NSAIDs resulted in apoptosis and reduced tumourigenicity in tumour xenografts (Baek *et al.*, 2001). Moreover, this induction of GDF15 expression can similarly be observed using a large variety of anti-tumourigenic dietary compounds (Bottone *et al.*, 2002; Ko and Auyeung, 2013; Shin *et al.*, 2012; Wilson *et al.*, 2003; Yang *et al.*, 2014; Zhong *et al.*, 2010). Ironically, the pro-tumourigenic activity of GDF15 is also supported by experimental evidence, albeit in other cancers. For instance, silencing of GDF15 inhibits melanoma growth in xenografts (Boyle *et al.*, 2009), while overexpression of GDF15 results in enhanced metastasis in both pancreatic and gastric cancer cells (Lee *et al.*, 2003; Senapati *et al.*, 2010). In an attempt to reconcile the contradictory data from both of the factions, Wang and colleagues (2013) suggested that the elevation of GDF15 in the serum is a response to stress signals produced as GDF15-resistant tumour cells proliferate in the developing cancer. Whether this may be true or not remains to be explored. However, regardless of the functional role that GDF15 may play, there is still very strong evidence to show that GDF15 is a highly promising serological biomarker for colorectal cancer prognosis and disease monitoring.

4.2.2. Differential secretion of proteins involved in adhesion and migration

During metastasis, the tumour cells must adhere to, interact with and degrade the various components of the basement membrane and the extracellular

matrix (ECM). In order to do so, the tumour cells would have to interfere with the release of a class of proteins known as matricellular proteins which, though non-structural by nature, are able to define and influence the composition of the ECM (Bornstein and Sage, 2002). In addition, degradation of the ECM is tightly controlled by secreted proteases, which are inevitably dysregulated as well in order for the tumour cells to invade into surrounding tissues and enter the circulation (Berger, 2002). Therefore, in this section, we will examine two such proteins that were shown to be oversecreted in the E1 secretome: secreted protein acidic and rich in cysteine (SPARC), and serpin peptidase inhibitor, clade E member 1 (SERPINE1).

SPARC, also known as osteonectin and BMP-40, is a multifunctional secreted protein that has a highly controversial role in cancer. Although SPARC mainly functions as a matricellular protein by binding to ECM proteins and regulating ECM reorganisation, it is also considered to be involved in signalling. This is because SPARC is known to be able to directly and indirectly influence several growth factor signalling pathways, such as the fibroblast growth factor (FGF), vascular endothelial growth factor (VEGF), platelet-derived growth factor (PDGF) and TGF-β pathways (reviewed in Arnold and Brekken, 2009). Therefore, since SPARC is poised to influence multiple cancer hallmarks, including angiogenesis, migration, proliferation and survival, it is easily foreseen that perturbation of SPARC expression would be frequently observed in many cancers (Arnold and Brekken, 2009; Tai and Tang, 2008).

The data from our iTRAQ analysis indicated that SPARC was oversecreted in the CM of E1 cells, and this was further confirmed using SWATH-MS and western blot (Figure 14B). This would imply that SPARC may play a role in colorectal cancer progression and metastasis. However, similar to GDF15 that was discussed earlier, SPARC also appears to have contradictory roles in cancer. In many different kinds of cancers, including cancers in the brain, breast, bladder, colon, lung, liver, pancreas and prostate, there are conflicting reports on whether SPARC is a tumour promoter or suppressor in cancer (reviewed in Arnold and Brekken, 2009). In colorectal cancer, SPARC was first reported by Porter and colleagues (1995) to exhibit positive immunoreactivity in patient tissue sections. Using techniques including microarrays, western blot, northern blot, quantitative real-time PCR and IHC, others have also similarly shown that SPARC expression in the peritumour stroma was stronger than in tumour cells, while there was usually little or no SPARC expression in normal cells or their surrounding stroma (Chew *et al.*, 2011; Kim *et al.*, 2013; Madoz-Gurpide *et al.*, 2006; Porte *et al.*, 1995; Takemasa *et al.*, 2001; Viana Lde *et al.*, 2013; Wiese *et al.*, 2007). Increased SPARC expression could also be correlated to metastasis (Porte *et al.*, 1995) and poorer patient survival (Kim *et al.*, 2013). In addition, there are some studies which have used *in vitro* or mouse models to illustrate that SPARC could be a tumour promoter. Deficiency of SPARC in mice positive for multiple intestinal neoplasia appeared to suppress adenoma formation (Sansom *et al.*, 2007), while the absence of the tumour suppressor Smad4 in SW480 cells resulted in the oversecretion of SPARC (Volmer *et al.*, 2004).

On the other hand, there are other studies reporting that elevation of SPARC levels showed either no correlation to staging (Viana Lde *et al.*, 2013), or could be associated with better disease outcome and survival (Chew *et al.*, 2011; Liang *et al.*, 2010). Lower stromal SPARC expression appeared to be associated with lymph node invasion and recurrence (Yoshimura *et al.*, 2011), and the SPARC promoter region was found to be hypermethylated in various colorectal cancer cell lines and patient tissue samples (Cheetham *et al.*, 2008; Yang *et al.*, 2007). SPARC expression was also reportedly down-regulated in chemoresistant MIP101 colon cancer cells, and the re-expression of SPARC restored chemosensitivity and led to tumour regression in xenografts (Tai *et al.*, 2005). Further studies on MIP101 cells revealed that the chemosensitivity induced by SPARC could be enhanced by vitamin D (Taghizadeh, 2007), and was dependent on p53 and caspase 8 (Chan *et al.*, 2010; Rahman *et al.*, 2011; Tang and Tai, 2007).

However, it may be important to note that the above-mentioned functional studies were mainly focused on intracellular SPARC, and that all of the studies were performed on a single cell line model. Therefore, given that it is mainly the secreted form of SPARC that is elevated during colorectal cancer progression, it is likely that secreted SPARC may play a completely different role as compared to its intracellular counterpart. Furthermore, since the oversecretion of SPARC was one of the highest that we observed in the E1 secretome, SPARC holds very high promise as a potential serological biomarker for colorectal cancer, since SPARC has already been detected in

human plasma and serum using mass spectrometry methods (Nanjappa *et al.*, 2014).

SERPINE1, or usually known as plasminogen activator inhibitor-1 (PAI-1), is part of the plasminogen activator system (PAS), which is a well-known extracellular protease system involved in invasion and metastasis (Berger, 2002). The central member of the PAS is plasmin, an extracellular serine protease which can degrade, either directly or indirectly, a broad range of substrates including fibrin, fibrinogen and other ECM proteins such as laminin and fibronectin. The precursor of active plasmin is plasminogen, which is activated by urokinase-type plasminogen activator (uPA). Moreover, the activation of plasmin can be enhanced by the binding of uPA to its receptor (uPAR). Regulation of the PAS is controlled by SERPINE1, which inhibits uPA by forming an enzymatically inactive complex with uPA and uPAR. The resulting complex would then be internalised through endocytosis (reviewed in McMahon and Kwaan, 2008).

The PAS has been implicated in a wide range of metastasis-associated processes, including migration and invasion. When the uPA-uPAR complex activates plasmin, this in turn activates matrix metalloproteases, which result in degradation of the ECM, thus releasing the cell from its adhesion site. Furthermore, since the activation of plasmin is enhanced when uPA is bound to uPAR, the location of uPAR on the cell surface would determine the direction of the migration (Burridge *et al.*, 1988; McMahon and Kwaan, 2008). On the other hand, SERPINE1 has a high affinity for an ECM protein known

as vitronection, and the association of SERPINE1 and vitronectin serves to stabilise SERPINE1 (Berger, 2002). Moreover, this interaction is also involved in the internalisation of the SERPINE1-uPA-uPAR complex through endocytosis, which promotes cell adhesion and thereby allowing for propulsion (McMahon and Kwaan, 2008; Nykjaer *et al.*, 1997).

In addition to its role in migration and invasion, SERPINE1 may be involved in promoting proliferation as well. When exogenous SERPINE1 was introduced to tumour cells *in vitro*, this resulted in inhibition of both spontaneous and induced apoptosis (Kwaan *et al.*, 2000). Given the importance of SERPINE1 in these metastasis-related processes, we were hardly surprised to observe that SERPINE1 was oversecreted in the E1 secretome in the iTRAQ analysis, and this was verified in SWATH-MS and western blot analyses (Figure 14C). This finding was consistent with previous studies in the literature, where increased SERPINE1 expression has been observed in colorectal cancer tissues as compared to controls, and can also be correlated to disease staging and metastasis (Herszenyi *et al.*, 1999; Illemann *et al.*, 2009; Markl *et al.*, 2010; Sakakibara *et al.*, 2005). SERPINE1 serum or plasma levels has also been reported to be able to distinguish between colorectal cancer patients from healthy controls, as well as discriminate for different disease staging and post-operative recurrence (Herszenyi *et al.*, 2008; Yamada *et al.*, 2010).

Although SERPINE1 may appear to be a highly promising serological biomarker for colorectal cancer detection, prognosis and monitoring, it is

important to note that the time of collection for the patient blood samples may be a potential confounding factor in analysis of SERPINE1 levels. The *SERPINE1* gene is known to be under circadian control and SERPINE1 levels in the blood have been shown to be increased in the mornings (Oishi, 2009). Recently, a study by Mazzoccoli and colleagues (2012) investigated the mRNA levels of *SERPINE1* together with *ARNTL2*, which is a circadian gene that directly activates SERPINE1 transcription, in colorectal cancer tissue samples. Their results show that both *SERPINE1* and *ARNTL2* expression were increased in colorectal cancer, and could be associated with aggressive traits such as lymph node involvement and microsatellite instability. Therefore, this is an indication that perhaps more studies are needed to examine the fluctuations of serum/plasma SERPINE1 levels in colorectal cancer patients, before SERPINE1 can be translated for clinical use as a biomarker.

4.2.3. Differential secretion of proteins involved in post-translational modifications

Many secreted proteins would undergo some form of post-translational modification during their synthesis process. This is particularly so for proteins that are secreted through the classical secretory pathway, which are often glycosylated. However, the proteins that are involved in post-translational modifications would most often be localised in the endoplasmic reticulum or the Golgi apparatus. It was therefore rather unexpected to observe the oversecretion of two such proteins, procollagen-lysine, 2-oxoglutarate 5-

dioxygenase 3 (PLOD3) and mannosidase, alpha class 1A, member 1 (MAN1A1), in the E1 secretome.

PLOD3, or more commonly known as lysyl hydroxylase 3 (LH3), is a multifunctional enzyme that possesses lysyl hydroxlase, hydroxylysl galactosyltransferase, and galactosylhydroxylysl glucosyltransferase (GGT) activities (Risteli *et al.*, 2009). In corroboration with our results, PLOD3 has been shown to be present in the CM of several cell lines (Salo *et al.*, 2006). The authors also detected GGT activity in mouse and human serum, suggesting that PLOD3 could possibly be secreted *in vivo* as well. Although the PLOD3 has been recently shown to be oversecreted in the CM of pancreatic cancer cell lines (Schiarea *et al.*, 2010), there is currently no evidence of PLOD3 oversecretion in colorectal cancer.

The data from our iTRAQ analysis indicated that PLOD3 was oversecreted in the E1 secretome, and the same trend was observed in both SWATH-MS and western blot analyses (Figure 14D). This suggested that PLOD3 may possibly have some role in tumour progression. Indeed, the glycosyltransferase activities of PLOD3 has been shown to be important for proliferation and cell viability, as cells treated with glycosyltransferase-deficient PLOD3 fragments showed arrest in cell growth and eventual lethality (Wang *et al.*, 2009). Moreover, there is also evidence indicating that PLOD3 may be involved in the deposition and remodelling of the ECM (Risteli *et al.*, 2009; Salo *et al.*, 2006), of which the latter is a crucial process during metastasis. In addition, since PLOD3 has already been detected in human serum, it could be a

promising potential serological biomarker for colorectal cancer metastasis and deserves further validation in clinical samples.

MAN1A1 belongs to the family of class I α-mannosidases, which are enzymes that cleave α-1,2-mannose residues from oligosaccharide chains on proteins (Howard *et al.*, 1997; Lal *et al.*, 1998). MAN1A1 is a type II membrane enzyme localised to the Golgi, and is responsible for processing $Man_{8-9}GlcNAc_2$ to yield $Man_5GlcNAc_2$, which is essential for the formation of complex and hybrid N-glycans (Herscovics, 1999; Kornfeld and Kornfeld, 1985). However, a secreted form of MAN1A1 is suspected to exist, as an enzyme with class 1 α-1,2-mannosidase activity has been previously detected in human serum (Porwoll *et al.*, 1999). Furthermore, MAN1A1 has also been recently detected in human plasma using mass spectrometric methods (Nanjappa *et al.*, 2014). To the best of our knowledge, there is currently no evidence of the involvement of MAN1A1 in any cancer.

The activity of MAN1A1 in modifying cell surface N-glycans has been shown to have some influence on natural killer (NK) cell-mediated lysis (Ahrens, 1993). K-562 cells treated with kifunensine, an inhibitor of class I α-mannosidases, were shown to accumulate greater amounts of high mannose-type N-linked $Man_9GlcNAc_2$ glycans on the cell surface. This was confirmed by the observation that the treated cells had increased binding to Con A, which has an affinity for high mannose-type N-linked glycans. More importantly, the accumulation of $Man_9GlcNAc_2$ glycans resulted in increased binding with NK cells and an increase in cell lysis by NK cells. Since NK cells are known to

express receptors that show sequence similarities to lectin (Giorda *et al.*, 1990; Houchins *et al.*, 1991; Yokoyama *et al.*, 1990) which can interact with N-glycans (Higai *et al.*, 2011), it may be possible that the high mannose type $Man_9GlcNAc_2$ oligosaccharides could be recognised by lectin-like receptors, which in turn can mediate NK cell-dependent cytotoxicity. More recently, kifunensine treatments on T-cells were also shown to be able to reduce the activation threshold of naive T-cells (Gebuhr *et al.*, 2011). This could indicate that the complex N-glycans generated by MAN1A1 may have some inhibitory effects on the activation of naive T-cells.

Since we found that MAN1A1 was oversecreted in the CM of E1 cells, an observation which was confirmed by both SWATH-MS and western blot (Figure 14E), this suggests that MAN1A1 may be involved in the colorectal cancer metastatic cascade. One of the most important steps in the metastatic cascade is the ability of tumour cells to evade the host immune cells during circulation. Considering that the activity of MAN1A1 has some influence on the immune response, it could be likely that MAN1A1 is secreted by metastatic cells as a means to evade immune detection in the blood circulation. Moreover, since MAN1A1 has also been detected in both human serum and plasma, it may be worthwhile to pursue the possibility of using MAN1A1 as a serological biomarker for colorectal cancer detection and prognosis.

4.2.4. Limitations in clinical validation of target proteins

One of the major bottlenecks in translating from biomarker discovery studies to the clinic is the validation of the potential biomarkers in clinical samples. Although there are many biomarker discovery studies published regularly each year, few of the reported candidates arrive to the clinic. One likely possibility could be the unavailability of suitable antibodies for clinical assays. Unfortunately, we were faced with the same predicament in our study as well. Although a number of promising and potentially novel biomarker candidates, such as PLOD3 and MAN1A1, were identified, we were unable to further validate these proteins in colorectal cancer patient samples because of the unavailability of suitable ELISA kits. To overcome this, MS-based alternatives to ELISA may be a possible option, and this will be further discussed later in the section on future work.

