

# Subcellular proteomics analysis of different stages of colorectal cancer cell lines.

*Alex-Ane Mathieu, Emma Ohl-Séguy, Marie-Line Dubois, Dominique Jean, Christine Jones, François*

*Boudreau and François-Michel Boisvert<sup>1</sup>*

<sup>1</sup> Corresponding author: François-Michel Boisvert

Department of Anatomy and Cell Biology, Université de Sherbrooke, 3201 Jean-Mignault,

Sherbrooke, Québec, J1E 4K8, Canada

Phone: 819-821-8000 Ext.75430

Fax: 819-820-6831

Email: [Francois.Michel.Boisvert@USherbrooke.ca](mailto:Francois.Michel.Boisvert@USherbrooke.ca)

KEYWORDS: Colorectal cancer, Mass Spectrometry, SILAC, Subcellular proteomics

Received: 25/07/2016; Revised: 20/09/2016; Accepted: 23/09/2016

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/pmic.201600314](https://doi.org/10.1002/pmic.201600314).

This article is protected by copyright. All rights reserved.

**ABSTRACT**

Studying cell differentiation and transformation allows a better understanding of the mechanisms involved in the initiation and the evolution of cancer. The role of proteins which participate in these processes is dependent on their location within the cell. Determining the subcellular localization of proteins or the changes in localization is, therefore, paramount in elucidating their role. Using quantitative mass spectrometry, we characterized the protein expression and subcellular localization of nearly 5,000 proteins from seven different colorectal cancer (CRC) cell lines, as well as normal colon fibroblasts and intestinal epithelial cells. This cellular characterization allowed the identification of colon cancer-associated proteins with differential expression patterns as well as deregulated protein networks and pathways. Indeed, our results demonstrate differential expression of proteins involved in cell adhesion, cytoskeleton, and transcription in colon cancer cells compared to normal colon-derived cells. Pathway analyses identified different cellular functions, including endocytosis and eIF2 signaling, whose deregulation correlates with mutations found in the different CRC phenotypes. Our results provide an unbiased, quantitative and high-throughput approach to measure changes in protein expression and subcellular protein locations in different colorectal cancer cell lines.

**Statement of significance**

The subcellular proteomics analysis identified nearly 5,000 proteins in nine commonly used colon cell lines including HCT-116, CCL-233, HT-29, CCL-228, CCL-227, DLD-1, Caco-2/15, HIEC and CRL-1459. These data provide information on protein expression, but also on protein subcellular localization in the cytosol, membrane, nucleus and cytoskeleton. We identified several pathways that could potentially be representative for the different phenotypes of colon cancer such as caveolar-mediated endocytosis and signalling through eIF2.

## INTRODUCTION

Colorectal cancer (CRC) arises from three molecular phenotypes resulting in the acquisition of different forms of DNA instability [1, 2]. The most prevalent of these phenotypes is the chromosomal instability pathway (CIN) involved in over 85% of colorectal cancers. Tumors with CIN, typically characterized by imbalances in chromosome number (aneuploidy), display large-scale chromosomal rearrangements resulting from chromosomal segregation, telomere stabilization or defects in the DNA damage response. In approximately 15% of CRC, deficiencies in DNA mismatch repair mechanisms lead to an increase in genomic instability named microsatellite instability, or MSI [3, 4]. Of those, 3% are associated with Lynch syndrome (hereditary nonpolyposis colorectal cancer) [5], while the other 12% are caused by sporadic hypermethylation of the MLH1 gene, resulting in the CpG island methylator phenotype (CIMP) [6]. A precise molecular definition of these cancer types and the identification of markers for proper classification is thus essential in order to differentiate between different CRC phenotypes [7, 8]. Thus, the identification of protein markers and pathways affected in different CRC will be critical to eventually compare the outcome of treatments.

Proteins are found within the cell in specific subcellular compartments often associated with their function. Characterizing the localization of proteins within the cell allows the identification of their cellular role, and aberrant protein localization can help determine pathways that are deregulated in specific diseases [9]. For example, the APC protein, by shuttling between the nucleus and the cytoplasm, controls the nuclear accumulation of the oncogenic transcriptional activator  $\beta$ -catenin. Mutations in APC nuclear export sequences results in APC nuclear sequestering, thereby preventing  $\beta$ -catenin function [10]. The localization of proteins with unknown functions within subcellular compartments also enhances our understanding of their cellular role. Techniques used for the characterization of protein localization are mainly based on microscopy and cell fractionation [11]. Microscopy-based methods use specific antibodies for immunofluorescence microscopy or protein fusions with green fluorescent protein (GFP) [12, 13]. High-throughput approaches, such as the Human Protein Atlas project, are generating antibodies against all human proteins to characterize their

localization in human cells and tissues [14]. While this approach will eventually generate a map of protein localization, it will be restricted to the cell lines and tissues analyzed.

The combination of traditional biochemical fractionation coupled to mass spectrometry-based identification has been the next step in the characterization of subcellular organization [11]. After isolation of specific organelles or subcellular compartments, the protein content can thus be identified to provide an inventory of proteins localized in that compartment. This kind of high-throughput approach to identify protein localization by proteomic methods is proving to be an important way to elucidate the protein functions and to identify new roles for specific subcellular compartments [15]. Many studies on organelle proteomics have provided a detailed list of the protein contents of these organelles [16-20]. Such studies have also used quantitative proteomics for high throughput assignment of proteins to subcellular compartments by protein correlation profiling [18, 21] and by recording the number of ions detected per protein [16, 17], while localization of organelle proteins was assessed by isotope tagging (LOPIT) [22]. Although organelle-based approaches can provide valuable information about specific subcellular compartments in isolation, it is also important to measure the relative abundance of proteins in different locations to determine how modifications in cell growth and physiological conditions alter subcellular localization [23].

While thousands of studies report each year the differential expression of proteins under different conditions in cells and tissues, only a few reports have investigated the changes in protein localization. In this work, we developed a quantitative and high throughput mass spectrometry-based approach to measure the relative intracellular localization of proteins, and to compare changes in their subcellular localization, in different colorectal cancer cells. Such a simultaneous proteomic study of different human cellular compartments will thus provide novel and significant insights in protein function during differentiation and transformation.

## MATERIALS AND METHODS

### Cells and cell culture

Colorectal cancer cell lines HCT-116, Caco-2/15, CCL-228, CCL-227, CCL-233, DLD-1, HT-29, as well as the colon fibroblast cell line CRL-1459 were grown as adherent cells in Dulbecco's modified eagle medium (DMEM) depleted of arginine and lysine (Life Technologies A14431-01), and supplemented with 10% dialyzed fetal bovine serum (FBS) (Invitrogen, 26400-044), 100 U/ml penicillin/streptomycin and 2 mM GlutaMax. Arginine and lysine were added in either light (Arg0, Sigma, A5006; Lys0, Sigma, L5501), medium (Arg6, Cambridge Isotope Lab (CIL), CNM-2265; Lys4, CIL, DLM-2640), or heavy (Arg10, CIL, CNLM-539; Lys8, CIL, CNLM- 291) forms to a final concentration of 28 µg/ml for arginine and 49 µg/ml for lysine. L-proline was added to a final 10 µg/ml concentration to prevent arginine to proline conversion. Proteins were tested for >99% label incorporation after six passages by mass spectrometry (data not shown). Human intestinal epithelial crypt-like cells (HIEC-6) were grown as adherent cells in Opti-MEM (Gibco) containing 2.4 g/L of sodium bicarbonate, 2 mM GlutaMax and 10 mM HEPES, supplemented with 5% FBS (WISENT) and 5 ng/mL EGF (BD Biosciences).

