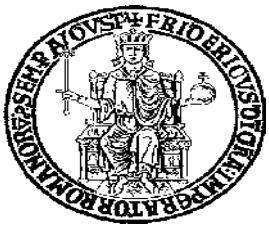
European School of Molecular Medicine
Università degli studi di Napoli Federico II

PhD in Molecular Medicine XXIII Cycle
Molecular Oncology

CHARACTERIZATION OF PROTEINS INVOLVED IN INTRACELLULAR PATHWAYS OF COLON CANCER STEM CELLS

Claudia Corbo



Academic years: 2008-2011

European School of Molecular Medicine
Università degli studi di Napoli Federico II

PhD in Molecular Medicine XXIII Cycle
Molecular Oncology

CHARACTERIZATION OF PROTEINS INVOLVED IN INTRACELLULAR PATHWAYS OF COLON CANCER STEM CELLS

Tutor:

Prof. Francesco Salvatore

Internal Supervisor:

Prof. Margherita Ruoppolo

External Supervisor:

Dr. Simona Francese

Coordinator:

Prof. Francesco Salvatore

PhD student:

Dr.Claudia Corbo

"Nella vita non c'è nulla da temere, solo da capire.
Sii meno curioso della gente e più curioso delle idee"
Marie Curie

LIST OF ABBREVIATIONS

ACN	Acetonitrile
AMBIC	Ammonium Bicarbonate
APC	Adenomatosis polyposis coli
APS	Ammonium persulphate
BVA	Biological Variation Analysis
CHAPS	3-[(3-Cholamidopropyl)Dimethylammonio]-1-Propanesulfonate Hydrate
CID	Collision-induced dissociation
CRC	Colorectal cancer
CSCs	Cancer stem cells
CSLCs	Cancer stem like cells
2DE	Bidimensional Electrophoresis
DIA	Differential Intra-gel Analysis
DIGE	Differential in Gel Electrophoresis
DMEM	Dulbecco Minimal Essential Medium
DMSO	Dimethylsulfoxide
DTT	Dithiothreitol
ECL	Enhanced Chemio-Luminescence
EDTA	Ethylenediamine-Tetraacetic Acid
ESI	Electrospray Ionization
FACS	Fluorescence-activated cell sorting
FBS	Fetal Bovine Serum
HPLC	High Performance Liquid Chromatography
IAA	Iodoacetamide
IEF	Isoelectrofocusing
IPG	Immobyline Polyacrilamide Gel
LC	Liquid Chromatography

MALDI	Matrix Assisted Laser Desorption Ionisation
MS	Mass Spectrometry
NCBI	National Center for Biotechnology Information
PAGE	Polyacrilamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
pI	Isoelectric point
PMSF	Phenylmethanesulfonyl Fluoride
Q	Quadrupole
SC	Stem cell
SDS	Sodium Dodecyl Sulphate
TOF	Time Of Flight

FIGURES INDEX

TABLES INDEX

ABSTRACT

Recent findings suggest that malignant neoplasms are derived from a small sub-population of cells that acts as the "root" of tumours. This conclusion comes from the observation that when neoplastic cells of different types were tested for their growth potential both by *in vitro* and *in vivo* experiments, only a restricted minority of them displayed extensive proliferation. These cells are called *cancer stem cells* (CSCs): both anti-cancer drugs and irradiation cause cancer cells to die by apoptosis, however CSCs might survive and regenerate cancer. At present, CSCs theory represents a breakthrough in cancer research. The aim of this project is to characterize the protein expression pattern of CSCs to obtain further insights into the mechanisms of this class of cells. The knowledge of deregulated proteins could be the first step into the accomplishment of novel therapies targeted directly against CSCs. Particularly, we studied colon CSCs by using as experimental model two different colon cancer cell line systems: CaCo-2 and HCT-116. Putative CSCs were separated from non-CSCs by flow cytometry using CD133 as stemness marker. Then, total protein extract of CD133+ cells was compared to protein extract of CD133- cells and differentially expressed proteins were identified by 2D DIGE coupled with tandem mass spectrometry. Forty-nine differentially expressed proteins in CaCo-2 CD133+ vs CD133- cells and thirty-six in HCT-116 CD133+ vs CD133- cells were identified. Bioinformatics analysis of the differentially expressed proteins by using GeneOnthology and Ingenuity Pathway Analysis (IPA) software showed an alteration of energy metabolism, furthermore the examination of this network showed that several proteins were

directly or indirectly connected to MCC (mutated in colorectal cancer), a negative regulator of Wnt pathway. Interestingly, among the identified proteins it has been observed a 2-fold change up-regulation of the splicing factor SRp20, newly identified target gene of the Wnt/β-catenin pathway and we demonstrated a direct cause-effect relationship between Wnt pathway activation and the increased level of SRp20 expression. Furthermore, the results of this work show that SRp20 influences cell proliferation thus suggesting a putative function of this protein in tumorigenicity of CD133+ cells. In conclusion, the activation of the Wnt pathway in CD133+ cells and the consequent up-regulation of SRp20, which is implicated in tumorigenesis, raises the possibility of a sequential series of molecular events occurring in connection with this process.

1. INTRODUCTION

1.1 Cancer stem cells

Stem cells are distinguished from other cell types by two important characteristics. First, they are unspecialized cells capable of renewing themselves through cell division, sometimes after long periods of inactivity. Second, under certain physiologic or experimental conditions, they can be induced to become tissue or organ specific cells with special functions. One strategy by which stem cells can accomplish these two tasks is asymmetric cell division (Fig.1), whereby each stem cell divides to generate one daughter with a stem-cell fate (self-renewal) and one daughter that differentiates.

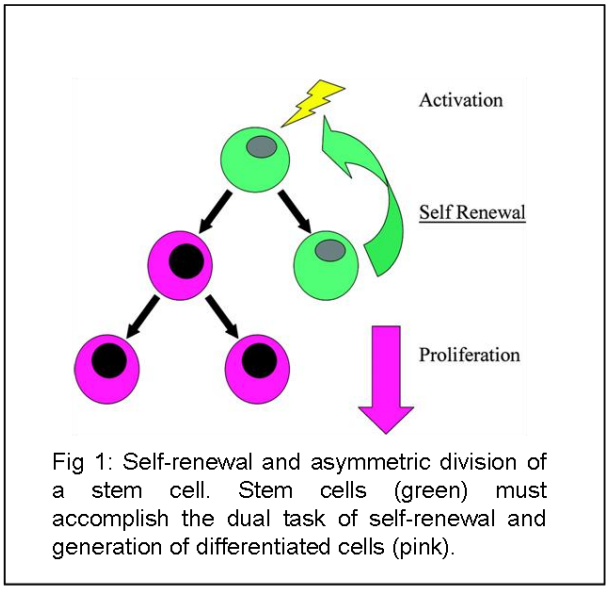


Fig 1: Self-renewal and asymmetric division of a stem cell. Stem cells (green) must accomplish the dual task of self-renewal and generation of differentiated cells (pink).

Recently, another class of stem cells have been identified: cancer stem cells (CSCs). Although the idea of CSCs has been proposed for a number of decades, demonstration of their existence has only occurred within the last ten

years. Advances in CSC isolation were first achieved in hematological malignancies, with the first CSC demonstrated in acute myeloid leukemia [1]. However, using similar strategies and technologies, and taking advantage of available surface markers, CSCs have been more recently demonstrated in a growing range of epithelial and other solid organ malignancies [2-4], including colorectal cancer [5-7] suggesting that the majority of malignancies are dependent on such a compartment. It is still unclear what is the origin of these cells: they may arise from normal stem cells, partially differentiated progenitor cells or fully differentiated cells (Fig.2) [8]. Certainly, normal stem cells, since they have a long lifespan, they have more opportunities to accumulate mutations leading to malignant transformation.

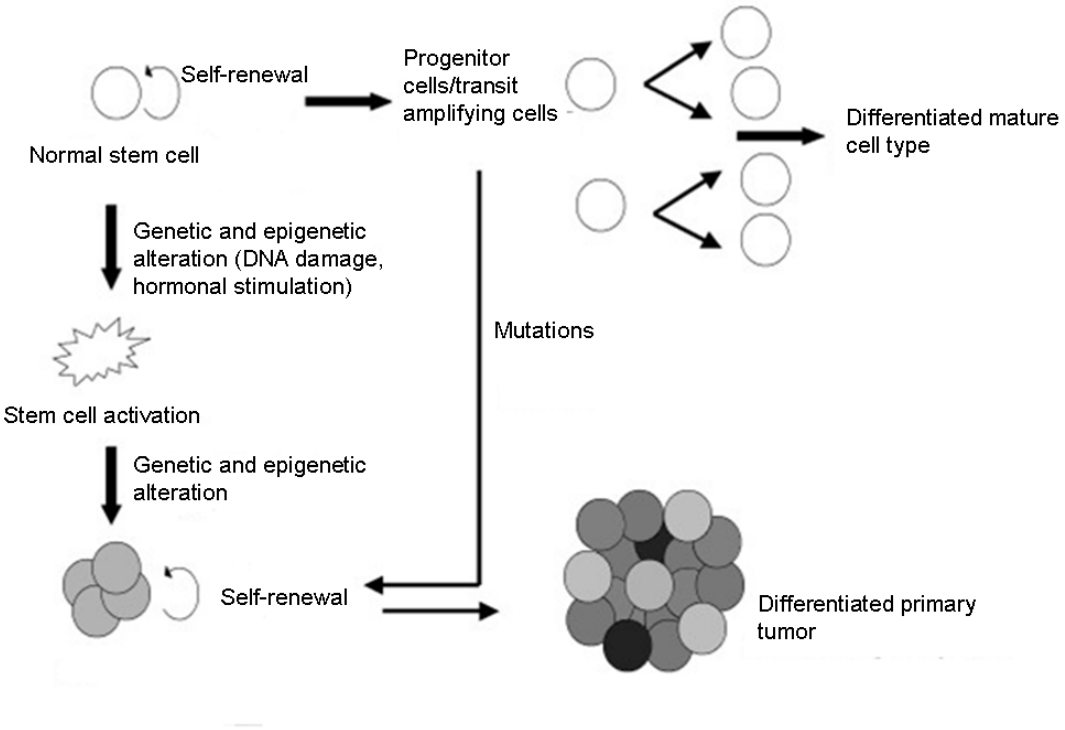


Fig.2: Development of cancer stem cells from the normal stem cells and progenitor cells.

Subsequently, after the CSCs isolation, a new cancer model has been proposed, in addition to the old stochastic model, to explain tumour development: the so called CSCs model or hierarchical model (Fig.3). According to the stochastic model, tumour cells are biologically equivalent but their behaviour is influenced by intrinsic and extrinsic factors and is both variable and unpredictable. Therefore, tumour-initiating activity cannot be enriched by sorting cells based on intrinsic characteristics. Hence the stochastic model (Fig.3A) assumes that every tumour cell is capable of initiating tumour growth and if it acquires a set of somatic mutations can develop metastatic capacity. In this case, all cancer cells must be eliminated to cure the patient. Conversely, the hierarchical model (Fig.3B) postulates the existence of biologically distinct classes of cells with differing functional abilities and behaviour and proposes that only few cells, the CSCs, are able to initiate and sustain tumour growth and to spread throughout the body, forming new tumours at distant sites [9].

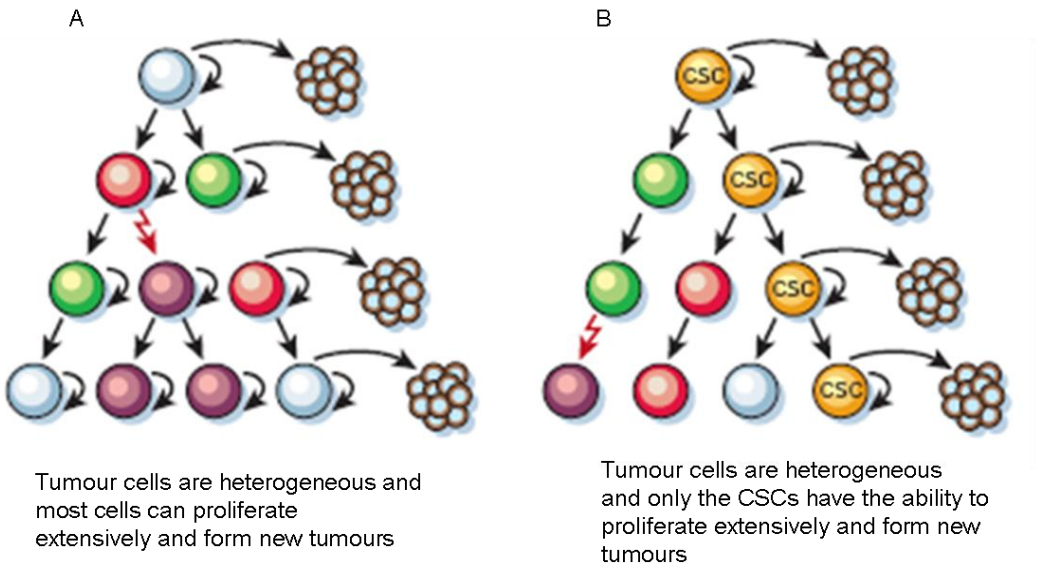


Fig. 3: Two general models of heterogeneity in solid cancer cells. A) Stochastic model and B) Hierarchical model (CSCs model)

Although CSCs are similar to normal adult stem cells, they have a set of several features that cause physiologic disarray: angiogenesis, invasion, metastasis and resistance to apoptosis with their cellular division that is driven by internal cellular events regardless of external stimuli [10]. Moreover these cells have a slower rate of division and greater ability to correct DNA defects than other cells. Therefore, CSCs are more resilient to adjuvant therapy promoting the evolution of resistant clones that persist and even if the bulk of the tumour is destroyed by chemotherapy or radiotherapy, the tumour will reoccur. Conventional therapies may shrink tumours by killing mainly cells with limited proliferative potential. If the putative cancer stem cells are less sensitive to these therapies, then they will remain viable after therapy and re-establish the tumour. By contrast, if therapies can be targeted against cancer stem cells, then they might more effectively kill the CSCs, rendering the tumours unable to

maintain themselves or grow (Fig. 4). Thus, even if CSCs-directed therapies do not shrink tumours initially, they may eventually lead to cures. To target CSCs, it is important to identify their regulatory mechanisms and signalling pathways.