4.2.5. Summary

In the second part of this study, we showed that the HFC system was indeed an attractive alternative for preparation of CM samples, as it effectively concentrates secreted proteins in a smaller volume and also provides better culture conditions for the cells. Using an iTRAQ-based comparative analysis of the CM samples collected from HCT-116 and E1 cells, we observed that 38 proteins were differentially secreted in E1 cells as compared to HCT-116. In addition, we also showed that the SWATH-MS technology could be a viable option for large-scale verification of iTRAQ data. Using SWATH-MS acquisition as an approach to perform SRM-like targeted quantitative analysis

on the 38 differentially secreted proteins in E1, we found that 25 of these were corroborated by the SWATH-MS analysis. Majority of these 25 differentially secreted proteins in E1 were found to be involved in processes such as signalling, cell adhesion and migration, as well as protein PTM. We validated the oversecretion of well-known proteins in colorectal cancer, such as GDF15, SPARC and SERPINE1, and also potentially novel secreted proteins, such as PLOD3 and MAN1A1, using western blot. However, we were unable to advance any further to perform clinical validation as there were no suitable commercially available ELISA kits.

4.3. Part Three. SWATH-MS analysis of the HCT-116 and E1 MLAC-enriched glycosecretomes: identification of LAMB1 as a potential serological biomarker for colorectal cancer

In addition to the several promising candidate proteins that were identified in the previous chapter, we were also interested in delving deeper into the HCT-116 and E1 secretomes to search for more low-abundance secreted proteins. Since majority of secreted proteins would be glycosylated in some form, we utilised the multi-lectin affinity chromatography (MLAC) approach to enrich for secreted glycoproteins from HCT-116 and E1 CM samples. In addition, we explored the use of SWATH-MS as a label-free quantitative proteomics approach to compare between the HCT-116 and E1 glycosecretomes. A total of 149 glycoproteins were identified to be differentially secreted in E1 cells as compared to HCT-116, with a large portion of these proteins mapped to processes important in metastasis, such as adhesion and ECM organisation, signalling, as well as migration and cytoskeletal organisation

4.3.1. Differential secretion of glycoproteins involved in adhesion and ECM organisation

The ECM is not a mere passive bystander but is actively involved in tumour progression, including directly promoting proliferation and metastasis, as well as influencing the behaviour of stromal cells, and facilitating angiogenesis and inflammation (Lu *et al.*, 2012). The composition of the ECM can be controlled by matricellular proteins and proteases which are secreted by the tumour cells and some of which have been discussed earlier in section 4.2.2. However, the tumour cells can also directly affect the composition of the ECM by expressing and secreting ECM proteins (Sherman-Baust *et al.*, 2003). Since many ECM proteins are glycoproteins (Gu *et al.*, 2012), it was entirely

expected that ECM proteins would be among the majority of the differentially secreted proteins in the MLAC-enriched glyco-secretome of E1 as compared to HCT-116. As such, in this section, we will examine more closely two such proteins which were oversecreted in E1 cells: collagen type VI, α-2 (COL6A2) and laminin, β-1 (LAMB1).

COL6A2 is a polypeptide chain that, together with COL6A1 and COL6A3, comprises the ECM protein, collagen VI. Collagen VI is widely distributed across various tissues, and is responsible for forming a network of beaded microfilaments that interacts with other ECM proteins and provide structural support for cells. In addition to its structural role, collagen VI has also been shown to be able to affect apoptosis, autophagy, proliferation, angiogenesis, and inflammation (reviewed in Chen *et al.*, 2013). Therefore, considering the involvement of collagen VI in these processes, it is not unexpected to find that increased expression of collagen VI has been observed in several cancer types, including breast cancer (Motrescu *et al.*, 2008), ovarian cancer (Sherman-Baust *et al.*, 2003), melanomas (Burchardt *et al.*, 2003), gliobastomas (Paulus *et al.*, 1988) and juvenile angiofibromas (Gramann *et al.*, 2009).

In our SWATH-MS analysis, we observed that COL6A2 was oversecreted in E1 cells, and this was confirmed using western blot (Figure 18A). Although there is currently no evidence of COL6A2 in colorectal cancer metastasis, similar findings have already been observed in other cancers. For instance, the *COL6A2* gene has been reported to be part of a ten-gene panel that was commonly up-regulated in metastatic head and neck square cell carcinoma as

compared to lung squamous cell carcinoma (Vachani *et al.*, 2007). Similarly, COL6A2 was identified to be a part of a ten-gene signature which was associated with poor overall survival in high-grade serous ovarian cancer patients (Cheon *et al.*, 2014). Using quantitative real-time PCR, Liu and colleagues (Liu *et al.*, 2010b) also found that expression levels of COL6A2 was up-regulated in primary and metastatic gliomas. These findings indicate that COL6A2 might have a role to play in cancer metastasis. Indeed, collagen VI has been shown to be able to promote cell process adhesion and extension in glioblastoma cells (Han and Daniel, 1995), and collagen VI treatment in malignant lung epithelial carcinoma cells was also found to enhance cell motility to a considerable extent (Wright *et al.*, 2008). More importantly, collagen VI levels were reportedly increased in serum samples from patients with melanoma, although there was no significance between Stage IV and I/II patients (Burchardt *et al.*, 2003). Since there is already evidence of the presence of COL6A2 in the human plasma (Nanjappa *et al.*, 2014), it might be interesting to pursue the validation of COL6A2 as a potential biomarker in colorectal cancer metastasis.

Laminins are the most abundant non-collagenous glycoproteins which comprise the basement membrane, and LAMB1 is one of the three β-subunits which assemble together with various α- and γ-subunits to form a variety of heterotrimeric laminin isoforms (Aumailley and Smyth, 1998). More than 14 laminin isoforms have been identified, and each have different tissue distributions and functions (Givant-Horwitz *et al.*, 2005). For instance, laminin-2 has been found to be localised to the basement membrane in

muscles, while laminins-8 and -10 have been described mostly at the endothelial basement membranes (Engbring and Kleinman, 2003). Laminins have also been implicated in many cancer-related processes, including proliferation, adhesion, migration, invasion and angiogenesis, and these are typically mediated through integrin signalling (reviewed in Engbring and Kleinman, 2003; Givant-Horwitz *et al.*, 2005).

The results of our SWATH-MS analysis indicated that LAMB1 was oversecreted in the E1 secretome, and this was confirmed using western blot (Figure 18B). Our findings were similar to several reports in the literature, where LAMB1 expression has been found to be increased in hepatocellular carcinoma tissues (Lim *et al.*, 2002), and co-expressed with keratin-19, a hepatic progenitor marker that has been frequently associated with poor prognosis hepatocellular carcinomas (Govaere *et al.*, 2014). The *LAMB1* gene has also been reportedly found to be up-regulated in chemoresistant ovarian cancer cell lines (Januchowski *et al.*, 2014), and associated with ulcerative colitis, which is known to lead to increased risk of colorectal cancer (UK IBD Genetics Consortium *et al.*, 2009). Currently, only Saito and Kameoka (2005) has performed an investigation of laminin levels in the serum of colorectal cancer patients. However, although not specified in their report, the anti-laminin enzyme-immunoassay (EIA) kit used in their study appeared to be targeted against the laminin γ-1 (LAMC1) chain. Therefore, since there are currently no studies on LAMB1 in any colorectal cancer clinical samples, we decided to further validate the levels of LAMB1 in colorectal cancer patient serum samples.

4.3.2. Clinical relevance of LAMB1 elevation in colorectal cancer patient serum samples

Using a commercially available ELISA kit, we observed that the serum levels of LAMB1 were significantly higher in colorectal cancer patients as compared to healthy controls (Figure 19A). Moreover, serum LAMB1 levels could be used to discriminate colorectal cancer patients from healthy controls with a sensitivity of 64% and specificity of 96%. This performance was better than that of CEA at the typical clinical cut-off value of 5.0 ng/ml, where the sensitivity and specificity were 53% and 89% respectively. When LAMB1 and CEA were used in combination, the sensitivity was improved to 80%, although the specificity was slightly reduced to 92% (Figure 20D). This indicated that LAMB1 could be a potential serological biomarker for diagnosis of colorectal cancer.

We further analysed the LAMB1 serum levels according to the respective Dukes' classification that the 45 colorectal cancer patient samples were grouped into, but there was no significant difference between the LAMB1 levels across the different Duke's stages (Figure 19C). This was similarly observed by Saito and Kameoka (2005) as well, when they assessed for association between serum LAMC1 levels and patients with different Duke's lesions. However, the number of patients in most of the stages in our analysis may be too small to be meaningful. Furthermore, as the clinical information available to us was only restricted to the Duke's staging, we were unable to

assess if LAMB1 could be used as a biomarker for prognosis or disease monitoring. To do so, we would require more clinical data on the patients, such as whether the serum samples were collected pre- or post-surgical resection, occurrence of disease relapse and metastasis, as well as the location of the metastases. If these data could be obtained, further analysis to explore the prognostic utility of LAMB1 would definitely be warranted in any future work.

Although our findings indicated that secreted LAMB1 could have some clinical involvement in colorectal cancer metastasis, the exact mechanism behind its involvement remains unclear. The YIGSR peptide of LAMB1 has been shown to bind to the 67 kDa laminin receptor (LamR), and this interaction was found to promote cell adhesion (Massia *et al.*, 1993). Subsequent studies have further shown that the LamR may also be involved in proliferation (Satoh *et al.*, 1999), angiogenesis (Tanaka *et al.*, 2000), migration (Vande Broek *et al.*, 2001), and invasion (Mafune and Ravikumar, 1992). Furthermore, overexpression of LamR has been previously shown to be associated with colorectal cancer progression and metastasis (Sanjuan *et al.*, 1996). As such, the oversecretion of LAMB1, in conjunction with the overexpression of its receptor LamR, may play an important role in colorectal cancer metastatic progression.

4.3.3. Summary

In this third part of the study, we enriched for secreted glycoproteins in HCT-116 and E1 CM samples using the MLAC approach. Out of the 421 proteins identified for quantitative analysis, 96% of these were predicted to be secreted proteins and 70% were shown to contain glycosylations, indicating that the MLAC enrichment for secreted glycoproteins was largely successful. Using SWATH-MS as a label-free quantitative proteomics approach to compare between the MLAC-enriched glycosecretomes from HCT-116 and E1, we identified LAMB1, a component of the basement membrane laminins, to be oversecreted in E1 cells. In addition, we have shown for the first time that LAMB1 could be a potential biomarker for colorectal cancer when used in conjunction with CEA. However, due to the limited clinical information on the patient samples, we were unable to assess further whether serum LAMB1 levels could be used for disease prognosis or monitoring. In terms of functional role, the evidence in the literature appears to indicate that extracellular LAMB1 could have an influence on metastatic-related processes through its receptor, LamR. However, the reason why the secretion of LAMB1 is enhanced in metastatic colorectal cancer cells is yet unknown, and more work will be required to elucidate the significance of the LAMB1-LamR interaction in colorectal cancer metastasis.

5. Conclusion and future work

In our study, we have attempted to address some of the clinical needs in colorectal cancer management by using various proteomics approaches. In the first part, to uncover novel proteins in colorectal cancer metastasis, we performed an iTRAQ-based comparison between the intracellular proteins in the HCT-116 cell line and its liver metastatic derivative, E1. We observed that DBN1 was overexpressed in E1 cells, as compared to HCT-116, and further showed that DBN1 overexpression was clinically relevant in colorectal cancer patient tissue samples. We also noticed that DBN1 showed some indication of enrichment along cell-cell membrane boundaries, suggesting that DBN1 could be involved in physical coupling of the actin cytoskeletal network to adherens junction or gap junction complexes. Considering this, potential future studies could include investigating whether manipulations of DBN1 expression in colorectal cancer cells would affect cell motility, morphology and adhesion. In addition, we could also explore the cellular localisation of DBN1 in colorectal cancer cells using immunofluorescence and whether it co-localises with actin, vinculin or connexin-43. Subsequently, we can confirm whether DBN1 indeed binds to these proteins by identifying the interaction partners of DBN1 using co-immunoprecipitation. In addition, this would enable us to map the possible pathways of the mechanism that DBN1 contributes to colorectal cancer metastasis.

In the second part of our study, we showed that the HFC system is an appealing and viable alternative for preparation of CM samples. The HFC system facilitates CM collection by concentrating secreted proteins in a small volume, and also provides better growth conditions for the cells. Subsequently,

to identify secreted proteins that could be potential biomarkers for colorectal cancer metastasis, we performed another iTRAQ-based comparison between the CM samples collected from HCT-116 and E1 cells. Before selecting candidate proteins for validation using western blot, we further improved the confidence in the list of differentially secreted proteins by performing a large-scale targeted quantitation verification using SWATH-MS analysis. Sixty-six percent of the differentially secreted proteins identified by iTRAQ were corroborated by the SWATH-MS analysis, and out of the five proteins selected for western blot validation, the differential secretion of all five were correctly reported. These findings showed that the SWATH-MS technology could be an attractive approach for performing large-scale verification of iTRAQ data before selecting any proteins for downstream validation and analysis.

Several interesting differentially secreted proteins in E1 were identified, including well-known secreted proteins in colorectal cancer such as GDF15, SPARC and SERPINE1. Potentially novel secreted proteins involved in colorectal cancer metastasis were identified as well, such as PLOD3 and MAN1A1. However, due to unavailability of suitable commercial ELISA kits, we were unable to further validate the feasibility of PLOD3 and MAN1A1 as potential biomarkers in patient serum samples. A possible alternative to overcome this in any subsequent studies would be the development of MS-based techniques, such as SRM assays, to detect and quantify the levels of the candidate proteins in clinical samples. Advances in SRM technology have been shown to potentially be able to address the problem of antibody

availability, as demonstrated recently by Ahn and colleagues (2010), who used a targeted LC-MS/MS workflow to validate eight potential breast cancer biomarkers in patient plasma samples. Therefore, the use of SRM-based assays could become an attractive alternative for ELISAs in the foreseeable future, given its multiplexing capability and the continually improving sensitivity of the mass spectrometers.

In addition, given the importance of immune evasion during metastasis, it would be interesting to examine the function of MAN1A1 in colorectal cancer more closely. Since the N-glycan modification activity of MAN1A1 could be linked to immune response, it may be possible that MAN1A1 might be secreted by metastatic colorectal cancer cells to process high mannose-type N-glycans on the cell surface to evade immune detection. To investigate this, possible future work could include examining if E1 cells elicit a lower NK cell response as compared to HCT-116 cells, and subsequently verifying if this could be overcome by inhibition of MAN1A1 activity. It would also be interesting to compare between the cell surface glycan profiles of HCT-116 and E1 cells to observe if E1 cell surface is composed of more processed and complex N-glycans due to the activity of MAN1A1.

In the third and final part of our study, we decided to probe deeper into the HCT-116 and E1 secretomes by performing a MLAC-based enrichment for secreted glycoproteins. We also adopted the SWATH-MS technology as a label-free quantitative approach to compare between the HCT-116 and E1 glycosecretomes. We found that LAMB1 was oversecreted in E1 cells, and

further validated that serum LAMB1 levels were significantly higher in colorectal cancer patients as compared to healthy volunteers. However, due to the limited clinical information on the patient samples, we were only able to conclude at the moment that LAMB1 could be a potential screening biomarker to distinguish between colorectal cancer patients and healthy individuals.

The clinical information available to us at the moment is restricted to only the Dukes' staging of the tumour assessed at the moment of surgical resection. However, in order to assess whether LAMB1 could be used for prognosis and disease monitoring, potential future work could include a prospective study where the pre-operative and post-operative LAMB1 levels would be measured in the serum, and the patients monitored for any subsequent recurrence and/or metastasis. Other possible studies could also include exploring further into the MLAC-enriched glycosecretomes from HCT-116 and E1 cells. For instance, in our current study, we focused only on secretion level differences between proteins which are known to be glycosylated. However, as altered glycosylation patterns has been long recognised to be one of the key events in cancer progression, the MLAC enrichment may have also captured for secreted proteins which are newly-glycosylated or hyper-glycosylated during colorectal cancer metastasis. Therefore, it could be meaningful to confirm if these secreted proteins are truly glycosylated, and to identify the nature and location of the glycosylation.

