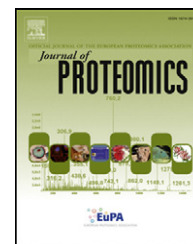


Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

In-depth proteomic delineation of the colorectal cancer exoproteome: Mechanistic insight and identification of potential biomarkers



George S. Karagiannis^{a,b}, Maria P. Pavlou^{a,b}, Punit Saraon^{a,b}, Natasha Musrap^{a,b}, Annie Xie^a, Ihor Batruch^b, Ioannis Prassas^b, Apostolos Dimitromanolakis^b, Constantina Petraki^c, Eleftherios P. Diamandis^{a,b,d,*}

^aDepartment of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada

^bDepartment of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario, Canada

^cPathology Department, Metropolitan Hospital, Athens, Greece

^dDepartment of Clinical Biochemistry, University Health Network, Toronto, Ontario, Canada

ARTICLE INFO

Article history:

Received 20 December 2013

Accepted 18 March 2014

Available online 27 March 2014

Keywords:

Colorectal cancer

Exoproteome

Kallikrein-related peptidase

Olfactomedin-4

Proteomics

Enrichment map analysis

ABSTRACT

Systemic mining of cancer exoproteome/secretome has emerged as a pivotal strategy for delineation of molecular pathways with mechanistic importance in cancer development, as well as the discovery of diagnostic/prognostic biomarkers. Although major advances in diagnostic and therapeutic management of colorectal cancer have been underscored in the last decade, this cancer still remains the second leading cause of cancer-related deaths in the developed world. Despite previous studies on deciphering the colorectal cancer exoproteome, such studies lack adequate depth and robustness due to technological limitations. Here, using a well-established LC-MS/MS method on an LTQ-Orbitrap mass spectrometer, we extensively delineated the exoproteome of 12 colon cancer cell lines. In total, 2979 non-redundant proteins were identified with a minimum of two peptides, of which ~62% were extracellular or cell membrane-bound, based on prediction software. To further characterize this dataset and identify clinical opportunities, first, we investigated overrepresented molecular concepts of interest via enrichment map analysis and second, we demonstrated translational importance of certain proteins, such as olfactomedin-4 and kallikrein-related peptidases-6 and -10, by investigating their expression levels in patient

Abbreviations: ATCC, American Type Culture Collection; AUC, Area Under the Curve; AZGP1, Zinc-alpha 2-glycoprotein; CA9, Carbonic Anhydrase IX; CDCHO, Chemically-defined Chinese Hamster Ovary; CEA, Carcinoembryonic Antigen; CM, Conditioned Media; CRC, Colorectal Cancer; DTT, Dithiothreitol; ECM, Extracellular Matrix; EGFR, Epithelial Growth Factor Receptor; EMEM, Eagle's Minimum Essential Medium; EMT, Epithelial-to-Mesenchymal Transition; FAP, Familial Adenomatous Polyposis; FBS, Fetal Bovine Serum; FDR, False Discovery Rate; GO, Gene Ontology; GREM1, Gremlin-1; HL, Hosmer-Lemeshow; (HP)LC, (High Performance) Liquid Chromatography; IBD, Inflammatory Bowel Disease; KLK, Kallikrein-related Peptidase; LOXL2, Lysyl-Oxidase Homolog-2; MS/MS, Tandem Mass Spectrometry; NME, Nucleoside Diphosphate Kinase A; OLFM4, Olfactomedin-4; PBS, Phosphate-buffered Saline; RER, Replication Error; ROC, Receiver Operating Characteristic; RPMI, Roswell Park Memorial Institute; SRPX2, sushi repeat-containing protein-2; VCAN, Versican.

* Corresponding author at: Mount Sinai Hospital, Joseph & Wolf Lebovic Ctr., 60 Murray St., Box 32, Floor 6, Room L6-201, Toronto, ON M5T 3L9, Canada. Tel.: +1 416 586 8443; fax: +1 416 619 5521.

E-mail addresses: gkaragiannis@mtsinai.on.ca (G.S. Karagiannis), mpavlou@mtsinai.on.ca (M.P. Pavlou), punit.saraon@utoronto.ca (P. Saraon), natasha.musrap@mtsinai.on.ca (N. Musrap), annixie31@gmail.com (A. Xie), ibatruch.acdclabs@gmail.com (I. Batruch), yprassas@gmail.com (I. Prassas), apostol@cs.toronto.edu (A. Dimitromanolakis), constantinapetraki@gmail.com (C. Petraki), ediamandis@mtsinai.on.ca (E.P. Diamandis).

tissues and/or fluids. Overall, the present study details a comprehensive colorectal cancer exoproteome dataset, and may be used as future platform for biomarker discovery, and hypothesis-generating studies.

Biological significance

This article represents one of the most extensive and comprehensive proteomic datasets regarding the secreted/extracellular proteome of colorectal cancer cell lines. The reported datasets may form a platform for a plethora of future, discovery-based or hypothesis-generating studies, attempting to either delineate putative cancer biomarkers for CRC, or elucidate questions of mechanistic importance (e.g. investigation of deregulated pathways for CRC progression).

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Colorectal cancer (CRC) represents one of the most important causes of cancer-related death, and is the second most frequent type of cancer after lung cancer [1–3]. While CRC is mostly identified in the sporadic form, a significant portion can also occur in the context on inflammatory bowel disease (IBD) [4,5] or genetic syndromes, such as familial adenomatous polyposis (FAP) [6,7] and Lynch syndrome [8]. The development of sporadic CRC is caused by the accumulation of genetic and epigenetic changes, which can be generally categorized into two types: (I) Approximately 80% of CRC patients undergo a well-characterized series of molecular events, described as *adenoma–carcinoma sequence* [9–11], involving chromosomal aberrations and mutations in several genes, such as APC, KRAS, P53 and DCC [12–17]. (II) The remaining 20% of CRC patients undergoes a secondary molecular pathway, which causes genetic instability in microsatellite loci attributable to alterations in the DNA mismatch repair genes, such as MLH1, MSH2, MSH6 and PMS2 [10,11,18]. These latter cancers are considered as replication error-deficient (RER+), while the former ones as replication error-proficient (RER–) [19].

Recently, high-throughput proteomic pipelines coupled to mass spectrometry have played a pivotal role in protein research, especially in the simultaneous identification, quantification and characterization of thousands of proteins in complex biological samples [20,21]. The emergence of these technologies has enabled the field of cancer research (i.e. oncoproteomics) with a plethora of opportunities, such as the diagnosis and therapeutic management of cancer [20,22]. An emerging subfield of oncoproteomics originates from the so-called ‘secretome analysis’, which attempts to delineate the extracellular proteome of cancer and/or other types of cells. The term ‘secretome’ was originally adapted by Tjalsma et al. [23] and Antelmann et al. [24] as a concept providing an integrated and global view of the protein secretion by considering both to the secretion systems and their protein substrates. It should be mentioned that proteins found in the extracellular milieu, i.e. the exoproteins, are not systematically secreted. Secreted proteins are defined as proteins actively transported across biological membrane by a secretion system (i.e. canonical or non-canonical) [25–28]. The term ‘exoproteome’ was later coined by Tjalsma et al. (2007) [29] to specifically refer to the subset of proteins present in the extracellular milieu, i.e. the extracellular proteome.

The indications, thus far, point to the fact that the exoproteome is a promising source of candidate biomarkers and therapeutic targets for various types of cancer, in the era of personalized medicine [30–34]. With the exception of a small number of studies, attempts to decipher the colorectal cancer exoproteome have been lacking. The aforementioned ones have yielded a set of candidate CRC biomarkers, of which a subset was selected for validation studies in human tissues and serum. Wu et al. [35] identified dataset of 325 unique proteins, of which collapsing response mediator-2 (CRMP-2) was validated by immunohistochemistry and its levels were significantly higher in CRC patients versus healthy controls. Xue et al. [36] performed differential proteomic analysis of the SW480/SW620 model, using label-free quantification. This study yielded a total of 910 proteins, of which 145 exhibited differential expressions. Trefoil factor 3 and growth/differentiation factor 15 were further validated in a large cohort of clinical tissues and serum, in which they could predict colorectal cancer metastasis. In an integrative approach, Wu et al. analyzed the secretomes of 23 human cancer cell lines derived from 11 cancer types including CRC, using one-dimensional SDS-PAGE and nano-LC-MS/MS, and proposed a list of candidate serological biomarkers [37]. Additional studies have quantitatively compared the extracellular proteomes between metastatic and primary cell lines [38], or coculture models to mimic the CRC microenvironment [39], as well as the CRC stem cell exoproteome [40,41] and identified key candidates of CRC development, progression and/or drug resistance.

To complement efforts for characterization of the CRC exoproteome, here, we performed in-depth proteomic analyses, integrating and comparing the proteomes of conditioned media (CM) from 12 different CRC cell lines (SW1116, SW480, LS174T, LS180, WiDR, SW620, RKO, LoVo, HCT116, DLD1, Colo320HSR and Colo205), which were chosen to recapitulate, as much as possible, the heterogeneity of the disease. As such, these cell lines represent individuals of varying ethnic backgrounds and age groups, mutational profiles and disease stage and/or differentiation status. All samples were analyzed in triplicate using strong cation exchange (SCX) chromatography followed by liquid chromatography (LC)–tandem mass spectrometry (MS/MS) on a linear trap quadrupole (LTQ)–Orbitrap mass spectrometer. The reported dataset may form a platform for a plethora of future, discovery-based or hypothesis-generating studies, attempting to either delineate

putative cancer biomarkers for CRC, or elucidate questions of mechanistic importance (e.g. investigation of deregulated pathways for CRC progression). As proof-of-concept, here we describe a few of these elements.

2. Materials and methods

2.1. Cell lines

A total of twelve colon cancer cell lines were obtained from the American Type Culture Collection (ATCC, Rockville, MD). These all belong to the sporadic CRC subtype and have the following characteristics: SW1116 (CCL-233), Duke's stage A, RER-; SW480 (CCL-228), Duke's stage B, RER-; LS174T (CL-188), Duke's stage B, RER+; LS180 (CL-187), Duke's stage B, RER+; WiDR (CCL-218), Duke's stage C, RER-; SW620 (CCL-227), Duke's stage C, RER-; RKO (CRL-2577), Duke's stage C, RER+; LoVo (CCL-229), Duke's stage C, RER+; HCT116 (CCL-247), Duke's stage C, RER+; DLD1 (CCL-221), Duke's stage C, RER+; Colo320HSR (CCL-220.1), Duke's stage C, RER-; Colo205 (CCL-222), Duke's stage D, RER-.

Cell culture media (conditioned media; CM) specified from ATCC for each of the twelve CRC cell lines were used and are as follows: Leibovitz L-15 Medium (Catalog No. 30-2008 from ATCC) with 10% fetal bovine serum (FBS) (Catalog No. 10091-148 from Invitrogen) was used for SW1116, SW480 and SW620; Eagle's Minimum Essential Medium (EMEM; Catalog No. 30-2003 from ATCC) with 10% FBS was used for LS174T, LS180, WiDr and RKO; F-12K Medium (Catalog No. 30-2004 from ATCC) with 10% FBS was used for LoVo; McCoy's 5a Medium Modified (Catalog No. 30-2007 from ATCC) with 10% FBS was used for HCT116; Roswell Park Memorial Institute-1640 (RPMI-1640; Catalog No. 30-2001 from ATCC) with 10% FBS was used for DLD1, ColoHSR320 and Colo205. All cells were cultured in an atmosphere of 5% CO₂ in air in a humidified incubator at 37 °C. All experiments were conducted within <5 passages.