### Subcellular fractionation

Cells at 80-95% confluence were harvested from petri dishes with an enzyme free cell dissociation buffer (Gibco, Life Technologies) followed by scraping. The cell pellet was then washed twice with PBS. Cells were then fractionated into four different subcellular compartments (cytosol, membrane, nucleus and cytoskeleton) using the QProteome Cell Compartment kit (Qiagen). Protein concentrations of each cellular compartment were determined by the bicinchoninic acid (BCA) protein assay (Pierce) prior to loading on SDS-PAGE.

### **Gel electrophoresis and in-gel digestion**

For each subcellular fraction, proteins were reduced in 10 mM DTT, alkylated with 50 mM iodoacetamide prior to boiling in loading buffer, and then separated by one-dimensional SDS-PAGE (4–12% Bis-Tris Novex mini-gel, Life Technologies) and visualized by Coomassie staining (Simply Blue Safe Stain, Life Technologies). The entire protein gel lanes were excised and cut into 8 slices each. Every gel slice was subjected to in-gel digestion with trypsin (Trypsin Gold, Mass Spectrometry Grade, Promega Corporation). The resulting tryptic peptides were extracted by 1% formic acid, then 100% acetonitrile, lyophilized by speed vacuum centrifugation, and resuspended in 1% formic acid.

### **Antibodies and western blotting**

The following antibodies were used: anti-GAPDH (rabbit monoclonal, Cell Signaling #2118S), anti-Claudin-1 (rabbit polyclonal, Invitrogen #51-9000), anti- $\gamma$ H2AX (rabbit polyclonal, Santa Cruz Biotechnology #sc-101696), anti-cytokeratin (mouse monoclonal, Sigma-Aldrich #C8541), anti-eIF2S1 (phosphorylated serine 51, Abcam #ab32157), anti eIF2a (Santa Cruz Biotechnology, Inc., sc-11386) and anti-actin (Millipore #MAB1501R). Secondary antibodies used were anti-mouse IgG-HRP (goat polyclonal, Santa Cruz Biotechnology #sc-2005) and anti-rabbit IgG-HRP (goat polyclonal, Santa Cruz Biotechnology #sc-2004).

### **LC-MS/MS**

Trypsin-digested peptides were separated using a Dionex Ultimate 3000 nanoHPLC system. 10  $\mu$ l of the sample (2  $\mu$ g) in 1% (vol/vol) formic acid were loaded with a constant flow of 4  $\mu$ l/min on an Acclaim PepMap100 C18 column (0.3 mm id x 5 mm, Dionex Corporation). After trap enrichment, peptides were eluted in a PepMap C18 nanocolumn (75  $\mu$ m x 50 cm, Dionex Corporation) with a linear gradient of 5-35% solvent B (90% acetonitrile with 0.1% formic acid) over 240 minutes with a constant flow of 200 nl/min. The HPLC system was coupled to an OrbiTrap QExactive mass spectrometer (Thermo Fisher Scientific) via an EasySpray source. The spray voltage was set to 2.0 kV

and the temperature of the column was set to 40°C. Full scan MS survey spectra ( $m/z$  350-1600) in profile mode were acquired in the Orbitrap with a resolution of 70,000 after the accumulation of 1,000,000 ions. The ten most intense peptide ions from the preview scan in the Orbitrap were fragmented by collision induced dissociation (normalized collision energy 35% and resolution of 17,500) after the accumulation of 50,000 ions. Maximal filling times were 250 ms for the full scans and 60 ms for the MS/MS scans. Precursor ion charge state screening was enabled and all unassigned charge states as well as singly, 7 and 8 charged species were rejected. The dynamic exclusion list was restricted to a maximum of 500 entries with a maximum retention period of 40 seconds and a relative mass window of 10 ppm. The lock mass option was enabled for survey scans to improve mass accuracy. Data were acquired using the Xcalibur software.

#### **Quantification and bioinformatics analysis**

Data were processed, searched and quantified using the MaxQuant software package version 1.5.2.8, as described previously (24), employing the Human Uniprot database (16/07/2013, 88,354 entries). The settings used for the MaxQuant analysis were: 2 miscleavages were allowed; fixed modification was carbamidomethylation on cysteine; enzymes were Trypsin (K/R not before P); variable modifications included in the analysis were methionine oxidation and protein N-terminal acetylation. A mass tolerance of 7 ppm was used for precursor ions and a tolerance of 20 ppm was used for fragment ions. The re-quantify option was selected to calculate the ratio for isotopic patterns not assembled in SILAC pairs as often observed during pulldown experiments (25). To achieve reliable identifications, all proteins were accepted based on the criteria that the number of forward hits in the database was at least 100-fold higher than the number of reverse database hits, thus resulting in a false discovery rate (FDR) of less than 1%. A minimum of 2 peptides was quantified for each protein. Protein isoforms and proteins that cannot be distinguished based on the peptides identified are grouped and displayed on a single line with multiple accession numbers (see Supplementary Tables). The mass spectrometry raw files have been deposited to the

ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository with the identifier PXD004879.

Ingenuity pathway analysis calculates a score for each network according to the list of supplied proteins. The score is derived with a  $p$ -value and adjusted following multiple hypothesis testing, which indicates the likelihood of proteins within a specific network of being identified together due to random chance. Graphical representations of the molecular relationships between protein products were generated with focus proteins represented as nodes with the intensity of the node color indicating either an increase (red) or decrease (green).

## RESULTS

### Subcellular fractionation of colorectal cancer cell lines

Seven commonly used colorectal cancer cell lines (HCT-116, CCL-233, HT-29, DLD-1, CCL-228, CCL-227 and Caco-2/15) and cells representing a more normal phenotype, the colon-derived fibroblast cell line CRL-1459 (CCD-18Co) and the human intestinal epithelial cell line HIEC [24], represented various colorectal cancer phenotypes and mutations. The CCL-228 (SW480) and CCL-227 (SW620) cell lines are from a primary adenocarcinoma and from a metastasis from the same patient one year apart [25], and the DLD-1 (HCT-15) and HT-29 (WiDr) cell lines are derived from the same cancer specimen and appear genetically identical, while harboring different mutations [26]. These cell lines contain the different molecular defects resulting in different forms of DNA instability. The different phenotypes for each cell line are shown in Table 1. The phenotypes represent the chromosomal instability (CIN), the microsatellite instability (MSI) and the CpG island methylator phenotype (CIMP) status for each cell line. Additionally, mutations in some of the most commonly altered genes such as KRAS, BRAF, PIK3CA, PTEN and TP53 are shown (Table 1).

In order to study protein expression and localization differences, four different subcellular fractions (cytosol, membranes, nucleus and cytoskeleton) were obtained from the different cell lines

grown in SILAC media (Supplementary Figure 1). A representative control of a Coomassie stained SDS-PAGE gel with extracts from each fraction shows a very different protein band pattern in each fraction (Figure 1A). The method is very reproducible, and the general pattern of protein expression is very similar between each cell line (Supplementary Figure 2). Indeed, proteins most likely to be histones appear in the nuclear fraction (Figure 1A, lane 3), but not in the other fractions (Figure 1A, lanes 1, 2 and 4). Using different proteins normally found in each compartment, we confirmed that each fraction contains proteins expected to be in each compartment (Figure 1B). Additionally, on the intensity for each proteins enriched in each subcellular compartment as measured by mass spectrometry is shown in (Figure 1C), coinciding with the Western Blot analysis of Figure 1B. The cytosol proteins were mostly annotated as either cytosolic or from vesicles. The proteins identified as enriched in the membrane fraction were mostly annotated as either from the mitochondrial membrane or the cellular membrane, with approximately 50% of the proteins containing a transmembrane domain. Proteins enriched in the nuclear fraction were either annotated as nucleoplasmic or nucleolar. Finally, the proteins in the cytoskeletal fraction were annotated as intermediate filaments, keratin filaments, cytoskeleton as well as some ribosomal proteins.