6. References

Abal, M., Obrador-Hevia, A., *et al.* (2007). APC inactivation associates with abnormal mitosis completion and concomitant BUB1B/MAD2L1 up-regulation. Gastroenterology *132*, 2448-2458.

Ahn, Y., Kang, U.B., *et al.* (2010). Mining of serum glycoproteins by an indirect approach using cell line secretome. Mol Cells *29*, 123-130.

Ahrens, P.B. (1993). Role of target cell glycoproteins in sensitivity to natural killer cell lysis. J Biol Chem *268*, 385-391.

Arnold, S.A., and Brekken, R.A. (2009). SPARC: a matricellular regulator of tumorigenesis. J Cell Commun Signal *3*, 255-273.

Aumailley, M., and Smyth, N. (1998). The role of laminins in basement membrane function. J Anat *193 (Pt 1)*, 1-21.

Baek, S.J., Kim, K.S., *et al.* (2001). Cyclooxygenase inhibitors regulate the expression of a TGF-beta superfamily member that has proapoptotic and antitumorigenic activities. Mol Pharmacol *59*, 901-908.

Barderas, R., Mendes, M., *et al.* (2013). In-depth characterization of the secretome of colorectal cancer metastatic cells identifies key proteins in cell adhesion, migration, and invasion. Mol Cell Proteomics *12*, 1602-1620.

Bendtsen, J.D., Jensen, L.J., *et al.* (2004). Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng Des Sel *17*, 349-356.

Berger, D.H. (2002). Plasmin/plasminogen system in colorectal cancer. World J Surg *26*, 767-771.

Besson, D., Pavageau, A.H., *et al.* (2011). A quantitative proteomic approach of the different stages of colorectal cancer establishes OLFM4 as a new nonmetastatic tumor marker. Mol Cell Proteomics *10*, M111 009712.

Bitarte, N., Bandres, E., *et al.* (2007). Moving forward in colorectal cancer research, what proteomics has to tell. World J Gastroenterol *13*, 5813-5821.

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

Bornstein, P., and Sage, E.H. (2002). Matricellular proteins: extracellular modulators of cell function. Curr Opin Cell Biol *14*, 608-616.

Bottone, F.G., Jr., Baek, S.J., *et al.* (2002). Diallyl disulfide (DADS) induces the antitumorigenic NSAID-activated gene (NAG-1) by a p53-dependent mechanism in human colorectal HCT 116 cells. J Nutr *132*, 773-778.

Boyle, G.M., Pedley, J., *et al.* (2009). Macrophage inhibitory cytokine-1 is overexpressed in malignant melanoma and is associated with tumorigenicity. J Invest Dermatol *129*, 383-391.

Brockhausen, I., Schachter, H., *et al.* (2009). O-GalNAc Glycans. In Essentials of Glycobiology 2nd edition A. Varki, R.D. Cummings, J.D. Esko, *et al.*, eds. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. Available from: http://www.ncbi.nlm.nih.gov/books/NBK1896/

Brown, D.A., Ward, R.L., *et al.* (2003). MIC-1 serum level and genotype: associations with progress and prognosis of colorectal carcinoma. Clin Cancer Res *9*, 2642-2650.

Burchardt, E.R., Hein, R., *et al.* (2003). Laminin, hyaluronan, tenascin-C and type VI collagen levels in sera from patients with malignant melanoma. Clin Exp Dermatol *28*, 515-520.

Burridge, K., Fath, K., *et al.* (1988). Focal adhesions: transmembrane junctions between the extracellular matrix and the cytoskeleton. Annu Rev Cell Biol *4*, 487-525.

Butkevich, E., Hulsmann, S., *et al.* (2004). Drebrin is a novel connexin-43 binding partner that links gap junctions to the submembrane cytoskeleton. Curr Biol *14*, 650-658.

Cancer Genome Atlas, N. (2012). Comprehensive molecular characterization of human colon and rectal cancer. Nature *487*, 330-337.

Carbon, S., Ireland, A., *et al.* (2009). AmiGO: online access to ontology and annotation data. Bioinformatics *25*, 288-289.

Carpelan-Holmstrom, M., Haglund, C., *et al.* (1996). Pre-operative serum levels of CA 242 and CEA predict outcome in colorectal cancer. Eur J Cancer *32A*, 1156-1161.

Carriquiry, L.A., and Pineyro, A. (1999). Should carcinoembryonic antigen be used in the management of patients with colorectal cancer? Dis Colon Rectum *42*, 921-929.

Chan, J.M., Ho, S.H., *et al.* (2010). Secreted protein acidic and rich in cysteine-induced cellular senescence in colorectal cancers in response to irinotecan is mediated by P53. Carcinogenesis *31*, 812-819.

Chang, H.J., Lee, M.R., *et al.* (2007). Identification of mitochondrial FoF1-ATP synthase involved in liver metastasis of colorectal cancer. Cancer Sci *98*, 1184-1191.

Chang, Y.H., Lee, S.H., *et al.* (2012). Comparative Secretome Analyses Using a Hollow Fiber Culture System with Label-Free Quantitative Proteomics Indicates the Influence of PARK7 on Cell Proliferation and

Migration/Invasion in Lung Adenocarcinoma. J Proteome Res *11*, 5167-5185.

Chang, Y.H., Wu, C.C., *et al.* (2009). Cell secretome analysis using hollow fiber culture system leads to the discovery of CLIC1 protein as a novel plasma marker for nasopharyngeal carcinoma. J Proteome Res *8*, 5465-5474.

Cheetham, S., Tang, M.J., *et al.* (2008). SPARC promoter hypermethylation in colorectal cancers can be reversed by 5-Aza-2'deoxycytidine to increase SPARC expression and improve therapy response. Br J Cancer *98*, 1810-1819.

Chen, E.I., and Yates, J.R., 3rd (2007). Cancer proteomics by quantitative shotgun proteomics. Molecular oncology *1*, 144-159.

Chen, P., Cescon, M., *et al.* (2013). Collagen VI in cancer and its biological mechanisms. Trends Mol Med *19*, 410-417.

Chen, Y.R., Juan, H.F., *et al.* (2006). Quantitative proteomic and genomic profiling reveals metastasis-related protein expression patterns in gastric cancer cells. J Proteome Res *5*, 2727-2742.

Cheon, D.J., Tong, Y., *et al.* (2014). A collagen-remodeling gene signature regulated by TGF-beta signaling is associated with metastasis and poor survival in serous ovarian cancer. Clin Cancer Res *20*, 711-723.

Chew, A., Salama, P., *et al.* (2011). SPARC, FOXP3, CD8 and CD45 correlation with disease recurrence and long-term disease-free survival in colorectal cancer. PLoS One *6*, e22047.

Chiu, K.H., Chang, Y.H., *et al.* (2013). Secretome analysis using a hollow fiber culture system for cancer biomarker discovery. Biochim Biophys Acta *1834*, 2285-2292.

Chung, S., Kim, M., *et al.* (2000). Expression of translationally controlled tumor protein mRNA in human colon cancer. Cancer Lett *156*, 185-190.

Colzani, M., Waridel, P., *et al.* (2009). Metabolic labeling and protein linearization technology allow the study of proteins secreted by cultured cells in serum-containing media. J Proteome Res *8*, 4779-4788.

Cooper, S. (2003). Reappraisal of serum starvation, the restriction point, G0, and G1 phase arrest points. FASEB J *17*, 333-340.

de Jong, M.C., Pulitano, C., *et al.* (2009). Rates and patterns of recurrence following curative intent surgery for colorectal liver metastasis: an international multi-institutional analysis of 1669 patients. Ann Surg *250*, 440-448.

de Wit, M., Fijneman, R.J., *et al.* (2013). Proteomics in colorectal cancer translational research: biomarker discovery for clinical applications. Clin Biochem *46*, 466-479.

Deans, G.T., Parks, T.G., *et al.* (1992). Prognostic factors in colorectal cancer. Br J Surg *79*, 608-613.

DeMali, K.A., and Burridge, K. (2003). Coupling membrane protrusion and cell adhesion. J Cell Sci *116*, 2389-2397.

Diehl, H.C., Stuhler, K., *et al.* (2007). A catalogue of proteins released by colorectal cancer cells in vitro as an alternative source for biomarker discovery. Proteomics Clin Appl *1*, 47-61.

Dukes, C.E. (1932). The classification of the cancer of the rectum. J Pathol Bacteriol *35*, 323-332.

Dun, X.P., and Chilton, J.K. (2010). Control of cell shape and plasticity during development and disease by the actin-binding protein Drebrin. Histol Histopathol *25*, 533-540.

Eichelbaum, K., Winter, M., *et al.* (2012). Selective enrichment of newly synthesized proteins for quantitative secretome analysis. Nat Biotechnol *30*, 984-990.

Engbring, J.A., and Kleinman, H.K. (2003). The basement membrane matrix in malignancy. J Pathol *200*, 465-470.

Fan, N.J., Gao, C.F., *et al.* (2012). Discovery and verification of gelsolin as a potential biomarker of colorectal adenocarcinoma in the Chinese population: Examining differential protein expression using an iTRAQ labelling-based proteomics approach. Can J Gastroenterol *26*, 41-47.

Fanayan, S., Hincapie, M., *et al.* (2012). Using lectins to harvest the plasma/serum glycoproteome. Electrophoresis *33*, 1746-1754.

Gabriel, W.B., Dukes, C., *et al.* (1932). Lymphatic spread in cancer of the rectum. Br J Surg *23*, 395-413.

Gebuhr, I., Keeren, K., *et al.* (2011). Differential expression and function of alpha-mannosidase I in stimulated naive and memory CD4+ T cells. J Immunother *34*, 428-437.

Gelman, I.H. (2002). The role of SSeCKS/gravin/AKAP12 scaffolding proteins in the spaciotemporal control of signaling pathways in oncogenesis and development. Front Biosci *7*, d1782-1797.

Gelman, I.H. (2010). Emerging Roles for SSeCKS/Gravin/AKAP12 in the Control of Cell Proliferation, Cancer Malignancy, and Barriergenesis. Genes Cancer *1*, 1147-1156.

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

Gillet, L.C., Navarro, P., *et al.* (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol Cell Proteomics *11*, O111 016717.

Giorda, R., Rudert, W.A., *et al.* (1990). NKR-P1, a signal transduction molecule on natural killer cells. Science *249*, 1298-1300.

Givant-Horwitz, V., Davidson, B., *et al.* (2005). Laminin-induced signaling in tumor cells. Cancer Lett *223*, 1-10.

Gordon, T., Grove, B., *et al.* (1992). Molecular cloning and preliminary characterization of a novel cytoplasmic antigen recognized by myasthenia gravis sera. J Clin Invest *90*, 992-999.

Govaere, O., Komuta, M., *et al.* (2014). Keratin 19: a key role player in the invasion of human hepatocellular carcinomas. Gut *63*, 674-685.

Gramann, M., Wendler, O., *et al.* (2009). Prominent collagen type VI expression in juvenile angiofibromas. Histochem Cell Biol *131*, 155-164.

Gu, J., Isaji, T., *et al.* (2012). Potential roles of N-glycosylation in cell adhesion. Glycoconj J *29*, 599-607.

Han, J., and Daniel, J.C. (1995). Biosynthesis of type VI collagen by glioblastoma cells and possible function in cell invasion of three-dimensional matrices. Connect Tissue Res *31*, 161-170.

Hardt, M., Witkowska, H.E., *et al.* (2005). Assessing the effects of diurnal variation on the composition of human parotid saliva: quantitative analysis of native peptides using iTRAQ reagents. Anal Chem *77*, 4947-4954.

Harrison, L.E., Guillem, J.G., *et al.* (1997). Preoperative carcinoembryonic antigen predicts outcomes in node-negative colon cancer patients: a multivariate analysis of 572 patients. J Am Coll Surg *185*, 55-59.

Hasan, N.M., Adams, G.E., *et al.* (1999). Effect of serum starvation on expression and phosphorylation of PKC-alpha and p53 in V79 cells: implications for cell death. Int J Cancer *80*, 400-405.

Heo, S.H., Lee, S.J., *et al.* (2007). Identification of putative serum glycoprotein biomarkers for human lung adenocarcinoma by multilectin affinity chromatography and LC-MS/MS. Proteomics *7*, 4292-4302.

Herscovics, A. (1999). Importance of glycosidases in mammalian glycoprotein biosynthesis. Biochim Biophys Acta *1473*, 96-107.

Herszenyi, L., Farinati, F., *et al.* (2008). Tumor marker utility and prognostic relevance of cathepsin B, cathepsin L, urokinase-type plasminogen activator, plasminogen activator inhibitor type-1, CEA and CA 19-9 in colorectal cancer. BMC Cancer *8*, 194.

Herszenyi, L., Plebani, M., *et al.* (1999). The role of cysteine and serine proteases in colorectal carcinoma. Cancer *86*, 1135-1142.

Higai, K., Suzuki, C., *et al.* (2011). Binding affinities of NKG2D and CD94 to sialyl Lewis X-expressing N-glycans and heparin. Biol Pharm Bull *34*, 8-12.

Honda, K., Yamada, T., *et al.* (2005). Actinin-4 increases cell motility and promotes lymph node metastasis of colorectal cancer. Gastroenterology *128*, 51-62.

Houchins, J.P., Yabe, T., *et al.* (1991). DNA sequence analysis of NKG2, a family of related cDNA clones encoding type II integral membrane proteins on human natural killer cells. J Exp Med *173*, 1017-1020.

Howard, S., Braun, C., *et al.* (1997). Human lysosomal and jack bean alpha-mannosidases are retaining glycosidases. Biochem Biophys Res Commun *238*, 896-898.

Hu, G., Wei, Y., *et al.* (2009). The multifaceted role of MTDH/AEG-1 in cancer progression. Clin Cancer Res *15*, 5615-5620.

Hurwitz, H., Fehrenbacher, L., *et al.* (2004). Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. N Engl J Med *350*, 2335-2342.

Ikeda, K., Kaub, P.A., *et al.* (1996). Stabilization of adhesion plaques by the expression of drebrin A in fibroblasts. Brain Res Dev Brain Res *91*, 227-236.

Illemann, M., Bird, N., *et al.* (2009). Two distinct expression patterns of urokinase, urokinase receptor and plasminogen activator inhibitor-1 in colon cancer liver metastases. Int J Cancer *124*, 1860-1870.

Januchowski, R., Zawierucha, P., *et al.* (2014). Extracellular matrix proteins expression profiling in chemoresistant variants of the a2780 ovarian cancer cell line. Biomed Res Int *2014*, 365867.

Jemal, A., Bray, F., *et al.* (2011). Global cancer statistics. CA Cancer J Clin *61*, 69-90.

Jeon, B.G., Shin, R., *et al.* (2013). Individualized Cutoff Value of the Preoperative Carcinoembryonic Antigen Level is Necessary for Optimal Use as a Prognostic Marker. Ann Coloproctol *29*, 106-114.

Jin, M., Tanaka, S., *et al.* (2002). A novel, brain-specific mouse drebrin: cDNA cloning, chromosomal mapping, genomic structure, expression, and functional characterization. Genomics *79*, 686-692.

Jones, S., Chen, W.D., *et al.* (2008). Comparative lesion sequencing provides insights into tumor evolution. Proc Natl Acad Sci U S A *105*, 4283-4288.

Jonker, D.J., O'Callaghan, C.J., *et al.* (2007). Cetuximab for the treatment of colorectal cancer. N Engl J Med *357*, 2040-2048.

Karagiannis, G.S., Pavlou, M.P., *et al.* (2010). Cancer secretomics reveal pathophysiological pathways in cancer molecular oncology. Molecular oncology *4*, 496-510.

Keon, B.H., Jedrzejewski, P.T., *et al.* (2000). Isoform specific expression of the neuronal F-actin binding protein, drebrin, in specialized cells of stomach and kidney epithelia. J Cell Sci *113 Pt 2*, 325-336.