2.2. Mass spectrometry-based proteomics

2.2.1. Sample preparation

All cell cultures were washed with phosphate-buffered Saline (PBS) after removal of their respective media and switched to Chemically-defined Chinese Hamster Ovary Medium (CDCHO) for 2 days. All CM were collected and normalized to ~1 µg of total protein (Coomassie colorimetric assay; Pierce biotechnology). The samples were dialyzed, using a 3.5 kDa molecular cut-off membrane (Spectrum Laboratories, Inc., CA, US) in 4 L of 1 mM ammonium bicarbonate solution (Fisher Scientific) (4 buffer changes). Dialyzed CM were frozen at -80 °C and subjected to lyophilization until total dryness. The samples were then denatured by 8 M urea (Fisher), reduced with dithiothreitol (DDT) (Sigma-Aldrich) to a final concentration of 13 mM at 50 °C for 30 min, alkylated with 500 mM iodoacetamide (Sigma-Aldrich) in the dark, at room temperature for 1 h, and finally desalted using a NAP5 column (GE Healthcare), using the manufacturer's instructions. The eluted 1 mL samples were lyophilized and trypsin-digested (Promega) in molar ratio of 1:50 (trypsin:protein concentration) for 8 h.

2.2.2. Strong cation exchange liquid chromatography (SCX-LC)

The resulting peptides were lyophilized to dryness and resuspended in 120 µL of 0.26 M formic acid in 10% acetonitrile (mobile phase A). The samples were fractionated using an Agilent high-performance liquid chromatography (HPLC) system connected to a PolySULFOETHYL A™ column with 200-Å pore size and a diameter of 5 µm (The Nest Group Inc.). A 1-h linear gradient was used with 1 mol/L ammonium formate and 0.26 mol/L formic acid in 10% acetonitrile (mobile phase B) at a flow rate of 200 µL/min. Fractions were collected every 5 min for the first 20 min of the run and every 2 min for the following 40 min, to a total number of 24 fractions/replicate. Of these, 16 fractions, corresponding to the highest concentration of eluted peptides, were kept for mass spectrometry. A peptide cation exchange standard (Bio-Rad), consisting of four peptides, was run at the beginning of each replicate to assess column performance.

2.2.3. Reverse phase liquid chromatography (RP-LC)

Sixteen HPLC fractions per cell line replicate were C₁₈-extracted using a ZipTipC₁₈ pipette tip (Millipore) and eluted in 5 µL of 90% acetonitrile, 0.1% formic acid, 10% water and 0.02% trifluoroacetic acid (Buffer B). 80 µL of 95% water, 0.1% formic acid, 5% acetonitrile and 0.02% trifluoroacetic acid (Buffer A) were added to this mixture, and 40 µL were injected via an autosampler on an Agilent HPLC. The peptides were separated in a 2-cm C₁₈ guard column (inner diameter 200 µm), and eluted in a resolving 5-cm analytical C₁₈ column (inner diameter 75 µm) with an 8-µm tip (New Objective).

2.2.4. Tandem mass spectrometry (MS/MS)

This HPLC system was coupled online to an LTQ-Orbitrap XL (Thermo Fisher Scientific) mass spectrometer, using nano-electrospray ionization (ESI) source (Proxeon Biosystems), in data-dependent mode. Each fraction was run with a 55-min gradient and eluted peptides were subjected to one full MS scan (450–1450 *m/z*) in the Orbitrap at 60,000 resolution, followed by six data-dependent MS/MS scans in the linear ion trap (LTQ Orbitrap). Unassigned charge states and charges 1+ and 4+ were all ignored, as depicted through the charge-state screening and preview mode.

2.2.5. Protein identification

Data files were created by use of Mascot Daemon (version 2.2.0) and extract_msn. The resulting mass spectra from each fraction were analyzed using Mascot (Matrix Science; version 2.2) and X!Tandem (Global Proteome Machine Manager; version 2005.06.01) search engines, using the International Protein Index human database (version 3.62, 167,894 protein sequences), which includes both forward and reverse sequences. The resulting Mascot and X!Tandem files were loaded into Scaffold (Proteome Software; version 2.6) to cross validate the data files from both engines. Detection of a minimum of two unique peptides was required to accept positive protein identification. Spectrum reports were exported from Scaffold for further analysis. False discovery rate (FDR) was calculated as previously described [42,43].

2.2.6. Label-free quantification

Relative quantification of proteins among the various cancer cell lines was ascertained, based on spectral counting [44–46]. All proteomic datasets were subjected to spectral counting normalization to avoid possible operator- or machine-specific biases during the experimental part of the pipeline [47]. We used the “Quantitative Value” function of Scaffold 2.6 software, which provides normalized spectral counts based on the total number of spectra identified in each sample. One Scaffold file containing all normalized spectral counts of each of the three replicates from 12 CRC cell lines was generated for proteins. Not all proteins were identified in all of the cell lines and unidentified proteins or missing values in a particular biological sample were assigned a normalized spectral count of 0.54487, which represented the lowest quantifiable value that could be assigned from “Quantitative Value” function in Scaffold. This method also assisted in keeping from dividing by zero or overestimating fold-changes. Quantification of multiple isoforms occurred only if the isoforms were identified as individual entries (i.e. with distinct gene names) through Scaffold software.

2.3. Bioinformatics

2.3.1. Dataset organization and visualization tools

To organize and visualize lists of protein subsets according to specific questions asked, the Venn diagram-generating tool VENNY (<http://bioinfogp.cnb.csic.es/tools/venny/index.html>) was utilized.

2.3.2. Exoproteome dendrogram

Proteins were examined for predicted secretion with presence of signal peptide (SignalP 4.0) (<http://www.cbs.dtu.dk/services/SignalP/>) [48,49] or without (SecretomeP 2.0) (<http://www.cbs.dtu.dk/services/SecretomeP/>) [50], and amino acids relevant to transmembrane helices (TMHMM 2.0) (<http://www.cbs.dtu.dk/services/TMHMM/>) [51]. The IPI human database v3.71 and the ENSEMBL human gene annotations (version 62) were used to map the protein identifiers to FASTA sequences and those sequences were used as input to all three prediction software. The following cutoffs were considered successful pass: >0.5 score for SignalP, >0.7 score for SecretomeP, and presence of at least one (i.e. >0) transmembrane helices for TMHMM. An automation tool, here termed as ‘exoproteome dendrogram’, was developed in-house in R and can be run on a Linux environment, facilitating the mapping of a list of protein identifiers (IPI number of ENSEMBL gene transcripts) to sequences, running the three prediction software mentioned above and creating a report file with consolidated predictions. The input can also be a list of gene names, in which case the IPI database is used to obtain the list of identifiers associated with each gene name. For efficiency reasons, a locally installed version of SignalP and TMHMM were used to obtain predictions, whereas for SecretomeP the web-based server was queried.

2.3.3. Gene ontology — enrichment map analysis

Enrichment map was conducted as previously described [52]. In brief, BINGO (v.2.44) was utilized to calculate over-representation of Gene Ontology (GO) “biological process” terms in the input gene list [53]. The hypergeometric test was

performed to assess the significance of the enrichment and resulting p-values were FDR corrected using the Benjamini & Hochberg method ($p < 0.05$). Data from the enriched GO annotations was exported. Functional enrichment visualization was constructed by the enrichment map plug-in in Cytoscape [54,55], by using the exported BINGO results as the input. The Jaccard coefficient was used at a cutoff of 0.5 to connect related GO biological process terms and create an enrichment network. Parameters for the selection of the enrichment results, which appear on the network, were set to a p-value cutoff of 0.005 and FDR of <10%. The generated clusters within the enrichment map were named after a commercially-available word cloud algorithm for word frequencies (Wordle®) (<http://www.wordle.net/>).

2.3.4. Protein–protein interaction analysis

Protein–protein interaction analysis was performed using the STRING software (Version 9.1) (<http://string-db.org/>) [56]. Protein datasets of interest were uploaded in the application using gene identifiers and complete lists of human orthologs were included. Protein–protein interaction networks were created based on neighborhood, gene-fusion, co-occurrence, co-expression, experiments, and databases. The text-mining algorithm was excluded from the analysis and STRING confidence scores of >0.9 were accepted to ensure high network quality. Proteins were shown as nodes and connectivity as connecting lines between them. Confidence view modes of the networks were exported for visualization purposes.

2.3.5. Gene expression meta-analysis

We mined the NCBI GEO [57] and ArrayExpress [58] repositories for experiments performed in 13 cancer types for in-house use. Only experiments performed on Affymetrix Human Genome U133 Plus 2.0 arrays were considered, in an effort to avoid biases resulting from the joint analysis of different array platforms. In total, we collected expression profiles and annotations from 4413 samples across 83 different experiments. Samples were filtered to include only primary human cancerous tissue and normal tissue from the site of origin, excluding xenografts, cell lines and other pathologies. The raw data from CEL files were jointly normalized using the RMA algorithm [59], following the methods described previously [60]. Normalization and data analysis were performed in the statistical software R version 2.15.2, using packaged under the BioConductor platform. Cancer vs normal ratios were calculated as the ratio of mean probe expression in cancer vs the mean probe expression in normal tissue from the same origin. In this meta-analysis, we calculated comparative gene expression differences for >40,000 gene identifiers, out of which 615 (which corresponded to 475 unique genes), were >2-fold upregulated in CRC compared to normal colon epithelia.

2.4. Immunoassays (ELISAs)

The concentration of kallikrein-related peptidases-6 (KLK6), -7 (KLK7), -10 (KLK10) and -11 (KLK11) in cell culture supernatants was determined by in-house developed sandwich-type immunoassays, as previously described [61,62]. Concentration

was determined by interpolation from a standard curve, using recombinant KLK proteins, as standards.

For verification studies, serum samples from a small cohort of 28 cases (10 CRC patients; 8 patients with benign lesions; 10 healthy controls) were used. Blood samples collected from these patients following a standardized protocol (age range of 48–85, median age of 64; 12 male and 16 female). All samples were collected with informed consent, and IRB approval, and were subjected to standardized protocols for preparation. All cases were histologically confirmed sporadic colorectal adenocarcinomas, and benign lesions involved seven patients with polyps and two patients with liver cirrhosis. Concentration of carbonic anhydrase IX (CA9), gremlin-1 (GREM1), lysyl oxidase-like 2 (LOXL2) and olfactomedin-4 (OLFM4) in these samples were determined by commercial ELISAs (USCN Life Sciences), according to the manufacturer's instructions.

2.5. Immunohistochemistry (IHC)

IHC analysis for KLK6 and KLK10 was performed, as previously described [63–66]. In brief, tissue sections from 15 paraffin-embedded archived cases of moderately to poorly differentiated human CRC were randomly selected. IHC staining was performed using the Bond automated IHC system (Bondmax, Leica Microsystems, UK)-pretreatment with epitope retrieval (pH 8). Primary antibodies against KLK6 and KLK10 were rabbit polyclonal, previously developed in our laboratory [62].