Fractions from different cell lines grown in SILAC media to incorporate different isotopically labeled arginines and lysines were combined in groups of three to generate different combinations (Figure 2), prior to trypsin digestion and mass spectrometry identification and quantification. In contrast to the HIEC cell line which can only be grown in OptiMEM and can only be labeled with normal, light isotopes, all other cell lines were grown in either light (L), medium (M) or heavy (H) SILAC media. Samples were either separated on SDS-PAGE and the whole lane excised and in-gel digested with trypsin, or the proteins were solubilized and digested using the filter-aided sample preparation (FASP) procedure [27]. Every cell type and fraction were included in at least three separate combinations to ensure a good representation and reproducibility for each protein (Supplementary Figure 3), generating a total of 224 mass spectrometry analyses. Additionally, HCT-

116 cell fractions were present in every combination for normalization and internal calibration. A total of 46,117 peptides mapping to 4,828 proteins were identified, with a minimum of two peptides for each protein with a false discovery rate of 0.01.

The number of proteins identified in each cell line and fraction, for each replicate, shows a similar distribution. Cytoskeletal fractions (F4) have a relatively low number of proteins as compared to the other fractions (Figure 3). However, the intensities of the few proteins identified are generally high, consistent with high abundant proteins such as actin and tubulin. Similarly, the distribution of intensities in each fraction is very reproducible and underlines the wide range of protein expression detected in each fraction (Figure 3B). The average intensities for each identified protein, in each fraction and every cell line, were then measured. Hierarchical clustering analysis classified proteins based on their subcellular localization, with each fraction clustering together, except for the cytoskeletal fraction of the CRL-1459 cell line, which clustered away from all the other fractions (Figure 4). Normal fibroblast cell protein fractions were largely different from the other cell line fractions (Figure 4). Interestingly, the CCL-228 and CCL-227 cell lines derived from the same patient often clustered close to each other, underlining the similarities between these fractions.

To obtain additional functional information into the pathways differentially regulated in normal cells compared to cancer cell lines, proteins from each fraction of the different cell lines were subjected to analysis using Ingenuity Pathway Analysis (IPA). This analysis allowed the identification of functions significantly increased or decreased in the different cancer cell lines, in comparison to the normal cell lines used in this study. For each fraction, the intensities for each proteins were first normalized to take into account the variability between each analysis. These normalized intensities were then averaged and then compared between similar fractions of each cell lines. The data for each protein was kept only if the protein was identified in at least two replicates, and with a minimum of three peptides. We use relative up or down-regulation of protein intensities in one fraction compared to the average of all the other fraction. We considered that a cellular function

was up or down-regulated when the calculated z-score was over 2 or -2, respectively. The IPA z-score is derived from the *p*-value and adjusted following multiple hypothesis testing (see methods). The horizontal orange lines on the different graph (Figure 5, 6 and supplementary figure 4) represent a *p*-value of 0.05. For the cytosolic fractions (Fraction F1), IPA analysis identified many proteins involved in cytoskeleton organization, which were downregulated in cancer cell lines compared to the two normal cell lines (Supplementary Figure 4): this is consistent with a larger, more elongated and organized cytoplasm in normal cells as opposed to cancer cells. Interestingly, we found that proteins in pathways involved in caveolar-mediated endocytosis signaling, integrin signaling and actin cytoskeleton signaling were highly increased in most cancer cell lines (Figure 5A). For example, proteins involved in caveolar-mediated endocytosis were among the 10% with the highest expression in cancer cell lines (Figure 5B), while being among the lowest 10% expression in normal HIEC and CRL1459 cells (Figure 5B). However, clathrin-mediated endocytosis and tight junction signaling did not show significant differences between cancer and normal cell lines (Figure 5A). Our analysis also showed important differences in eIF2 signaling (Figure 6A), which revealed an upregulation of eIF2 related proteins in cancer cell lines compared to normal cells (Figure 6B), indicating a difference in protein synthesis. Interestingly, this upregulation was also observed in the cytoskeletal fraction (Supplementary Figure 5). To confirm the activation of the eIF2 pathway, equal amount of proteins from whole cell lysates of the different cell lines were separated by SDS-PAGE and immunoblotted with either an antibody recognizing eIF2, or eIF2 phosphorylated on serine 51 (Figure 6C). Consistent with the cell lines showing the highest activation of the eIF2 pathway (Figure 6B), cancer cell lines showed a much higher activation of eIF2 compared to normal cell lines, with Caco-2/15 and CCL-228 displaying the highest increase.

## DISCUSSION AND CONCLUSIONS

In this study, we provide a comprehensive proteomic characterization of seven different colorectal cancer cell lines as well as one normal colon epithelial cell line and one colon fibroblast cell line. Using a quantitative mass spectrometry approach through stable isotope labeling (SILAC), we identified nearly 5,000 proteins from each cell line, and characterized their subcellular distribution in four different cellular compartments, namely the cytosol, the membrane, the nucleus and the cytoskeleton. Through different mixing and replication for every cell type and fraction, we reached a reliable representation and reproducibility pattern in the identification and quantification of each protein. Moreover, fractions of HCT-116 cells were used in every analysis as a control to compare expression patterns. Thus, averaging these different combinations lowered the variability of the mass spectrometry analysis. Clustering analysis identified subgroups that are more similar between the different cancer cell lines according to protein expression and localization. Protein expression patterns in DLD-1 and Caco2/15 cells were always the closest in every fraction, whereas protein expression in HCT-116 cells and in cell lines derived from patients with KRAS mutations (CCL-227, -228 and -233) was closely related. Interestingly, the two normal cell lines, either human intestinal epithelial or colon fibroblast cell line, were very different from all cancer cell lines for protein expression and localization, again underlining that the transformation process has a profound effect on protein levels. Further data mining should allow the detection of distinct subtypes of colorectal cancer.

Analysis of the pathways that differ between different cancer and normal cell lines revealed that several cellular canonical pathways were over-represented. For example, one of the most significantly represented function in membrane fractions was the signaling pathway connected to caveolin-mediated endocytosis (Caveolar-mediated endocytosis signaling), with a p-value of  $10^{-8}$ . This pathway was not as significantly altered in the other compartments of the subcellular fractionation. Combining or doing a whole cell analysis often dilute these changes that are more obvious when looking at a specific fraction, so it proves useful to look at changes within each

fractions. Additionally, it is possible to identify proteins whose expression in one subcellular compartment is decreases, while increasing in another, often suggesting possible regulation or activation of a specific pathway. To investigate the causes of this over-representation, the calculated ratios of proteins belonging to other endocytic pathways were compared between the different cell lines. In general, proteins from normal HIEC and CRL-1459 cells had a lower intensity than the average, with most proteins present in the 10% proteins showing the highest change. On the other hand, the intensity of proteins related to endocytosis in cancer cell lines was higher than the average. Interestingly, this decrease in caveolar-mediated endocytosis was also observed in a systematic quantification of fixed colorectal cancer tissues from different stages [28], and caveolin is deregulated in several types of cancer [29].

Another major canonical pathway differentially regulated was eIF2 signaling mechanisms. Among the proteins regulating eIF2 (eukaryotic initiation factor 2) were a family of four kinases necessary for initiation of protein translation. The p-value for this pathway for several cell lines was among the highest of all enriched pathways in nuclear and cytoskeletal fractions. In nuclear fractions, proteins associated with eIF2 signaling were predominantly over-expressed in three cancer cell lines (CCL-228, Caco-2/15 and HT-29) while being predominantly under-expressed in normal cells. On the other hand, in cytoskeletal fractions, the p-values were even more significant (Supplementary Figure 5) in most cancer cell lines. Thus, it appears that the initiation factor eIF2 plays an important role in colorectal cancer cells. In fact, several studies have demonstrated the importance of eIF2 activation through phosphorylation in cancer therapy [30, 31]. It would thus be interesting to analyze the phosphoproteome in the different colorectal cancer cell lines.