Kim, J.Y., Jeong, D., *et al.* (2013). Expression of Secreted Protein Acidic and Rich in Cysteine in the Stroma of a Colorectal Carcinoma is Associated With Patient Prognosis. Ann Coloproctol *29*, 93-99.

Kinzler, K.W., and Vogelstein, B. (1996). Lessons from hereditary colorectal cancer. Cell *87*, 159-170.

Ko, J.K., and Auyeung, K.K. (2013). Target-oriented mechanisms of novel herbal therapeutics in the chemotherapy of gastrointestinal cancer and inflammation. Curr Pharm Des *19*, 48-66.

Kojima, N., Shirao, T., *et al.* (1993). Molecular cloning of a developmentally regulated brain protein, chicken drebrin A and its expression by alternative splicing of the drebrin gene. Brain Res Mol Brain Res *19*, 101-114.

Kornfeld, R., and Kornfeld, S. (1985). Assembly of asparagine-linked oligosaccharides. Annu Rev Biochem *54*, 631-664.

Koziol, M.J., Garrett, N., *et al.* (2007). Tpt1 activates transcription of oct4 and nanog in transplanted somatic nuclei. Curr Biol *17*, 801-807.

Kwaan, H.C., Wang, J., *et al.* (2000). Plasminogen activator inhibitor 1 may promote tumour growth through inhibition of apoptosis. Br J Cancer *82*, 1702-1708.

Labianca, R., Nordlinger, B., *et al.* (2013). Early colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol *24 Suppl 6*, vi64-72.

Lal, A., Pang, P., *et al.* (1998). Substrate specificities of recombinant murine Golgi alpha1, 2-mannosidases IA and IB and comparison with endoplasmic

reticulum and Golgi processing alpha1,2-mannosidases. Glycobiology *8*, 981-995.

Lee, D.H., Yang, Y., *et al.* (2003). Macrophage inhibitory cytokine-1 induces the invasiveness of gastric cancer cells by up-regulating the urokinase-type plasminogen activator system. Cancer Res *63*, 4648-4655.

Lee, H.P., Chew, L., *et al.* (2014). Singapore Cancer Registry Interim Annual Registry Report. Trends in Cancer Incidence in Singapore 2009-2013. National Registry of Diseases Office, https://www.nrdo.gov.sg/uploadedFiles/NRDO/Publications/Cancer%20Trends%20Report%200913%200920FINAL.pdf.

Leufkens, A.M., van Oijen, M.G., *et al.* (2012). Factors influencing the miss rate of polyps in a back-to-back colonoscopy study. Endoscopy *44*, 470-475.

Liang, J.F., Wang, H.K., *et al.* (2010). Relationship and prognostic significance of SPARC and VEGF protein expression in colon cancer. J Exp Clin Cancer Res *29*, 71.

Lim, S.O., Park, S.J., *et al.* (2002). Proteome analysis of hepatocellular carcinoma. Biochem Biophys Res Commun *291*, 1031-1037.

Lin, Q., Tan, H. T., *et al.* (2014). iTRAQ analysis of colorectal cancer cell lines suggests Drebrin (DBN1) is overexpressed during liver metastasis. *Proteomics*, 14: 1434-1443.

Liu, H., Sadygov, R.G., *et al.* (2004). A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem *76*, 4193-4201.

Liu, L., Zhao, L., *et al.* (2007). Proteomic analysis of Tiam1-mediated metastasis in colorectal cancer. Cell Biol Int *31*, 805-814.

Liu, W., Guan, M., *et al.* (2011). Re-expression of AKAP12 inhibits progression and metastasis potential of colorectal carcinoma in vivo and in vitro. PLoS One *6*, e24015.

Liu, W., Guan, M., *et al.* (2010a). Quantitative assessment of AKAP12 promoter methylation in colorectal cancer using methylation-sensitive high resolution melting: Correlation with Duke's stage. Cancer Biol Ther *9*, 862-871.

Liu, Y., Carson-Walter, E.B., *et al.* (2010b). Vascular gene expression patterns are conserved in primary and metastatic brain tumors. J Neurooncol *99*, 13-24.

Liu, Y., Huttenhain, R., *et al.* (2013). Quantitative measurements of N-linked glycoproteins in human plasma by SWATH-MS. Proteomics *13*, 1247-1256.

Loei, H., Tan, H.T., *et al.* (2012). Mining the gastric cancer secretome: identification of GRN as a potential diagnostic marker for early gastric cancer. J Proteome Res *11*, 1759-1772.

Longley, D.B., Harkin, D.P., *et al.* (2003). 5-fluorouracil: mechanisms of action and clinical strategies. Nat Rev Cancer *3*, 330-338.

Lu, P., Weaver, V.M., *et al.* (2012). The extracellular matrix: a dynamic niche in cancer progression. J Cell Biol *196*, 395-406.

Lundquist, J.J., and Toone, E.J. (2002). The cluster glycoside effect. Chem Rev *102*, 555-578.

Ma, Q., Geng, Y., *et al.* (2010). The role of translationally controlled tumor protein in tumor growth and metastasis of colon adenocarcinoma cells. J Proteome Res *9*, 40-49.

Ma, Y.L., Peng, J.Y., *et al.* (2009). Heterogeneous nuclear ribonucleoprotein A1 is identified as a potential biomarker for colorectal cancer based on differential proteomics technology. J Proteome Res *8*, 4525-4535.

Madoz-Gurpide, J., Lopez-Serra, P., *et al.* (2006). Proteomics-based validation of genomic data: applications in colorectal cancer diagnosis. Mol Cell Proteomics *5*, 1471-1483.

Mafune, K., and Ravikumar, T.S. (1992). Anti-sense RNA of 32-kDa laminin-binding protein inhibits attachment and invasion of a human colon carcinoma cell line. J Surg Res *52*, 340-346.

Makridakis, M., and Vlahou, A. (2010). Secretome proteomics for discovery of cancer biomarkers. Journal of proteomics *73*, 2291-2305.

Mallawaaratchy, D.M., Mactier, S., *et al.* (2012). The phosphoinositide 3-kinase inhibitor LY294002, decreases aminoacyl-tRNA synthetases, chaperones and glycolytic enzymes in human HT-29 colorectal cancer cells. J Proteomics *75*, 1590-1599.

Markl, B., Renk, I., *et al.* (2010). Tumour budding, uPA and PAI-1 are associated with aggressive behaviour in colon cancer. J Surg Oncol *102*, 235-241.

Massia, S.P., Rao, S.S., *et al.* (1993). Covalently immobilized laminin peptide Tyr-Ile-Gly-Ser-Arg (YIGSR) supports cell spreading and co-localization of the 67-kilodalton laminin receptor with alpha-actinin and vinculin. J Biol Chem *268*, 8053-8059.

Mathivanan, S., Fahner, C.J., *et al.* (2012). ExoCarta 2012: database of exosomal proteins, RNA and lipids. Nucleic Acids Res *40*, D1241-1244.

Mazzoccoli, G., Pazienza, V., *et al.* (2012). ARNTL2 and SERPINE1: potential biomarkers for tumor aggressiveness in colorectal cancer. J Cancer Res Clin Oncol *138*, 501-511.

Mbeunkui, F., Fodstad, O., *et al.* (2006). Secretory protein enrichment and analysis: an optimized approach applied on cancer cell lines using 2D LC-MS/MS. J Proteome Res *5*, 899-906.

McMahon, B., and Kwaan, H.C. (2008). The plasminogen activator system and cancer. Pathophysiol Haemost Thromb *36*, 184-194.

Mimeault, M., and Batra, S.K. (2010). Divergent molecular mechanisms underlying the pleiotropic functions of macrophage inhibitory cytokine-1 in cancer. J Cell Physiol *224*, 626-635.

Mitra, R., Lee, J., *et al.* (2011). Prediction of postoperative recurrence-free survival in non-small cell lung cancer by using an internationally validated gene expression model. Clin Cancer Res *17*, 2934-2946.

Moertel, C.G., O'Fallon, J.R., *et al.* (1986). The preoperative carcinoembryonic antigen test in the diagnosis, staging, and prognosis of colorectal cancer. Cancer *58*, 603-610.

Mori, Y., Cai, K., *et al.* (2006). A genome-wide search identifies epigenetic silencing of somatostatin, tachykinin-1, and 5 other genes in colon cancer. Gastroenterology *131*, 797-808.

Motrescu, E.R., Blaise, S., *et al.* (2008). Matrix metalloproteinase-11/stromelysin-3 exhibits collagenolytic function against collagen VI under normal and malignant conditions. Oncogene *27*, 6347-6355.

Na, K., Lee, E.Y., *et al.* (2009). Human plasma carboxylesterase 1, a novel serologic biomarker candidate for hepatocellular carcinoma. Proteomics *9*, 3989-3999.

Nanjappa, V., Thomas, J.K., *et al.* (2014). Plasma Proteome Database as a resource for proteomics research: 2014 update. Nucleic Acids Res *42*, D959-965.

Neilson, K.A., Ali, N.A., *et al.* (2011). Less label, more free: approaches in label-free quantitative mass spectrometry. Proteomics *11*, 535-553.

Nickel, W. (2003). The mystery of nonclassical protein secretion. A current view on cargo proteins and potential export routes. Eur J Biochem *270*, 2109-2119.

Nykjaer, A., Conese, M., *et al.* (1997). Recycling of the urokinase receptor upon internalization of the uPA:serpin complexes. EMBO J *16*, 2610-2620.

O'Connell, J.B., Maggard, M.A., *et al.* (2004). Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. J Natl Cancer Inst *96*, 1420-1425.

O'Farrell, P.H. (1975). High resolution two-dimensional electrophoresis of proteins. J Biol Chem *250*, 4007-4021.

Oishi, K. (2009). Plasminogen activator inhibitor-1 and the circadian clock in metabolic disorders. Clin Exp Hypertens *31*, 208-219.

Ong, S.E., Blagoev, B., *et al.* (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics *1*, 376-386.

Pahlsson, P., Blackall, D.P., *et al.* (1994). Biochemical characterization of the O-glycans on recombinant glycophorin A expressed in Chinese hamster ovary cells. Glycoconj J *11*, 43-50.

Pan, S., Chen, R., *et al.* (2011). Mass spectrometry based glycoproteomics--from a proteomics perspective. Mol Cell Proteomics *10*, R110 003251.

Paschos, K.A., and Bird, N. (2008). Current diagnostic and therapeutic approaches for colorectal cancer liver metastasis. Hippokratia *12*, 132-138.

Patel, V.J., Thalassinos, K., *et al.* (2009). A comparison of labeling and label-free mass spectrometry-based proteomics approaches. J Proteome Res *8*, 3752-3759.

Paulus, W., Roggendorf, W., *et al.* (1988). Immunohistochemical investigation of collagen subtypes in human glioblastomas. Virchows Arch A Pathol Anat Histopathol *413*, 325-332.

Pei, H., Zhu, H., *et al.* (2007). Proteome analysis and tissue microarray for profiling protein markers associated with lymph node metastasis in colorectal cancer. J Proteome Res *6*, 2495-2501.

Peitsch, W.K., Hofmann, I., *et al.* (2005). Drebrin, an actin-binding, cell-type characteristic protein: induction and localization in epithelial skin tumors and cultured keratinocytes. J Invest Dermatol *125*, 761-774.

Peitsch, W.K., Hofmann, I., *et al.* (2003). Cell biological and biochemical characterization of drebrin complexes in mesangial cells and podocytes of renal glomeruli. J Am Soc Nephrol *14*, 1452-1463.

Pino, M.S., and Chung, D.C. (2010). The chromosomal instability pathway in colon cancer. Gastroenterology *138*, 2059-2072.

Podwojski, K., Eisenacher, M., *et al.* (2010). Peek a peak: a glance at statistics for quantitative label-free proteomics. Expert review of proteomics *7*, 249-261.

Porte, H., Chastre, E., *et al.* (1995). Neoplastic progression of human colorectal cancer is associated with overexpression of the stromelysin-3 and BM-40/SPARC genes. Int J Cancer *64*, 70-75.

Porter, P.L., Sage, E.H., *et al.* (1995). Distribution of SPARC in normal and neoplastic human tissue. J Histochem Cytochem *43*, 791-800.

Porwoll, S., Fuchs, H., *et al.* (1999). Characterization of a soluble class I alpha-mannosidase in human serum. FEBS Lett *449*, 175-178.

Qi, Y.J., Ward, D.G., *et al.* (2014). Proteomic profiling of N-linked glycoproteins identifies ConA-binding procathepsin D as a novel serum biomarker for hepatocellular carcinoma. Proteomics *14*, 186-195.

Rahman, M., Chan, A.P., *et al.* (2011). A peptide of SPARC interferes with the interaction between caspase8 and Bcl2 to resensitize chemoresistant tumors and enhance their regression in vivo. PLoS One *6*, e26390.

Righetti, P.G., Castagna, A., *et al.* (2004). Critical survey of quantitative proteomics in two-dimensional electrophoretic approaches. J Chromatogr A *1051*, 3-17.

Risteli, M., Ruotsalainen, H., *et al.* (2009). Reduction of lysyl hydroxylase 3 causes deleterious changes in the deposition and organization of extracellular matrix. J Biol Chem *284*, 28204-28211.

Ross, P.L., Huang, Y.N., *et al.* (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics *3*, 1154-1169.

Saito, N., and Kameoka, S. (2005). Serum laminin is an independent prognostic factor in colorectal cancer. Int J Colorectal Dis *20*, 238-244.

Sakakibara, T., Hibi, K., *et al.* (2005). Plasminogen activator inhibitor-1 as a potential marker for the malignancy of colorectal cancer. Br J Cancer *93*, 799-803.

Salo, A.M., Wang, C., *et al.* (2006). Lysyl hydroxylase 3 (LH3) modifies proteins in the extracellular space, a novel mechanism for matrix remodeling. J Cell Physiol *207*, 644-653.

Sanjuan, X., Fernandez, P.L., *et al.* (1996). Overexpression of the 67-kD laminin receptor correlates with tumour progression in human colorectal carcinoma. J Pathol *179*, 376-380.

Sansom, O.J., Mansergh, F.C., *et al.* (2007). Deficiency of SPARC suppresses intestinal tumorigenesis in APCMin/+ mice. Gut *56*, 1410-1414.

Sasaki, H., Kunimatsu, M., *et al.* (2001). Autoantibody to gravin is expressed more strongly in younger and nonthymomatous patients with myasthenia gravis. Surg Today *31*, 1036-1037.

Satoh, K., Narumi, K., *et al.* (1999). Diminution of 37-kDa laminin binding protein expression reduces tumour formation of murine lung cancer cells. Br J Cancer *80*, 1115-1122.

Schaaij-Visser, T.B., de Wit, M., *et al.* (2013). The cancer secretome, current status and opportunities in the lung, breast and colorectal cancer context. Biochim Biophys Acta *1834*, 2242-2258.

Schiarea, S., Solinas, G., *et al.* (2010). Secretome analysis of multiple pancreatic cancer cell lines reveals perturbations of key functional networks. J Proteome Res *9*, 4376-4392.

Senapati, S., Rachagani, S., *et al.* (2010). Overexpression of macrophage inhibitory cytokine-1 induces metastasis of human prostate cancer cells through the FAK-RhoA signaling pathway. Oncogene *29*, 1293-1302.

Sherman-Baust, C.A., Weeraratna, A.T., *et al.* (2003). Remodeling of the extracellular matrix through overexpression of collagen VI contributes to cisplatin resistance in ovarian cancer cells. Cancer Cell *3*, 377-386.

Shin, J.S., Hong, S.W., *et al.* (2008). Serum starvation induces G1 arrest through suppression of Skp2-CDK2 and CDK4 in SK-OV-3 cells. Int J Oncol *32*, 435-439.

Shin, S.Y., Kim, J.H., *et al.* (2012). 2'-Hydroxyflavanone induces apoptosis through Egr-1 involving expression of Bax, p21, and NAG-1 in colon cancer cells. Mol Nutr Food Res *56*, 761-774.