2.6. Statistical analysis

The following statistical analyses were performed with the SPSS (version 20) software: (A) All correlations between KLK spectral counting and their actual quantity ($\mu\text{g/L}$) as measured by specific KLK immunoassays were tested by the non-parametric Spearman's ranked correlation coefficient. Statistical significance was shown at the 0.01 level. These data were presented as scatter plots with p-values and coefficient of determination (r^2). (B) The comparison of mean spectral counting levels for the identified KLKs among the four distinct Duke's stages was tested by the non-parametric Jonckheere–Terpstra statistic. Holms correction was performed for multiple comparisons (and as such, significance was shown at the 0.0125 level). These data were presented as bars representing means with standard errors. (C) The comparison of serum concentrations of CA9, LXL2, GREM1 and OLFM4 among the three different groups of individuals (i.e. cancer patients, benign controls, healthy controls) was tested by the non-parametric Jonckheere–Terpstra statistic. Holms correction was performed for multiple comparisons (and as such significance was shown at the 0.0125 level). Post-hoc analysis was performed using the non-parametric Mann–Whitney U test (significance demonstrated with $p < 0.05$). These data were presented as box-plots with median lines and error bars. (D) The possibility that (a) OLFM4 alone, or (b) OLFM4 combined with CEA serum levels could predict colon cancer incidence was tested using logistic regression models. In both cases, prediction criterion was a dichotomous variable indicating incidence of cancer (i.e. cancer patients) or absence of cancer (i.e. healthy controls and benign controls). Factors

included in the analysis were the serum levels of OLFM4 for question (a) or the combination of OLFM4 and CEA serum levels for question (b). Significance was demonstrated with $p < 0.05$. Goodness-of-fit was a priori performed with the Hosmer–Lemeshow (HL) statistic; a p-value of less than 0.05 was considered significant to demonstrate improper data calibration. Receiver operating characteristic (ROC) curves were used to assess the adequacy of the predictive power of the factor(s).

3. Results

3.1. Delineation of CRC exoproteome via LC–MS/MS analysis in 12 cell line conditioned media

Twelve commercially-available cell lines, all belonging to the sporadic CRC subtype, were grown as described in experimental procedures and their 2-day CM were harvested and analyzed in triplicate through LC–MS/MS (Fig. 1A). Using both MASCOT and X!Tandem search engines, between 551 and 1868 proteins were identified in the CM, with a minimum of two unique peptide hits and $\text{FDR} < 1.0\%$ for protein identification (Fig. 1B). Complete lists of proteins identified, including molecular weight, corresponding gene symbols, international protein index accession number, and normalized values of assigned spectra in each replicate for HCT116, LoVo, Colo320HSR, WiDr, SW1116, SW620, RKO, DLD1, LS174T, LS180, Colo205 and SW480 cell lines, are summarized in Supplementary Table 1. The reproducibility among the triplicates ranged between ~60% and ~90% (Fig. 1B) and was considered satisfactory, based on previous experience [67–73]. From our combined analysis, a total of 13,849 proteins were identified with two or more unique peptides. The majority of these proteins were common to multiple biological samples and our study resulted in the identification of 2979 non-redundant proteins (Supplementary Table 2).

One major obstacle during the experimental design of proteomic studies, which are focused on the delineation of the exoproteome subset through mass spectrometry is the difficulty in establishing a rationalized method to reliably differentiate this subset from intracellular contaminants, especially in in-vitro conditions [20]. To address this issue, we deployed a three-step dendrogram based on predictions from canonical (SignalP) or non-canonical (SecretomeP) protein secretion and presence of transmembrane helices (TMHMM). Through this process, we categorized all proteins into one of the five generated protein groups (I–V), depending on the scores that each protein received throughout the dendrogram (Fig. 1C). For additional details on the structuring of this dendrogram, here, briefly termed as “exoproteome dendrogram”, refer to details in [Materials and methods](#). The exoproteome dendrogram categorized the 2979 identified proteins into the following groups: non-membranous proteins, but secreted through canonical (group I) or non-canonical (group II) pathways; non-secreted proteins, but membrane-harbored in external (group III) or internal (group IV) membranes; non-membranous, non-secreted proteins (group V) (Fig. 1C & D). In an attempt to decipher the exoproteome subset, we reasoned that proteins of groups IV

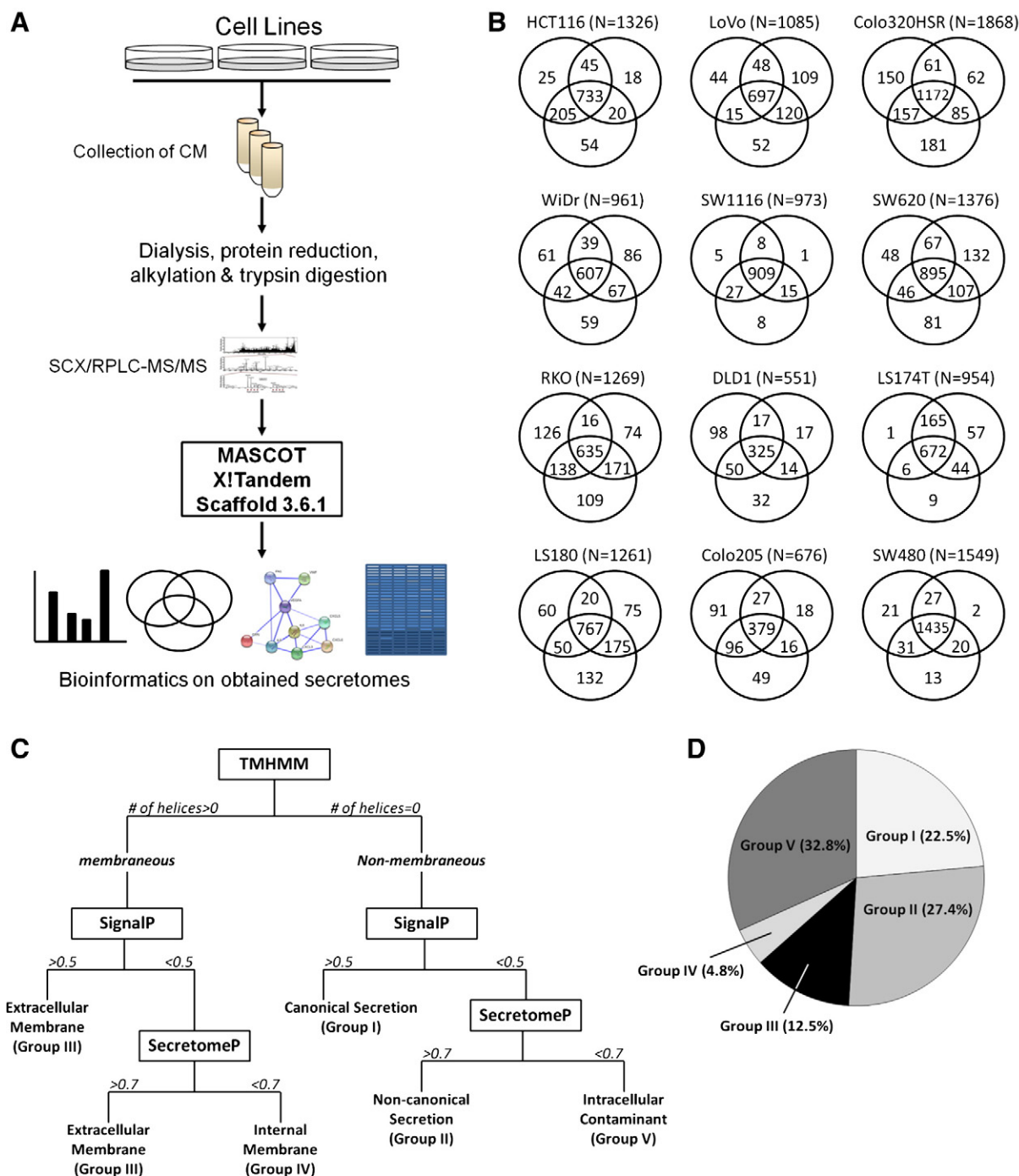


Fig. 1 – Delineation of the CRC exoproteome. (A) Schematic outline of proteomic analysis methodology. Sample processing, followed by strong cation exchange (SCX) and reverse-phase liquid chromatography (RPLC) coupled online to an LTQ-Orbitrap mass spectrometer, and subsequent data analysis is outlined. **(B)** Total non-redundant proteins identified. Venn diagrams (VENNY software), depict total proteins identified with a minimum of two peptides in the three replicates of each CRC cell line supernatant. **(C)** Exoproteome dendrogram. Each protein is sequentially checked through prediction software (i.e. TMHMM, SignalP, SecretomeP) and may move along the decision tree according to the following cutoffs for a successful pass: presence of at least one (i.e. >0) transmembrane helices for TMHMM, >0.5 score for SignalP, and >0.7 score for SecretomeP. The decision tree automatically assigns each protein to only one of the proposed protein groups (i.e. I–V). **(D)** Result summary of exoproteome dendrogram. The pie-chart indicates the distribution of 2979 proteins into groups I–V.

and V were “intracellular contaminants”, and as such, we only kept proteins of groups I–III for further consideration (Supplementary Table 3). Interestingly, 37.6% of the proteins in our dataset corresponded to group IV and V proteins, which indicates that our analysis provided enrichment, with 62.4% being exoproteome (Fig. 1D). This is satisfactory based on previous experience [67–72], as well as literature evidence [20]. In conclusion, 1867 out of the 2979 proteins composed our proposed CRC exoproteome, which, to our knowledge, is the most complete proteomic dataset for this type of cancer (Supplementary Table 3).

3.2. The CRC exoproteome identifies a kallikrein-related peptidase expression signature

A major challenge in high-throughput proteomic experiments is the robustness of these analyses and the accuracy of protein quantification. We have previously proposed the utilization of secreted-protein panels as internal controls, to verify the integrity of protein identification and quantification [74]. For such purpose, here, we utilized a small panel of extracellular proteases, namely KLK6, KLK7, KLK10 and KLK11, which are all members of the kallikrein family of serine peptidases [75,76]. Before the cell line CM were subjected to proteomic analysis, a 1 mL-aliquot from each replicate (N = 36) was kept for measuring protein concentrations with highly sensitive and specific KLK immunoassays. Immunoassays (ELISAs) for these KLKs have been previously developed in our laboratory [62]. When we performed correlation studies between KLK spectral counts and their quantities (ELISA), we found statistically significant ($p < 0.01$; Spearman's ranked correlation coefficient) associations for all four KLKs (Fig. 2A), an observation that validated our label-free quantification method.

As indicated from the label-free quantification data (Fig. 2B), KLK6 was the most abundantly secreted KLK and was identified in all but RKO and Colo320HSR cell line CM. On the other hand, KLK7 depicted very low but highly-selective secretion, as only few spectral counts were found in the SW1116, SW480 and HCT116 cell line CM (Fig. 2B). Finally, KLK10 and KLK11 were secreted by approximately 50% of the cell lines tested in variable amounts (Fig. 2B). Generally, all KLKs had tendency of reduced expression levels with increased tumor staging. To better demonstrate this notion, we compared the mean expression levels of each individual KLK across the various tumor stages, after combining cell lines that belonged to the same Duke's stage in one pool. This analysis demonstrated a statistically significant difference ($p < 0.125$; Holm's corrected; Jonckheere–Terpstra test) in the ranked expression score for KLK6 and KLK10 (Fig. 2C). These KLKs indeed exhibited a progressive decrease of their mean expression levels with increased tumor stage (Fig. 2C). As a proof-of-concept, when we investigated KLK6 and KLK10 immunohistochemical expression in a small cohort (N = 15) of CRC patients with varying degrees of differentiation, we observed that both KLKs presented with lower expression in low-grade compared to high-grade neoplastic lesions (Fig. 2D, Supplementary Fig. 1). Therefore, given our recent observations that KLK6 and KLK10 may hold prognostic significance in CRC patients [63,66], we expect that the newly-described

CRC exoproteome may hold crucial information of translational importance for CRC patients.