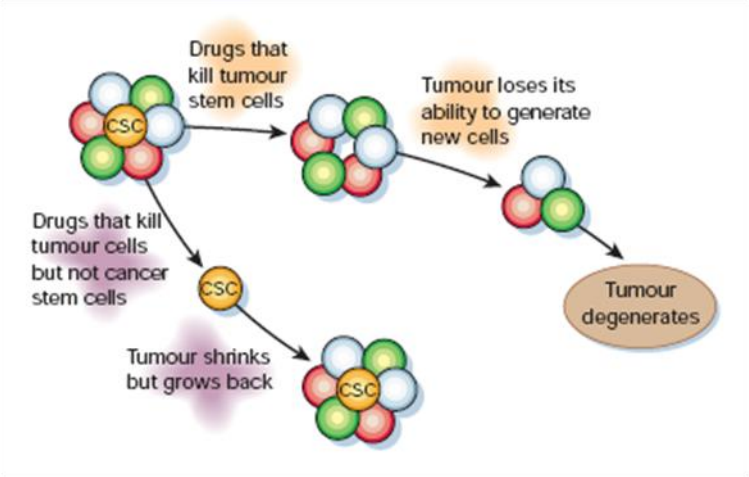


Fig. 4: Conventional therapy vs CSCs targeted therapy

1.2 Colon cancer stem cells

Colorectal cancer (CRC) is the third most common form of cancer and the second cause of cancer related death in the Western world, leading to 655,000 deaths worldwide per year [11]. It derives from an imbalance in proliferation and differentiation of the epithelium and apoptosis. When proliferation is no longer balanced from apoptosis, benign protrusions, called polyps, arise into the colon lumen. These polyps may develop into malignant cancers as CRC. The intestinal tract consists of the small intestine (duodenum, jejunum and ileum) and the large intestine or colon. The absorptive epithelium of the small intestine is ordered into villi and crypts of Lieberkühn. Differentiated cells (enterocytes, enteroendocrine cells and goblet cells) occupy the villi. A fourth differentiated type, the Paneth cells, functionally similar to neutrophils, resides

at the bottom of crypts only in the small intestine epithelium and secretes antimicrobial agents. The remainder of the crypts constitutes the stem/progenitor compartment. As shown in figure 5, putative stem cells (dark blue) reside immediately above the Paneth cells (yellow) near the crypt bottom, proliferating progenitor cells occupy the remainder of the crypt, differentiated cells (green) populate the villus, and include goblet cells, enterocytes and enter-endocrine cells [12]. As shown in figure 6, the mucosa of colon has a flat surface epithelium instead of villi and putative stem cells (dark blue) reside at the crypt bottom, proliferating progenitor cells occupy two-thirds of the crypt, differentiated cells (green) populate the remainder of the crypt and the flat surface epithelium. The terminally differentiated cells, which are found in the top third of the crypt, are continually extruded into the lumen. In fact, within 4-8 days, these differentiated cells die of apoptosis [13] and are replaced by other differentiated cells that derive from multipotent stem cells located at the bottom of the crypt in a ‘niche’ encased by intestinal subepithelial myofibroblasts (ISEMFs) separated only by the basal lamina. The ISEMFs form a syncytium within the lamina propria that extends along the length of the intestinal tract (Mc Donald et al., 2006; Radtke et al., 2005) [14,15]. This syncytium secretes a variety of cytokines that are important for wound healing and immune cell function. Furthermore, they control the proliferation and differentiation of the epithelial cells and play a role in electrolyte and water absorption. In pathological states, ISEMFs contribute to fibrosis, desmoplastic reactions and neoplasia [16].

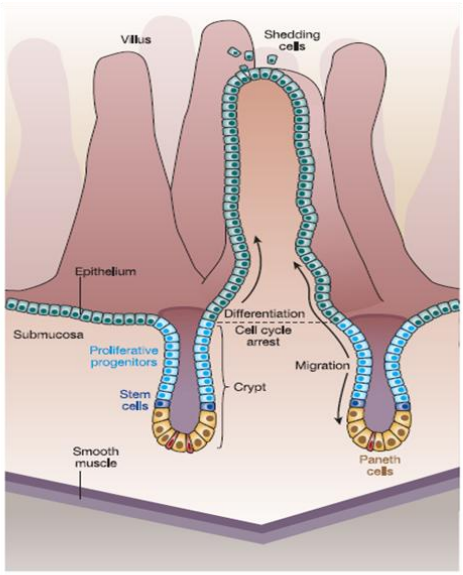


Fig. 5: Tissue anatomy of the adult small intestine

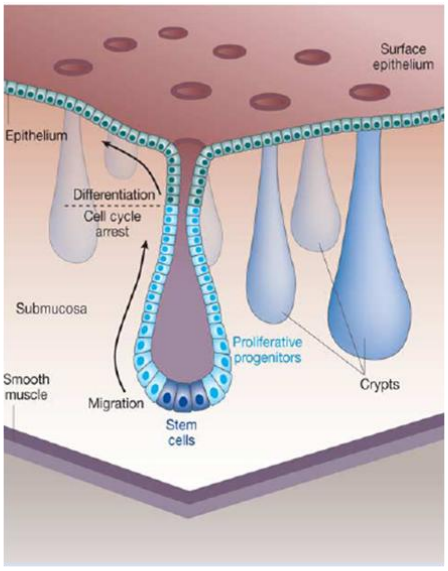


Fig. 6: Tissue anatomy of the colonic epithelium

During asymmetric division, the stem cells undergo self-renewal and generate a population of transit-amplifying cells that, upon migration upward the crypt proliferate and differentiate into one of the epithelial cell types of the intestinal wall and occupy the lower two thirds of the crypt. Although crypts are monoclonal, each villus receives cells from multiple crypts and is therefore polyclonal [17,18]. Existence of CSCs in colon cancer has been reported in 2007 by different research groups [5-7].

Symmetric division of colon CSCs is essential in achieving exponential numbers of tumour cells [19]. As mentioned above, the differentiated cells of colonic mucosa have a short lifespan of few days, whereas normal intestinal stem cells have a long lifespan and capacity to self-renew and therefore they have more opportunity to accumulate tumorigenic mutations [15]. These mutations are accompanied by phenotypic change in the mucosa. A new evidence suggests that the long-lived adult colonic stem cells can accumulate mutations for a prolonged period of time before phenotypic change becomes

apparent [20, 21]. The hypothesis of stem cell-driven tumourigenesis in CRC has received substantial support from the recent identification and phenotypic characterization of a subpopulation of colon cancer cells able to initiate tumour growth and to reproduce human colon carcinomas faithfully in mice.

The identification, isolation, and characterization of colon stem cells (SCs) is very difficult. Many obstacles have interfered with the identification of intestinal SCs among which the complexity of the crypt structure that limits the retrieval of putative SCs from their niche where they are interspersed among more differentiated daughter cells. Several studies have attempted to identify intestinal SCs within colonic crypts by using indirect techniques based on biological features restricted to the stem cell compartment. Long term retention of label DNA has been exploited as surrogate marker of stemness based on the observation that SCs in adult tissues usually divide at a slow rate when compared to the progenitor population [22]. This functional difference is highlighted by labelling the genetic material of proliferating cells in mouse intestinal crypts with tritiated thymidine [23] or by the administration of the DNA-labelling dye bromodeoxyuridine to rats in drinking water [24]. These approaches have allowed the identification of low mitotic index cells that undergo only limited dilution of label over time and are located at the bottom of the crypts. Only recently new methods have arisen to aid in the identification and isolation of CSCs. The most important of these has been the identification of surface markers by immuno-histochemistry. Other methods of identification include morphological features such as ‘bell shaped’ nuclei and their position at the base of the crypt [25].

The CSC population is defined by fluorescence-activated cell sorting (FACS) of tumour cells according to the expression of ‘signature’ cell surface biomarkers. Proposed biomarkers include CD133 [26-28], CD44 [29-31], CD34 [1,32], CD24 and epithelial-specific antigen (ESA) [29,30]. Injection of CSC-enriched populations into immunodeficient mice at low concentrations results in the formation of tumours with equivalent histology and phenotypic heterogeneity to the original neoplasm, whereas injection of non-CSCs, even at high concentrations, results in the growth of few or no tumours [26, 29-31]. Research so far suggests that the molecular ‘signature’ which specifically identifies CSCs is likely to constitute a combination of cell surface proteins that are co-ordinately expressed or repressed. The CD44, CD166, CD133 and EpCAM (epithelial cells adhesion molecule) are markers of tumorigenic cell population of colorectal cancer.

1.3 CD133

One of the main CSC markers identified was CD133, a pentaspan transmembrane glycoprotein also known in humans as Prominin 1 [33]. The CD133+ population is enriched in cancer-initiating cells in many tissues, including retinoblastoma [34], brain tumor [26,35], kidney cancer [36], prostate tumor [27], hepatocellular [37] and colon carcinomas [6,7]. Nonetheless, use of CD133 as a marker for identification and isolation of colon CSCs is a subject of debate; despite its use in isolating cell populations with cancer-initiating ability, studies have shown that CD133 is expressed by stem cells (SCs) and more differentiated progenitor cells [38]. CD133’s function is unclear, although it is

believed to have a role in asymmetric division and self-renewal. Bauer et al proposed that the polarized localization of CD133 indicates its role in regulating proliferation [39]. CD133 is concentrated in cell surface domains that correspond to the spindle pole region during metaphase. In telophase and cytokinesis, it is either equally or unequally distributed between the 2 nascent daughter cells. Studies have indicated a role for CD133 in tumor angiogenesis. CD133+ glioma cells produce proangiogenic factors that can directly modify endothelial cell behaviour [40]. Other data indicate that the CD133+ cell population can itself give rise to endothelial cells that promote vascularization and tumor growth, like renal progenitor cells do [41]. Within the intestine, CD133 would mark SCs susceptible to neoplastic transformation. These cells would be in fact prone to aberrantly activate Wnt signaling and such event would disrupt normal tissue maintenance leading to their aberrant expansion, resulting ultimately in neoplastic transformation of the intestinal mucosa [42].

The existence of colon CSCs was first reported by the research groups of John Dick and Ruggero De Maria [6,7] which independently described a small population of cancer cells capable of initiating tumor growth in immunodeficient mice. By implanting limiting dilutions of human colon cancer cell suspensions into pre-irradiated non-obese diabetic severe combined immunodeficient mice, O'Brien *et al* demonstrated that only a small subset of colon cancer cells ($1/5.7x10^4$ total cells) initiated tumor growth. Using flow cytometry, Ricci- Vitiani *et al* detected a rare population of CD133-/cytokeratin CK20- cells in colon tumor samples (2.5% ±1.4% of total cells). CK20 is considered a colonic epithelial terminal differentiation marker and therefore to be absent in the SC compartment. Based on immunohistochemical analyses, these cells were

present in areas of high cell density. The tumorigenic potential of colon CD133- cells was next analyzed by comparing the ability of CD133+ and CD133- populations to engraft and give rise to subcutaneous tumors in severe combined immunodeficient mice (Fig. 5). Low numbers of CD133+, but not high numbers of CD133- engrafted and formed tumors; high numbers of unsorted cells gave rise to tumors but, despite the high number of CD133- among them, tumor formation took more time.

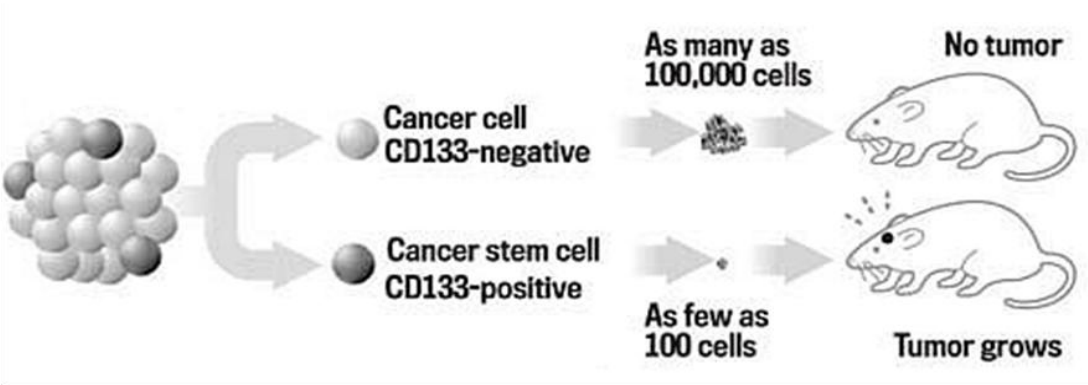


Fig. 7: Injection of CD133+ and CD133- tumor cells in mice

1.4 Proteomics and its impact on the life sciences

Genomics is the comprehensive analysis of the genetic content of an organism. It also often refers to genome wide studies of mRNA expression. Already during the “genomic era” that ended with the sequencing of Human Genome in the year 2003, the scientific community realized that the identification of coding sequences is insufficient to understand the molecular mechanisms of cell activity. Therefore, the attention increasingly focused on the products of the genome: the proteins and enzymes that determine cellular

architecture and function. The Proteome is the **prote**in complement of the gen**ome.** According to the current annotation, the human genome consists of about 25000 genes, scattered among 3 billions nucleotides of chromosome-based DNA code [43]. This represents a huge amount of static information, which needs to be correlated with dynamic information coming from gene products and their interactions. In contrast to the genome, the proteome is dynamic and is constantly modulated because of a combination of factors, which include mRNAs differential splicing, post translational modifications (PTMs), temporal and functional regulation of gene expression as well as the formation of multi-protein complexes. More than 100 modification types are recorded and additional ones are yet to be discovered [44]. All modified forms from one protein can vary in abundance, activity or location inside a cell. Indeed, cellular proteins are not invariant products of genes, but are subject to a high degree of interdependent processing at the protein level that is a critical component of cellular function and regulation. In addition, protein expression is dynamically regulated in response to external and internal perturbations under developmental, physiological, pathological, pharmacological and aging conditions. In fact, in contrast to the static genome, where all information could in theory be obtained from the DNA of a single cell, the proteome is considered dynamic because highly dependent not only on the type of cell, but also on the state of the cell [45]. Proteomics provides methods for correlating the vast amount of genomics information that is becoming available with the equally vast protein information that is being produced through analysis of cells under normal versus altered states [46]. In the last few years proteomics has become a powerful tool for the investigation of complex biochemical processes and

protein-protein interactions [47-48]. In medical sciences, proteomics has manifested significant impact on various aspects of clinical research, including understanding of disease pathogenesis, discovery of novel biomarkers for early disease diagnosis as well as identification of new drug and vaccine targets [49–52]. The biomarkers are biomolecules that are used to aid in monitoring disease progression and following prognosis in response to the therapeutic interventions. Identification of protein biomarkers is useful for early detection of various fatal diseases such as cancer or autoimmune disorders and has significant impact on human health [53, 54].