Shirao, T., Kojima, N., *et al.* (1988). Molecular cloning of a cDNA for the developmentally regulated brain protein, drebrin. Brain Res *464*, 71-74.

Simpson, R.J., Jensen, S.S., *et al.* (2008). Proteomic profiling of exosomes: current perspectives. Proteomics *8*, 4083-4099.

Simpson, W.C., and Mayo, C.W. (1939). The mural penetration of the carcinoma cell in the colon: anatomic and clinical study. Surg Gynecol Obstet *68*, 872-877.

Stanley, P., Schachter, H., *et al.* (2009). N-Glycans. In Essentials of Glycobiology 2nd edition, A. Varki, R.D. Cummings, J.D. Esko, *et al.*, eds. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. Available from: http://www.ncbi.nlm.nih.gov/books/NBK1917/

Susini, L., Besse, S., *et al.* (2008). TCTP protects from apoptotic cell death by antagonizing bax function. Cell Death Differ *15*, 1211-1220.

Tai, I.T., Dai, M., *et al.* (2005). Genome-wide expression analysis of therapy-resistant tumors reveals SPARC as a novel target for cancer therapy. J Clin Invest *115*, 1492-1502.

Tai, I.T., and Tang, M.J. (2008). SPARC in cancer biology: its role in cancer progression and potential for therapy. Drug Resist Updat *11*, 231-246.

Takemasa, I., Higuchi, H., *et al.* (2001). Construction of preferential cDNA microarray specialized for human colorectal carcinoma: molecular sketch of colorectal cancer. Biochem Biophys Res Commun *285*, 1244-1249.

Tan, H.T., Tan, S., *et al.* (2008). Quantitative and temporal proteome analysis of butyrate-treated colorectal cancer cells. Mol Cell Proteomics *7*, 1174-1185.

Tan, H.T., Wu, W., *et al.* (2012). Proteomic analysis of colorectal cancer metastasis: stathmin-1 revealed as a player in cancer cell migration and prognostic marker. J Proteome Res *11*, 1433-1445.

Tanaka, M., Narumi, K., *et al.* (2000). Expression of the 37-kDa laminin binding protein in murine lung tumor cell correlates with tumor angiogenesis. Cancer Lett *153*, 161-168.

Tang, M.J., and Tai, I.T. (2007). A novel interaction between procaspase 8 and SPARC enhances apoptosis and potentiates chemotherapy sensitivity in colorectal cancers. J Biol Chem *282*, 34457-34467.

Tay, P.N., Tan, P., *et al.* (2010). Palladin, an actin-associated protein, is required for adherens junction formation and intercellular adhesion in HCT116 colorectal cancer cells. Int J Oncol *37*, 909-926.

Taylor, A.D., Hancock, W.S., *et al.* (2009). Towards an integrated proteomic and glycomic approach to finding cancer biomarkers. Genome Med *1*, 57.

Telerman, A., and Amson, R. (2009). The molecular programme of tumour reversion: the steps beyond malignant transformation. Nat Rev Cancer *9*, 206-216.

Telikicherla, D., Marimuthu, A., *et al.* (2012). Overexpression of ribosome binding protein 1 (RRBP1) in breast cancer. Clin Proteomics *9*, 7.

Tjalsma, H., Bolhuis, A., *et al.* (2000). Signal peptide-dependent protein transport in Bacillus subtilis: a genome-based survey of the secretome. Microbiol Mol Biol Rev *64*, 515-547.

Tsai, H.Y., Yang, Y.F., *et al.* (2013). Endoplasmic reticulum ribosome-binding protein 1 (RRBP1) overexpression is frequently found in lung

cancer patients and alleviates intracellular stress-induced apoptosis through the enhancement of GRP78. Oncogene.

Tuynder, M., Fiucci, G., *et al.* (2004). Translationally controlled tumor protein is a target of tumor reversion. Proc Natl Acad Sci U S A *101*, 15364-15369.

UK IBD Genetics Consortium, Barrett, J.C., *et al.* (2009). Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. Nat Genet *41*, 1330-1334.

Vachani, A., Nebozhyn, M., *et al.* (2007). A 10-gene classifier for distinguishing head and neck squamous cell carcinoma and lung squamous cell carcinoma. Clin Cancer Res *13*, 2905-2915.

Valastyan, S., and Weinberg, R.A. (2011). Tumor metastasis: molecular insights and evolving paradigms. Cell *147*, 275-292.

Van Cutsem, E., Nordlinger, B., *et al.* (2010). Advanced colorectal cancer: ESMO Clinical Practice Guidelines for treatment. Ann Oncol *21 Suppl 5*, v93-97.

van der Schouw, Y.T., Verbeek, A.L., *et al.* (1992). Comparison of four serum tumour markers in the diagnosis of colorectal carcinoma. Br J Cancer *66*, 148-154.

Vande Broek, I., Vanderkerken, K., *et al.* (2001). Laminin-1-induced migration of multiple myeloma cells involves the high-affinity 67 kD laminin receptor. Br J Cancer *85*, 1387-1395.

Vaskova, M., Kovac, M., *et al.* (2011). High expression of cytoskeletal protein drebrin in TEL/AML1pos B-cell precursor acute lymphoblastic leukemia identified by a novel monoclonal antibody. Leuk Res *35*, 1111-1113.

Viana Lde, S., Affonso, R.J., Jr., *et al.* (2013). Relationship between the expression of the extracellular matrix genes SPARC, SPP1, FN1, ITGA5 and ITGAV and clinicopathological parameters of tumor progression and colorectal cancer dissemination. Oncology *84*, 81-91.

Volmer, M.W., Radacz, Y., *et al.* (2004). Tumor suppressor Smad4 mediates downregulation of the anti-adhesive invasion-promoting matricellular protein SPARC: Landscaping activity of Smad4 as revealed by a "secretome" analysis. Proteomics *4*, 1324-1334.

Wallin, U., Glimelius, B., *et al.* (2011). Growth differentiation factor 15: a prognostic marker for recurrence in colorectal cancer. Br J Cancer *104*, 1619-1627.

Wang, C., Kovanen, V., *et al.* (2009). The glycosyltransferase activities of lysyl hydroxylase 3 (LH3) in the extracellular space are important for cell growth and viability. J Cell Mol Med *13*, 508-521.

Wang, H., and Hanash, S. (2003). Multi-dimensional liquid phase based separations in proteomics. Journal of chromatography B, Analytical technologies in the biomedical and life sciences *787*, 11-18.

Wang, J.Y., Tang, R., *et al.* (1994). Value of carcinoembryonic antigen in the management of colorectal cancer. Dis Colon Rectum *37*, 272-277.

Wang, X., Baek, S.J., *et al.* (2013). The diverse roles of nonsteroidal anti-inflammatory drug activated gene (NAG-1/GDF15) in cancer. Biochem Pharmacol *85*, 597-606.

Weeraphan, C., Diskul-Na-Ayudthaya, P., *et al.* (2012). Effective enrichment of cholangiocarcinoma secretomes using the hollow fiber bioreactor culture system. Talanta *99*, 294-301.

Wei, X., and Li, L. (2009). Comparative glycoproteomics: approaches and applications. Brief Funct Genomic Proteomic *8*, 104-113.

Weiss, J.V., Klein-Scory, S., *et al.* (2011). Soluble E-cadherin as a serum biomarker candidate: elevated levels in patients with late-stage colorectal carcinoma and FAP. Int J Cancer *128*, 1384-1392.

Wen, Y.T., Chang, Y.C., *et al.* (2011). Collection of in vivo-like liver cell secretome with alternative sample enrichment method using a hollow fiber bioreactor culture system combined with tangential flow filtration for secretomics analysis. Anal Chim Acta *684*, 72-79.

West, I., and Goldring, O. (2004). Lectin affinity chromatography. Methods Mol Biol *244*, 159-166.

Wiese, A.H., Auer, J., *et al.* (2007). Identification of gene signatures for invasive colorectal tumor cells. Cancer Detect Prev *31*, 282-295.

Willis, N.D., Cox, T.R., *et al.* (2008). Lamin A/C is a risk biomarker in colorectal cancer. PLoS One *3*, e2988.

Wilson, L.C., Baek, S.J., *et al.* (2003). Nonsteroidal anti-inflammatory drug-activated gene (NAG-1) is induced by genistein through the expression of p53 in colorectal cancer cells. Int J Cancer *105*, 747-753.

Wright, A., Li, Y.H., *et al.* (2008). The differential effect of endothelial cell factors on in vitro motility of malignant and non-malignant cells. Ann Biomed Eng *36*, 958-969.

Wu, C.C., Chen, H.C., *et al.* (2008). Identification of collapsin response mediator protein-2 as a potential marker of colorectal carcinoma by comparative analysis of cancer cell secretomes. Proteomics *8*, 316-332.

Wu, H.Y., Chang, Y.H., *et al.* (2009). Proteomics analysis of nasopharyngeal carcinoma cell secretome using a hollow fiber culture system and mass spectrometry. J Proteome Res *8*, 380-389.

Wu, W.W., Wang, G., *et al.* (2006). Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel- or LC-MALDI TOF/TOF. J Proteome Res *5*, 651-658.

Xue, G., Hao, L.Q., *et al.* (2009). Expression of annexin a5 is associated with higher tumor stage and poor prognosis in colorectal adenocarcinomas. J Clin Gastroenterol *43*, 831-837.

Xue, H., Lu, B., *et al.* (2010). Identification of serum biomarkers for colorectal cancer metastasis using a differential secretome approach. J Proteome Res *9*, 545-555.

Yamada, Y., Arao, T., *et al.* (2010). Plasma concentrations of VCAM-1 and PAI-1: a predictive biomarker for post-operative recurrence in colorectal cancer. Cancer Sci *101*, 1886-1890.

Yang, E., Kang, H.J., *et al.* (2007). Frequent inactivation of SPARC by promoter hypermethylation in colon cancers. Int J Cancer *121*, 567-575.

Yang, M.H., Kim, J., *et al.* (2014). Nonsteroidal anti-inflammatory drug activated gene-1 (NAG-1) modulators from natural products as anti-cancer agents. Life Sci *100*, 75-84.

Yang, Z., and Hancock, W.S. (2004). Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column. J Chromatogr A *1053*, 79-88.

Yang, Z., Harris, L.E., *et al.* (2006). Multilectin affinity chromatography for characterization of multiple glycoprotein biomarker candidates in serum from breast cancer patients. Clin Chem *52*, 1897-1905.

Yao, L., Lao, W., *et al.* (2012). Identification of EFEMP2 as a serum biomarker for the early detection of colorectal cancer with lectin affinity capture assisted secretome analysis of cultured fresh tissues. J Proteome Res *11*, 3281-3294.

Yokoyama, W.M., Kehn, P.J., *et al.* (1990). Chromosomal location of the Ly-49 (A1, YE1/48) multigene family. Genetic association with the NK 1.1 antigen. J Immunol *145*, 2353-2358.

Yoshimura, T., Nagahara, M., *et al.* (2011). Lymphovascular invasion of colorectal cancer is correlated to SPARC expression in the tumor stromal microenvironment. Epigenetics *6*, 1001-1011.

Zander, L., and Bemark, M. (2008). Identification of genes deregulated during serum-free medium adaptation of a Burkitt's lymphoma cell line. Cell Prolif *41*, 136-155.

Zeng, Z., Hincapie, M., *et al.* (2011). A proteomics platform combining depletion, multi-lectin affinity chromatography (M-LAC), and isoelectric focusing to study the breast cancer proteome. Anal Chem *83*, 4845-4854.

Zhang, H., and Chan, D.W. (2007). Cancer biomarker discovery in plasma using a tissue-targeted proteomic approach. Cancer Epidemiol Biomarkers Prev *16*, 1915-1917.

Zhang, H., Li, X.J., *et al.* (2003). Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. Nat Biotechnol *21*, 660-666.

Zhang, X., Xiao, Z., *et al.* (2012a). The potential role of ORM2 in the development of colorectal cancer. PLoS One *7*, e31868.

Zhang, Y., Tang, X., *et al.* (2012b). Lectin capture strategy for effective analysis of cell secretome. Proteomics *12*, 32-36.

Zhao, L., Liu, L., *et al.* (2007). Differential proteomic analysis of human colorectal carcinoma cell lines metastasis-associated proteins. J Cancer Res Clin Oncol *133*, 771-782.

Zhao, L., Wang, H., et al. (2008). Overexpression of Rho GDP-dissociation inhibitor alpha is associated with tumor progression and poor prognosis of colorectal cancer. J Proteome Res 7, 3994-4003.

Zheng, P., Liu, Y.X., *et al.* (2010). Stathmin, a new target of PRL-3 identified by proteomic methods, plays a key role in progression and metastasis of colorectal cancer. J Proteome Res *9*, 4897-4905.

Zhong, Y., Krisanapun, C., *et al.* (2010). Molecular targets of apigenin in colorectal cancer cells: involvement of p21, NAG-1 and p53. Eur J Cancer *46*, 3365-3374.

7. Appendix

Supplementary Table 1. Immunoblotting conditions for all antibodies utilised for western blot validation experiments

Protein	Primary antibody	Incubation conditions	Secondary antibody	Incubation conditions	Study
TCTP	Mouse anti-TCTP Santa Cruz Biotechnology Inc. (SC-133131)	1:500 in 1% milk 2 hours	HRP-conjugated anti-mouse IgG GE-Healthcare	1:5000 in 1% milk, 1 hour	Part 1: iTRAQ analysis of HCT-116 and E1 whole cell lysates
AKAP12	Mouse anti-AKAP12 Abnova (H00009590-M01)	1:500 in 1% milk, 2 hours	HRP-conjugated anti-mouse IgG GE-Healthcare	1:5000 in 1% milk, 1 hour	
DBN1	Mouse anti-DBN1 Abcam (ab12350)	1:1000 in 5% milk, 16 hours	HRP-conjugated anti-mouse IgG GE-Healthcare	1:5000 in 1% milk, 1 hour	
Actin	Rabbit anti-actin Sigma-aldrich (A2066)	1:1000 in 1% milk, 2 hours	HRP-conjugated anti-rabbit IgG Santa-Cruz	1:5000 in 1% milk, 1 hour	
SPARC	Rabbit anti-SPARC Proteintech Group (15274-1-AP)	1:1000 in 5% milk 1 hour	HRP-conjugated anti-rabbit IgG Pierce Biotechnology	1:5000 in 5% milk 1 hour	Part 2: iTRAQ analysis of HCT-116 and E1 secretomes
GDF15	Rabbit anti-GDF15 Abcam (ab14586)	1:1000 in 1% milk 2 hours	HRP-conjugated anti-rabbit IgG Pierce Biotechnology	1:5000 in 5% milk 1 hour	
PLOD3	Mouse anti-PLOD3 Abcam (ab89263)	1:500 in 1% milk 16 hours	HRP-conjugated anti-mouse IgG GE-Healthcare	1:5000 in 1% milk, 1 hour	
PAI1	Mouse anti-PAI Santa Cruz Biotechnology Inc. (SC-5297)	1:1000 in 5% milk 2 hours	HRP-conjugated anti-mouse IgG GE-Healthcare	1:5000 in 1% milk, 1 hour	
MAN1A1	Rabbit anti-MAN1A1 Sigma-Aldrich (M3694)	1:5000 in 1% milk 2 hours	HRP-conjugated anti-rabbit IgG Pierce Biotechnology	1:5000 in 5% milk 1 hour	
COL6A2	Rabbit anti-COL6A2 Proteintech Group (14853-1-AP)	1:500 in 1% milk 16 hours	HRP-conjugated anti-rabbit IgG Pierce Biotechnology	1:5000 in 5% milk 1 hour	Part 3: SWATH-MS analysis of MLAC-enriched HCT-116 and E1 glycosecretomes
LAMB1	Mouse anti-LAMB1 Santa Cruz Biotechnology Inc. (sc-17810)	1:500 in 1% milk 16 hours	HRP-conjugated anti-mouse IgG GE-Healthcare	1:5000 in 1% milk, 1 hour	