We also examined potential transcriptional regulators of upstream signaling pathways (upstream regulators) that take into account protein expression and the known direction of the effect regulator on downstream proteins. We identified three regulators significantly altered in cancer cell lines as opposed to the normal cells, namely MYC, TP53, and TGFB1. For each regulator,

the direction of the effect was assessed by considering z scores whose value exceeded the cut-off of 2 or -2. Thus, the Myc transcription factor seemed to negatively regulate normal HIEC and CRL-1459 cell lines: transcribed proteins were found consistently in the cytosolic, membrane and nuclear fractions. In addition, the two CCL-228 and HCT-116 cancer cell lines showed significant upregulation of Myc target genes, suggesting that Myc may be an important regulator in these two cell lines. Similar changes were observed for TP53 and TGFB1. Indeed, TP53 was a positive regulatory effector in normal cells, but a negative effector of CCL-227, CCL-228 and Caco-2 cancer cells. Interestingly, these three cell lines have mutations in TP53. Regulation by TGFB1 was increased in normal cells and decreased in several cell lines including CCL-227, CCL-233, DLD-1 and HT-29. Interestingly, these cell lines have mutations in TP53 (see Table 1), and mutant p53 generally subverts suppressive TGF-B responses by inhibiting transcriptional activation of several TGF-B target genes [32].

In summary, we have provided a comprehensive proteomic analysis of commonly used colorectal cancer cell lines. We have identified several proteins and pathways upregulated in colorectal cancer cell lines. The information gained from this study generated a large amount of data useful for determining proteins potentially involved in colon epithelial cell differentiation and transformation processes. Our study goes a step beyond conventional studies by providing subcellular localization measurements of proteins, which enables a direct comparison between compartments, with high-throughput studies using antibodies and other large-scale methods to study protein localization. Further comparison with samples from patients should eventually allow a better classification and identification of deregulated pathways, thus ensuring a better diagnostic and an increased ability to choose the best treatment option.

#### **CONFLICT OF INTEREST**

None to declare.

## ACKNOWLEDGEMENT

Funding to FMB is from the National Sciences and Engineering Research Council of Canada (418404-2012). F.M.B. and F.B. are members of the FRQS-funded “Centre de Recherche du CHUS”. We thank Claude Asselin for his careful reading of the manuscript.

## Supporting Information.

### Supplementary Figure Legends

### Supplementary Table I

Proteomic analysis of all fractions for all cell lines. Fractions are represented as F1 (cytosol), F2 (membrane), F3 (nucleus) and F4 (cytoskeleton). The nine cell lines are named CRL-1459 (1459), CCL-227 (227), CCL-228 (228), CCL-233 (233), Caco-2/15 (CACO), DLD-1 (DLD), HCT-116 (HCT), HIEC and HT-29 (HT).

### Supplementary Figure 1. Subcellular fractionation overview.

This protocol allows the subcellular fractionation of proteins from different cellular compartments. The different solubilization buffers 1, 2, 3 and 4 are sequentially added to the pellet and supplemented with a protease inhibitor solution (x100). After each extraction, the cellular pellet is resuspended in the next buffer: the supernatant constitutes one subcellular fraction. The DNA nuclease Benzonase is used to solubilize nuclear proteins. Four fractions are obtained (cytosol, membranes, nucleus, cytoskeleton). The cell images (left) and the color of the fractions illustrate the location of the extracted proteins.

### Supplementary Figure 2: Validation of the subcellular fractionation protocol for each cell line.

Cells were fractionated in four different subcellular fractions (Cs: cytosol; M: membrane; N: nucleus; Ct: Cytoskeleton). Each protein fraction was resolved by SDS-PAGE and visualized by Coomassie blue.

### Supplementary Figure 3: Combinations used for mass spectrometry analysis.

Each combination analyzed by mass spectrometry is shown and labeled A to M.

**Supplementary Figure 4: Regulated functions identified in cytosolic fractions.**

**A.** Some of the most important cellular functions altered positively or negatively in the cytosol of the selected cell lines, according to the  $-\log_{10}$  (p-value), are represented on the y-axis. The horizontal orange line represents a significant p-value of 0.05. **B.** For the cytosolic fraction (F1) of each cell line, the z-score and p-value of the biological function named "cytoskeletal organization" are represented. The positive z-score is shown in orange while the negative z-score is shown in blue. Color intensity is greater for values far from zero. The p-value is represented in shades of purple, paler p-values representing decreasing significance value, but still under 0.05. **C.** Each protein involved in cytoskeleton organization and found in the cytosol of the DLD-1 cell line is represented (39 proteins). Proteins increased, in comparison to other cell lines, are shown in red and proteins decreased are shown in green. Color intensity is related to the magnitude of over- or under-expression. The effect of each protein on cytoskeletal organization and function is shown by a colored dotted line (see Supplementary Figure 4 legend).

**Supplementary Figure 5: The eIF2 signaling pathway is over-represented in the cytoskeletal fraction of cancer cell lines.**

**A.** Canonical pathways significantly over-represented in the cytoskeletal compartment. Each strip of a different color (see Figure legend) illustrates the  $-\log_{10}$  (p-value) of the significance of a cellular pathway based on the increase or decrease in protein intensities in this fraction. **B.** Percentage of increased (red) or decreased (green) proteins involved in the EIF2 signaling pathway (185 proteins) that are identified in cytoskeletal fractions from 4 different cell lines. Only data from cell lines with significant p-value have been used.