3.3. Enrichment map profiling of the CRC exoproteome

We wished to investigate whether our proposed CRC exoproteome holds key factors and/or mediators that may regulate malignancy during CRC development and progression. To accomplish this, we performed enrichment map profiling in our dataset, as previously described [52,77]. In this systemic approach, we unraveled overrepresented themes in our 1867-protein dataset through enrichment analysis in Gene Ontology (GO) annotations for biological processes. This analysis resulted in the significant ($p < 0.05$) overrepresentation of 684 GO terms (Supplementary Table 4). Since several annotations are branched together, we visualized the analysis as an enrichment network, which algorithmically clustered GO terms with highly similar content, using the enrichment map plug-in in the cytoscape environment. A few of the resulting clusters corresponded to: “proliferation”, “cell cycle”, “mitosis”, “motility”, “angiogenesis”, “apoptosis”, and “inflammation” (Fig. 3), which are considered as traditional hallmarks of cancer development and progression [78]. Interestingly, several other clusters corresponded to specific biological programs, such as: “epithelial-to-mesenchymal transition (EMT)”, “response to irradiation”, “cell adhesion”, “chemotaxis and locomotion”, “collagen organization and assembly”, “smooth muscle differentiation”, “blood coagulation” and “bone development” (Fig. 3) whose implication in cancer development and progression is also significant or, at least, is beginning to emerge [78,79]. Finally, a rather impressive observation is the overrepresentation of the Wnt/ β -catenin signal transduction pathway (Fig. 3; upper left corner) whose importance in CRC progression has been extensively underscored [79].

3.4. Generation of a protein subset with putative markers of colon cancer progression

Manipulation of large-scale data and their translation into proper clinical settings represent a major challenge in high-throughput proteomics. To address this, we have previously underscored reproducible and biologically-relevant criteria for candidate selection [31]. As a proof-of-concept, here, we will exclusively focus on the “cell adhesion” cluster of the enrichment map to select for cell adhesion regulatory proteins that could be potentially associated with development and progression of CRC and serve as markers of CRC progression. Our rationale relies on the observed deregulation of pertinent elements of the cell adhesion machinery during cancer progression, especially during phenotypic reprogramming (i.e. EMT) [78,79]. In addition, a variety of membranous proteins involved in the regulation of cell-to-cell or cell-to-matrix adhesion, are directly affected by phenotypic reprogramming or by overrepresented proteolytic cascades [80,81]. These are often captured as markers of cancer progression with the most prominent paradigm that of carcinoembryonic antigen (CEA), a cell-surface glycoprotein and a broadly-used CRC biomarker [82–84]. Thus, by focusing on the “cell adhesion” cluster, we narrowed the 1867-protein dataset of the enrichment map down to 179 cell

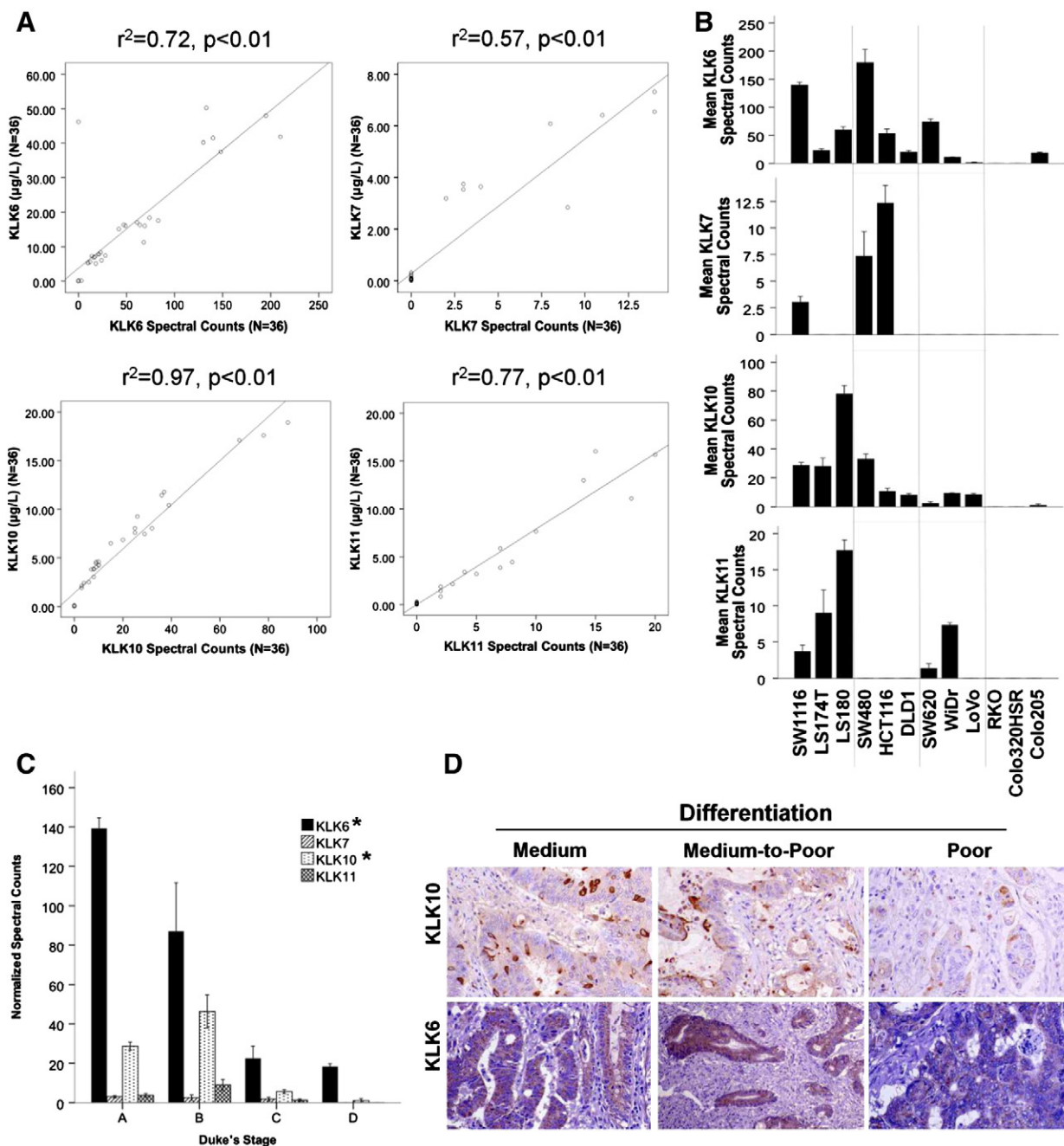


Fig. 2 – The kallikrein-related peptidase expression signature in CRC. (A) Scatter plots demonstrate the correlation between spectral counts and concentration levels as measured by specific immunoassays, for each KLK separately. All correlations depict statistical significance, $p < 0.01$, Spearman's ranked correlation coefficient. (B) Mean spectral counts of each KLK in each of the 12 CRC cell lines. (C) Mean spectral counts of each KLK in each of the tumor stages, according to Duke's classification system. Differences in the mean scores occur only for KLK6 and KLK10. Asterisks demonstrate statistical significance, $p < 0.0125$, Holms-corrected Jonckheere–Terpstra test. (D) Representative IHC snapshots from three selected CRC patients, showing sections from medium-differentiated and poorly-differentiated neoplastic lesions, demonstrating gradual loss of KLK6 and KLK10 with loss of differentiation and progression of the disease. All magnifications $\times 200$.

adhesion-specific proteins (Fig. 4A, Supplementary Table 5), where CEA was also present.

The 179-protein dataset was further subjected to protein–protein interaction analysis using STRING, in an effort to identify the most robust cell adhesion networks within. This analysis revealed 59 proteins, interconnected in a vast

network (confidence >0.9), bearing extracellular matrix-(ECM-) interconnected hubs, such as the collagen family and the focal adhesion machinery, as well as signaling pathways, such as the epithelial growth factor receptor (EGFR) system, and the catenin family (Fig. 4B). Since all these protein relays harbor well-established causative associations with various

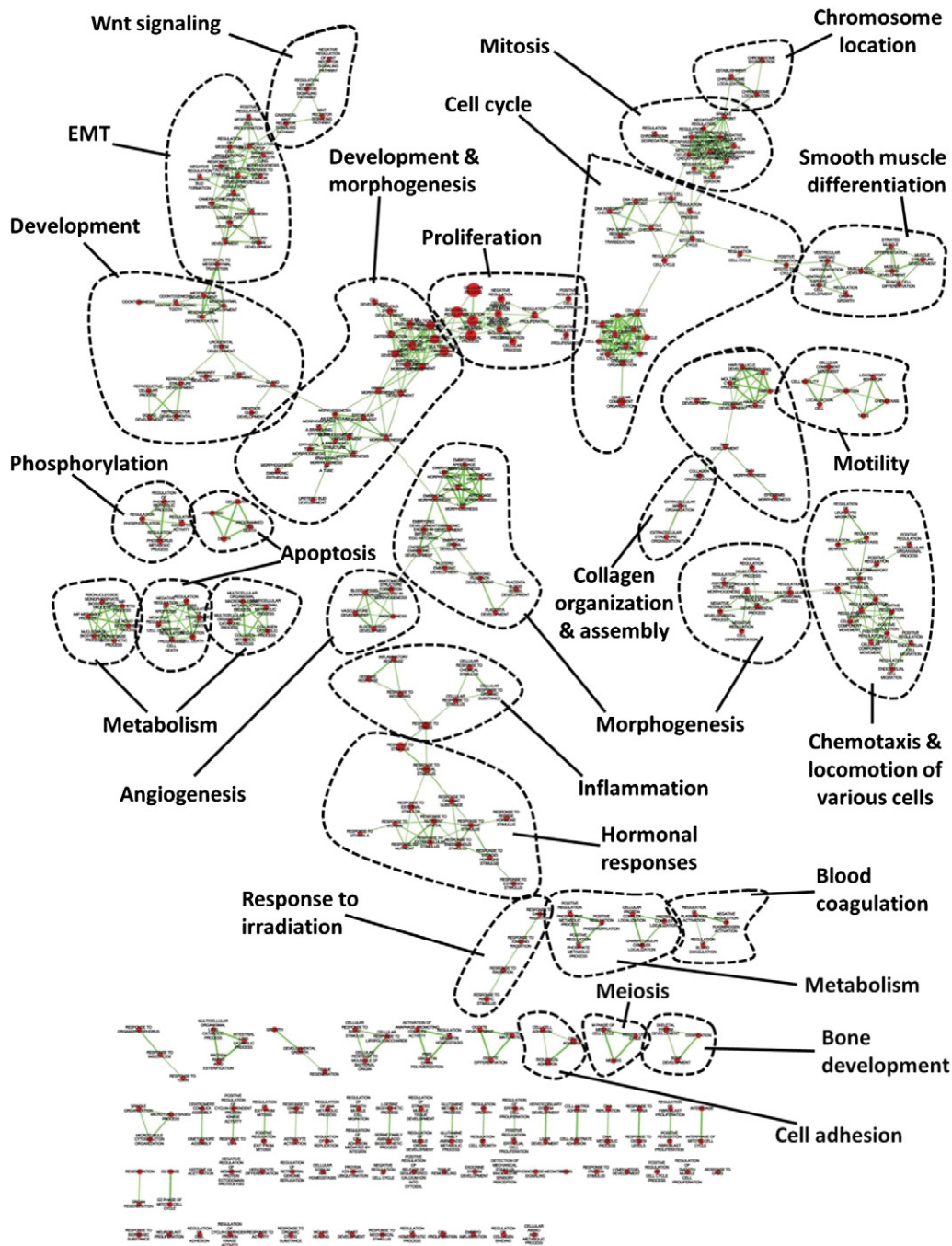


Fig. 3 – Enrichment map profiling of the CRC exoproteome showed overrepresentation of functional clusters associated with cancer development and progression. Nodes represent GO terms and lines demonstrate connectivity. Dashed lines encircle groups of relevant GO terms into functional clusters, which are named with the assistance of a word frequency algorithm.

cancers, including CRC [78,79], we focused on the remaining 120 proteins, which comprised a group of proteins with medium-to-low abundance and rather underexplored in cancer settings (Fig. 4A, Supplementary Table 5).