Supplementary Table 2. Clinical data of colorectal cancer patient tissue section samples used in DBN1 IHC analyses

Parameters	Total (n=7)
Age (years)	
≥ 60	2
< 60	5
Gender	
Male	4
Female	3
Histological Grade	
Moderately differentiated	5
Poorly differentiated	1
Unknown	1
Site of tumour	
Colon	5
Rectum	2
TNM Staging	
T stage	
T1	0
T2	1
T3	4
T4	2
N stage	
N0	0
N1	4
N2	4
M stage	
M0	4
M1	3
Liver metastases	
Synchronous	3
Metachronous	4

Supplementary Table 3. Clinical data of colorectal cancer patient serum samples used for ELISA analyses

Parameters	Total (n=45)
Age (years)	9
> 60	25
50 - 60	11
< 50	
Gender	
Male	
Female	15
	30
Ethnicity	
Chinese	35
Malay	5
Indian	2
Others	3
Modified Dukes' staging	
A	6
B	9
C	22
D	8

Supplementary Table 4. Differentially expressed in HCT-116 cells cultured in HFC system, as compared to culture flasks

Unused	Total	% Cov	UniProt Accession	Name	Peptides(95%)	HFC:Flask Replicate 1	HFC:Flask Replicate 2
3.23	3.36	11.5	Q08209	PPP3CA Serine/threonine-protein phosphatase 2B catalytic subunit alpha isoform	3	9.5095	1.5336
29.99	30.03	75.6	P16403	HIST1H1C Histone H1.2	19	3.2815	3.7936
2.5	29.3	76.7	P10412	HIST1H1E Histone H1.4	21	2.7074	3.0307
5.03	5.03	9.8	Q08J23	NSUN2 tRNA (cytosine(34)-C(5))-methyltransferase	3	2.1835	1.6223
4.42	4.42	36.7	P28070	PSMB4 Proteasome subunit beta type-4	4	2.1746	1.8203
6.85	10.34	53.7	P84103	SRSF3 Serine/arginine-rich splicing factor 3	6	1.8371	1.8293
6	6	20.6	Q13595	TRA2A Transformer-2 protein homolog alpha	4	1.7883	1.5012
4.72	4.82	13.6	P23381	WARS Tryptophan--tRNA ligase, cytoplasmic	3	1.7641	1.6441
31.05	31.05	34	P17844	DDX5 Probable ATP-dependent RNA helicase DDX5	18	1.7468	2.1189
4	4	18.8	Q9P0J0	NDUFA13 NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 13	2	1.7319	1.5976
3.07	3.07	16.4	Q15287	RNPS1 RNA-binding protein with serine-rich domain 1	2	1.7059	1.741
2	2	2.5	Q5T8P6	RBM26 RNA-binding protein 26	2	1.6673	1.8222
7.27	7.27	29.8	P33316	DUT Deoxyuridine 5'-triphosphate nucleotidohydrolase, mitochondrial	5	1.6373	1.8227
12.99	12.99	28.6	Q16629	SRSF7 Serine/arginine-rich splicing factor 7	8	1.6139	1.8175
2	2.01	6.1	Q96CS3	FAF2 FAS-associated factor 2	3	1.6006	2.7971
31.1	31.1	56	P49411	TUFM Elongation factor Tu, mitochondrial	19	1.6004	1.718
12.03	12.13	18	P33991	MCM4 DNA replication licensing factor MCM4	7	1.5861	2.0499
3.24	3.28	30.5	P60900	PSMA6 Proteasome subunit alpha type-6	4	1.5498	1.645
4	4	21	Q9Y5Y2	NUBP2 Cytosolic Fe-S cluster assembly factor NUBP2	2	1.5405	1.7131
2	2	17.1	P62308	SNRPG Small nuclear ribonucleoprotein G	2	1.8277	1.4012
5.48	5.58	9.7	P33993	MCM7 DNA replication licensing factor MCM7	3	1.7763	1.3538
8.36	8.36	19.4	O75367	H2AFY Core histone macro-H2A.1	5	1.7749	1.4916
4	4	22.5	Q01081	U2AF1 Splicing factor U2AF 35 kDa subunit	4	1.7513	1.3851
4	4.04	30.1	P63208	SKP1 S-phase kinase-associated protein 1	2	1.7444	1.3036
7.06	7.06	21.7	P61289	PSME3 Proteasome activator complex subunit 3	4	1.7182	1.3308
3.82	3.82	20.2	O75607	NPM3 Nucleoplasmin-3	3	1.6303	1.4407
2.1	4.4	24.3	Q9H2U2	PPA2 Inorganic pyrophosphatase 2, mitochondrial	4	1.6263	1.4327
6.98	6.98	21	Q12849	GRSF1 G-rich sequence factor 1	5	1.552	1.4479
9.26	9.68	22.5	O76021	RSL1D1 Ribosomal L1 domain-containing protein 1	6	1.5399	1.3204
5.52	5.52	31.6	P62847	RPS24 40S ribosomal protein S24	3	1.5369	1.3893
7.47	7.47	26.3	P06493	CDK1 Cyclin-dependent kinase 1	4	1.5293	1.467
10.1	10.11	37.1	Q13126	MTAP S-methyl-5'-thioadenosine phosphorylase	6	1.5246	1.4058
4	4	12.8	P26368	U2AF2 Splicing factor U2AF 65 kDa subunit	2	1.5203	1.3921
10.44	10.51	21.6	Q9NVI7	ATAD3A ATPase family AAA domain-containing protein 3A	6	1.4972	1.3474

Unused	Total	% Cov	UniProt Accession	Name	Peptides(95%)	HFC:Flask Replicate 1	HFC:Flask Replicate 2
5.64	5.64	16.7	P62995	TRA2B Transformer-2 protein homolog beta	3	1.4892	1.7095
9.65	9.65	31.9	P63244	GNB2L1 Guanine nucleotide-binding protein subunit beta-2-like 1	6	1.4872	1.879
19.57	19.57	62.5	P12004	PCNA Proliferating cell nuclear antigen	13	1.4679	1.7957
4.06	4.11	8.4	Q96TA2	YME1L1 ATP-dependent zinc metalloprotease YME1L1	2	1.4661	1.3256
5.51	5.51	14.2	P25205	MCM3 DNA replication licensing factor MCM3	4	1.4603	1.4317
44.46	44.46	30.4	P42704	LRPPRC Leucine-rich PPR motif-containing protein, mitochondrial	27	1.4574	1.523
5.11	6.61	28.5	Q02878	RPL6 60S ribosomal protein L6	4	1.4415	1.4008
4.68	4.68	10.9	Q9NYF8	BCLAF1 Bcl-2-associated transcription factor 1	4	1.4333	1.5943
3.45	3.45	9.6	P49736	MCM2 DNA replication licensing factor MCM2	2	1.4136	1.3536
3.19	3.19	27.3	P35244	RPA3 Replication protein A 14 kDa subunit	2	1.4085	1.3538
6.05	6.07	20	Q9BVP2	GNL3 Guanine nucleotide-binding protein-like 3	3	1.4075	1.4306
9.66	9.74	24.3	O95433	AHSA1 Activator of 90 kDa heat shock protein ATPase homolog 1	5	1.383	1.4443
2.01	2.02	24	Q9Y3B4	SF3B14 Pre-mRNA branch site protein p14	2	1.3685	1.5941
4.18	4.37	14.5	O60313	OPA1 Dynamin-like 120 kDa protein, mitochondrial	3	1.3652	1.4442
9.62	9.62	43.1	P62244	RPS15A 40S ribosomal protein S15a	5	1.3602	1.3813
16.91	16.91	19.6	Q08945	SSRP1 FACT complex subunit SSRP1	8	1.3498	1.6101
6.21	6.28	19.2	P07339	CTSD Cathepsin D	5	1.3441	1.7922
4.48	4.48	13.4	Q3LXA3	DAK Bifunctional ATP-dependent dihydroxyacetone kinase/FAD-AMP lyase (cyclizing)	2	1.3412	1.5571
4.01	4.01	9.6	Q7L2H7	EIF3M Eukaryotic translation initiation factor 3 subunit M	2	1.3313	1.4335
5.89	5.89	34.2	P25787	PSMA2 Proteasome subunit alpha type-2	4	1.3172	1.7267
2.43	2.43	26.8	P46781	RPS9 40S ribosomal protein S9	2	1.3142	1.3214
6	6	31.8	P62318	SNRPD3 Small nuclear ribonucleoprotein Sm D3	5	1.3122	1.7327
7.35	7.52	13.5	Q14566	MCM6 DNA replication licensing factor MCM6	4	1.3121	1.3521
2	2.11	4.2	O43143	DHX15 Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	2	0.7666	0.7093
9	22.66	62	P09104	ENO2 Gamma-enolase	16	0.7575	0.6991
211.45	211.45	61.9	Q09666	AHNAK Neuroblast differentiation-associated protein AHNAK	127	0.7565	0.6716
15.76	15.78	25.3	P55209	NAP1L1 Nucleosome assembly protein 1-like 1	11	0.75	0.7321
10.59	10.59	33.3	P40121	CAPG Macrophage-capping protein	7	0.7446	0.6726
21.02	21.2	66.5	P54819	AK2 Adenylate kinase 2, mitochondrial	17	0.7364	0.767
61.07	61.07	37.6	Q02952	AKAP12 A-kinase anchor protein 12	37	0.7349	0.7624
14	14.02	34.7	O00273	DFFA DNA fragmentation factor subunit alpha	7	0.7304	0.7254
4.71	4.71	69.5	P61956	SUMO2 Small ubiquitin-related modifier 2	4	0.7276	0.7523
12	12	58.8	Q04837	SSBP1 Single-stranded DNA-binding protein, mitochondrial	7	0.7247	0.7365
4.12	4.12	13.6	P29966	MARCKS Myristoylated alanine-rich C-kinase substrate	2	0.7222	0.7326
4	4	19.5	Q8WZA0	LZIC Protein LZIC	2	0.7208	0.6282
14.22	14.32	47.3	P10606	COX5B Cytochrome c oxidase subunit 5B, mitochondrial	7	0.7192	0.763
1.7	1.7	16.4	P09234	SNRPC U1 small nuclear ribonucleoprotein C	2	0.7187	0.5482
7.9	8.79	47.5	O15347	HMGB3 High mobility group protein B3	5	0.7167	0.729
6.01	6.01	26.1	O75940	SMNDC1 Survival of motor neuron-related-splicing factor 30	3	0.7157	0.6685
24.2	24.2	53.8	P12277	CKB Creatine kinase B-type	23	0.7143	0.7255

Unused	Total	% Cov	UniProt Accession	Name	Peptides(95%)	HFC:Flask Replicate 1	HFC:Flask Replicate 2
13.64	13.64	58.6	Q9UIJ7	AK3 GTP:AMP phosphotransferase AK3, mitochondrial	8	0.7117	0.7658
14.26	14.26	21.3	Q14247	CTTN Src substrate cortactin	7	0.7109	0.7224
32.17	32.18	48.5	Q07065	CKAP4 Cytoskeleton-associated protein 4	19	0.7032	0.6397
2.62	2.63	8.5	P29590	PML Protein PML	2	0.6984	0.7541
7.95	7.95	75.6	P05204	HMGN2 Non-histone chromosomal protein HMG-17	12	0.6939	0.7343
4.88	4.94	13.5	P07686	HEXB Beta-hexosaminidase subunit beta	3	0.692	0.6477
16.02	16.46	69.1	P16949	STMN1 Stathmin	10	0.6887	0.7156
22.27	22.36	43.3	P13674	P4HA1 Prolyl 4-hydroxylase subunit alpha-1	12	0.6884	0.42
13.53	13.54	19	Q14126	DSG2 Desmoglein-2	7	0.6861	0.66
4	4	25.2	Q9Y6H1	CHCHD2 Coiled-coil-helix-coiled-coil-helix domain-containing protein 2, mitochondrial	2	0.678	0.7229
6.82	6.82	34.4	O00264	PGRMC1 Membrane-associated progesterone receptor component 1	4	0.6707	0.6101
8.03	8.03	22.2	O94903	PROSC Proline synthase co-transcribed bacterial homolog protein	6	0.6703	0.6343
6.88	6.94	28.5	O95881	TXNDC12 Thioredoxin domain-containing protein 12	4	0.6643	0.7667
38.63	38.86	65.1	P04075	ALDOA Fructose-bisphosphate aldolase A	39	0.6633	0.7244
10	10	69.1	P30049	ATP5D ATP synthase subunit delta, mitochondrial	6	0.6554	0.7399
10.04	10.04	30.4	P43487	RANBP1 Ran-specific GTPase-activating protein	5	0.6538	0.7064
6.66	7.06	24.9	P09497	CLTB Clathrin light chain B	4	0.6478	0.7495
14	14	31.6	Q9Y224	C14orf166 UPF0568 protein C14orf166	7	0.6439	0.6836
14.22	14.22	54.4	Q04760	GLO1 Lactoylglutathione lyase	8	0.6333	0.7446
23.84	24.71	39.8	Q8NBS9	TXNDC5 Thioredoxin domain-containing protein 5	13	0.6301	0.7123
5.96	6	32.5	P49773	HINT1 Histidine triad nucleotide-binding protein 1	4	0.6131	0.6945
8	8	47.7	P25398	RPS12 40S ribosomal protein S12	4	0.6094	0.7362
8.67	8.67	16.3	P22307	SCP2 Non-specific lipid-transfer protein	5	0.6068	0.7032
100.23	100.28	71.1	P02545	LMNA Prelamin-A/C	79	0.5849	0.686
47.85	47.85	72.5	P06748	NPM1 Nucleophosmin	64	0.5839	0.6975
6.97	6.98	22.5	P20042	EIF2S2 Eukaryotic translation initiation factor 2 subunit 2	4	0.5814	0.7192
5.49	5.49	26.7	P49006	MARCKSL1 MARCKS-related protein	3	0.5752	0.7398
16	16	44.9	P30044	PRDX5 Peroxiredoxin-5, mitochondrial	10	0.5745	0.7246
55.55	70.11	69	P08107	HSPA1A Heat shock 70 kDa protein 1A/1B	51	0.5717	0.6915
9.75	16.11	62.7	P26583	HMGB2 High mobility group protein B2	10	0.5717	0.6879
5.4	5.4	33.2	Q8WW12	PCNP PEST proteolytic signal-containing nuclear protein	4	0.5678	0.7513
3.44	3.45	13.7	P78417	GSTO1 Glutathione S-transferase omega-1	3	0.5666	0.7036
18.21	18.21	58.2	Q9UJZ1	STOML2 Stomatin-like protein 2, mitochondrial	12	0.566	0.7436
11.04	11.04	29	P50502	ST13 Hsc70-interacting protein	5	0.5623	0.7317
16.19	16.19	80.8	O75947	ATP5H ATP synthase subunit d, mitochondrial	9	0.5522	0.6803
5.29	5.29	65.5	P07108	DBI Acyl-CoA-binding protein	3	0.5519	0.7224
6.01	6.01	15.3	P19404	NDUFV2 NADH dehydrogenase [ubiquinone] flavoprotein 2, mitochondrial	3	0.5499	0.729
4.82	4.82	91.8	P04080	CSTB Cystatin-B	3	0.5497	0.6676
22.24	22.24	47.3	Q15084	PDIA6 Protein disulfide-isomerase A6	16	0.5481	0.6826
12.3	12.55	36.5	P05455	SSB Lupus La protein	6	0.5456	0.7453