## REFERENCES

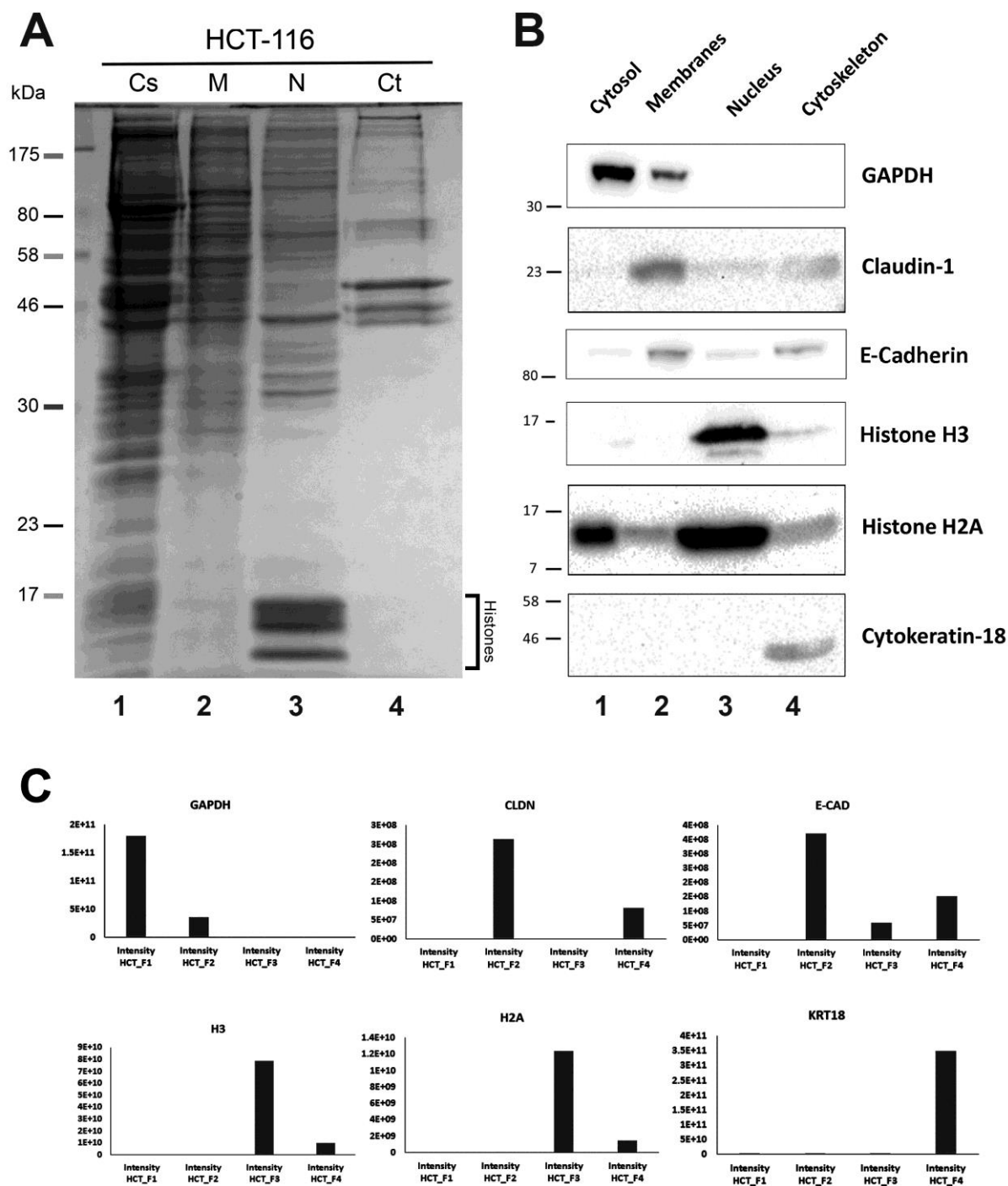
- [1] Nowak, M. A., Komarova, N. L., Sengupta, A., Jallepalli, P. V., *et al.*, The role of chromosomal instability in tumor initiation. *Proc Natl Acad Sci U S A* 2002, *99*, 16226-16231.
- [2] Pino, M. S., Chung, D. C., The chromosomal instability pathway in colon cancer. *Gastroenterology* 2010, *138*, 2059-2072.
- [3] Boland, C. R., Goel, A., Microsatellite instability in colorectal cancer. *Gastroenterology* 2010, *138*, 2073-2087 e2073.
- [4] Bellizzi, A. M., Frankel, W. L., Colorectal cancer due to deficiency in DNA mismatch repair function: a review. *Adv Anat Pathol* 2009, *16*, 405-417.
- [5] Lynch, H. T., Rozen, P., Schuelke, G. S., Hereditary colon cancer: polyposis and nonpolyposis variants. *CA Cancer J Clin* 1985, *35*, 95-114.
- [6] Nazemalhosseini Mojarad, E., Kuppen, P. J., Aghdaei, H. A., Zali, M. R., The CpG island methylator phenotype (CIMP) in colorectal cancer. *Gastroenterol Hepatol Bed Bench* 2013, *6*, 120-128.
- [7] Sinicrope, F. A., Okamoto, K., Kasi, P. M., Kawakami, H., Molecular Biomarkers in the Personalized Treatment of Colorectal Cancer. *Clin Gastroenterol Hepatol* 2016, *14*, 651-658.
- [8] Zhang, B., Wang, J., Wang, X., Zhu, J., *et al.*, Proteogenomic characterization of human colon and rectal cancer. *Nature* 2014, *513*, 382-387.
- [9] Hung, M. C., Link, W., Protein localization in disease and therapy. *J Cell Sci* 2011, *124*, 3381-3392.
- [10] MacDonald, B. T., Tamai, K., He, X., Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Dev Cell* 2009, *17*, 9-26.
- [11] Drissi, R., Dubois, M. L., Boisvert, F. M., Proteomics methods for subcellular proteome analysis. *FEBS J* 2013, *280*, 5626-5634.
- [12] Fagerberg, L., Stadler, C., Skogs, M., Hjelmare, M., *et al.*, Mapping the subcellular protein distribution in three human cell lines. *J Proteome Res* 2011, *10*, 3766-3777.
- [13] Kumar, A., Agarwal, S., Heyman, J. A., Matson, S., *et al.*, Subcellular localization of the yeast proteome. *Genes Dev* 2002, *16*, 707-719.
- [14] Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., *et al.*, Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol* 2010, *28*, 1248-1250.
- [15] Walther, T. C., Mann, M., Mass spectrometry-based proteomics in cell biology. *The Journal of cell biology* 2010, *190*, 491-500.
- [16] Gilchrist, A., Au, C. E., Hiding, J., Bell, A. W., *et al.*, Quantitative proteomics analysis of the secretory pathway. *Cell* 2006, *127*, 1265-1281.
- [17] Kislinger, T., Cox, B., Kannan, A., Chung, C., *et al.*, Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. *Cell* 2006, *125*, 173-186.
- [18] Wiese, S., Gronemeyer, T., Ofman, R., Kunze, M., *et al.*, Proteomics characterization of mouse kidney peroxisomes by tandem mass spectrometry and protein correlation profiling. *Mol Cell Proteomics* 2007, *6*, 2045-2057.
- [19] Andersen, J. S., Lyon, C. E., Fox, A. H., Leung, A. K., *et al.*, Directed proteomic analysis of the human nucleolus. *Curr Biol* 2002, *12*, 1-11.
- [20] Andersen, J. S., Mann, M., Organellar proteomics: turning inventories into insights. *EMBO Rep* 2006, *7*, 874-879.
- [21] Foster, L. J., de Hoog, C. L., Zhang, Y., Xie, X., *et al.*, A mammalian organelle map by protein correlation profiling. *Cell* 2006, *125*, 187-199.

- [22] Dunkley, T. P., Watson, R., Griffin, J. L., Dupree, P., Lilley, K. S., Localization of organelle proteins by isotope tagging (LOPIT). *Mol Cell Proteomics* 2004, 3, 1128-1134.
- [23] Boisvert, F. M., Lam, Y. W., Lamont, D., Lamond, A. I., A quantitative proteomics analysis of subcellular proteome localization and changes induced by DNA damage. *Mol Cell Proteomics* 2010, 9, 457-470.
- [24] Perreault, N., Beaulieu, J. F., Use of the dissociating enzyme thermolysin to generate viable human normal intestinal epithelial cell cultures. *Exp Cell Res* 1996, 224, 354-364.
- [25] Leibovitz, A., Stinson, J. C., McCombs, W. B., 3rd, McCoy, C. E., *et al.*, Classification of human colorectal adenocarcinoma cell lines. *Cancer Res* 1976, 36, 4562-4569.
- [26] Chen, T. R., Dorotinsky, C. S., McGuire, L. J., Macy, M. L., Hay, R. J., DLD-1 and HCT-15 cell lines derived separately from colorectal carcinomas have totally different chromosome changes but the same genetic origin. *Cancer Genet Cytogenet* 1995, 81, 103-108.
- [27] Wisniewski, J. R., Zougman, A., Nagaraj, N., Mann, M., Universal sample preparation method for proteome analysis. *Nat Methods* 2009, 6, 359-362.
- [28] Yin, X., Zhang, Y., Guo, S., Jin, H., *et al.*, Large scale systematic proteomic quantification from non-metastatic to metastatic colorectal cancer. *Sci Rep* 2015, 5, 12120.
- [29] Razani, B., Lisanti, M. P., Caveolin-deficient mice: insights into caveolar function human disease. *J Clin Invest* 2001, 108, 1553-1561.
- [30] Schewe, D. M., Aguirre-Ghiso, J. A., Inhibition of eIF2alpha dephosphorylation maximizes bortezomib efficiency and eliminates quiescent multiple myeloma cells surviving proteasome inhibitor therapy. *Cancer Res* 2009, 69, 1545-1552.
- [31] Zhu, K., Chan, W., Heymach, J., Wilkinson, M., McConkey, D. J., Control of HIF-1alpha expression by eIF2 alpha phosphorylation-mediated translational repression. *Cancer Res* 2009, 69, 1836-1843.
- [32] Elston, R., Inman, G. J., Crosstalk between p53 and TGF-beta Signalling. *J Signal Transduct* 2012, 2012, 294097.