1.5 Proteomics analysis

The term Proteomics is associated to the set of analytical tools used to depict the protein compartment of a cell. It is the natural continuation of Genomics approach and it moves away from classical Protein Chemistry, taking advantages of all the heritage of knowledge and methods developed from the latter. The great innovation of the Proteomics analysis is in fact the idea that to study the cellular molecular mechanisms, in which proteins play a key role, it is necessary to study the entire proteome as a “single analyte” this means that, the proteins target of the analysis are no more purified and isolated from their highly complexity context found in living systems. Indeed, they are analyzed all together in order to obtain a real snapshot of the proteome, related to a particular cell state. From this general definition it is possible to distinguish two main areas of interest: i) “expression proteomics” that is focused on the characterization of the change in protein expression levels and eventually the

definition of PTMs [55], ii) “functional proteomics” aimed to understand protein-protein interactions [56], signal pathways [57] and structure function relations. However, the proteome analysis is hindered by several analytical problems. First of all is the large range of protein concentration present in the samples. For example, in the human serum the 50 most abundant proteins represent about 99% of the total amount of protein mass but only less than 0.1% in number [58]. Another important challenge is surely the detection of PTMs. In fact only a minor part of the proteins of interest are post translationally modified. The high sample complexity, in terms of number of analytes, is also a feature that has to be taken in account, in fact for about 21000 human protein encoding genes are estimated around 106 human proteins [59]. For these reasons the Proteomics analysis needs a pool of methodologies and technologies that are high throughput, sensitive, selective toward the proteins target of the analysis and with large dynamic range effectiveness. Currently, Proteomics may rely on many chromatographic and electrophoresis tools to fractionate the analytes. However, if different approaches are in relation to these techniques of separation, all the strategies have a common essential final step: the mass spectrometry (MS) analysis of peptides or proteins. In the following section the expression proteomics employed for the purposes of this PhD programme will be covered in more detail.

1.5.1 Expression proteomics

Mammalian cells contain thousands of different proteins and only a small number of them changes under such circumstances. The primary goal of expression proteomics is to detect differences in protein expression patterns between normal and diseased tissue. Proteome analysis bases on two essential components: protein separation and protein identification. Historically, the tool of choice for maximal separation of proteins was two-dimensional electrophoresis (2-DE) that relies on separating proteins based on their isoelectric point (pI) and molecular weight (MW) (Fig. 8).

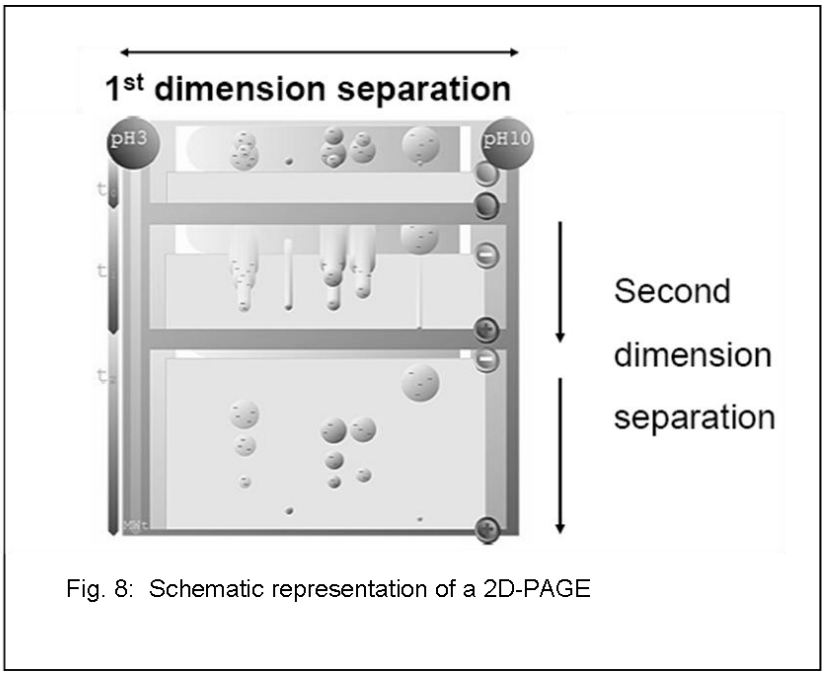


Fig. 8: Schematic representation of a 2D-PAGE

Proteins carry a negative, positive or zero net charge depending on their amino acid composition and covalent modification (such as phosphorylation, nitrosylation, sulphation and glycosylation), and the pH of the environment. The

pI of a protein is the pH at which the protein carries no net charge. If the proteins migrate in a pH-gradient, they will move until they reach a position in the pH-gradient where their overall net charge is zero, i.e. the pH is equal to the pI of the protein. The original 2DE method, described by O'Farrell (1975) [60], used carrier ampholytes in tube gels to create and maintain a pH-gradient. Carrier ampholytes are small amphoteric molecules with high buffering capacity near their isoelectric points and are usually employed as mixtures covering a set pH range. When an electric field is applied across a mixture of carrier ampholytes the most negatively charged proteins move towards the anode and the most positive ones towards the cathode. In this way it is possible to form a continuous pH-gradient within a gel, which is suitable for focusing larger amphoteric molecules such as proteins. However, this method had limitations in the resolving power and in the pH gradient stability. Several innovations significantly improved reproducibility and performance of the first dimension focusing step (IEF) in 2D-PAGE such as commercially available immobilized pH gradient strips, in which the carrier ampholytes are co-polymerised into the gel matrix. Furthermore, the gel is cast onto a plastic support strip making the system much more robust and easier to handle. The second dimension uses the traditional SDS-PAGE technique; the IEF strip replaces the stacking gel. Despite being a well-established technique for protein analysis, traditional 2D gel electrophoresis is time-consuming and labour-intensive. Nevertheless, in 2-DE there are sources of variability that can distort the real difference in protein expression: a) analytical variations due to treatment of the sample, to procedures for staining or to image acquisition; b) biological variations due to environment in which the sample was produced,

processed and preserved. To minimize these variations, it is useful to work with multiple biological and analytical replicates thus increasing the difficulty of analysis (Fig.9).

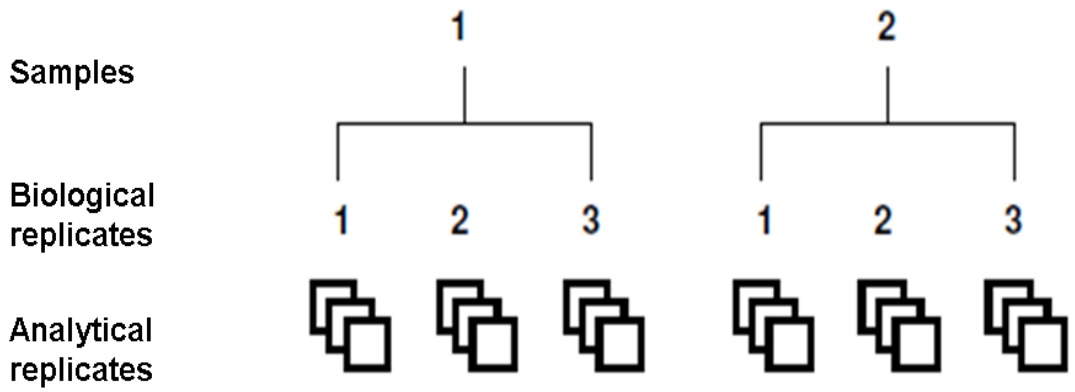


Fig. 9: Hierarchical structure for differential studies.

1.5.2 2D-DIGE

The development of 2-D differential in-gel electrophoresis (2D- DIGE) in 1997 overcame this limitation by allowing up to three distinct protein mixtures to be separated within a single 2D-PAGE gel [61]. In a typical 2D-DIGE experiment, proteins extracted from three different samples, healthy, diseased and internal control (a pooled sample formed from mixing equal amounts of the proteins extracted from the healthy and diseased samples), are covalently labeled, each with a cyanine fluorescent dye that has a different excitation and emission wavelength. The ability to multiplex different cyanine dye labelled samples on the same gel means that the different samples will be subject to

exactly the same 1st and 2nd dimension running conditions. Consequently the same protein labeled with any of the dyes and separated on the same gel will migrate to the same position on the 2D gel and overlay. This limits experimental variation and ensures accurate within-gel matching. The used cyanine dyes are: 1-(5-carboxypentyl)- 1′-propylindocarbocyanine halide N-hydroxysuccinimidyl ester (Cy3); N 1-(5-carboxypentyl)-1′-methylindodicarbocyanine halide N-hydroxysuccinimidyl ester (Cy5); and 3-(4-carboxymethyl)phenylmethyl- 3′-ethyloxacarbocyanine halide N-hydroxysuccinimidyl ester (Cy2) (Fig. 10). Equal concentrations of the differentially labeled proteomes and of the control sample are mixed, applied to a single gel plate, and separated using 2D-PAGE. The control sample serves as an internal standard, enabling both inter- and intra-gel matching. The control sample should contain every protein present across all samples in an experiment. This means that every protein in the experiment has a unique signal in the internal standard, which is used for direct quantitative comparisons within each gel and to normalize quantitative abundance values for each protein between gels. Scanning the gel at the specific excitation wavelengths of each dye, using a fluorescence imager, allows visualization of the differentially labeled proteins. The images are then merged and analyzed using imaging software, which enables differences between the abundance levels of proteins to be compared.

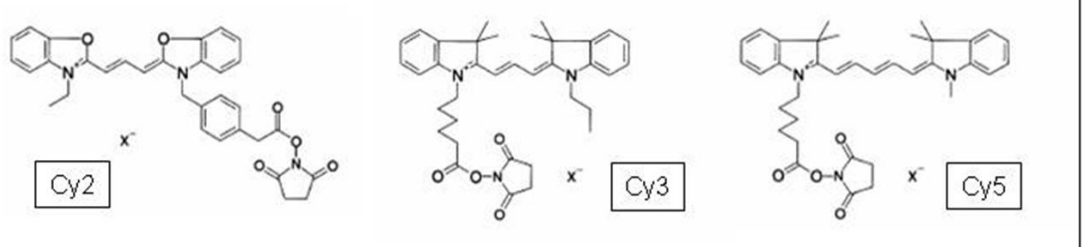


Fig.10: Cy2, Cy3 and Cy5 chemical structures

To compare protein expression across a range of experimental samples and gels, two distinct steps are required during the image analysis (Figure 11): a) intra-gel co-detection of sample and internal standard protein spots and b) inter-gel matching of internal standard samples across all gels within the experiment. In the first step three scans are made of each gel, Cy2, Cy3 and Cy5 scans. Scanned images of each sample and the internal standard are overlaid by a software. The algorithms within the software co-detect the spots present in each scan-image, effectively identifying the position of each spot within the gel (Figure 11a). In this way every protein in the sample is intrinsically linked to the corresponding protein spot in the internal standard sample. In the second step, the inter-gel comparisons of spot abundance are carried out. Following co-detection, each image has a spot map species. The internal standard image with the most detected spots is assigned as the 'Master'. The spot map species for the internal standard assigned as the Master, is used as a template to which all remaining spot map species for the other internal standards (intrinsically linked to their co-detected sample

images) are matched (Figure 11b). Once the protein spots have been matched, the ratio of protein abundance between samples can be determined.

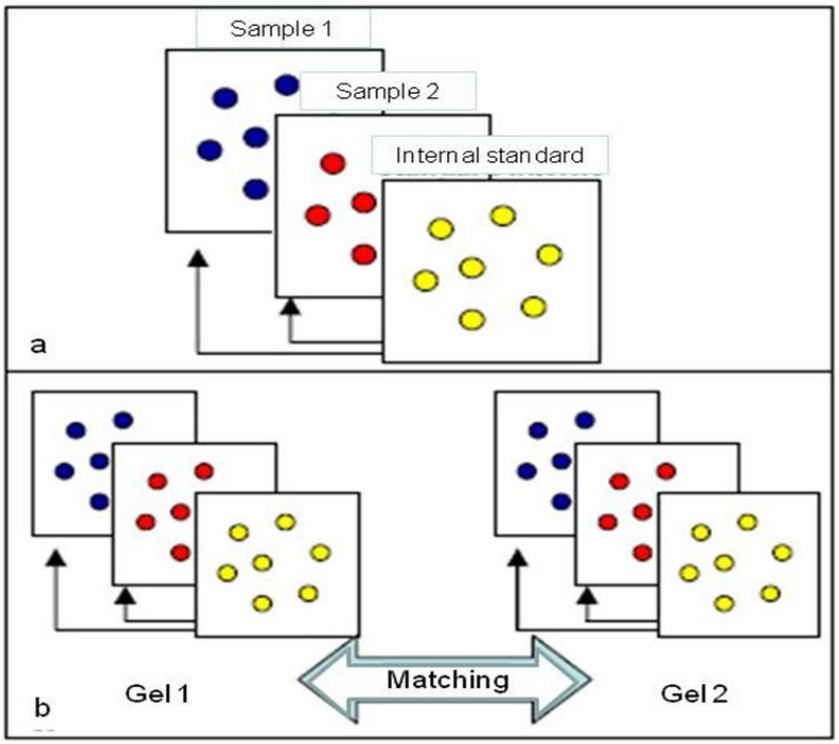


Fig.11: 2D-DIGE image analysis: a) intra-gel co-detection of sample and internal standard protein spots; b) inter-gel matching of internal standard samples across all gels within the experiment.

Spot volume (i.e. the sum of the pixel values within a spot minus background) for each experimental sample is compared directly to the internal standard by the software. Spot volume ratios are calculated indicating the change in spot volume between the two images. The protein abundance for each spot in each sample is then expressed as a (normalized) ratio relative to the internal standard. Statistical tests such as the Student's t-Test can then be applied to the data-software. The statistical tests verify that any change between the groups is significant and give the user a level of confidence by taking into

account the inherent biological variation within a group compared to the induced difference between groups. It assigns a confidence rating as to whether this change is above the biological variation. Concluding, the introduction of 2D-DIGE contributed immensely to solving problems of reproducibility and quantization. The use of imagers and computers allows not only fast data mining, acquisition, and analysis but also spot detection, normalization, protein profiling, background correction, reporting and exporting of data. As a separation, detection and quantization technique, 2D-DIGE is an important tool especially for clinical laboratories involved in the determination of protein expression levels and disease biomarker discovery. When absolute biological variation between samples is the main objective, as in biomarker discovery, 2D-DIGE is the method of choice.