Unused	Total	% Cov	UniProt Accession	Name	Peptides(95%)	HFC:Flask Replicate 1	HFC:Flask Replicate 2
10.31	10.38	43.3	Q9HAV7	GRPEL1 GrpE protein homolog 1, mitochondrial	5	0.5397	0.6788
5.25	5.25	32.6	Q8N5K1	CISD2 CDGSH iron-sulfur domain-containing protein 2	3	0.5374	0.7445
3.53	3.54	63	P05114	HMGN1 Non-histone chromosomal protein HMG-14	3	0.5313	0.6691
17	19.32	54.6	P32119	PRDX2 Peroxiredoxin-2	12	0.5209	0.7331
21.22	21.22	38.9	P06744	GPI Glucose-6-phosphate isomerase	11	0.5159	0.7626
5.04	5.04	47.1	Q9HCY8	S100A14 Protein S100-A14	3	0.4999	0.6755
6.22	6.22	56.2	O75531	BANF1 Barrier-to-autointegration factor	3	0.4896	0.6821
21.61	21.64	54	P09429	HMGB1 High mobility group protein B1	12	0.4458	0.673
11.05	11.05	45.1	P02792	FTL Ferritin light chain	10	0.3443	0.7051
4.03	4.03	25.5	P62841	RPS15 40S ribosomal protein S15	6	0.3366	0.7115
31.48	31.48	63.4	O60664	PLIN3 Perilipin-3	23	0.6607	0.557
11.93	11.93	42.4	Q00688	FKBP3 Peptidyl-prolyl cis-trans isomerase FKBP3	6	0.6591	0.6404
8.28	8.29	32.7	P09496	CLTA Clathrin light chain A	5	0.6591	0.6546
10.28	12.4	50	P30085	CMPK1 UMP-CMP kinase	6	0.6565	0.626
2.8	2.82	9.5	Q9H930	SP140L Nuclear body protein SP140-like protein	2	0.6443	0.3929
9.54	9.54	25.1	P20810	CAST Calpastatin	5	0.6414	0.5837
6.42	6.75	94.7	Q15847	ADIRF Adipogenesis regulatory factor	5	0.6263	0.4467
14.99	14.99	25.9	Q14257	RCN2 Reticulocalbin-2	9	0.6218	0.6272
4.32	4.69	49.2	Q9UKY7	CDV3 Protein CDV3 homolog	2	0.6185	0.6248
4	4.01	15	Q92882	OSTF1 Osteoclast-stimulating factor 1	2	0.6122	0.5712
7.6	7.61	22.4	P43307	SSR1 Translocon-associated protein subunit alpha	5	0.5932	0.5059
1.69	1.72	11	O14562	UBFD1 Ubiquitin domain-containing protein UBFD1	3	0.581	0.6112
16.34	16.34	80.1	P60660	MYL6 Myosin light polypeptide 6	14	0.5808	0.4922
15.77	15.78	44.1	O43852	CALU Calumenin	8	0.5805	0.6306
3.79	3.8	28.1	Q00059	TFAM Transcription factor A, mitochondrial	3	0.5735	0.6062
26.3	26.3	42.7	P27824	CANX Calnexin	15	0.5667	0.6449
22.78	22.91	97.4	P05387	RPLP2 60S acidic ribosomal protein P2	28	0.5607	0.6058
12.02	12.02	55.9	P14927	UQCRB Cytochrome b-c1 complex subunit 7	7	0.5589	0.6153
26.24	26.24	44.7	P08758	ANXA5 Annexin A5	16	0.5551	0.5555
45.22	45.22	84.3	P67936	TPM4 Tropomyosin alpha-4 chain	29	0.5544	0.545
4.22	4.22	27	P04179	SOD2 Superoxide dismutase [Mn], mitochondrial	3	0.5536	0.6085
21.25	21.25	57.5	P51858	HDGF Hepatoma-derived growth factor	11	0.5519	0.6033
10.92	10.92	45.3	Q15293	RCN1 Reticulocalbin-1	8	0.5437	0.5589
5.5	5.5	36.5	O14561	NDUFAB1 Acyl carrier protein, mitochondrial	7	0.5429	0.6538
3.03	3.07	12.6	Q9BT09	CNPY3 Protein canopy homolog 3	2	0.5404	0.5417
16.02	16.02	60.2	P19105	MYL12A Myosin regulatory light chain 12A	8	0.5376	0.5675
18.35	18.35	34.1	Q8NC51	SERBP1 Plasminogen activator inhibitor 1 RNA-binding protein	13	0.5324	0.5894
17.57	17.57	70.7	P20674	COX5A Cytochrome c oxidase subunit 5A, mitochondrial	12	0.518	0.6051
65.51	65.96	72.1	P30101	PDIA3 Protein disulfide-isomerase A3	42	0.5167	0.5744
41.12	41.12	72.2	P27797	CALR Calreticulin	36	0.5072	0.6616

Unused	Total	% Cov	UniProt Accession	Name	Peptides(95%)	HFC:Flask Replicate 1	HFC:Flask Replicate 2
16.56	25.17	52.5	P09972	ALDOC Fructose-bisphosphate aldolase C	18	0.5024	0.6047
24.28	24.33	28	P14314	PRKCSH Glucosidase 2 subunit beta	14	0.5009	0.5951
39.15	39.17	79.7	P60174	TPI1 Triosephosphate isomerase	31	0.4948	0.5436
8.05	8.05	15.3	Q8WU90	ZC3H15 Zinc finger CCCH domain-containing protein 15	4	0.4945	0.6519
7.02	7.02	24.7	Q96A26	FAM162A Protein FAM162A	4	0.4914	0.6343
15.43	15.45	56.4	P24534	EEF1B2 Elongation factor 1-beta	9	0.4874	0.6003
3.47	33.65	57.4	P07951	TPM2 Tropomyosin beta chain	23	0.4857	0.5239
25.59	25.67	83.9	P62158	CALM1 Calmodulin	19	0.4838	0.5197
23.79	23.79	90.2	P61604	HSPE1 10 kDa heat shock protein, mitochondrial	18	0.4823	0.6624
17.67	17.84	97.7	P63313	TMSB10 Thymosin beta-10	11	0.4813	0.3627
20.01	24.67	63.7	P31947	SFN 14-3-3 protein sigma	24	0.4719	0.4399
20.72	20.72	80.5	P04792	HSPB1 Heat shock protein beta-1	15	0.4684	0.6077
3.02	3.02	36.8	P62851	RPS25 40S ribosomal protein S25	2	0.4673	0.5904
6.62	6.63	20.1	P68402	PAFAH1B2 Platelet-activating factor acetylhydrolase IB subunit beta	5	0.4652	0.5434
7.01	7.01	28.2	O15479	MAGEB2 Melanoma-associated antigen B2	5	0.4633	0.5065
5.08	5.08	8.9	Q12765	SCRN1 Secernin-1	3	0.4624	0.5567
13.37	13.84	56.2	P10599	TXN Thioredoxin	12	0.4604	0.5474
7.7	13.23	93.2	P62328	TMSB4X Thymosin beta-4	6	0.4597	0.457
20.09	20.09	74.8	P09382	LGALS1 Galectin-1	22	0.4552	0.6159
52.94	52.94	66.3	P31948	STIP1 Stress-induced-phosphoprotein 1	30	0.4546	0.5591
9.77	9.77	40.7	P55145	MANF Mesencephalic astrocyte-derived neurotrophic factor	5	0.453	0.591
95.37	95.65	67.3	P11021	HSPA5 78 kDa glucose-regulated protein	76	0.4491	0.5275
10	10	67.3	P17096	HMGA1 High mobility group protein HMG-I/HMG-Y	5	0.4469	0.569
76.34	77.63	59.2	P14625	HSP90B1 Endoplasmin	55	0.4463	0.5808
25.08	25.1	55	Q9BTM1	H2AFJ Histone H2A.J	25	0.4391	0.5459
6	6.36	33.3	P20962	PTMS Parathymosin	3	0.4356	0.5322
15.84	15.85	38.3	Q07021	C1QBP Complement component 1 Q subcomponent-binding protein, mitochondrial	16	0.4355	0.5332
67.62	67.62	79.5	P07237	P4HB Protein disulfide-isomerase	51	0.4344	0.6053
15.33	15.33	90.9	P00441	SOD1 Superoxide dismutase [Cu-Zn]	8	0.4335	0.59
1.71	1.71	26.4	Q9UII2	ATPIF1 ATPase inhibitor, mitochondrial	2	0.4324	0.5593
43.95	44.95	79.9	P00558	PGK1 Phosphoglycerate kinase 1	35	0.4297	0.571
6.02	8.03	41.4	Q71UI9	H2AFV Histone H2A.V	5	0.421	0.5661
8.05	8.05	74.1	P18859	ATP5J ATP synthase-coupling factor 6, mitochondrial	4	0.4192	0.5414
7.13	18.49	46.9	P16104	H2AFX Histone H2A.x	24	0.4158	0.5701
4.66	4.66	39.1	P14174	MIF Macrophage migration inhibitory factor	7	0.3829	0.5501
4.21	4.21	44.8	P31949	S100A11 Protein S100-A11	4	0.3703	0.4721
12.76	12.76	31.7	Q9BTT0	ANP32E Acidic leucine-rich nuclear phosphoprotein 32 family member E	8	0.3675	0.5163
4.39	4.39	49	Q99584	S100A13 Protein S100-A13	4	0.3651	0.4098
7.72	7.72	36.1	P60903	S100A10 Protein S100-A10	5	0.3568	0.5037
23.48	23.48	76	P18669	PGAM1 Phosphoglycerate mutase 1	17	0.354	0.5524

Unused	Total	% Cov	UniProt Accession	Name	Peptides(95%)	HFC:Flask Replicate 1	HFC:Flask Replicate 2
9.47	9.71	51.1	P06703	S100A6 Protein S100-A6	5	0.347	0.4754
4	6.03	71.1	P05386	RPLP1 60S acidic ribosomal protein P1	4	0.3409	0.4921
8.88	8.96	37.4	P39687	ANP32A Acidic leucine-rich nuclear phosphoprotein 32 family member A	12	0.3045	0.5461
11.79	11.79	51.4	P06454	PTMA Prothymosin alpha	16	0.3036	0.3669
13.27	13.27	52.3	Q92597	NDRG1 Protein NDRG1	12	0.2952	0.3225
14.12	14.12	61.8	P02794	FTH1 Ferritin heavy chain	10	0.2854	0.6296
4.01	4.01	42.5	O43169	CYB5B Cytochrome b5 type B	3	0.2574	0.4639

Supplementary Table 5. Differentially expressed proteins in E1 cells cultured in the HFC system, as compared to culture flasks

Unused	Total	% Cov	UniProt Accession	Name	Peptides(95%)	HFC:Flask Replicate 1	HFC:Flask Replicate 2
29.82	29.99	77.7	P62805	HIST1H4A Histone H4	40	1.7015	1.5553
2.5	29.3	76.7	P10412	HIST1H1E Histone H1.4	21	3.4622	3.6859
29.99	30.03	75.6	P16403	HIST1H1C Histone H1.2	19	5.3845	4.846
31.05	31.05	34	P17844	DDX5 Probable ATP-dependent RNA helicase DDX5	18	2.1861	1.7266
13.45	13.5	48.5	Q71DI3	HIST2H3A Histone H3.2	9	2.4898	1.8119
12.99	12.99	28.6	Q16629	SRSF7 Serine/arginine-rich splicing factor 7	8	1.8191	1.8217
16.91	16.91	19.6	Q08945	SSRP1 FACT complex subunit SSRP1	8	1.758	1.6358
6.85	10.34	53.7	P84103	SRSF3 Serine/arginine-rich splicing factor 3	6	2.635	1.9422
10	10	30.4	P00403	MT-CO2 Cytochrome c oxidase subunit 2	6	1.7931	1.7345
9.65	9.65	31.9	P63244	GNB2L1 Guanine nucleotide-binding protein subunit beta-2-like 1	6	1.6324	1.6806
6	6	31.8	P62318	SNRPD3 Small nuclear ribonucleoprotein Sm D3	5	1.5723	1.6516
8.31	8.31	23.6	Q05519	SRSF11 Serine/arginine-rich splicing factor 11	5	1.5677	1.5358
6.21	6.28	19.2	P07339	CTSD Cathepsin D	5	1.5488	1.5455
4	4	22.5	Q01081	U2AF1 Splicing factor U2AF 35 kDa subunit	4	1.8778	1.5808
3.62	3.62	29.1	Q14011	CIRBP Cold-inducible RNA-binding protein	3	2.1753	1.8882
5.64	5.64	16.7	P62995	TRA2B Transformer-2 protein homolog beta	3	1.5886	1.5706
2	2	2.5	Q5T8P6	RBM26 RNA-binding protein 26	2	3.4458	5.6566
2	2	17.1	P62308	SNRPG Small nuclear ribonucleoprotein G	2	2.2135	1.9608
2.12	2.68	4.8	P55060	CSE1L Exportin-2	2	1.956	1.6074
2	2.02	9.8	Q14344	GNA13 Guanine nucleotide-binding protein subunit alpha-13	2	1.8923	2.9428
4	4.02	8.8	Q3ZCQ8	TIMM50 Mitochondrial import inner membrane translocase subunit TIM50	2	1.6977	1.5444
40.28	40.28	66.3	P68363	TUBA1B Tubulin alpha-1B chain	50	1.4488	1.3241
31.1	31.1	56	P49411	TUFM Elongation factor Tu, mitochondrial	19	1.323	1.4016
30.91	30.92	36.9	Q00839	HNRNPU Heterogeneous nuclear ribonucleoprotein U	16	1.4463	1.3249
19.57	19.57	62.5	P12004	PCNA Proliferating cell nuclear antigen	13	1.6955	1.4595
22.76	22.78	43.1	P07195	LDHB L-lactate dehydrogenase B chain	12	1.3036	1.3368
21.01	21.05	22	Q12906	ILF3 Interleukin enhancer-binding factor 3	11	1.3075	1.445
12.27	12.31	49.7	P45880	VDAC2 Voltage-dependent anion-selective channel protein 2	10	1.472	1.3994
12.19	14.02	25.9	Q09028	RBBP4 Histone-binding protein RBBP4	9	1.3423	1.3049
12.88	12.93	24.4	P04181	OAT Ornithine aminotransferase, mitochondrial	8	1.5425	1.408
10.18	14.29	22.1	O14979	HNRPDL Heterogeneous nuclear ribonucleoprotein D-like	8	1.4159	1.5627
13.55	13.72	38.7	P18124	RPL7 60S ribosomal protein L7	8	1.3055	1.3126
5.96	9.33	18.4	Q15366	PCBP2 Poly(rC)-binding protein 2	7	1.4195	1.6702
12.03	12.13	18	P33991	MCM4 DNA replication licensing factor MCM4	7	1.3491	1.5924