Our previous studies on breast cancer subtyping [73] have demonstrated that integration of transcriptomic data into proteomic pipelines significantly increases the chances of delineating promising candidates. We thus proceeded to comprehensive mining of the NCBI GEO and ArrayExpress repositories for experiments performed in 13 cancer types,

which collectively depicted transcriptional perturbations of >40,000 gene identifiers in CRC (Supplementary Table 6). Of these, 615 gene identifiers corresponding to 475 genes, were found to be significantly upregulated (>2-fold threshold; red line of Fig. 4C) in CRC compared to normal colon (Fig. 4C, Supplementary Table 6). Subsequently, we compared this robust 475-gene dataset to our 120-protein dataset and tracked down 13 common elements (Fig. 4D, Supplementary Table 5).

Conclusively, a 3-step prioritization process (Fig. 4A), brought the 1867 proteins of the CRC exoproteome down to

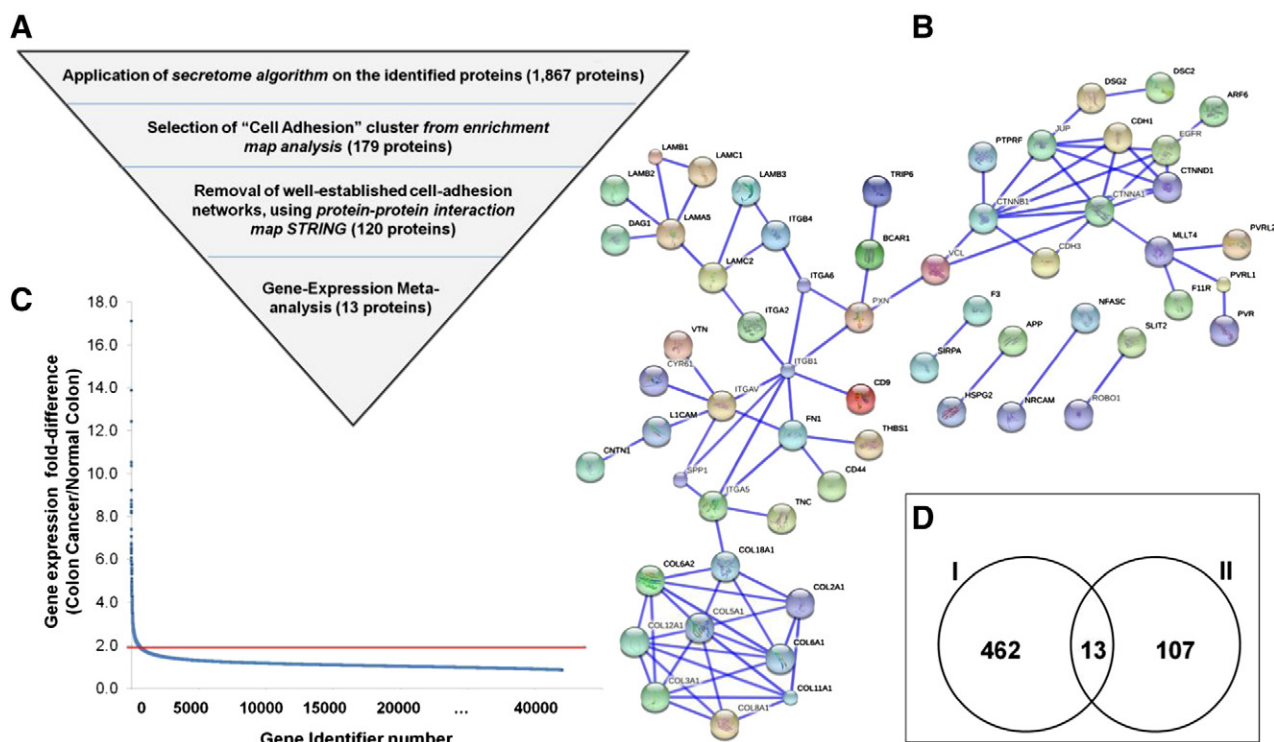


Fig. 4 – Prioritization criteria for identifying cell-adhesion regulators, as putative markers of CRC progression. (A) Sequential application of filtering criteria narrowed the 1867 secreted proteins of the original CRC dataset down to 13 cell-adhesion regulators. (B) STRING network of the 179-proteins belonging to the “cell adhesion” cluster of the enrichment map revealed a total of 59 proteins, interconnected in large cell adhesion networks; the remaining 120-proteins (dataset II) were filtered out of this well-established network and prioritized for further analysis. (C) Gene-expression meta-analysis using publicly-available microarray data yielded 475 deregulated genes (dataset I) in CRC compared to normal colonic epithelium. The red-line demonstrates the 2-fold upregulation threshold, adopted for further analysis. (D) Venn diagram revealed 13 proteins commonly present in datasets I and II.

13 cell-adhesion regulators, not belonging in well-established cell adhesion networks, and having portrayed significant overexpression in CRC at the mRNA level.

3.5. Expression levels of four cell adhesion mediators in serum of colon cancer patients

Out of these 13 proteins, we randomly selected 4 [i.e. gremlin-1 (GREM1), lysyl oxidase homolog-2 (LOXL2), olfactomedin-4 (OLFM4), and carbonic anhydrase-9 (CA9)] for further investigations. All these proteins have been previously linked to the development and progression of various cancers, including CRC [52,79,85–94]. Also, two of these, CA9 and LOXL2 have been investigated through IHC in a variety of cancer tissues, including CRC, and they have Human Protein Atlas entries (Supplementary Fig. 2). Expression levels of GREM1, LOXL2, OLFM4 and CA9 were measured with enzyme-linked immunosorbent assays in the serum of a small cohort of 28 individuals (10 patients with sporadic CRC; 8 patients with benign masses; 10 healthy individuals). Out of the 4 tested proteins, only OLFM4 demonstrated statistically significant ($p < 0.0125$; Holms-corrected Jonckheere–Terpstra test) differential secretion among the three groups. Specifically, OLFM4 showed a significant ($p = 0.003$; Mann–Whitney U test) elevation in cancer patients (mean = 336.7 pg/mL) in

comparison to healthy controls (mean = 119.2 pg/mL) (Fig. 5A). OLFM4 could not discriminate between cancer patients and patients with benign conditions (mean = 211.1 pg/mL) (Fig. 5A). However, there was a statistically significant ($p = 0.013$; Mann–Whitney U test) elevation in patients with benign conditions in comparison to healthy controls (Fig. 5A). In CM, OLFM4 was identified in the LS174T, LS180, Colo320HSR and Colo205 cell lines with >10 spectra (Supplementary Table 2).

3.6. Olfactomedin-4 serum levels may predict colon cancer incidence

At present, CEA is the most widely-used serological CRC biomarker [82], and it was also assessed in this screening cohort. The AUC of CEA alone was 0.694, while the AUC of OLFM4 was 0.822 (Fig. 5B). Based on this, we exploited the possibility that OLFM4 held any information for predicting disease progression in the case of CRC. An a priori goodness-of-fit test demonstrated that our ELISA data were well-calibrated ($p = 0.362$; Hosmer–Lemeshow test), and subsequent logistic regression showed a statistically significant ($p = 0.02$) relationship between OLFM4 serum levels and cancer occurrence (Fig. 5B). As detailed in the scatter-plot, OLFM4 serum levels could identify two patients with CRC, which did not show any CEA elevation (Fig. 5C). Thus, we

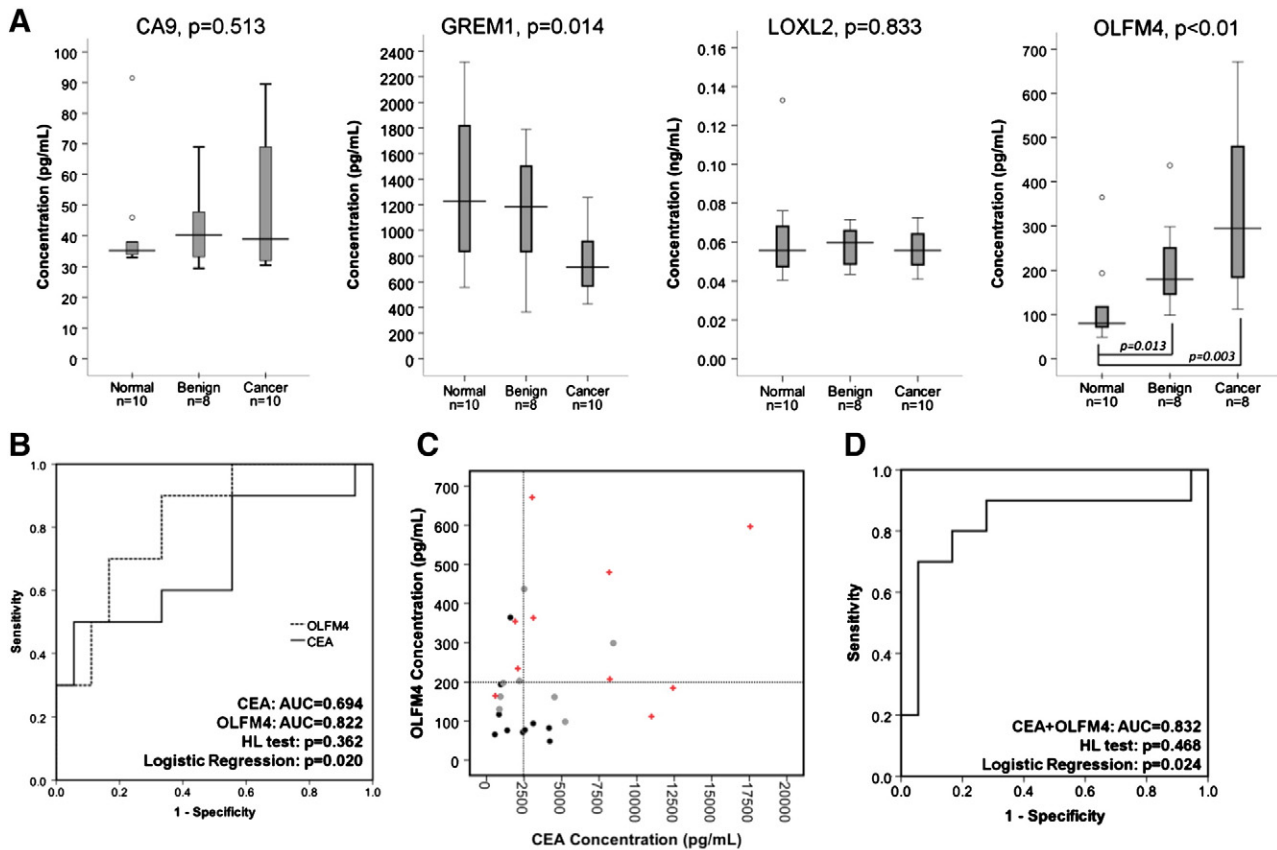


Fig. 5 – Biomarker verification experiments. (A) Investigation of CA9, LOXL2, GREM1 and OLFM4 in serum from CRC patients of the sporadic subtype, healthy controls and patients with benign lesions. The p-value after the protein's name refers to the Holms-corrected, Jonckheere–Terpstra test for comparison of mean protein levels across all three groups of individuals; n, number of individuals in each group. **(B)** ROC curve demonstrating the potential of CEA or OLFM4 plasma levels alone, in discriminating between colon cancer and healthy individuals. Area under the curve (AUC); p-values from Hosmer–Lemeshow (HL) test and logistic regression are described within the box. **(C)** Scatter-plot demonstrating the efficiency of combination of OLFM4 and CEA in discriminating cancer cases from controls. Red crosses, cancer patients; black circles, healthy individuals; gray circles, benign individuals. **(D)** ROC curve demonstrating the increased potential of CEA and OLFM4 in combination, in discriminating between colon cancer and healthy individuals. Area under the curve (AUC); p-values from Hosmer–Lemeshow (HL) test and logistic regression are described within the box.

speculated that OLFM4 could potentially complement CEA in discriminating CRC patients from healthy and/or benign controls. To directly test this, we repeated logistic regression analysis ($p = 0.024$), and found that, in contrast to OLFM4 alone, the combination of OLFM4 and CEA serum levels could more accurately discriminate between cases and controls (AUC = 0.832) (Fig. 5D).