### Figure 1. Validation of the subcellular fractionation protocol.

HCT-116 cells were fractionated in four different subcellular fractions (Cs: Cytosol; M: Membrane; N: Nucleus; Ct: Cytoskeleton). Each fraction was resolved by SDS-PAGE and visualized by Coomassie blue staining (A). Different control proteins were visualized by Western blot (B) with antibodies against GAPDH (Cytosol), Claudin-1 (Membranes), E-Cadherin (Membranes), Histone H3 (Nucleus), Histone H2A (Nucleus) or Cytokeratin-18 (Cytoskeleton). Molecular weight markers are indicated. (C) The intensities for each proteins in each of the fractions as measured by mass spectrometry is shown as a histogram..

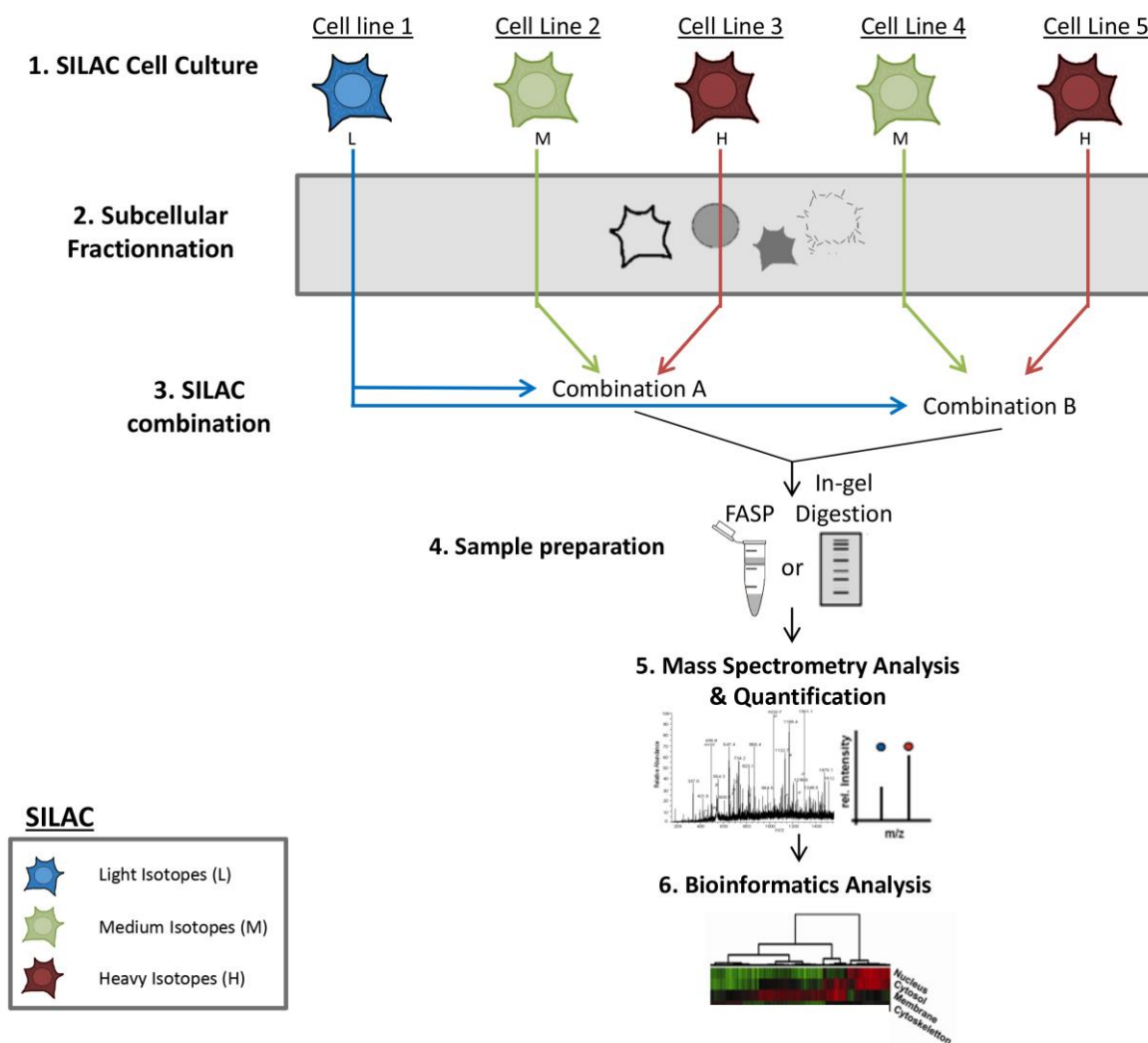
Figure 1



**Figure 2. Experimental SILAC approach to label and combine different fractions isolated from various cell lines for protein identification and quantification.**

Nine cell lines, with five appearing on the figure, were grown in light (L), medium (M) or heavy (H) SILAC medium (Step 1). The protein content of each cell line was then fractionated separately (Step 2) before being combined in different mixes (Step 3). Sample preparation, including the tryptic digest, was carried out by filter aided sample prep (FASP) or in-gel digestion (Step 4) and peptides were analyzed by mass spectrometry (Step 5). Different bioinformatics analyses were performed on the data (Step 6).