Unused	Total	% Cov	UniProt Accession	Name	Peptides(95%)	HFC:Flask Replicate 1	HFC:Flask Replicate 2
8.44	8.44	31.8	P27635	RPL10 60S ribosomal protein L10	6	1.4359	1.3226
8.44	8.44	40.7	P62081	RPS7 40S ribosomal protein S7	6	1.4215	1.4218
9.66	9.66	47.2	P53999	SUB1 Activated RNA polymerase II transcriptional coactivator p15	6	1.3574	1.3255
8.36	8.36	19.4	O75367	H2AFY Core histone macro-H2A.1	5	1.4808	1.3127
9.62	9.62	43.1	P62244	RPS15A 40S ribosomal protein S15a	5	1.4597	1.4722
8.25	8.31	14.3	O75400	PRPF40A Pre-mRNA-processing factor 40 homolog A	5	1.457	1.3289
8.62	8.69	17.7	Q9Y2W1	THRAP3 Thyroid hormone receptor-associated protein 3	5	1.406	1.6268
5.93	5.93	40.4	P62277	RPS13 40S ribosomal protein S13	5	1.3747	1.5018
4.68	4.68	10.9	Q9NYF8	BCLAF1 Bcl-2-associated transcription factor 1	4	1.6567	1.3696
4.66	4.7	16.1	P33240	CSTF2 Cleavage stimulation factor subunit 2	4	1.5083	1.4374
7.66	7.66	21.4	O75494	SRSF10 Serine/arginine-rich splicing factor 10	4	1.4993	1.5418
6.94	6.98	36.6	P07305	H1F0 Histone H1.0	4	1.3789	1.6063
7.35	7.52	13.5	Q14566	MCM6 DNA replication licensing factor MCM6	4	1.3465	1.3445
8.19	8.28	29.6	Q9Y383	LUC7L2 Putative RNA-binding protein Luc7-like 2	4	1.3353	1.4293
3.24	3.28	30.5	P60900	PSMA6 Proteasome subunit alpha type-6	4	1.3331	1.37
7.47	7.47	26.3	P06493	CDK1 Cyclin-dependent kinase 1	4	1.3262	1.3121
6.64	6.64	38.6	P62829	RPL23 60S ribosomal protein L23	4	1.3176	1.3908
8	8	32.6	P62834	RAP1A Ras-related protein Rap-1A	4	1.3166	1.4181
6.38	6.51	8.9	O15042	U2SURP U2 snRNP-associated SURP motif-containing protein	4	1.3076	1.3042
6.04	6.11	13	Q9GZR7	DDX24 ATP-dependent RNA helicase DDX24	3	2.0212	1.3704
5.52	5.52	31.6	P62847	RPS24 40S ribosomal protein S24	3	1.9284	1.3645
3.82	3.82	20.2	O75607	NPM3 Nucleoplasmin-3	3	1.6215	1.4543
6	6	28.6	P60866	RPS20 40S ribosomal protein S20	3	1.5963	1.3011
4	4	25	P62304	SNRPE Small nuclear ribonucleoprotein E	3	1.5614	1.3848
5.39	5.84	9	Q14839	CHD4 Chromodomain-helicase-DNA-binding protein 4	3	1.455	1.4232
3.43	5.6	11.4	Q13363	CTBP1 C-terminal-binding protein 1	3	1.4063	1.4502
6.05	6.07	20	Q9BVP2	GNL3 Guanine nucleotide-binding protein-like 3	3	1.3593	1.4905
6	6	15.6	P62140	PPP1CB Serine/threonine-protein phosphatase PP1-beta catalytic subunit	3	1.3584	1.5047
6.51	6.51	36.5	P62701	RPS4X 40S ribosomal protein S4, X isoform	3	1.3517	1.39
2	2.01	10.6	O60232	SSSCA1 Sjoegren syndrome/scleroderma autoantigen 1	2	2.6983	1.4781
4.12	4.12	9.6	O75150	RNF40 E3 ubiquitin-protein ligase BRE1B	2	1.5942	1.4857
2.01	2.02	24	Q9Y3B4	SF3B14 Pre-mRNA branch site protein p14	2	1.482	1.7053
3.96	3.97	26.9	P61457	PCBD1 Pterin-4-alpha-carbinolamine dehydratase	2	1.4665	1.3106
4	4.03	20.4	Q01130	SRSF2 Serine/arginine-rich splicing factor 2	2	1.4082	1.5835
4	4.1	27.3	P62820	RAB1A Ras-related protein Rab-1A	2	1.3964	1.4774
2.04	2.04	17.7	Q8N5N7	MRPL50 39S ribosomal protein L50, mitochondrial	2	1.37	1.4144
76.34	77.63	59.2	P14625	HSP90B1 Endoplasmin	55	0.6952	0.742
55.55	70.11	69	P08107	HSPA1A Heat shock 70 kDa protein 1A/1B	51	0.7481	0.7624
67.62	67.62	79.5	P07237	P4HB Protein disulfide-isomerase	51	0.6631	0.7573
30.97	30.97	90.3	P62937	PPIA Peptidyl-prolyl cis-trans isomerase A	42	0.7666	0.5937

Unused	Total	% Cov	UniProt Accession	Name	Peptides(95%)	HFC:Flask Replicate 1	HFC:Flask Replicate 2
65.51	65.96	72.1	P30101	PDIA3 Protein disulfide-isomerase A3	42	0.7346	0.7056
38.63	38.86	65.1	P04075	ALDOA Fructose-bisphosphate aldolase A	39	0.5773	0.6883
61.07	61.07	37.6	Q02952	AKAP12 A-kinase anchor protein 12	37	0.5423	0.7394
24.2	24.2	53.8	P12277	CKB Creatine kinase B-type	23	0.6843	0.6375
6.92	25.99	28.6	P35241	RDX Radixin	20	0.7555	0.6793
23.79	23.79	90.2	P61604	HSPE1 10 kDa heat shock protein, mitochondrial	18	0.5529	0.7225
21.02	21.2	66.5	P54819	AK2 Adenylate kinase 2, mitochondrial	17	0.6435	0.6829
23.48	23.48	76	P18669	PGAM1 Phosphoglycerate mutase 1	17	0.5812	0.6882
20.72	20.72	80.5	P04792	HSPB1 Heat shock protein beta-1	15	0.6182	0.7214
17	19.32	54.6	P32119	PRDX2 Peroxiredoxin-2	12	0.6841	0.7308
22.27	22.36	43.3	P13674	P4HA1 Prolyl 4-hydroxylase subunit alpha-1	12	0.6736	0.6273
21.61	21.64	54	P09429	HMGB1 High mobility group protein B1	12	0.5646	0.6777
21.22	21.22	38.9	P06744	GPI Glucose-6-phosphate isomerase	11	0.6104	0.7469
15.21	15.21	59.3	O75347	TBCA Tubulin-specific chaperone A	9	0.6379	0.7617
15.43	15.45	56.4	P24534	EEF1B2 Elongation factor 1-beta	9	0.5978	0.6971
12.76	12.76	31.7	Q9BTT0	ANP32E Acidic leucine-rich nuclear phosphoprotein 32 family member E	8	0.7366	0.7099
14.43	14.43	43.1	Q15185	PTGES3 Prostaglandin E synthase 3	8	0.6856	0.683
15.77	15.78	44.1	O43852	CALU Calumenin	8	0.5716	0.764
13.53	13.54	19	Q14126	DSG2 Desmoglein-2	7	0.709	0.6893
14.26	14.26	21.3	Q14247	CTTN Src substrate cortactin	7	0.6636	0.7512
5.5	5.5	36.5	O14561	NDUFAB1 Acyl carrier protein, mitochondrial	7	0.6189	0.738
10.05	10.06	48.8	O14737	PDCD5 Programmed cell death protein 5	6	0.7309	0.6753
10.28	12.4	50	P30085	CMPK1 UMP-CMP kinase	6	0.6485	0.6959
8.03	8.03	22.2	O94903	PROSC Proline synthase co-transcribed bacterial homolog protein	6	0.5601	0.7256
6.07	6.13	19.1	Q8NFV4	ABHD11 Alpha/beta hydrolase domain-containing protein 11	5	0.7644	0.7022
6	8.4	20.4	P30483	HLA-B HLA class I histocompatibility antigen, B-45 alpha chain	5	0.6904	0.5969
10.04	10.04	30.4	P43487	RANBP1 Ran-specific GTPase-activating protein	5	0.6611	0.7604
7.6	7.61	22.4	P43307	SSR1 Translocon-associated protein subunit alpha	5	0.659	0.6674
6.88	6.94	28.5	O95881	TXNDC12 Thioredoxin domain-containing protein 12	4	0.7634	0.7169
5.4	5.4	33.2	Q8WW12	PCNP PEST proteolytic signal-containing nuclear protein	4	0.6961	0.7588
5.96	6	32.5	P49773	HINT1 Histidine triad nucleotide-binding protein 1	4	0.6643	0.6903
4	6.03	71.1	P05386	RPLP1 60S acidic ribosomal protein P1	4	0.6611	0.7019
6.82	6.82	34.4	O00264	PGRMC1 Membrane-associated progesterone receptor component 1	4	0.6499	0.74
8.05	8.05	74.1	P18859	ATP5J ATP synthase-coupling factor 6, mitochondrial	4	0.5108	0.681
5.08	5.08	8.9	Q12765	SCRN1 Secernin-1	3	0.7373	0.75
3.79	3.8	28.1	Q00059	TFAM Transcription factor A, mitochondrial	3	0.6341	0.6942
3.03	3.07	12.6	Q9BT09	CNPY3 Protein canopy homolog 3	2	0.7441	0.6796
4	4	19.5	Q8WZA0	LZIC Protein LZIC	2	0.7393	0.7373
5.16	5.16	16.2	Q16658	FSCN1 Fascin	2	0.6973	0.7562
2	2	9.7	P55957	BID BH3-interacting domain death agonist	2	0.4754	0.7179

Unused	Total	% Cov	UniProt Accession	Name	Peptides(95%)	HFC:Flask Replicate 1	HFC:Flask Replicate 2
2.8	2.82	9.5	Q9H930	SP140L Nuclear body protein SP140-like protein	2	0.4411	0.7142
4	4	25.2	Q9Y6H1	CHCHD2 Coiled-coil-helix-coiled-coil-helix domain-containing protein 2, mitochondrial	2	0.4007	0.7293
95.37	95.65	67.3	P11021	HSPA5 78 kDa glucose-regulated protein	76	0.6348	0.6588
43.95	44.95	79.9	P00558	PGK1 Phosphoglycerate kinase 1	35	0.4241	0.4836
39.15	39.17	79.7	P60174	TPI1 Triosephosphate isomerase	31	0.5239	0.616
52.94	52.94	66.3	P31948	STIP1 Stress-induced-phosphoprotein 1	30	0.6093	0.6444
45.22	45.22	84.3	P67936	TPM4 Tropomyosin alpha-4 chain	29	0.3779	0.5533
20.01	24.67	63.7	P31947	SFN 14-3-3 protein sigma	24	0.364	0.3587
31.48	31.48	63.4	O60664	PLIN3 Perilipin-3	23	0.4447	0.5899
3.47	33.65	57.4	P07951	TPM2 Tropomyosin beta chain	23	0.3973	0.5429
25.59	25.67	83.9	P62158	CALM1 Calmodulin	19	0.4461	0.5535
16.56	25.17	52.5	P09972	ALDOC Fructose-bisphosphate aldolase C	18	0.4801	0.6074
26.24	26.24	44.7	P08758	ANXA5 Annexin A5	16	0.6062	0.6428
15.84	15.85	38.3	Q07021	C1QBP Complement component 1 Q subcomponent-binding protein, mitochondrial	16	0.5477	0.4953
9	22.66	62	P09104	ENO2 Gamma-enolase	16	0.4488	0.5865
11.79	11.79	51.4	P06454	PTMA Prothymosin alpha	16	0.2873	0.4655
16.34	16.34	80.1	P60660	MYL6 Myosin light polypeptide 6	14	0.6092	0.6439
13.37	13.84	56.2	P10599	TXN Thioredoxin	12	0.4328	0.4668
13.27	13.27	52.3	Q92597	NDRG1 Protein NDRG1	12	0.2781	0.4632
21.25	21.25	57.5	P51858	HDGF Hepatoma-derived growth factor	11	0.5151	0.6338
17.67	17.84	97.7	P63313	TMSB10 Thymosin beta-10	11	0.3357	0.4218
16	16	44.9	P30044	PRDX5 Peroxiredoxin-5, mitochondrial	10	0.6491	0.6149
14.12	14.12	61.8	P02794	FTH1 Ferritin heavy chain	10	0.4035	0.5433
11.05	11.05	45.1	P02792	FTL Ferritin light chain	10	0.3883	0.5433
16.02	16.02	60.2	P19105	MYL12A Myosin regulatory light chain 12A	8	0.6158	0.6267
10.92	10.92	45.3	Q15293	RCN1 Reticulocalbin-1	8	0.6111	0.634
15.33	15.33	90.9	P00441	SOD1 Superoxide dismutase [Cu-Zn]	8	0.3376	0.6091
4.66	4.66	39.1	P14174	MIF Macrophage migration inhibitory factor	7	0.4565	0.6381
7.7	13.23	93.2	P62328	TMSB4X Thymosin beta-4	6	0.2944	0.5376
9.54	9.54	25.1	P20810	CAST Calpastatin	5	0.5184	0.6514
10.31	10.31	45.1	P55327	TPD52 Tumor protein D52	5	0.5002	0.6483
6.62	6.63	20.1	P68402	PAFAH1B2 Platelet-activating factor acetylhydrolase IB subunit beta	5	0.4973	0.5689
9.47	9.71	51.1	P06703	S100A6 Protein S100-A6	5	0.4958	0.592
6.42	6.75	94.7	Q15847	ADIRF Adipogenesis regulatory factor	5	0.4925	0.5014
4.21	4.21	44.8	P31949	S100A11 Protein S100-A11	4	0.6044	0.6306
5.9	5.95	34.7	Q96HY6	DDRGK1 DDRGK domain-containing protein 1	4	0.5406	0.6514
4.39	4.39	49	Q99584	S100A13 Protein S100-A13	4	0.4212	0.5444
5.17	5.17	33.3	Q9BRA2	TXNDC17 Thioredoxin domain-containing protein 17	3	0.6655	0.5187
5.04	5.04	47.1	Q9HCY8	S100A14 Protein S100-A14	3	0.6014	0.4955
5.29	5.29	65.5	P07108	DBI Acyl-CoA-binding protein	3	0.4735	0.6584

Unused	Total	% Cov	UniProt Accession	Name	Peptides(95%)	HFC:Flask Replicate 1	HFC:Flask Replicate 2
4.82	4.82	91.8	P04080	CSTB Cystatin-B	3	0.444	0.4606
6	6.36	33.3	P20962	PTMS Parathymosin	3	0.3804	0.528
4.32	4.69	49.2	Q9UKY7	CDV3 Protein CDV3 homolog	2	0.5048	0.4637
1.71	1.71	26.4	Q9UII2	ATPIF1 ATPase inhibitor, mitochondrial	2	0.4485	0.6281
2.99	2.99	32.8	P80297	MT1X Metallothionein-1X	2	0.2524	0.6491

Publications and awards

Publications

Lin, Q., Tan, H.T., Lim, H.S., Chung, M.C. (2013) Sieving through the cancer secretome. *Biochimica et Biophysica Acta* 1834: 2360-2371

Lin, Q., Tan, H. T., Lim, T. K., Khoo, A., Lim, K. H. and Chung, M. C. M. (2014), iTRAQ analysis of colorectal cancer cell lines suggests Drebrin (DBN1) is overexpressed during liver metastasis. *Proteomics*, 14: 1434-1443.

Posters at International Conferences

6th International Conference on Structural Biology and Functional Genomics, Singapore (2010)

Identifying novel proteins involved in colorectal cancer metastasis using iTRAQ technology.

Singapore Society for Mass Spectrometry (SSMS) Conference,
Singapore (2010)

Identifying novel proteins involved in colorectal cancer metastasis using iTRAQ technology.

Human Proteome Organisation (HUPO) 10th Annual World Congress,
Geneva, Switzerland (2011)

iTRAQ analysis reveals Drebrin (DBN1) as a novel player involved in colorectal cancer liver metastasis.

Asia Oceania Human Proteome Organisation (AOHUPO) 6th Congress,
Beijing, China (2012)

Identification of metastatic biomarkers in the colorectal cancer secretome using the hollow fibre culture system.

Separation Science Mass Spec 2013,
Singapore (2013)

Sifting through the colorectal cancer secretome for metastatic biomarkers using the iTRAQ and SWATH-MS technologies.

Oral Presentations at International Conferences

Asia Oceania Human Proteome Organisation (AOHUPO) 6th Congress, Beijing, China (2012)

"Hollow fibre culture system - an alternative in sample preparation in secretome studies."
Invited presentation at Education and Training Workshop "Proteomics - its principles and applications"

Awards

Asia Oceania Human Proteome Organisation (AOHUPO) 6th Congress, Beijing, China (2012)

Young Scientist Travel Fellowship

Separation Science Mass Spec 2013, Singapore (2013)

Silver Poster Award