4. Discussion

The proteome presented in this article is, to our knowledge, one of the largest and most comprehensive secreted proteomes to date, for colorectal cancer in a single study. Our group has contributed in the field with many studies on breast, lung, ovarian, prostate and pancreatic [67–73] cancers, and this study provides further contribution for colon cancer. Of note, the progressive technological advancement of mass spectrometers over the past years has allowed for particularly

thorough and comprehensive proteomic analyses. In this study, we identified approximately 10 times higher number of proteins compared to two relevant studies on CRC exoproteome, published less than half a decade ago [35,36], which clearly demonstrates that technological advancements in the field might result in more comprehensive outcomes.

Although certain differences exist, the genomic and transcriptional profiles of cancer cell lines have been indicated to recapitulate the most leading features of primary tumors [95–97]. The identification of many known cancer biomarkers in the conditioned media of cancer cell lines supports that it is a representative source to mine, despite the fact that it lacks the contribution of the tumor microenvironment [31]. For instance, the most well-known cancer biomarker for CRC, carcinoembryonic antigen (CEA), was successfully identified and quantified (with up to 50 spectra in certain cell lines; Supplementary Tables 1 & 2) in our analysis. A more direct validation for the accuracy of our approach came through our correlation experiments, using a panel of kallikrein-related

peptidases (KLK6, KLK7, KLK10 and KLK11) as internal controls. It was efficiently shown that the actual concentration levels for all these KLKs, as assessed through highly-sensitive and -specific ELISAs, correlated well with the spectral counts obtained from the respective MS/MS. In addition, two of these KLKs, KLK6 and KLK10, showed good correlation in their expression levels, when investigated in clinical cases of CRC patients. Thus, despite the fact that cancer cell lines have generally received considerable criticism regarding their validity in clinical investigations, our data suggest that there is a certain degree of reliability, when they are used as a discovery platform.

Currently, pathway analysis is pivotal for conceptualizing molecular pathways in large proteomic or transcriptomic datasets, especially when derived from mass spectrometry-based and microarray experiments [98–105]. However, network analysis needs to be performed with vigilance to ensure data reproducibility, robustness and enhanced sensitivity. Our previous experience on such analyses [52,77,106,107] has allowed us to propose an integrative 2-step model when using such software in these pipelines. We first propose the application of generic tools, capable of providing dataset enrichment in specific molecular concepts and/or pathways, always coupled to a visualization map (e.g. molecular concept maps, gene set enrichment map). Subsequent pathway analysis tools, particularly those based on protein–protein interaction algorithms (such as Ingenuity or STRING), will further allow for an in-depth analysis and the identification of molecular hubs, quite promising for holding translational or clinical importance.

Using a 3-step prioritization process outlined in the **Results** section (Fig. 4A), 13 cell-adhesion regulators were found as pertinent markers for CRC progression. Notably, included in this list were: gremlin-1 (GREM1), a bone-morphogenetic protein antagonist that has been previously associated with the maintenance and self-renewal of cancer stem cell niches [94], as well as with CRC progression [52]; nucleotide diphosphate kinase A (NME1), a proposed metastasis suppressor [108]; insulin growth factor binding protein-7 (IGFBP7), a member of the IGF superfamily of proteins that binds to and inhibits IGFs, and which has been epigenetically linked to CRC metastasis [109–111]; carbonic anhydrase IX (CA9), an enzyme belonging to the family of carbonic anhydrases (zinc metalloenzymes that catalyze the reversible hydration of carbon dioxide), which has been suggested as prominent marker of hypoxia [86] and tumor progression [112], including CRC [85]; lysyl-oxidase like-2 (LOXL2), a LOX family gene that serves as collagen organizer during ECM remodeling and progression of cancer [79]; versican (VCAN), a large ECM proteoglycan involved in cell adhesion [113]; Zn-alpha 2-glycoprotein (AZGP1), a protein physiologically involved in lipolysis and reduction of body fat stores [114] that has also been linked to cancer development and progression, especially prostate cancer [115]; sushi repeat-containing protein 2 (SRPX2), a mediator originally identified in speech-related disorders [116], that has been recently linked to gastrointestinal cancer progression [117] and angiogenesis [118]; olfactomedin-4 (OLFM4), a member of the olfactomedin family of proteins, originally found in inflamed colonic epithelia [119–121].

Of these, four proteins (CA9, GREM1, LOXL2, OLFM4) were investigated in a clinical cohort and OLFM4 showed significant perturbation in CRC patients. OLFM4 mainly has an antiapoptotic function that promotes tumor growth and is an extracellular matrix glycoprotein that facilitates cell adhesion [122]. It has been shown that specific overexpression of OLFM4 mRNA level has been reported in colon, breast and lung cancers [123], and, in this study, we have additionally confirmed that such overexpression is also reflected systematically in CRC patients at the protein level. OLFM4 could be potentially utilized along with a panel of CRC-associated proteins to complement the TNM system for classification of colorectal cancers [124], as it has been shown that it is a marker for the differentiation and progression of colorectal cancers [125,126]. Another recent study has established OLFM4 as a novel non-metastatic tumor marker in CRC [126]. In a complementary fashion, we have shown that OLFM4 could successfully discriminate between cancerous and healthy, as well as between benign and healthy states in colorectal cancer, using a logistic regression model. OLFM4 has captured additional interest as a putative marker in a few other settings, such as: (I) marker for defining subsets of neutrophils [127], (II) serum-based biomarker for diagnosis [128] or prognosis [129] of gastric cancer, (III) biomarker for pancreatic [67] and cervical [130] cancer, as well as (IV) stem cell marker [122]. Given that OLFM4 may play mechanistic and regulatory roles in the development and progression of a variety of cancers [89,131–133], these observations collectively demonstrate that efforts should continue towards investigating the expression of OLFM4 in larger patient cohorts.

In this study, OLFM4 performed better alone, when compared to the already established marker CEA in the same patient cohort (AUC: 0.822 vs 0.694). A growing consensus in the field of biomarker discovery is towards development of panels of biomarkers, as the co-assessment of multiple molecules can result in increased sensitivity and specificity, in comparison to molecules individually [67]. Interestingly, co-assessment of CEA and OLFM4 presented with slightly better AUC compared to OLFM4 alone (AUC: 0.832 vs 0.822). However, it should be noted that this verification study was performed in only a limited number of randomly-selected cases of CRC. We used a few samples, since the goal of our current study was to determine the utility of our approach to identify proteins, whose increase is reflected in serum/plasma of CRC patients. Following the paradigm of Makawita et al. [67], we conducted this study in CRC to preliminary examine if such candidates could be informative. Our marker panel requires further validation with samples that may have normal or even low CEA levels and include patients with early-stage disease.

The current state of oncoproteomics offers a large number of discovery studies resulting in a vast plethora of proposed biomarkers for diagnostic, prognostic and therapeutic leads. However, most of them are poorly reproduced or validated, partly as a result of lacking quantitative aspects, and partly as a result of lacking standardized protocols (i.e. most of such proteomic pipelines tend to be in-house optimized, thus considered laboratory-dependent). In this view, our comprehensive CRC exoproteome should be extensively — but carefully — mined in the future and especially upon combination with

existing repositories or compendiums, further prioritization of promising candidates might be more safely pursued.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.jprot.2014.03.018>.

Conflict of interest

The authors would like to report no conflicts of interest.

Author contributions

Conceptualized study: G.S.K. and E.P.D. Designed, performed and interpreted proteomics analyses: G.S.K., M.P.P., P.S., N.M., I.B. and I.P. Designed, performed and interpreted bioinformatics analyses: G.S.K., A.X. and A.D. Designed, performed and interpreted IHC studies: G.S.K. and C.P. Wrote manuscript: G.S.K. Revised manuscript: G.S.K., M.P., A.X., P.S., N.M., I.B., I.P., A.D., C.P. and E.P.D.

Acknowledgments

We would like to thank Antoninus Soosaipillai, Shalini Makawita and Chris Smith for advice and technical assistance.

REFERENCES

- [1] Haggard FA, Boushey RP. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clin Colon Rectal Surg* 2009;22:191–7.
- [2] Jemal A, Bray F, Center MM, Ferlay J, et al. Global cancer statistics. *CA Cancer J Clin* 2011;61:69–90.
- [3] Zavoral M, Suchanek S, Zavada F, Dusek L, et al. Colorectal cancer screening in Europe. *World J Gastroenterol* 2009;15:5907–15.
- [4] Erdman SE, Poutahidis T. Roles for inflammation and regulatory T cells in colon cancer. *Toxicol Pathol* 2009;38:76–87.
- [5] Fox JG, Ge Z, Whary MT, Erdman SE, Horwitz BH. *Helicobacter hepaticus* infection in mice: models for understanding lower bowel inflammation and cancer. *Mucosal Immunol* 2010;4:22–30.
- [6] Galiatsatos P, Foulkes WD. Familial adenomatous polyposis. *Am J Gastroenterol* 2006;101:385–98.
- [7] Half E, Bercovich D, Rozen P. Familial adenomatous polyposis. *Orphanet J Rare Dis* 2009;4:22.
- [8] Lynch HT, Smyrk T. Hereditary nonpolyposis colorectal cancer (Lynch syndrome). An updated review. *Cancer* 1996;78:1149–67.
- [9] Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell* 1990;61:759–67.
- [10] Markowitz SD, Bertagnolli MM. Molecular origins of cancer: molecular basis of colorectal cancer. *N Engl J Med* 2009;361:2449–60.
- [11] Ilyas M, Straub J, Tomlinson IP, Bodmer WF. Genetic pathways in colorectal and other cancers. *Eur J Cancer* 1999;35:335–51.
- [12] Al-Sohaily S, Biankin A, Leong R, Kohonen-Corish M, Warusavitarne J. Molecular pathways in colorectal cancer. *J Gastroenterol Hepatol* 2012;27:1423–31.
- [13] Smith G, Carey FA, Beattie J, Wilkie MJ, et al. Mutations in APC, Kirsten-ras, and p53-alternative genetic pathways to colorectal cancer. *Proc Natl Acad Sci U S A* 2002;99:9433–8.
- [14] Gryfe R, Swallow C, Bapat B, Redston M, et al. Molecular biology of colorectal cancer. *Curr Probl Cancer* 1997;21:233–300.
- [15] Bright-Thomas RM, Hargest R. APC, beta-Catenin and hTCF-4; an unholy trinity in the genesis of colorectal cancer. *Eur J Surg Oncol* 2003;29:107–17.
- [16] Hedrick L, Cho KR, Fearon ER, Wu TC, et al. The DCC gene product in cellular differentiation and colorectal tumorigenesis. *Genes Dev* 1994;8:1174–83.
- [17] Russo A, Bazan V, Iacopetta B, Kerr D, et al. The TP53 colorectal cancer international collaborative study on the prognostic and predictive significance of p53 mutation: influence of tumor site, type of mutation, and adjuvant treatment. *J Clin Oncol* 2005;23:7518–28.
- [18] Boland CR, Goel A. Microsatellite instability in colorectal cancer. *Gastroenterology* 2010;138:2073–87 [e2073].
- [19] Stone JG, Tomlinson IP, Houlston RS. Optimising methods for determining RER status in colorectal cancers. *Cancer Lett* 2000;149:15–20.
- [20] Karagiannis GS, Pavlou MP, Diamandis EP. Cancer secretomics reveal pathophysiological pathways in cancer molecular oncology. *Mol Oncol* 2010;4:496–510.
- [21] Kulasingam V, Pavlou MP, Diamandis EP. Integrating high-throughput technologies in the quest for effective biomarkers for ovarian cancer. *Nat Rev Cancer* 2010;10:371–8.
- [22] Jain KK. Innovations, challenges and future prospects of oncoproteomics. *Mol Oncol* 2008;2:153–60.
- [23] Tjalsma H, Bolhuis A, Jongbloed JD, Bron S, van Dijk JM. Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. *Microbiol Mol Biol Rev* 2000;64:515–47.
- [24] Antelmann H, Tjalsma H, Voigt B, Ohlmeier S, et al. A proteomic view on genome-based signal peptide predictions. *Genome Res* 2001;11:1484–502.
- [25] Desvaux M, Hebraud M, Talon R, Henderson IR. Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue. *Trends Microbiol* 2009;17:139–45.
- [26] Desvaux M, Hebraud M, Talon R, Henderson IR. Outer membrane translocation: numerical protein secretion nomenclature in question in mycobacteria. *Trends Microbiol* 2009;17:338–40.
- [27] Armengaud J, Christie-Olea JA, Clair G, Malard V, Duport C. Exoproteomics: exploring the world around biological systems. *Expert Rev Proteomics* 2012;9:561–75.
- [28] Chagnot C, Zorgani MA, Astruc T, Desvaux M. Proteinaceous determinants of surface colonization in bacteria: bacterial adhesion and biofilm formation from a protein secretion perspective. *Front Microbiol* 2013;4:303.
- [29] Tjalsma H. Feature-based reappraisal of the *Bacillus subtilis* exoproteome. *Proteomics* 2007;7:73–81.
- [30] Pavlou MP, Diamandis EP. The cancer cell secretome: a good source for discovering biomarkers? *J Proteomics* 1896–1906;2010:73.
- [31] Kulasingam V, Diamandis EP. Tissue culture-based breast cancer biomarker discovery platform. *Int J Cancer* 2008;123:2007–12.
- [32] Piersma SR, Fiedler U, Span S, Lingnau A, et al. Workflow comparison for label-free, quantitative secretome proteomics for cancer biomarker discovery: method evaluation, differential analysis, and verification in serum. *J Proteome Res* 1913–1922;2010:9.
- [33] Jimenez CR, Piersma S, Pham TV. High-throughput and targeted in-depth mass spectrometry-based approaches for biofluid profiling and biomarker discovery. *Biomark Med* 2007;1:541–65.