1.6 Mass spectrometry tools for proteomic analysis

-Principles and instrumentation-

Whatever the used proteomic approach, the common essential final step of a proteomic experimental workflow is the mass spectrometry (MS) analysis. Mass spectrometers consist of three basic components: an ion source, a mass analyser, and an ion detector. Molecular mass measurements are carried out indirectly, calculating the mass charge ratio (m/z), by the analysis of kinetics behaviour of the ionized analytes in the gas phase and in electromagnetic fields. The need of a method that transfers efficiently intact molecules from solution or solid phase into the gas phase, was the limit that for many years

excluded proteins and peptides from this powerful technique. With the development and the commercialization of mass spectrometers based on the ion source technologies of “Matrix-Assisted Laser Desorption/Ionisation” (MALDI) [62] and “Electrospray Ionisation” (ESI) [63] the gas phase ionization of polar analytes has also become possible. Both MALDI and ESI are considered as soft ionisation techniques with which the generated ions undergo little or none fragmentation. In MALDI technique, samples are co-crystallised with a weak aromatic acid matrix on a metal target. A pulsed laser is used to excite the matrix, which causes rapid thermal heating of the molecules and desorption of ions into the gas phase in a pulsed beam fashion. After the laser activation, the weak acid nature of the matrix drives to generate single-charged ions in the form of MH+ adduct (Fig.12). The one-one analyte-MS signal relation in MALDI-MS makes possible to analyze also multi-component mixtures without any interpretation difficulties caused by the peaks overlapping.

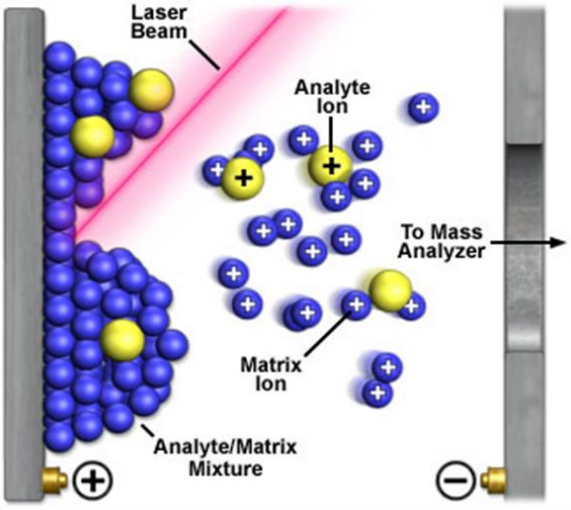


Fig.12: Schematic representation of the processes occurring in a MALDI source

ESI creates ions by spraying an electrically generated fine mist of ions into the inlet of a mass spectrometer at atmospheric pressure. By creating a potential difference between the capillary, through which the liquid flows, and the inlet of the mass spectrometer, small droplets of liquid are formed. These are transferred into a heated device to induce efficient evaporation of solvent. Once the droplets have reached the Rayleigh limit, ions are desorbed from the droplet generating gas-phase ions in a continuous beam fashion (Fig.13). The ions produced by ESI sources are multiply charged adducts MH_n^{n+}, where the number of protons incorporated depends on the statistical acid/basic equilibrium of the analytes. This situation complicates the spectra interpretation because each analyte may give rise to many signals in the spectrum. For this reason it is not possible to analyze complex mixture without a previous fractionation. However, because the ionization is carried on a liquid flow, ESI source may be simply coupled to liquid chromatography systems.

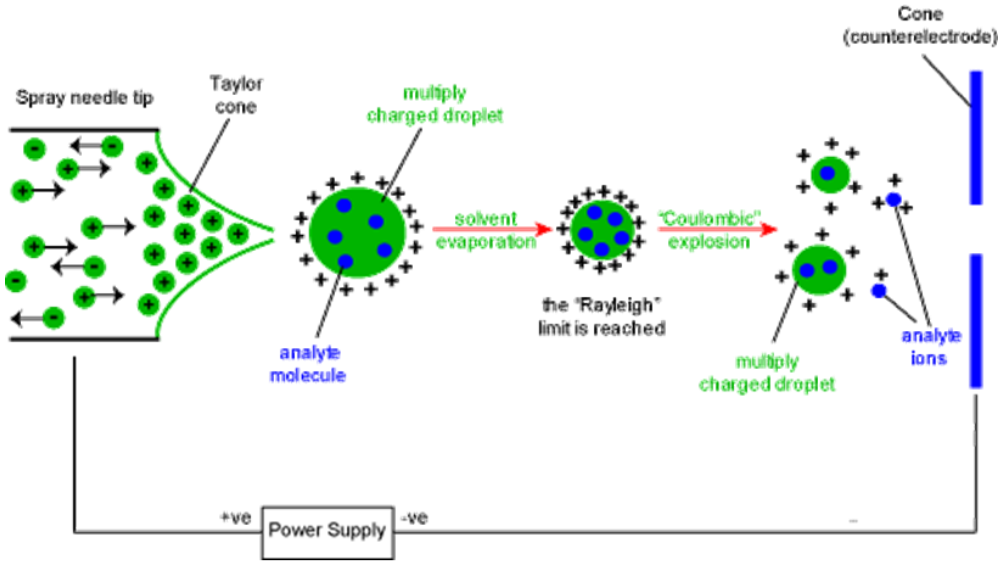


Fig.13: Schematic representation of the processes occurring in a ESI source

The most notable improvement of ESI technique has come from the reduction of the liquid flow rate used to create the electrospray to nano-scale level. Nano-sources create ions more efficiently [64] because the charge density at the Rayleigh limit increases significantly with decreasing droplet size. Another advantage of using separation techniques with nano-ESI is the increase in the concentration of the analyte as it elutes off the column. After ionisation, the analytes reach the mass analyzer, which separates ions by their mass-to-charge (m/z) ratios. Ion motion in the mass analyser can be manipulated by electric or magnetic fields to direct ions to a detector, usually an electro-multiplier, which records the numbers of ions at each individual m/z value converting the signals in current. In Proteomic research, four basic kinds of mass analysers are currently used: time-of-flight (TOF), ion trap (IT), quadrupole (Q), and Fourier transform ion cyclotron resonance (FTICR) analysers (Fig.14). All four differ considerably in sensitivity, resolution, mass accuracy and the possibility to fragment peptide ions. They are very different in design and performance, each with its own strength and weakness. These analysers can be stand alone or, in some cases, put together in tandem to take advantage of the strengths of each.

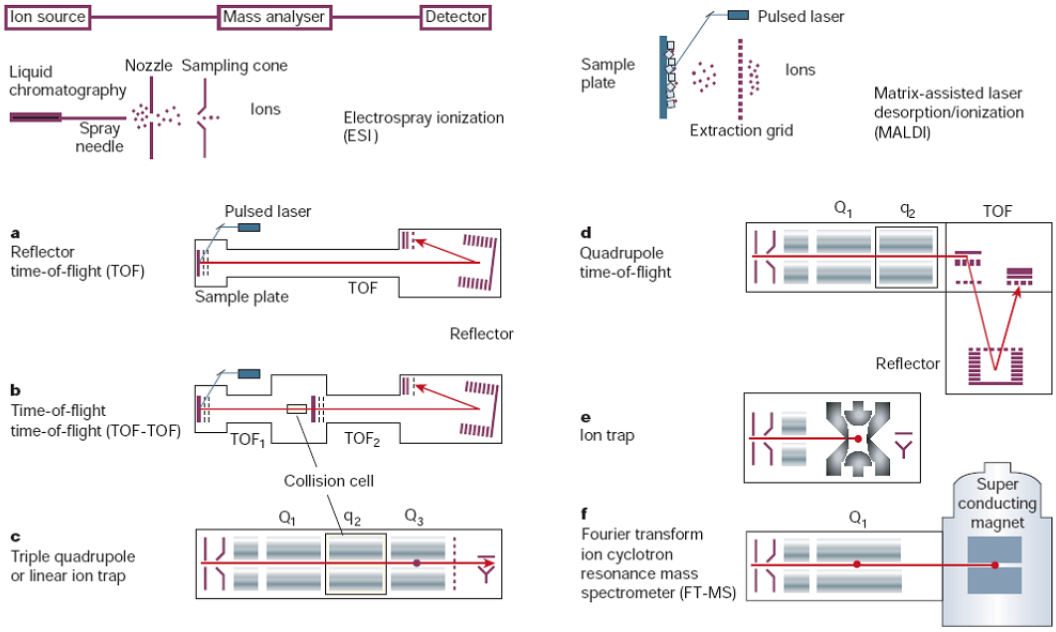


Fig.14: Main components of Proteomics commonly used mass spectrometers

The quadrupole mass filter consists of a linear array of four symmetrically arranged rods to which radio frequency (RF) and DC voltages are supplied. Forces are exerted in a plane normal to the direction (z-direction) in which the ions drift. The RF potential gives rise to a field which alternatively reinforces and then dominates the DC field. Ions oscillate in the x,y-plane with frequencies which depend on their m/z values. If the oscillations of an ion in this plane are stable, the ion will continue to drift down the rod assembly and reach the detector. Stable oscillations are only achieved by ions of given m/z value for a given rod assembly, oscillation frequency, RF voltages and DC voltages. Commercially available instruments usually have mass/charge limits ranging from 0 to 4000 m/z and at best are normally set to resolve the various ^{13}C isotope peaks for a singly charged ion, although the resolution may be intentionally degraded to improve sensitivity. In ESI, multiple charging enables quadrupole mass measurement of molecules >100,000 Da, if the molecule can be charged sufficiently.

The ion trap consists of three electrodes, the central ring electrode and two end-cap electrodes of hyperbolic cross-section. In this device too, ions are subjected to forces applied by an RF field but the forces occur in all three, instead of just two, dimensions. In ion-trap analysers, ions are resonantly activated and ejected by electronic manipulation of this field. Ion traps are robust, sensitive and relatively inexpensive, and so have produced much of the proteomics data reported in the literature. A disadvantage of ion traps is their relatively low mass accuracy, due in part to the limited number of ions that can be accumulated at their point-like centre before space-charging effects distort their distribution and thus the accuracy of the mass measurement.

Usually, in the Time-of flight (TOF) analyzers the ions are accelerated through a fixed potential into a drift tube. As all the ions with same charge obtain the same kinetic energy after acceleration, the lower m/z ions achieve higher velocities than higher m/z ions. Moreover, ion velocities are inversely related to the square root of the m/z. Thus, by measuring the time it takes to reach the detector, the m/z of the ion can be determined. Thanks to several strategies, such as delayed extraction of ions from the source, two stage sources with complex voltage gradients, and reflectron detectors, a commercial TOF instrument can typically achieve resolution of 10,000 or greater.

A Fourier-transform ion-cyclotron resonance (FT-ICR) mass spectrometer (also referred to as a Fourier-transform mass spectrometer, or FTMS) uses a magnetic field to determine the m/z of an ion. In an FT-ICR ions have kinetic energies, at most, of a few tens of electron volts (eV). At low kinetic energies, ions are actually trapped under high vacuum in the magnetic field. For a

constant magnetic field, ions oscillate around the magnetic field with a cyclotron frequency that is inversely related to the m/z. In a very simplified view of FT-ICR, the cyclotron frequencies of the ions trapped in the FT-ICR are measured and then converted into m/z. State-of-the-art electronic equipment is capable of measuring frequencies with extremely high precision. This translates to a very high mass resolution, which is the property FT-ICR is most widely known for. Mass resolving power in the hundreds of thousands are fairly easy to obtain on instruments with large magnetic field strengths (that is, > 7 Tesla), and resolutions in the millions have been demonstrated. Very high mass accuracies, down to the ppm level, can also be obtained (resolution >2,000,000).

- *Tandem MS* -

The power of mass spectrometry can be dramatically increased by employing methods of tandem mass spectrometry. Conventional MS produces ions that are separated by m/z and analyzed directly. If a soft ionization method is used, the mass spectrum will lead to calculate the molecular weight values of compounds present in the analyte but with little or no structural information. In a tandem MS (MS/MS) experiment, the first mass analyzer (MS1) is used to selectively pass an ion (precursor ion) into another reaction region where excitation and dissociation take place. The second mass analyzer (MS2) is used to record the m/z values of the dissociation products (daughter ions). Low-energy collision-induced (activated) dissociation (CID or CAD) [65] tandem mass spectrometry has been, by far, the most common method used to dissociate peptide ions for subsequent sequence analysis. Upon collisional

activation with a non-reactive gas, such as argon or helium, the amide bond of the peptide backbone will fragment to produce, ideally, a homologous series of **b** and **y**-type fragment ions (Fig.15). The observed fragmentation pattern depends on various parameters including the amino acid composition and size of the peptide, excitation method, time scale of the instrument, the charge state of the ion, etc. Peptide precursor ions dissociated under the most usual low-energy collision conditions fragment along the backbone at the amide bonds forming structurally informative sequence ions and less useful non-sequence ions by losing small neutrals like water, ammonia, etc.

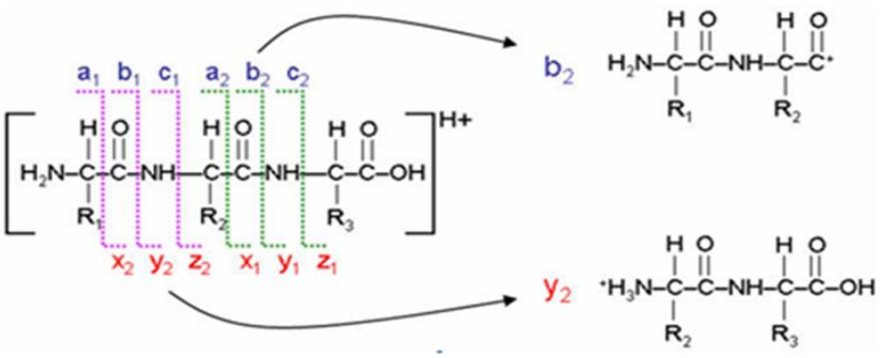


Fig.15: Peptide fragmentation and product ion generation Although **b** and **y** ions are considered to be the most useful sequence ion types, because they correspond to cleavage of the amide bond, other ion types are observed and used in spectral interpretation. These include **a** ions which correspond formally to loss of CO from **b** ion; a Δm= 28 Da between two peaks suggests an **a-b** ion pair and is useful in identifying the ion series to which the peaks belong. The **y** series is sometimes accompanied by satellite peaks formally corresponding to loss from the **y** ions, allowing designation of **y** ion series. Ions that correspond to NH_3 immonium ions, or fragments of immonium ions, of individual amino acid residues in a peptide are often detected, even for residues from the internal portion of the sequence.