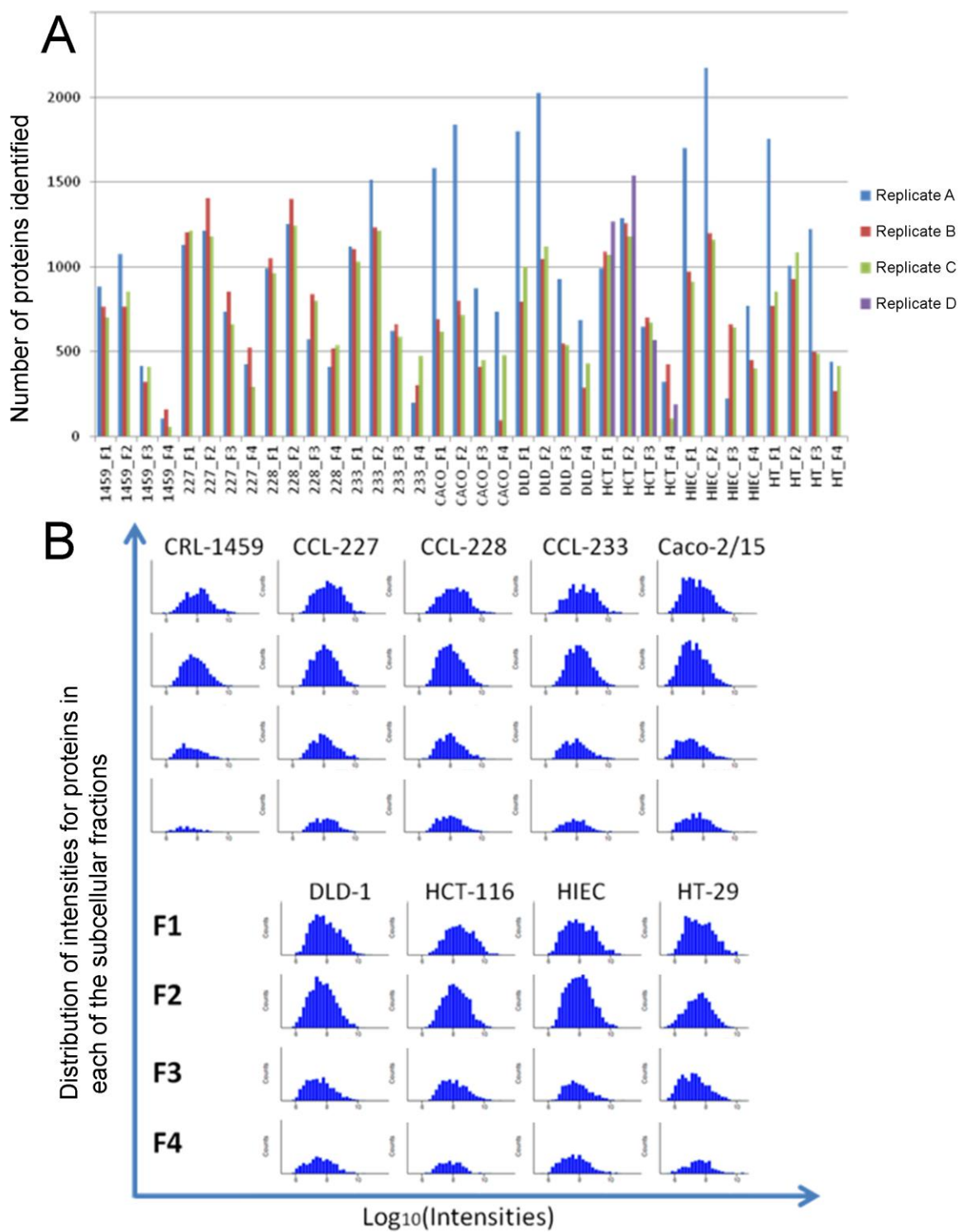
Figure 2



**Figure 3. Distribution of the number of proteins identified and intensities of the combined analysis of all replicates for each cell lines.**

- A.** The number of proteins identified for each cell line (9 cell lines) and fraction (F1: cytosol; F2: membrane; F3: nucleus; F4: cytoskeleton), for up to four replicates (A, B, C and D) is indicated (y-axis). Among 13 HCT-116 replicates, four were selected so that each combination is composed of different lineages. Only FASP digestions were used for HCT-116 isolated proteins. **B.** Histogram of the distribution of proteins by their average intensities in each fraction and each cell line. Proteins were separated by intervals in accordance with their intensities. The y-axis corresponds to the number of proteins contained in each interval. F1: cytosol; F2: membrane; F3: nucleus; F4: cytoskeleton.

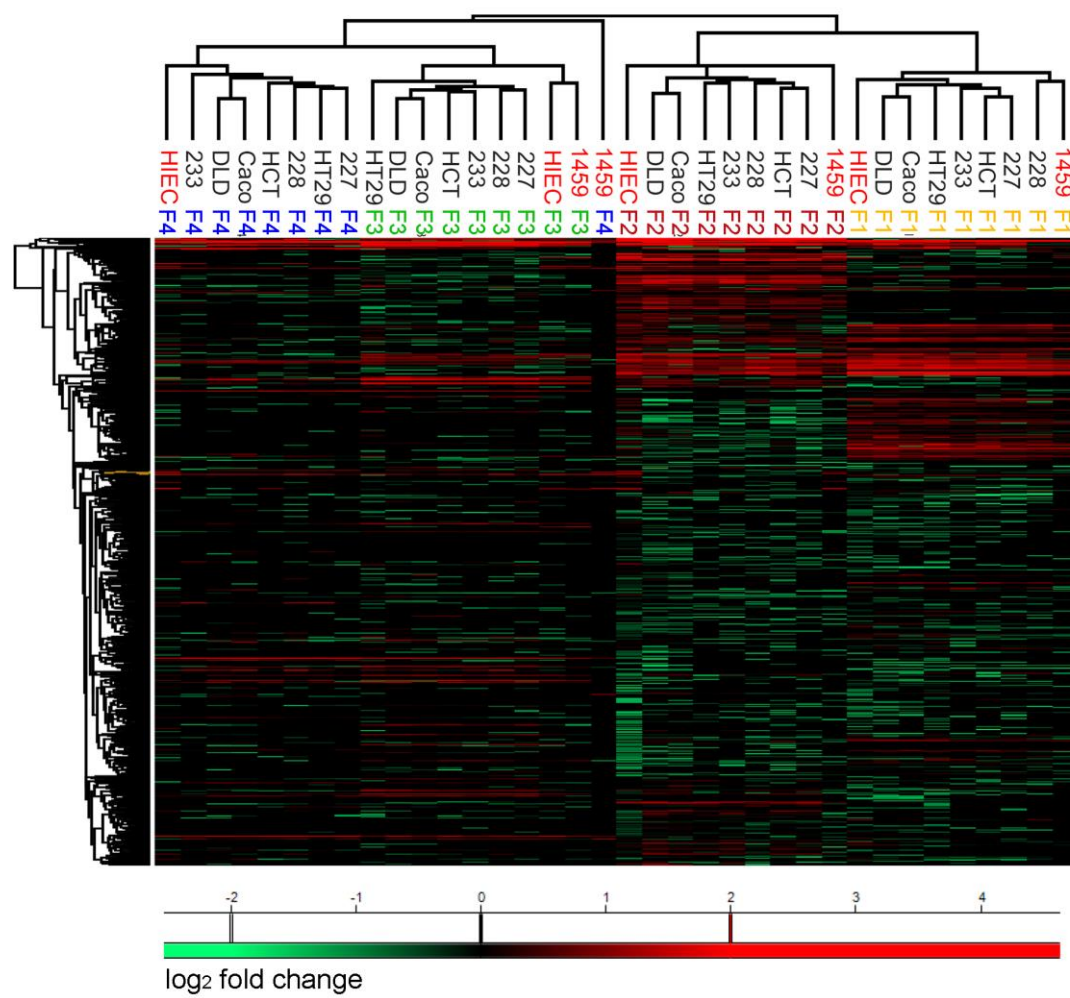
Figure 3



**Figure 4. Hierarchical clustering of the four different fractions for each cell line.**

Protein intensities for each subcellular fraction (F1: cytosol; F2: membrane; F3: nucleus; F4: cytoskeleton) for nine cell lines (CRL-1459 (1459), CCL-227 (227), CCL-228 (228), CCL-233 (233), Caco-2/15 (CACO), DLD-1 (DLD), HCT -116 (HCT), HIEC and HT-29 (HT)) are shown. The average intensities of the three proteomic replicates were used to group cell fractions by similarity, according to the Euclidean distance. No protein was eliminated for hierarchical clustering. The color scale ranges from green (protein intensities under the average) to red (protein intensities above the average) based on the fold change ( $\log_2$ ).