- [34] Schaaij-Visser TB, de Wit M, Lam SW, Jimenez CR. The cancer secretome, current status and opportunities in the lung, breast and colorectal cancer context. *Biochim Biophys Acta* 1834;2013:2242–58.
- [35] Wu CC, Chen HC, Chen SJ, Liu HP, et al. Identification of collapsin response mediator protein-2 as a potential marker of colorectal carcinoma by comparative analysis of cancer cell secretomes. *Proteomics* 2008;8:316–32.
- [36] Xue H, Lu B, Zhang J, Wu M, et al. Identification of serum biomarkers for colorectal cancer metastasis using a differential secretome approach. *J Proteome Res* 2009;9:545–55.
- [37] Wu CC, Hsu CW, Chen CD, Yu CJ, et al. Candidate serological biomarkers for cancer identified from the secretomes of 23 cancer cell lines and the human protein atlas. *Mol Cell Proteomics* 2010;9:1100–17.
- [38] Barderas R, Mendes M, Torres S, Bartolome RA, et al. In-depth characterization of the secretome of colorectal cancer metastatic cells identifies key proteins in cell adhesion, migration, and invasion. *Mol Cell Proteomics* 2013;12:1602–20.
- [39] Zeng X, Yang P, Chen B, Jin X, et al. Quantitative secretome analysis reveals the interactions between epithelia and tumor cells by in vitro modulating colon cancer microenvironment. *J Proteomics* 2013;89:51–70.
- [40] Emmink BL, Verheem A, Van Houdt WJ, Steller EJ, et al. The secretome of colon cancer stem cells contains drug-metabolizing enzymes. *J Proteomics* 2013;91:84–96.
- [41] Van Houdt WJ, Emmink BL, Pham TV, Piersma SR, et al. Comparative proteomics of colon cancer stem cells and differentiated tumor cells identifies BIRC6 as a potential therapeutic target. *Mol Cell Proteomics* 2011;10 [M111 011353].
- [42] Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* 2007;4:207–14.
- [43] Choi H, Nesvizhskii AI. False discovery rates and related statistical concepts in mass spectrometry-based proteomics. *J Proteome Res* 2008;7:47–50.
- [44] Zhu W, Smith JW, Huang CM. Mass spectrometry-based label-free quantitative proteomics. *J Biomed Biotechnol* 2010;840518.
- [45] Liu H, Sadygov RG, Yates III JR. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* 2004;76:4193–201.
- [46] Bachi A, Bonaldi T. Quantitative proteomics as a new piece of the systems biology puzzle. *J Proteomics* 2008;71:357–67.
- [47] Collier TS, Randall SM, Sarkar P, Rao BM, et al. Comparison of stable-isotope labeling with amino acids in cell culture and spectral counting for relative quantification of protein expression. *Rapid Commun Mass Spectrom* 2011;25:2524–32.
- [48] Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004;340:783–95.
- [49] Nielsen H, Engelbrecht J, Brunak S, von Heijne G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* 1997;10:1–6.
- [50] Bendtsen JD, Jensen LJ, Blom N, Von Heijne G, Brunak S. Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng Des Sel* 2004;17:349–56.
- [51] Moller S, Croning MD, Apweiler R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* 2001;17:646–53.
- [52] Karagiannis GS, Berk A, Dimitromanolakis A, Diamandis EP. Enrichment map profiling of the cancer invasion front suggests regulation of colorectal cancer progression by the bone morphogenetic protein antagonist, gremlin-1. *Mol Oncol* 2013;7:826–39.
- [53] Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 2005;21:3448–9.
- [54] Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* 2010;5:e13984.
- [55] Merico D, Isserlin R, Bader GD. Visualizing gene-set enrichment results using the Cytoscape plug-in enrichment map. *Methods Mol Biol* 2011;781:257–77.
- [56] Franceschini A, Szklarczyk D, Frankild S, Kuhn M, et al. STRING v9.1: protein–protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* 2012;41:D808–15.
- [57] Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 2002;30:207–10.
- [58] Parkinson H, Kapushesky M, Kolesnikov N, Rustici G, et al. ArrayExpress update—from an archive of functional genomics experiments to the atlas of gene expression. *Nucleic Acids Res* 2009;37:D868–72.
- [59] Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003;4:249–64.
- [60] Lukk M, Kapushesky M, Nikkila J, Parkinson H, et al. A global map of human gene expression. *Nat Biotechnol* 2010;28:322–4.
- [61] Luo LY, Soosaipillai A, Grass L, Diamandis EP. Characterization of human kallikreins 6 and 10 in ascites fluid from ovarian cancer patients. *Tumour Biol* 2006;27:227–34.
- [62] Shaw JL, Diamandis EP. Distribution of 15 human kallikreins in tissues and biological fluids. *Clin Chem* 2007;53:1423–32.
- [63] Petraki C, Dubinski W, Scorilas A, Saleh C, et al. Evaluation and prognostic significance of human tissue kallikrein-related peptidase 6 (KLK6) in colorectal cancer. *Pathol Res Pract* 2012;208:104–8.
- [64] Petraki CD, Karavana VN, Luo LY, Diamandis EP. Human kallikrein 10 expression in normal tissues by immunohistochemistry. *J Histochem Cytochem* 2002;50:1247–61.
- [65] Ruckert F, Hennig M, Petraki CD, Wehrum D, et al. Co-expression of KLK6 and KLK10 as prognostic factors for survival in pancreatic ductal adenocarcinoma. *Br J Cancer* 2008;99:1484–92.
- [66] Petraki C, Youssef YM, Dubinski W, Lichner Z, et al. Evaluation and prognostic significance of human tissue kallikrein-related peptidase 10 (KLK10) in colorectal cancer. *Tumour Biol* 2012;33:1209–14.
- [67] Makawita S, Smith C, Batruch I, Zheng Y, et al. Integrated proteomic profiling of cell line conditioned media and pancreatic juice for the identification of pancreatic cancer biomarkers. *Mol Cell Proteomics* 2011;10 [M111 008599].
- [68] Sardana G, Jung K, Stephan C, Diamandis EP. Proteomic analysis of conditioned media from the PC3, LNCaP, and 22Rv1 prostate cancer cell lines: discovery and validation of candidate prostate cancer biomarkers. *J Proteome Res* 2008;7:3329–38.
- [69] Gunawardana CG, Kuk C, Smith CR, Batruch I, et al. Comprehensive analysis of conditioned media from ovarian cancer cell lines identifies novel candidate markers of epithelial ovarian cancer. *J Proteome Res* 2009;8:4705–13.
- [70] Kulasingam V, Diamandis EP. Proteomics analysis of conditioned media from three breast cancer cell lines: a mine for biomarkers and therapeutic targets. *Mol Cell Proteomics* 1997–2011;2007:6.
- [71] Planque C, Kulasingam V, Smith CR, Reckamp K, et al. Identification of five candidate lung cancer biomarkers by