Therefore, to achieve a multi-stage MS analysis many mass spectrometers equipped with hybrid combinations of analyzer have been built and are now

commercialized. Triple quadrupole (QqQ), quadrupole/ time of flight (QqTOF) and time of flight/ time of flight (TOFTOF) may perform “space separated” tandem MS. Ion trap analyzer by itself may realize tandem MS experiments “time separated”. This is accomplished by the expulsion of the ions trapped in MS mode, isolating the parent ion. This one is further fragmented by collision and the produced ions are first accumulated, then scanned and detected. This “time separated” operative mode allows to realize very sensitive MS^n experiments with simple equipments. By using the above described MS tools, there are two different methods to identify proteins. The first method, usually obtained by MALDI-MS, is the *“peptide mass fingerprinting”* (PMF) [66] based on the idea that a protein digested by an enzyme with known specificity produces a peptides pool that may be used as discriminatory for its identification. MALDI-TOF analysis produces a unique spectrum giving the accurate monoisotopic mass of all the peptides produced by the protein digestion, producing peculiar molecular masses map for each protein. This map is then compared to the ones generated *in silico* by the virtual digestion of all the protein sequences present in a given interrogated database. The PMF method works well for isolated proteins, but the resulting protein identifications are not sufficiently specific for protein mixtures (e.g. for co-migrating proteins). The addition of sequencing capability to the MALDI method should make protein identifications more specific than those obtained by simple peptide-mass mapping. In this case the second identification method “MSMS ion search” should be implied. This method uses the MS tandem methodology and thanks to the peptide fragmentation, gives information about the amino acidic

sequence of the peptides (and so, of the proteins) in analysis. It is sufficient a sequence of 5-6 amminoacids to identify univocally a protein.

1.7 Aim of the PhD thesis

The aim of this study was to obtain insights into the colon cancer stem cells (CSCs) molecular mechanisms, in fact CSCs theory represents a breakthrough in the recent cancer research. As experimental model, we used a combined approach of flow-cytometry, differential proteomic and mass spectrometry on two different colon cancer cell line systems i.e. CaCo-2 and HCT-116 derived from tumors taken directly from patients. Flow-cytometry was employed to separate putative colon CSCs from non-CSCs according the expression of the universal recognized stem cell marker CD133. Then, a differential proteomic approach was used to compare total protein extracts of CD133+ cells (putative CSCs) to total protein extracts of CD133- cells (non-CSCs). Particularly, the innovative and powerful methodology of 2D-DIGE was applied to our samples. The protein spots differentially expressed in the two sub-population of cells were identified by mass spectrometry. Several selected proteins were validated by western blotting, the whole group of identified proteins for each cell line was submitted to bio-informatic analysis and finally, functional assays on selected proteins of interest were performed. In the following sections the results obtained during the four years of PhD programme will be illustrated in detail.

2. MATERIALS AND METHODS

2.1 Cell cultures

Three different colon adenocarcinoma-derived cell lines were used: CaCo-2, HCT-116 and HT-29, all available from the CEINGE Cell Bank (Naples, Italy). The CaCo-2 cell line was grown in Essential Minimum Eagle's Medium (EMEM, Sigma-Aldrich, Oakville, ON) supplemented with 10% Fetal Bovine Serum (FBS, Gibco, Carlsbad, CA), 1% ultraglutammine (Cambrex, East Rutherford, NJ), 1% non essential amino acids and 1% sodium pyruvate. HCT-116 and HT-29 were propagated in McCoy'S 5A (Sigma-Aldrich) medium supplemented with 10% FBS and 1% ultraglutammine. Adherent cells were detached using Trypsin-EDTA (Sigma-Aldrich) solution, floating cells were collected and Trypsin-EDTA was inactivated using complete culture medium.

2.2 Immunophenotyping and flow cytometry experiment

After detaching, cells were washed and reconstituted to a final concentration of 10 x 106 cells/mL in 2% FBS/PBS. As preliminary step, we analyzed the intrinsic fluorescence of CaCo-2, HCT-116 and HT-29 by using unlabelled samples that act as control. Cells were subdivided in 5ml polystyrene tubes (Falcon, Becton Dickinson, San Jose, CA). Cell suspensions (50µL) were incubated with 5µL of each antibody for 30 minutes at 4°C. Cells were then washed with 1mL 2% FBS/PBS, re-suspended in 500µL 2% FBS/PBS and analyzed by flow cytometry. FACSAria cell sorter (Becton Dickinson), equipped

with four excitation lines (633nm, 488nm, 407nm and 375nm lasers), was employed for the analysis. FITC, PE, Per-CP, PE-Cy7 were analyzed on the 488nm line while allophyco-cyanin was excited by the 633nm laser. The antibodies used in this study included CD133-PE and CD133-allophyco-cyanin (AC133 clone; Miltenyi Biotec, Auburn,CA).

2.3 Gating strategies

Cells were first gated on physical parameters (FSC and SSC) to exclude the majority of dead and apoptotic cells. Then, FSC-Area versus FSC-height profiles were used to identify single cells and to exclude doublets. CD133+ cells were gated in CD133 vs “empty channel” dot plot and then were sorted and collected. The isolated cell pellets were stored at -80°C until further analysis.

2.4 Sample preparation for 2D-DIGE analysis

To obtain total protein extracts, cells were washed twice with cold PBS, centrifugated at 1000 rpm for 5 min and resuspended with a lysis buffer containing 7M urea, 2M thiourea, 30 mM Tris-HCl pH 8.5, 4% CHAPS (w/v), 1x Complete® EDTA free, containing a cocktail of protease inhibitors (Roche Applied Science, Indianapolis, IN, USA). Protein extracts were incubated at 4°C for 5 min and than sonicated to disrupt the cells and to shear the DNA and RNA in the cell. Protein samples were cleared from cell debris by centrifugation at 14000 rpm at 4°C for 20 min and then purified using the 2-D Clean-up Kit

(GE Healthcare, Piscataway, NJ, USA) following supplier's instructions. Protein samples were then resuspended in lysis buffer. In order to perform the reaction between the N-hydroxysuccinimidyl ester reactive group of the CyDye fluorochrome and the epsilon amino group of lysine residues of proteins, protein solution pH was adjusted to the value of 8.5. Protein quantification was performed with the 2-D Quant Kit (GE Healthcare) by reading protein absorbance at 480 nm. To perform a successful CyDyes labeling using the Ettan DIGE Manual, protein concentrations was adjusted between 5 and 10 mg/ml of lysis buffer.

2.5 Labeling efficiency and same same same tests

The labeling efficiency of the samples with the CyDye DIGE Fluors was tested, before DIGE experiment, by performing the following reactions:

- 50 µg of total protein extract for each cell line was labeled with 400 pmol of Cydye Cy5.

- 50 µg of total *E. coli* protein extract, used as control, was labeled with 400 pmol of Cydye Cy5.

E. coli was chosen because its protein extract has already been labeled successfully. Herein, the procedure applied on CaCo-2 is described, but the same experiments were performed on HCT-116. Labeling reactions were carried out in the dark on ice for 30 min before quenching with 1 µl of a 10 mM L-Lysine solution for 10 min. Serial dilutions of 25 µg, 12.5 µg and 6.25 µg of CaCo-2 and *E.coli* protein lysates were made. Proteins were then resolved on

a one-dimensional SDS gel with a concentration of 12.5% polyacrylamide. The gel was than acquired at the Cy5 wavelength using the Typhoon 9400 imager (GE Healthcare) and processed and analyzed with Image Quant Analysis software (GE Healthcare) to verify that the labeling efficiency of the protein sample is comparable to the control.

The “same same same” test was performed to verify that none of the three Cydyes labels the same test sample preferentially. 50 µg of total CaCo-2 protein extract was labeled with 400 pmol of each Cydye. Proteins were resolved on a 2D polyacrylamide gel (26x20 cm) by using an Ettan DaltTwelve system (GE Healthcare). After electrophoretic separation, gels were scanned using the Typhoon 9400 imager (GE Healthcare). Fluorescence-labeled proteins were visualized at the appropriate wavelength for Cy3, for Cy5 and for Cy2. Images were acquired with Image Quant Analysis software (GE Healthcare). The images were processed and analyzed by DeCyder v5.02 software (GE Healthcare). The comparison of the volume of fluorescence for all spots allowed us to define a threshold. All variations under this threshold were not considerated.

2.6 Labeling of protein extracts

Protein extracts (50µg) from 4 different biological replicates of CD133+ and CD133- cells were separately labeled at pH 8.5 with 400pmol of Cy3 and Cy5, according to manufacturer’s protocol. We found no appreciable differences between spot patterns and volumes of Cy3 and Cy5 labeled proteins (data not

shown). However, to avoid any possible differences due to staining effectiveness we crossed the dyes between the pairs of analyzed samples (Table 1). As internal standard we used a mixture containing equal amount of the eight lysates in analysis labeled with Cy2. Labeling reactions were stopped with 1mM lysine. Each Cy3/Cy5-labeled sample pair was mixed with a Cy2-labeled pooled standard sample. The Cy2/Cy3/Cy5-labeled samples were run together on the same gel (Gels 1-4 in Table 1).

Table 1. 2D-DIGE experimental design

Gel	**Cy3[a] (50 µg)**	**Cy5[a] (50 µg)**	**Cy2[a] (50 µg)**
1	CD133+ replicate 1	CD133- replicate 2	Pool standard[b]
2	CD133+ replicate 2	CD133- replicate 1	Pool standard[b]
3	CD133- replicate 3	CD133+ replicate 4	Pool standard[b]
4	CD133- replicate 4	CD133+ replicate 3	Pool standard[b]

2.7 2D separation of CD133+ and CD133- protein samples

Protein samples, mixed as described in Table 1, were separated on 24-cm IPG strips with a 3-10 non linear pH range (GE Healthcare). Strips were rehydrated before use with 450µl of DeStreak rehydration solution, 0.5% pharmalyte and 0.5% IPG overnight at room temperature. Usually 50µg of protein lysates from CD133+ cells, 50µg from CD133- cells (Cy3 or Cy5 labeled) and 50µg of

pooled standard (Cy2 labeled) were mixed and the final volume was adjusted to 450µL with sample buffer containing 7M urea, 2M thiourea, 4% CHAPS, 1% DTT and 1% pharmalyte (GE Healthcare). The strips were then transferred to the Ettan IPGphor system (GE Healthcare) for IEF. Samples were loaded on the strips by anodic cup-loading. The IEF was carried out for 18h for a total of 60kV/h, 50µA/strip at 20°C, using the following protocol:

Step 1: 300 V for 4h

Step 2: from 300 V to 1000 V for 6 h

Step 3: from 1000 V to 8000 V for 3 h

Step 4: 8000 V for 5 h

Then, proteins were reduced by equilibrating IPG strips for 15min in 0.5% DTT (w/v), 100mM Tris pH 8.0, 6M urea, 30% glycerol (v/v), 2% SDS (w/v). Proteins were then alkylated for 15min using the same buffer containing 4.5% (w/v) IAA. After the equilibration steps, the strips were over layered onto 10% polyacrylamide gels (20 x 24cm) and the second dimension was carried out for 18h at 2W per gel using an Ettan Dalt Twelve system (GE Healthcare) until the bromophenol blue reached the bottom of the gel.

An independent two-dimensional preparative gel for each cell line was run with the same condition applied for the analytical gels, using for each gel, 0.5 mg of protein extract from Caco-2 and HCT-116 cells, respectively. Preparative gels were washed with a fixing solution of 40% methanol, 10% acetic acid, 50% water for 3 h before overnight staining in SYPRO Ruby (Molecular Probes, USA) with gently agitation, in the dark.

2.8 Analysis of gel images

The gels were scanned using a fluorescence scanner (Typhoon 9400, GE Healthcare) at 100 µm resolution. Fluorescence-labeled proteins were visualized at the appropriate excitation/emission wavelengths: 532/580 nm for Cy3, 633/670 nm for Cy5 and 488/520 nm for Cy2. Preparative gel images were acquired using the Typhoon imager at excitation/emission wavelengths of 457/610 nm. All gels were scanned by using the same parameters, selected to prevent pixel saturation.

The fluorescent images of the 2D-DIGE gels were analyzed using the DeCyder software suite, version 5.02 (GE Healthcare). For each gel, the overlapped images (Cy3, Cy5, Cy2) were imported into the DeCyder DIA (Difference In-gel Analysis) module to detect differentially expressed protein spots in each gel. In fact, protein spots were detected and quantified with the DIA module. The maximum number of estimated spots was fixed at 5000. In addition, DIA was used to detect spot boundaries and calculate spot volumes, normalized versus the volume of the corresponding spot present in the pool standard of the same gel. Protein spots that matched among the four gels were obtained using the BVA (Biological Variation Analysis) module. The Cy2 image containing the highest number of spots, was assigned as “master gel”. The spot boundary maps of the master image were used as template. Matching of protein spots across gels was performed automatically.

A standard abundance for each spot was thus calculated thereby allowing inter-gel variations. Each spot intensity was then expressed as mean of 4 standard abundances calculated for the four gels described in Table 1. Spot

intensities were then compared in the two used conditions: CD133+ and CD133- cells. Statistical significance of differences in spot intensity was determined by Student's t-test. Only protein spots with at least 1.20-fold change in volume ($p \leq 0.05$) after normalization were considered significantly altered. We verified the validity of these changes and accuracy of spot matching by manual investigation.