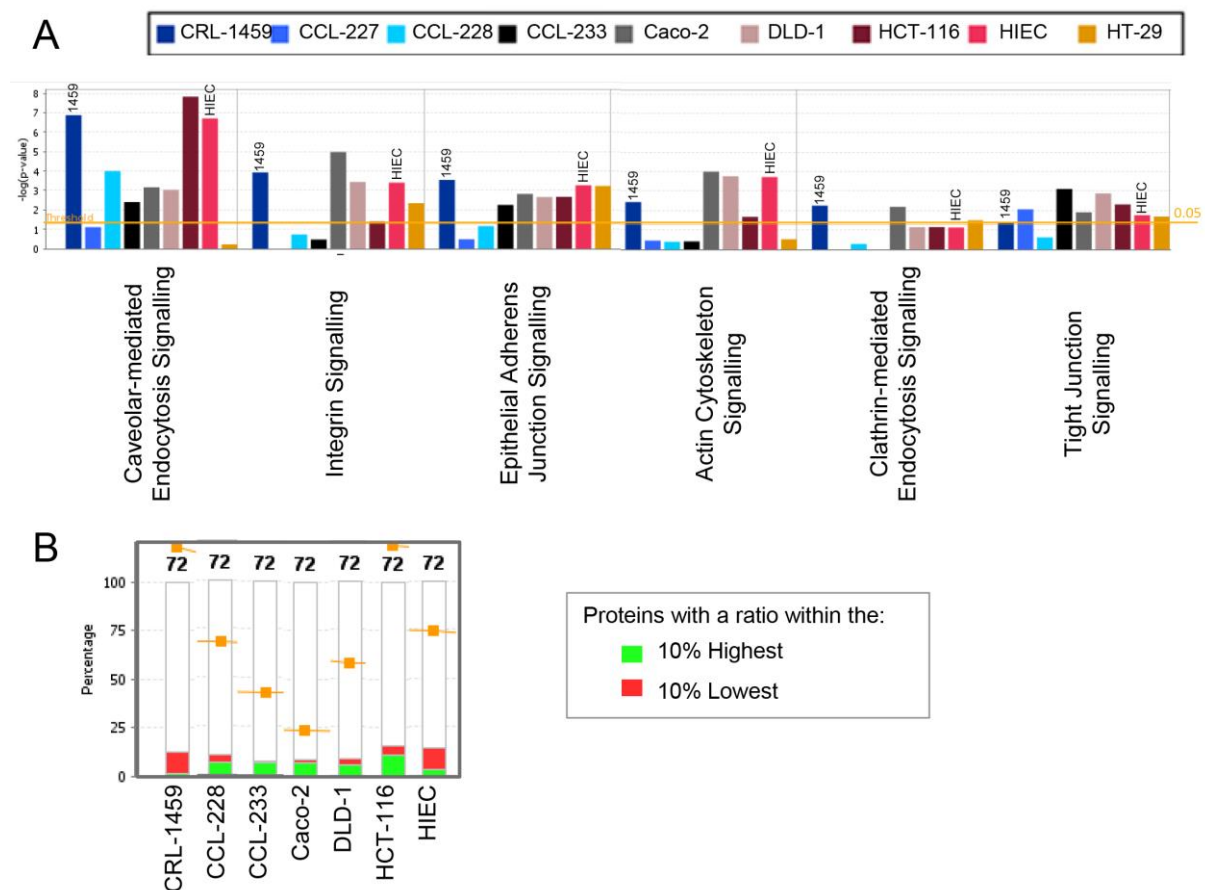
Figure 4



**Figure 5. Canonical cellular pathways significantly over-represented in membrane fractions of cancer cell lines.**

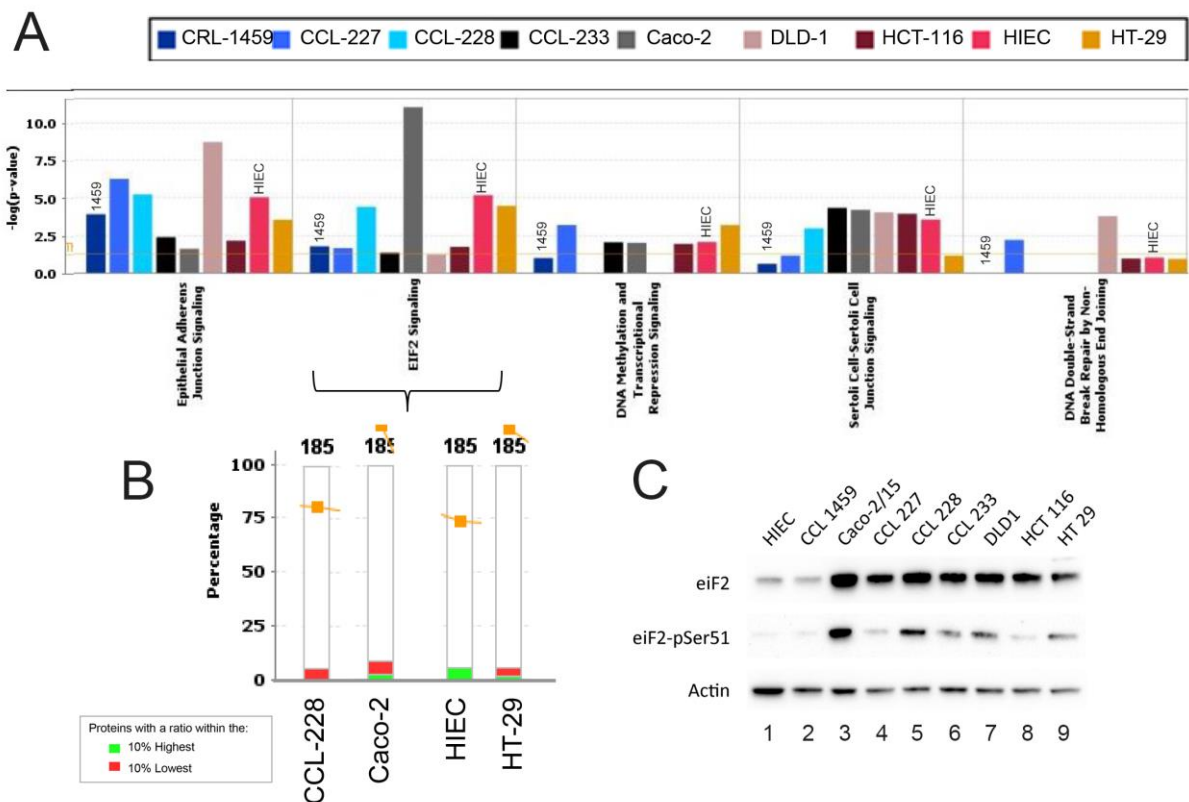
**A.** Canonical cellular pathways were chosen for their highly significant p-value and importance compared to other membrane processes. Each strip of a different color (see legend) illustrates the  $-\log_{10}$  (p-value) of the significance of a cellular pathway based on the increase or decrease in protein intensities in this fraction. The orange horizontal line represents a p-value of 0.05. **B.** Percentage of over-expressed (red) or under-expressed (green) proteins involved in the signaling pathway for caveolin-mediated endocytosis (72 proteins) that are identified in membrane fractions from 7 different cell lines. Only data from cell lines with significant p-value have been used.

**Figure 5**



**Figure 6. Cell signaling pathway by EIF2 is over-represented in cancer cell lines**

**A.** Canonical pathways significantly over-represented in the nuclear compartment. Each strip of a different color (see legend) illustrates the  $-\log_{10}$  (p-value) of the significance of a cellular pathway based on the increase or decrease in protein intensities in this fraction. **B.** Percentage of over-expressed (red) or under-expressed (green) proteins involved in the EIF2 signaling pathway (185 proteins) that are identified in nuclear fractions from 4 different cell lines. Only data from cell lines with significant p-value have been used. **C.** Equal amount of proteins from whole cell lysates of the different cell lines were separated by SDS-PAGE and immunoblotted with either an antibody recognizing eIF2, or eIF2 phosphorylated on serine 51. Actin was used as a loading control.

**Figure 6**

**Table 1. Cell lines used for proteomic analysis.**

The table lists the name of the different cell lines used in this study, the Duke's stage, the genetic profile and the associated mutations. The (+) sign indicates the presence of the phenotype/mutation.

<b>Cells:</b>	<b>HCT116</b>	<b>CCL-233</b>	<b>HT-29</b>	<b>CCL-228</b>	<b>CCL-227</b>	<b>DLD-1</b>	<b>Caco-2/15</b>	<b>HIEC</b>	<b>CRL-1459</b>
<b>Type:</b>	Colorectal Carcinoma							Human Intestinal Epithelial Cells	Normal Colon Fibroblasts
<b>Duke's Stage:</b>	D/A	A	C/B	B	C	C	Unknown	Normal	Normal
<b>Genetic Profile and Mutations</b>									
<b>CIN</b>		+	+	+	+		+		
<b>MSI</b>	+					+			
<b>CIMP</b>	+		+			+			
<b>KRAS</b>	G13D	G12A		G12V	G12V	G13D			
<b>BRAF</b>			V600E						
<b>PIK3CA</b>	H104R		P449T			E545K;D549N			

Received: 25/07/2016; Revised: 20/09/2016; Accepted: 23/09/2016

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/pmic.201600314](#).

This article is protected by copyright. All rights reserved.

**PTEN**

**TP53**

A159D

R273H

R273H;P309S

R273H;P309S

S241F

E204X

Accepted Article