- proteomics analysis of conditioned media of four lung cancer cell lines. *Mol Cell Proteomics* 2009;8:2746–58.
- [72] Saraon P, Musrap N, Cretu D, Karagiannis GS, et al. Proteomic profiling of androgen-independent prostate cancer cell lines reveals a role for protein S during the development of high grade and castration-resistant prostate cancer. *J Biol Chem* 2012;287:34019–31.
- [73] Pavlou MP, Dimitromanolakis A, Diamandis EP. Coupling proteomics and transcriptomics in the quest of subtype-specific proteins in breast cancer. *Proteomics* 2013;13:1083–95.
- [74] Karagiannis GS, Petraki C, Prassas I, Saraon P, et al. Proteomic signatures of the desmoplastic invasion front reveal collagen type XII as a marker of myofibroblastic differentiation during colorectal cancer metastasis. *Oncotarget* 2012;3:267–85.
- [75] Borgono CA, Diamandis EP. The emerging roles of human tissue kallikreins in cancer. *Nat Rev Cancer* 2004;4:876–90.
- [76] Borgono CA, Michael IP, Diamandis EP. Human tissue kallikreins: physiologic roles and applications in cancer. *Mol Cancer Res* 2004;2:257–80.
- [77] Karagiannis GS, Weile J, Bader GD, Minta J. Integrative pathway dissection of molecular mechanisms of moxLDL-induced vascular smooth muscle phenotype transformation. *BMC Cardiovasc Disord* 2013;13:4.
- [78] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011;144:646–74.
- [79] Karagiannis GS, Poutahidis T, Erdman SE, Kirsch R, et al. Cancer-associated fibroblasts drive the progression of metastasis through both paracrine and mechanical pressure on cancer tissue. *Mol Cancer Res* 2012;10:1403–18.
- [80] Doucet A, Butler GS, Rodríguez D, Prudova A, Overall CM. Metadegradomics: toward in vivo quantitative degradomics of proteolytic post-translational modifications of the cancer proteome. *Mol Cell Proteomics* 2008;7:1925–51.
- [81] Doucet A, Overall CM. Protease proteomics: revealing protease in vivo functions using systems biology approaches. *Mol Aspects Med* 2008;29:339–58.
- [82] Collins JJ, Black PH. Specificity of the carcinoembryonic antigen (CEA). *N Engl J Med* 1971;285:175–6.
- [83] Dallas MR, Liu G, Chen WC, Thomas SN, et al. Divergent roles of CD44 and carcinoembryonic antigen in colon cancer metastasis. *FASEB J* 2012;26:2648–56.
- [84] Thomas SN, Zhu F, Schnaar RL, Alves CS, Konstantopoulos K. Carcinoembryonic antigen and CD44 variant isoforms cooperate to mediate colon carcinoma cell adhesion to E- and L-selectin in shear flow. *J Biol Chem* 2008;283:15647–55.
- [85] Kivela AJ, Parkkila S, Saarnio J, Karttunen TJ, et al. Expression of von Hippel–Lindau tumor suppressor and tumor-associated carbonic anhydrases IX and XII in normal and neoplastic colorectal mucosa. *World J Gastroenterol* 2005;11:2616–25.
- [86] Cleven AH, Wouters BG, Schutte B, Spiertz AJ, et al. Poorer outcome in stromal HIF-2 alpha- and CA9-positive colorectal adenocarcinomas is associated with wild-type TP53 but not with BNIP3 promoter hypermethylation or apoptosis. *Br J Cancer* 2008;99:727–33.
- [87] Sansone P, Piazza G, Paterini P, Strillacci A, et al. Cyclooxygenase-2/carbonic anhydrase-IX up-regulation promotes invasive potential and hypoxia survival in colorectal cancer cells. *J Cell Mol Med* 2009;13:3876–87.
- [88] Smyth LG, O'Hurley G, O'Grady A, Fitzpatrick JM, et al. Carbonic anhydrase IX expression in prostate cancer. *Prostate Cancer Prostatic Dis* 2009;13:178–81.
- [89] Kobayashi D, Koshida S, Moriai R, Tsuji N, Watanabe N. Olfactomedin 4 promotes S-phase transition in proliferation of pancreatic cancer cells. *Cancer Sci* 2007;98:334–40.
- [90] Fong SF, Dietzsch E, Fong KS, Hollosi P, et al. Lysyl oxidase-like 2 expression is increased in colon and esophageal tumors and associated with less differentiated colon tumors. *Genes Chromosomes Cancer* 2007;46:644–55.
- [91] Hollosi P, Yakushiji JK, Fong KS, Csiszar K, Fong SF. Lysyl oxidase-like 2 promotes migration in noninvasive breast cancer cells but not in normal breast epithelial cells. *Int J Cancer* 2009;125:318–27.
- [92] Peinado H, Portillo F, Cano A. Switching on–off Snail: LOXL2 versus GSK3beta. *Cell Cycle* 2005;4:1749–52.
- [93] Mulvihill MS, Kwon YW, Lee S, Fang LT, et al. Gremlin is overexpressed in lung adenocarcinoma and increases cell growth and proliferation in normal lung cells. *PLoS One* 2012;7:e42264.
- [94] Sneddon JB, Zhen HH, Montgomery K, van de Rijn M, et al. Bone morphogenetic protein antagonist gremlin 1 is widely expressed by cancer-associated stromal cells and can promote tumor cell proliferation. *Proc Natl Acad Sci U S A* 2006;103:14842–7.
- [95] Wistuba II, Behrens C, Milchgrub S, Syed S, et al. Comparison of features of human breast cancer cell lines and their corresponding tumors. *Clin Cancer Res* 1998;4:2931–8.
- [96] Wistuba II, Bryant D, Behrens C, Milchgrub S, et al. Comparison of features of human lung cancer cell lines and their corresponding tumors. *Clin Cancer Res* 1999;5:991–1000.
- [97] Saraon P, Cretu D, Musrap N, Karagiannis GS, et al. Quantitative proteomics reveals that enzymes of the ketogenic pathway are associated with prostate cancer progression. *Mol Cell Proteomics* 2013;12:1589–601.
- [98] Tiger CF, Krause F, Cedersund G, Palmer R, et al. A framework for mapping, visualisation and automatic model creation of signal-transduction networks. *Mol Syst Biol* 2012;8:578.
- [99] Hanauer DA, Rhodes DR, Chinnaiyan AM. Exploring clinical associations using ‘-omics’ based enrichment analyses. *PLoS One* 2009;4:e5203.
- [100] Bode C, Kovacs IA, Szalay MS, Palotai R, et al. Network analysis of protein dynamics. *FEBS Lett* 2007;581:2776–82.
- [101] Chu LH, Chen BS. Construction of a cancer-perturbed protein–protein interaction network for discovery of apoptosis drug targets. *BMC Syst Biol* 2008;2:56.
- [102] Csermely P, Korcsmaros T, Kovacs IA, Szalay MS, Soti C. Systems biology of molecular chaperone networks. *Novartis Found Symp* 2008;291:45–54 [discussion 54–48, 137–140].
- [103] He X, Zhang J. Why do hubs tend to be essential in protein networks? *PLoS Genet* 2006;2:e88.
- [104] Levy SF, Siegal ML. Network hubs buffer environmental variation in *Saccharomyces cerevisiae*. *PLoS Biol* 2008;6:e264.
- [105] Wang Y, Zhang XS, Chen L. A network biology study on circadian rhythm by integrating various omics data. *OMICS* 2009;13:313–24.
- [106] Beck SE, Jung BH, Fiorino A, Gomez J, et al. Bone morphogenetic protein signaling and growth suppression in colon cancer. *Am J Physiol Gastrointest Liver Physiol* 2006;291:G135–45.
- [107] Prassas I, Karagiannis GS, Batruch I, Dimitromanolakis A, et al. Digitoxin-induced cytotoxicity in cancer cells is mediated through distinct kinase and interferon signaling networks. *Mol Cancer Ther* 2011;10:2083–93.
- [108] Dooley S, Seib T, Engel M, Theisinger B, et al. Isolation and characterization of the human genomic locus coding for the putative metastasis control gene nm23-H1. *Hum Genet* 1994;93:63–6.
- [109] Georges RB, Adwan H, Hamdi H, Hielscher T, et al. The insulin-like growth factor binding proteins 3 and 7 are associated with colorectal cancer and liver metastasis. *Cancer Biol Ther* 2011;12:69–79.

- [110] Ruan W, Xu E, Xu F, Ma Y, et al. IGFBP7 plays a potential tumor suppressor role in colorectal carcinogenesis. *Cancer Biol Ther* 2007;6:354–9.
- [111] Ruan WJ, Lin J, Xu EP, Xu FY, et al. IGFBP7 plays a potential tumor suppressor role against colorectal carcinogenesis with its expression associated with DNA hypomethylation of exon 1. *J Zhejiang Univ Sci B* 2006;7:929–32.
- [112] Kirkpatrick JP, Rabbani ZN, Bentley RC, Hardee ME, et al. Elevated CAIX expression is associated with an increased risk of distant failure in early-stage cervical cancer. *Biomark Insights* 2008;3:45–55.
- [113] Zimmermann DR, Ruoslahti E. Multiple domains of the large fibroblast proteoglycan, versican. *EMBO J* 1989;8:2975–81.
- [114] Tisdale MJ. Zinc-alpha2-glycoprotein in cachexia and obesity. *Curr Opin Support Palliat Care* 2009;3:288–93.
- [115] Hassan MI, Waheed A, Yadav S, Singh TP, Ahmad F. Zinc alpha 2-glycoprotein: a multidisciplinary protein. *Mol Cancer Res* 2008;6:892–906.
- [116] Roll P, Vernes SC, Bruneau N, Cillario J, et al. Molecular networks implicated in speech-related disorders: FOXP2 regulates the SRPX2/uPAR complex. *Hum Mol Genet* 2010;19:4848–60.
- [117] Tanaka K, Arai T, Maegawa M, Matsumoto K, et al. SRPX2 is overexpressed in gastric cancer and promotes cellular migration and adhesion. *Int J Cancer* 2009;124:1072–80.
- [118] Miljkovic-Licina M, Hammel P, Garrido-Urbani S, Bradfield PF, et al. Sushi repeat protein X-linked 2, a novel mediator of angiogenesis. *FASEB J* 2009;23:4105–16.
- [119] Gersemann M, Becker S, Nuding S, Antoni L, et al. Olfactomedin-4 is a glycoprotein secreted into mucus in active IBD. *J Crohns Colitis* 2012;6:425–34.
- [120] Liu W, Yan M, Liu Y, Wang R, et al. Olfactomedin 4 down-regulates innate immunity against *Helicobacter pylori* infection. *Proc Natl Acad Sci U S A* 2010;107:11056–61.
- [121] Liu W, Yan M, Sugui JA, Li H, et al. Olfm4 deletion enhances defense against *Staphylococcus aureus* in chronic granulomatous disease. *J Clin Invest* 2013;123:3751–5.
- [122] Grover PK, Hardingham JE, Cummins AG. Stem cell marker olfactomedin 4: critical appraisal of its characteristics and role in tumorigenesis. *Cancer Metastasis Rev* 2010;29:761–75.
- [123] Koshida S, Kobayashi D, Moriai R, Tsuji N, Watanabe N. Specific overexpression of OLFM4(GW112/HGC-1) mRNA in colon, breast and lung cancer tissues detected using quantitative analysis. *Cancer Sci* 2007;98:315–20.
- [124] Sawada T, Yashiro M, Sentani K, Oue N, et al. New molecular staging with G-factors (VEGF-C and Reg IV) by supplementing TNM classification in colorectal cancers. *Oncol Rep* 2013;30:2609–16.
- [125] Yu L, Wang L, Chen S. Olfactomedin 4, a novel marker for the differentiation and progression of gastrointestinal cancers. *Neoplasma* 2011;58:9–13.
- [126] Besson D, Pavageau AH, Valo I, Bourreau A, et al. A quantitative proteomic approach of the different stages of colorectal cancer establishes OLFM4 as a new nonmetastatic tumor marker. *Mol Cell Proteomics* 2011;10 [M111 009712].
- [127] Clemmensen SN, Bohr CT, Rorvig S, Glenthoj A, et al. Olfactomedin 4 defines a subset of human neutrophils. *J Leukoc Biol* 2012;91:495–500.
- [128] Oue N, Sentani K, Noguchi T, Ohara S, et al. Serum olfactomedin 4 (GW112, hGC-1) in combination with Reg IV is a highly sensitive biomarker for gastric cancer patients. *Int J Cancer* 2009;125:2383–92.
- [129] Luo Z, Zhang Q, Zhao Z, Li B, et al. OLFM4 is associated with lymph node metastasis and poor prognosis in patients with gastric cancer. *J Cancer Res Clin Oncol* 2011;137:1713–20.
- [130] Yu L, He M, Yang Z, Chen G, et al. Olfactomedin 4 is a marker for progression of cervical neoplasia. *Int J Gynecol Cancer* 2011;21:367–72.
- [131] Navab R, Strumpf D, Bandarchi B, Zhu CQ, et al. Prognostic gene-expression signature of carcinoma-associated fibroblasts in non-small cell lung cancer. *Proc Natl Acad Sci U S A* 2011;108:7160–5.
- [132] Chen L, Li H, Liu W, Zhu J, et al. Olfactomedin 4 suppresses prostate cancer cell growth and metastasis via negative interaction with cathepsin D and SDF-1. *Carcinogenesis* 2011;32:986–94.
- [133] Park KS, Kim KK, Piao ZH, Kim MK, et al. Olfactomedin 4 suppresses tumor growth and metastasis of mouse melanoma cells through downregulation of integrin and MMP genes. *Mol Cells* 2012;34:555–61.