2.9 Protein identification by MS

Protein spots on preparative gels were chosen for excision on comparison with the analytical gel. Spots of interest were picked using an Ettan Spot Picker (GE Healthcare). Gel pieces were washed in 50mM ammonium bicarbonate and 50% ACN and subsequently rehydrated in a modified trypsin (Sigma) solution (10 ng/µl) in 50 mM ammonium bicarbonate pH 8.5, at 4°C for 1 h. The enzymatic solution was then removed. A new aliquot of buffer solution was added to the gel particles and incubated at 37°C overnight. The supernatant was collected whereas gel pieces were subjected to another extraction in ACN at 37°C for 15 min. The supernatant fraction and samples obtained from extraction steps were pooled and dried in a vacuum centrifuge. The resulting peptide mixtures were resuspended in TFA 0,1% and analyzed on a Voyager DE-STR MALDI-TOF (Applied Biosystems, Framingham, MA) in positive ion reflectron mode, 20kV and a pulsed nitrogen laser (337nm). Samples were co-mixed to an equal volume of 10mg/mL CHCA (70:30 ACN:0.2% TFA), spotted onto a stainless steel MALDI target plate and air-dried. The mass spectra (m/z 600-4000) appropriately calibrated were analyzed using the Data Explorer

software v4.0 (Applied Biosystems). Mass values were then used for database searching using the MASCOT (version 2.1) PMF search program (http://www.matrixscience.com) selecting NCBInr sept09 database (9656715 sequences) (http://www.ncbi.nlm.nih.gov) and *Homo sapiens* as taxonomic origin of the samples. The search was performed using trypsin as proteolytic enzyme, one missed cleavage, cysteine as S-carboxyamidomethylcysteine, unmodified N- and C-terminal ends, partial methionine oxidation, putative pyro-Glu formation by Gln and a peptide maximum mass tolerance of 100ppm.

Protein spots not identified during the PMF, were further analyzed by µLC-MS/MS using the LC/MSD Trap XCT Ultra (Agilent Technologies, Palo Alto, CA, USA) equipped with a 1100 HPLC system and a chip cube (Agilent Technologies). After loading, the peptide mixture (7µl in 0.2% HCOOH) was concentrated and washed at 4 µl/min in 40nl enrichment column (Agilent Technologies chip), with 0.1% HCOOH as the eluent. The sample was then fractionated on a C18 reverse phase capillary column (75µm x 43mm) at 200nl/min with a linear gradient of eluent B (2% HCOOH in ACN) in eluent A (2% HCOOH) from 5% to 60% in 50min. Peptides were analyzed using data-dependent acquisition of one MS scan (mass range 400-2000m/z) followed by MS/MS scans of the three most abundant ions. Dynamic exclusion was used to acquire a more complete range of the peptides by automatic recognition and temporary exclusion (2 min) of ions from which definitive mass spectral data had previously been acquired. Moreover a permanent exclusion list of the most frequent peptide contaminants (keratins and trypsin doubly and triply charged peptides) was used. For MS/MS data, the search was performed by using MASCOT with a peptide tolerance of 300 ppm and the following additional

criteria: maximum fragment mass tolerance of 0.6 Da, 2+ and 3+ charged peptides. Identifications were accepted taking in account three issues: significant MASCOT Mowse scores, spectrum annotation and expected migration on 2D gel. A MASCOT score of 64 corresponds to $p<0.05$ for PMF experiments, while a MASCOT score of 41 corresponds to $p<0.05$ for MS/MS sequencing. These thresholds were chosen as cut-off for a significant hit.

2.10 Western Blot Analysis

CD133+ and CD133- protein extracts (25µg) from three independent cultures were resolved on a 10% SDS-PAGE gel and then transferred onto nitrocellulose membranes (GE Healthcare). The membranes were blocked in 5% non-fat milk in PBS pH 7.5 for 2h and incubated over night at 4°C with 1% milk/PBS pH 7.5 and 0.05% TWEEN containing specific commercial primary antibodies: mouse anti-SRp20 (1:200), anti-casein Kinase II (CKII) (1:200), anti-cytokeratin 8 (1:200), anti-catenin γ (1:200), anti-catenin β (1:200), anti-annexinA2 (AnxA2) (1:200), anti-lamin A/C (1:200); rabbit anti-annexinA1 (AnxA1) (1:200) and anti-Hsp27 (1:100) (Santa Cruz Biotechnology, Heidelberg, Germany). A mouse anti-Gapdh (1:1000, Sigma-Aldrich) antibody was used as loading control. The horseradish peroxidase (HRP)-conjugated anti-mouse (1:5000) and anti-rabbit (1:10000) secondary antibodies (GE Healthcare) were employed. Immunoblots were detected using the ECL-Advance Western Blotting Detection kit (GE Healthcare) by chemiluminescence. Band volumes were normalized by using Gapdh as

control, visualized on the same film. Densitometric measurements were made using the Quantity One 4.5 tool (Biorad).

2.11 Bioinformatic analysis

Data were analyzed through the use of Ingenuity Pathway Analysis (IPA) software 7.0 (http://www.ingenuity.com). Drawing on published, peer-reviewed literature, IPA constructs networks of direct and indirect interactions between orthologous mammalian genes, proteins and endogenous chemicals. These relationships include those that occur due to disease and/or environmental input. This system can generate a set of networks with a maximum size of 35 genes/gene products. Each network is characterized by a score computed according to the fit of the user's set of focus genes/gene products with all the genes/gene products stored in the knowledge base. The score is derived from a p-value (equal to or smaller than 0.05, Fisher's exact test) and indicates the likelihood of the focus genes/gene products in a network being found together due to random chance. Biological functions were then assigned to each network. Our data set was also analyzed by the GOstat software (http://gostat.wehi.edu.au/cgi-bin/goStat.pl) to identify significant biological pathways (GO terms). Searches were performed as follows: minimal length of the considered GO paths was 1, the maximal p-value was 0.01, no GO clustering was applied and Benjamini false discovering testing was used to correct for multiple testing.

2.12 Wnt/β catenin pathway stimulation

CaCo-2 cell line was grown as described in paragraph 2.1. CaCo-2 cells were cultured with 100ng/ml recombinant human Wnt3a (R&D Systems, Minneapolis, MN). After 48h and 72h, cells were collected and protein extracts were obtained as described under section 2.4. The assay was performed on three independent replicates.

2.13 RNAi and cell proliferation assay

The SRp20 siRNA s12732, targeting a splice junction of SRp20 exon 2 and exon 3, was purchased from Ambion (Applied Biosystems). RNAi was conducted by transfection with 20 nM siRNA in the presence of Lipofectamine 2000. The cells with silenced SRp20 were analyzed for protein expression by Western blotting and for tumor induction capability by proliferation assay.

CaCo-2 siRNA transfected cells were detached from culture flask after 24h of incubation and seeded in flat bottom 96 well plate (BD-Falcon) at 5000 cells /well in DMEM:F12 (Sigma-Aldrich) w/o Phenol Red supplemented with 10% FBS, 2% Ultraglutammine. Cells were incubated at 37°C and 5% CO2 until 120h after siRNA transfection. Cellular proliferation was measured daily by the use of WST-1 assay and colorimetric reaction was quantitated at 450–655nm with the Ultramark Plate Reader (Biorad) reagent according to manufacturer's instructions. The assay was performed on three independent biological replicates and for each point, the measure was performed on three technical replicates.

3. RESULTS

3.1 Gating and sorting of CD133+ cells by flow-cytometry

CD133+ cells were gated in CD133 versus an “empty channel” dot plot for each cell line. As a preliminary step, we evaluated the intrinsic fluorescence of each cell line. Therefore, as control we used an unlabelled sample to evaluate the baseline fluorescence specific for the HCT-116, CaCo-2 and HT-29 cells (Figs. 16 A*, B* and C*). Finally, we selected CD133- cells, i.e., cells with the same level of fluorescence as unlabelled cells (control cells) and CD133+ cells, i.e., the cells with the highest level of fluorescence. In the HCT-116 cells, 80% are CD133+ cells and 20% are CD133- cells, which were spontaneously separated. In this case our strategy was to gate the brightest CD133+ cells (2% of total cells) and, as negative counterpart, we gated the whole population lacking CD133 (15-20%) (Fig. 16A). In the CaCo-2 and HT-29 cells, in which a large homogeneous population of cells uniformly expressed the antigen, we gated 2-5% of cells with the most intense staining. As negative counterpart we used 15-20% of the cells with the less intense CD133 staining (Figs. 16B and 16C). Finally, we isolated and collected the selected cells.

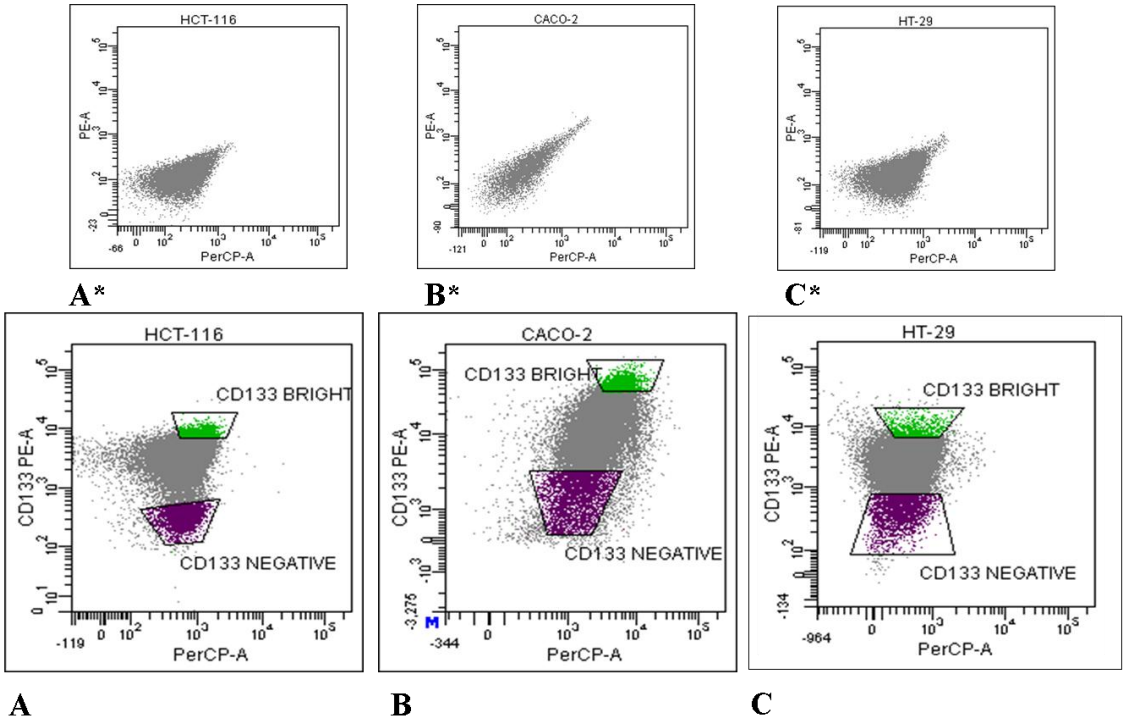


Fig. 16: Gating strategies for the selection of CD133+ and CD133- subsets in A) HCT-116 B) CaCo-2 and C) HT-29 cells. As a preliminary step, the auto-fluorescence of unlabelled A*) HCT-116 B*) CaCo-2 and C*) HT-29 cells was analyzed and used as negative control for sorting.

3.2 2-DIGE analysis: labeling efficiency and same-same-same tests

Prior to the DIGE experiment the labeling efficiency of the samples with the Cy5 dye was tested. To this aim, the labeling efficiency of CaCo-2 and HCT-116 protein extract were compared with that of a control which has already been labeled successfully. The control was constituted by *E. coli* total protein extract. As shown in Figure 17 proteins were resolved on a one-dimensional SDS-PAGE gel, as described in the “Materials and methods” section (paragraph 2.5). The gel was acquired at the Cy5 wavelength using the Typhoon imager and the relative images were processed and analyzed with the Image Quant Analysis software. The labeling efficiency was calculated as ratio between the average volumes of fluorescence of the sample with that of the control.

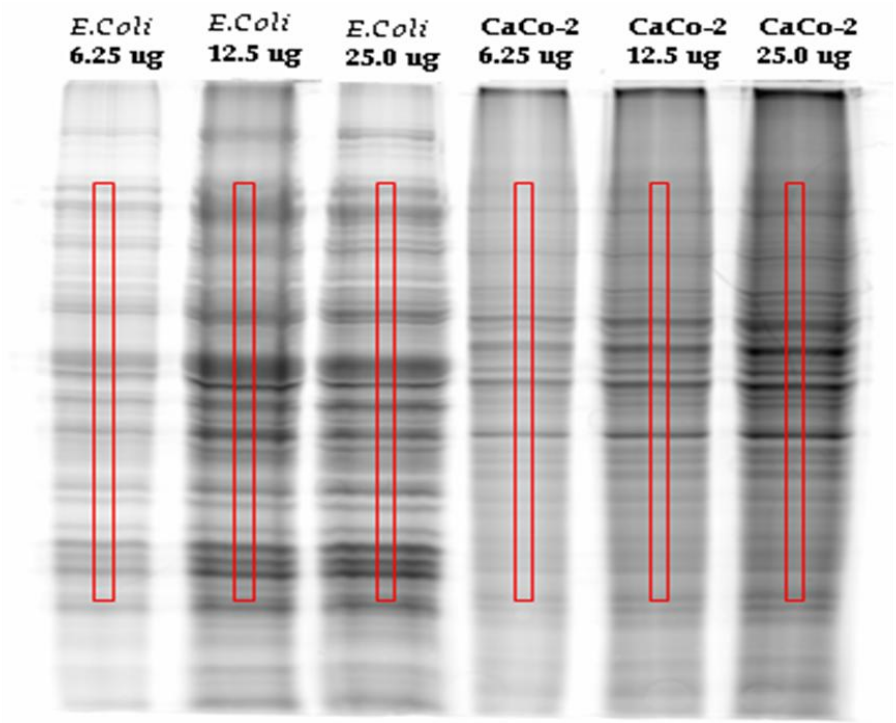


Figure 17: Cy5-gel images of CaCo-2 and *E.coli* protein lysates. Cy5-labeled proteins were loaded and resolved on a 12.5% SDS-PAGE gel using a serial dilution of 25 µg, 12.5 µg and 6.25 µg. The gel was acquired using the Typhoon 9400 imager.

In addition, to verify that none of the three CyDyes labels the same test sample preferentially, the “same same same” test was performed. By DeCyder software, the volumes of fluorescence for all spots of the sample labeled with Cy2 were compared with those labeled with Cy3 and Cy5, obtaining a similarity of labeling of 98% (Figure 18). The threshold was then set to 1.2 fold-change. All variations under this threshold were not considerate.

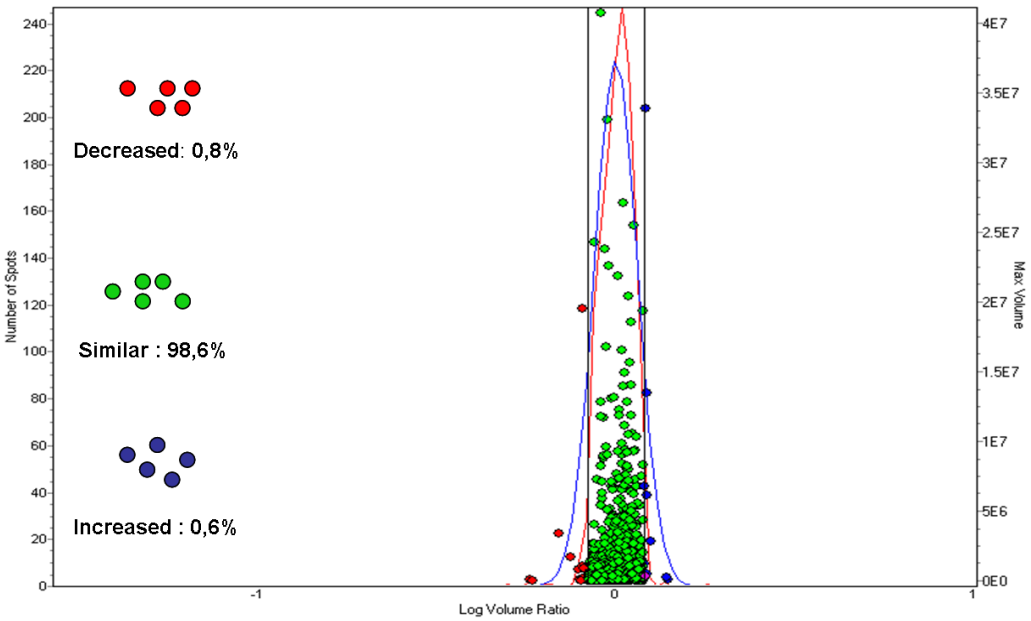


Figure 18: Statistical representation of the overall protein distribution in individual samples, in the same same same test. The red line indicates the number of spots with a specific abundance ratio (log scale). The blue line represents the calculated Gaussian distribution. Red dots represent decreased and blue increased protein content in individual spots detected in both samples (considered from Cy3 → Cy5). Green dots represent equal protein content between the samples. The threshold was set to 1.2 fold-change.

3.3 Identification of differentially expressed proteins by DIGE

To learn more about colon CSCs, we measured the differential protein expression pattern of CaCo-2 and HCT-116 CD133+ and CD133- cells, using 2-D DIGE. To increase biological and statistical significance of our results, we considered CD133+ and CD133- cells obtained from four independent cultures. About 2000 protein spots were detected in the analytical 2D gels for both cell lines. Fluorescence-labeled proteins in the 2D analytical gels were acquired at different wavelengths using an imager to generate an image specific for each CyDye. In this way three scans were made of each gel, Cy2, Cy3 and Cy5 scans. In the co-detection step scanned images of each sample were overlaid with the internal standard by the DIA module of the DeCyder software. By this way every protein in the sample was linked to the

corresponding protein spot in the internal standard sample. Following co-detection step, the matching of protein spots across the gels was performed using the BVA module of the DeCyder software. The spot map for the internal standard, with the most detected spots, was used as a template to which all remaining spot maps, for the other internal standards (intrinsically linked to their co-detected sample images), were matched. The spot volume, i.e. the sum of the pixel values within a spot minus background, for each experimental sample was directly compared, by the software, to the internal standard. By this way changes in the expression level of individual protein spots, expressed as ratio of protein abundance between CD133+ and CD133- cells, normalized to the internal standard, were identified. Each spot intensity was then expressed as mean of 4 standard abundances calculated for the four gels. Statistical significance of differences in spot intensity was determined by Student's t-test. Only protein spots with at least 1.2 fold-changes in volume ($p < 0.05$), after normalization, were considered significantly altered. DeCyder software displayed in a graphic the relative abundance of each protein spot in CD133+ and CD133- cells, and defined a fold change expression of each protein, above a biological variation. An example of the DeCyder analysis is reported in Figure 19.

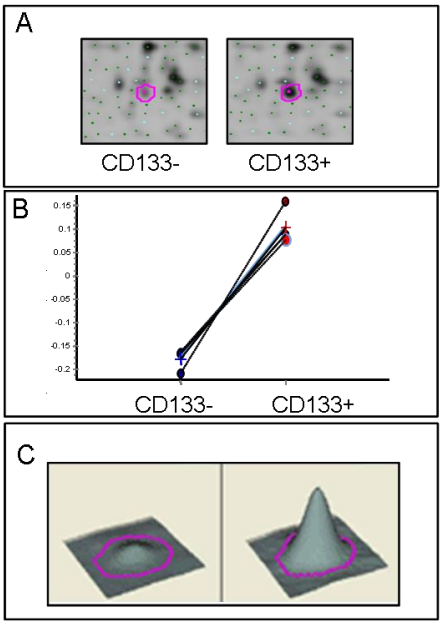


Figure 19: DeCyder software analysis. An example of a protein spot up-regulated in CD133+ cells. A: enlargement of 2D gel region of the spot. B: DeCyder graphic displays the relative abundance of the protein spot in the two sub-populations of cells. C: three-dimensional representation of the relative abundance the protein spot in CD133- and CD133+ cells.

The image analysis revealed that 61 spots in CaCo-2 CD133+ cells and 52 in HCT-116 CD133+ cells were differentially expressed compared to CD133- cells. From two Sypro Ruby-stained preparative gels we matched and picked 52 of 61 spots in CaCo-2 cells and 44 of 52 spots in HCT-116 cells. Excised spots were subjected to enzymatic digestion with trypsin and the resulting peptide mixtures were analyzed by mass spectrometry, MALDI MS or µLC-MS/MS, depending on spot intensity. Figures 20A and 20B show the differentially expressed spots identified on preparative gels: under-expressed spots and over-expressed spots in CD133+ *vs* CD133- cells are reported in red and green, respectively.

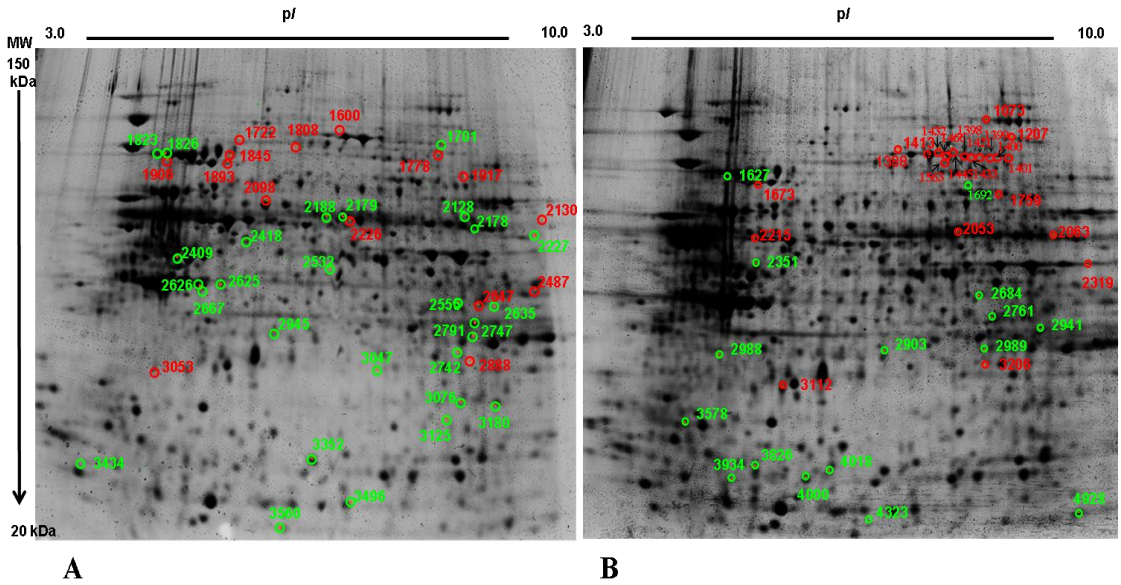


Fig. 20: Preparative gels. A) CaCo-2 cells B) HCT-116 cells. The red circles indicate spots down-regulated in CD133+ *vs* CD133- cells, the green circles indicate spots up-regulated in CD133+ *vs* CD133- cells. The corresponding identifications are listed in Table 2 and Table 3 for CaCo-2 and HCT 116 cells, respectively.

Forty-three of the picked spots were identified in CaCo-2 cells and 38 of the picked spots were identified in HCT-116 cells. The differentially expressed proteins identified on a representative DIGE gel of CaCo-2 cells and HCT-116 cells, respectively are shown in Figs 21A and 21B.

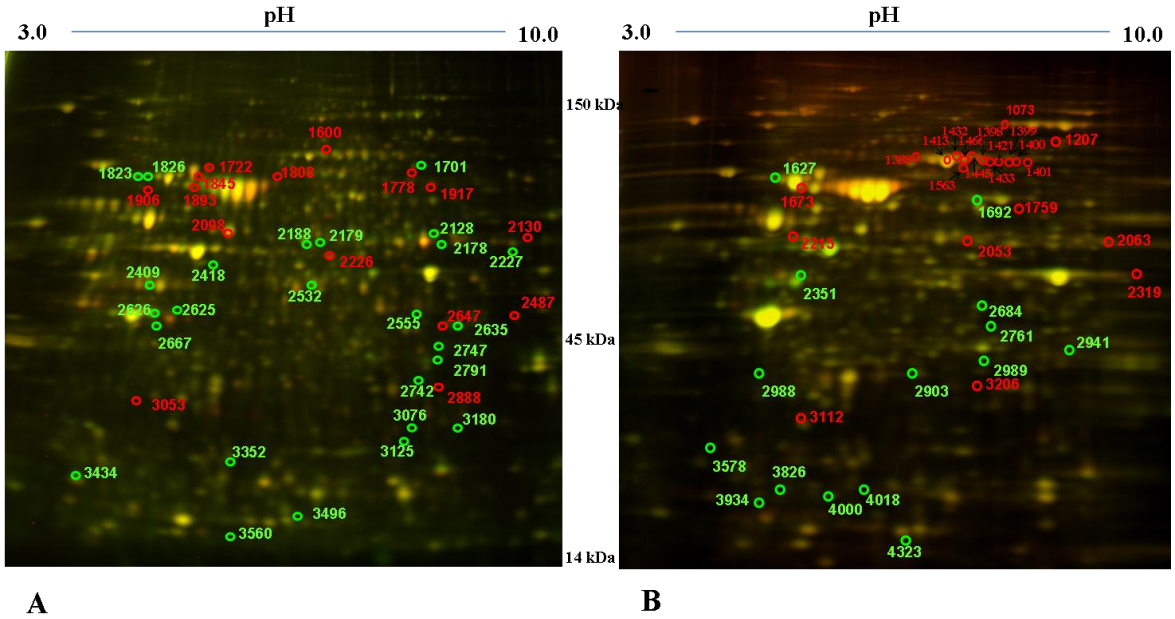
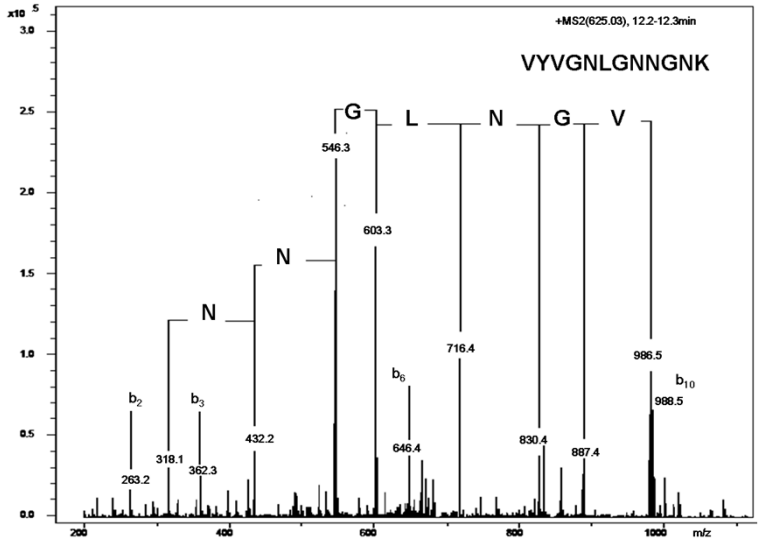


Fig. 21: DIGE experiment. A representative picture of DIGE analytical gels images for A) CaCo-2 CD133+ *vs* CD133- cells and B) HCT-116 CD133+ *vs* CD133- cells. For both panels, the overlay images of Cy3-Cy5 images are shown. Proteins that are down-regulated in CD133+ cells appear in red and those up-regulated appear in green, the unchanged or slightly changed proteins are in yellow. The spots of interest identified are indicated with their spot number.

Tables 2 and 3 show the proteins identified in CaCo-2 and HCT-116 cells. For each protein, the spot number, gene symbol, *p*-value, fold change, NCBI accession code, protein name and theoretical pI /MW are reported. The lists are ranked from the highest fold change value to the lowest. This range of fold changes (not greater than ± 2.45) seems to be congruent with the systems analyzed. In fact, we studied the protein expression profiles of two subpopulations of cells within two “homogeneous” colon cancer cell lines. Details of the MS analysis of the proteins identified in each cell line are reported in Tables 4 and 5 (Appendix, section 6). Most spots were unambiguously identified as single proteins: 32 in the CaCo-2 2D gel (21 over-expressed and 11 under-expressed) and 31 in the HCT-116 2D gel (11 over-expressed and 20 under-expressed). In total, 11 spots (2625, 2532, 2791, 2418, 2188, 3434, 1778, 1823, 1917, 1722, 1600) in CaCo-2 gel, and 7 spots (spots 4928, 2989, 4000, 2941, 3934, 2053, 3112) in the HCT-116 gel were associated to more than one protein. In these cases, the measured fold change could not be directly assigned to a single protein species. Thus, these proteins would require further validation analysis. The HCT-116 2D preparative gel, contained two series of spots with the same molecular weight and different p*I*. A series of 6 spots (no 1421, 1399, 1433, 1400, 1398 and 1401) of 75 kDa was identified as lamin A/C, and a series of 5 spots (no 1466, 1413, 1563, 1432 and 1445) of 70 kDa was identified as ezrin. These series of spots could be due to alternative post-translational modifications, namely phosphorylation or deamidation, or to alternative splicing protein products. Similarly, we attribute the increase of the p*I* value for Hspd1 (1388) to post translational

modifications, while Hsp90ab1 (2988) which appears in a spot at a lower molecular weight than the theoretical value, may have been extensively fragmented.

Among the differentially expressed proteins of CaCo-2 CD133+ cells, we detected SRp20 in spot 3560 at a p*I* of around 6.0, while the theoretical p*I* value of SRp20 is 11.64. This discrepancy may be due to phosphorylation; in fact, SRp20 is known to be extensively phosphorylated on the serine residues in the SR domain. Furthermore, the MS/MS spectrum of the peptide [12-23] of SRp20 at m/z=625.03 is reported (Fig. 22).



MHRDSCPLDCK**VYVGNLGNNGNK**TELERAFGYYGPLRSVWVARNPPGFAFVEFEDPR
DAADAVRELDGRTLCGCRVRVELSNGEKRSRNRGPPPSWGRRPRDDYRRRSPPPRR
RSPSFSRSRSRSLSRDRRRERSLSRERNHKPSRSFSRSRSRSRSNERK

Figure 22: MS/MS spectrum of peptide [16-23] of SRp20. The mass spectrum shows the reconstructed peptide sequence and the tryptic peptide used for the identification by Mascot. In the protein sequences, the identified peptide is in red

AnxA2 was detected in two contiguous spots (2747 and 2791), both of which were over-expressed at two slightly different molecular weights.

Creatine kinase B was detected in 3 over-expressed spots (2626, 2667 and 2625) at the expected molecular weight. Hspa8 was identified in two spots (1823 and 1826) both over-expressed at the expected molecular weight and p*I*. Hspa1a was detected in two spots (1906 and 1893) both down-regulated at two slightly different p*I* values, thus suggesting the possibility of post-translational modifications of this protein. Finally, Krt8 and Jup were detected in spots at molecular weights lower than the theoretical value which is suggestive of protein degradation. In conclusion, by considering only once the spots containing the same protein, the total number of differentially expressed proteins was 49 in the case of CaCo-2 CD133+ *vs* CD133- cells and 36 in the case of HCT-116 CD133+ *vs* CD133- cells, while the number of differentially expressed proteins identified as single proteins was 32 in CaCo-2 CD133+ cells and 22 in HCT-116 CD133+ cells.

There were 9 common differentially expressed proteins in CaCo-2 CD133+ cells (Table 2) and HCT-116 CD133+ cells (Table 3): 7 of them (Hspa1a, Krt8, Krt18, Hspd1, Hspa8, Hsp90ab1 and Impdh2) had a similar fold change in the two systems and 2 (Ugdh and Lmna) had a different trend, i.e., both are up-regulated in CaCo -2 CD133+ cells and down-regulated in HCT-116 CD133+cells. The fold-changes of Ugdh were very close to the threshold value (+1.23 in CaCo-2 and -1.22 in HCT-116), consequently, we did not investigate its role in CaCo-2 and HCT-116 CD133+ cells.

In the case of Lmna this is not surprising. In fact, Willis and colleagues reported that the expression of A-type lamins gives rise to increased tumor invasiveness, thus suggesting that reduced Lmna expression could be associated with a more favourable prognosis for CRC patients [67].

Conversely, Belt and collaborators subsequently showed that Lmna low expression is associated with an increased disease recurrence and, as a consequence, with a worse prognosis [68]. In our case we analyzed CD133+ and CD133- cells from two different colon cancer cell lines to look for common points of interest while being aware that we could also find some differences. Furthermore, in HCT-116, Lmna was identified in different spots, while in CaCo-2 Lmna was identified only in one spot. As described above, the presence of different p*I* isoforms of Lmna in HCT-116 gel could be due to post-translational modifications and to the existence of different alternative splicing products of the LMNA gene. Probably the CD133+ cells of CaCo-2 and HCT-116 express different isoforms of the same protein to fulfill different functions.

Table 2: Differentially expressed proteins in CaCo-2 CD133+ *vs* Cd133-cells

Spot No.	Gene name	*p*-value	Fold change	Gene ID	Protein name	Theoretical pI	Theoretical MW
3560	SFRS3	0.018	1.93	gi|4506901	Splicing factor, arginine/serine-rich 3 (SRp20)	11.64	19329
2635	CSNK2A1	0.04	1.92	gi|33358120	Casein kinase II subunit alpha (CKII)	7.29	45143
3076	GNPDA1	0.024	1.77	gi|13027378	Glucosamine-6-phosphate deaminase 1	6.42	32668
2747	ANXA2	0.043	1.76	gi|16306978	Annexin A2	7.56	38472
2626	CKB	0.0044	1.76	gi:49457530	Creatine kinase B-type	5.35	42513
2625	CKB	0.0013	1.73	gi:49457530	Creatine kinase B-type	5.35	42513
2625	MRPS27	0.0013	1.73	gi|186928850	28S ribosomal protein S27, mitochondrial	5.83	47611
2555	METAP1	0.024	1.64	gi|577315	Methionine aminopeptidase 1	6.8	43083
3125	LACTB2	0.0031	1.61	gi|7705793	Lactamase, beta 2	6.32	32805
2667	CKB	0.0067	1.59	gi:49457530	Creatine kinase B-type	5.35	42513
1701	LMNA	0.021	1.55	gi|57014047	Lamin A/C	6.44	72224
2532	TUFM	0.033	1.53	gi|704416	Elongation factor Tu (p43)	6.31	45045
2532	PDHA1	0.011	1.53	gi:12803199	Pyruvate dehydrogenase E1-alpha precursor	6.51	40228
3047	ESD	0.011	1.53	gi:182265	Esterase D	6.54	31462
2945	LDHB	0.023	1.52	gi|4557032	L-lactate dehydrogenase	5.72	36507
2791	ANXA2	0.045	1.46	gi|16306978	Annexin A2	7.56	38472
2791	GPD2	0.045	1.46	gi:18043793	Glycerol-3-phosphate dehydrogenase	6.51	40228
2418	ACTR3	0.039	1.4	gi|62088286	ARP3 actin-related protein 3 homolog variant	5.61	47239
2418	DDX39B	0.039	1.4	gi|4758112	Spliceosome RNA helicase BAT1	5.44	49416
2128	CAT	0.0055	1.39	gi|4557014	Catalase	6.95	59624
3180	PROSC	0.009	1.37	gi:6005842	Proline synthetase co-transcribed homolog	7.09	30343
3496	PCNP	0.043	1.36	gi|9966827	PEST proteolytic signal-containing nuclear protein	6.86	18924
2188	TUBB2C	0.041	1.34	gi|5174735	Tubulin beta-2C chain	4.79	49831
2188	TUBA1A	0.041	1.34	gi|37492	Tubulin alpha	5.02	50810
3434	TPT1	0.015	1.31	gi|4507669	Fortilin (p23)	4.84	19595
3434	JUP	0.015	1.31	gi|194373749	Catenin gamma	5.7	81744
1823	HSPA8	0.0071	1.28	gi|5729877	Heat shock cognate 71 kDa protein isoform 1	5.37	70766
1823	HSP90AB1	0.0071	1.28	gi|306891	Heat shock protein 90kDa alpha (cytosolic)	4.96	83133
3352	ECHS1	0.011	1.28	gi|1922287	Enoyl-CoA hydratase	5.88	28342
1826	HSPA8	0.013	1.25	gi|5729877	Heat shock cognate 71 kDa protein isoform 1	5.37	70766
2179	ALDH1A1	0.05	1.25	gi|114625020	Aldehyde dehydrogenase 1A1	6.29	54730
2178	UGDH	0.05	1.23	gi|4507813	UDP-glucose-dehydrogenase	6.73	55024

2409	KRT18	0.017	1.23	gi|30311	Cytokeratin 18	5.34	47926
2742	POLDIP2	0.043	1.23	gi|7661672	DNA polymerase delta interacting protein 2 (p38)	8.8	42033
2227	ATP5A1	0.018	1.22	gi|158259937	ATP synthase, H+ transporting, mitochondrial	8.28	55209
1808	TCP1	0.0053	-1.2	gi|36796	t-complex polypeptide 1	5.8	60343
2130	FARSA	0.0072	-1.21	gi|119604734	Phenylalanyl-tRNA synthetase	7.46	57432
3053	KRT8	0.036	-1.23	gi|181573	Cytokeratin 8	5.52	53573
2487	ALDOA	0.05	-1.25	gi|28614	Aldolase A	8.39	39288
1917	SYNCRIP	0.024	-1.27	gi|228008398	Heterogeneous nuclear ribonucleoprotein Q (hnRNP Q) isoform3	7.18	62656
1917	PCK2	0.024	-1.27	gi|2661752	Phosphoenolpyruvate carboxykinase (GTP)	8.4	47563
1778	DDX3X	0.013	-1.27	gi|2580550	Dead box, X isoform	6.73	73625
1778	NCL	0.013	-1.27	gi|189306	Nucleolin	4.59	76355
2647	PSMC6	0.015	-1.33	gi|195539395	Proteasome 26S ATPase subunit 6 (Proteasome subunit p42)	7.25	44041
1722	HSPD1	0.0016	-1.35	gi|306890	Chaperonin (HSP60)	5.24	57962
1722	PDIA4	0.0016	-1.35	gi|4758304	Protein disulfide-isomerase A4 precursor	4.9	70102
2226	IMPDH2	0.033	-1.35	gi|66933016	Inosine monophosphate dehydrogenase 2	6.44	55804
2098	TAPBP	0.041	-1.38	gi|220702506	Tapasin	6.53	45714
2888	HS3ST3A1	0.05	-1.4	gi|52695687	Heparan sulfate glucosamine 3-O-sulfotransferase 3A1	9.54	44899
1906	HSPA1A	0.015	-1.63	gi|6729803	Heat shock 70 kDa protein 1A/1B	5.48	69921
1600	FLNA	0.0053	-1.86	gi|18676444	Filamin A	5.7	280607
1600	TARS	0.0053	-1.86	gi:52632425	Threonyl-tRNA synthetase	6.2	83435
1893	HSPA1A	0.0087	-2.19	gi|194388088	Heat shock 70 kDa protein 1A/1B	5.48	69921
1845	ANXA6	0.0074	-2.32	gi|119582091	Annexin A6	5.46	75276

Table 3: Differentially expressed proteins in HCT-116 CD133+ *vs* Cd133- cells

Spot No.	Gene name	*p*-value	Fold Change	Gene ID	Protein name	Theoretical pI	Theoretical MW
4928	PRDX1	0.035	2.45	gi|4505591	Peroxiredoxin-1	8.27	22110
4928	PEBP1	0.035	2.45	gi|913159	Neuropolypeptide h3	7.42	21027
2988	HSP90AB1	0.014	2.13	gi|306891	Heat shock protein 90kDa alpha (cytosolic)	4.96	83133
3826	HSPB1	0.0073	1.92	gi|662841	Heat shock protein 27 (Hsp27)	5.98	22782
4323	CDC42	0.001	1.83	gi|4757952	Cell division control protein 42 homolog isoform 1	6.16	20933
4018	PRTFDC1	0.015	1.68	gi|9910262	Phosphoribosyltransferase domain-containing protein 1	5.77	25542
2989	ANXA1	0.013	1.44	gi|4502101	Annexin A1	6.64	38583
2989	HNRNPH3	0.013	1.44	gi 14141157	Heterogeneous nuclear ribonucleoprotein H3 isoform a	6.8	36925
4000	PRDX4	0.034	1.41	gi|5453549	Peroxiredoxin-4	5.54	26572
4000	PRDX6	0.034	1.41	gi|3318841	Peroxiredoxin 6	6.02	24903
2941	CPOX	0.02	1.37	gi|840693	Coproporphyrinogen oxidase	6.32	39248
2941	HNRNPC	0.02	1.37	gi|306875	Heterogeneous nuclear ribonucleoprotein C (C1/C2)	5.01	33538
3934	ACTB	0.025	1.33	gi|15277503	Actin beta	5.29	41736
3934	ENO	0.025	1.33	gi|2661039	Enolase	6.5	36628
2903	TALDO1	0.0015	1.3	gi|5803187	Transaldolase	6.36	37540
3578	CLIC1	0.031	1.28	gi|895845	Chloride intracellular channel protein 1	5.09	26791
1627	HSPA8	0.01	1.26	gi|5729877	Heat shock cognate 71 kDa protein isoform 1	5.37	70766
2684	SEPT2 (DIFF6)	0.048	1.23	gi|4758158	Septin 2	6.15	41487
1692	SDHA	0.04	1.22	gi|119571367	Succinate dehydrogenase complex, subunit A, flavoprotein (Fp)	6.25	68012
2351	KRT18	0.018	1.22	gi|30311	Keratin 18	5.34	47926
2761	ALDR1	0.016	1.21	gi|5174391	Alcohol dehydrogenase [NADP+]	6.35	36441
1759	STIP1	0.031	-1.21	gi|114638255	Hsp70/hsp90 organizing protein(HOP)	6.4	62639
2053	PHGDH	0.0067	-1.21	gi|23308577	Phosphoglycerate dehydrogenase	6.31	56519
2053	IMPDH2	0.0067	-1.21	gi:66933016	Inosine monophosphate dehydrogenase 2	6.44	55804
2319	FH	0.0021	-1.21	gi|19743875	Fumarate hydratase precursor	7.0	50081
1388	HSPD1	0.024	-1.22	gi|306890	Chaperonin (HSP60)	5.24	57962
2063	UGDH	0.043	-1.22	gi|4507813	UDP-glucose dehydrogenase	6.73	55024
1421	LMNA - LMN1	0.013	-1.24	gi|57014047	Lamin A/C	6.44	72224
2215	KRT8	0.049	-1.33	gi|119617057	Keratin8	5.52	53573
1466	EZR	0.036	-1.36	gi|46249758	Ezrin	5.95	69281
1207	KHSRP	0.0037	-1.37	gi|2055427	KSRP /KH type-splicing regulatory protein (p75)	6.85	73115
1401	LMNA-LMN1	0.0099	-1.4	gi|27436946	Lamin-A/C	6.44	72224
3206	MDH1	0.02	-1.4	gi|5174539	Cytosolic malate dehydrogenase	6.89	36294
1413	EZR	0.0036	-1.43	gi|46249758	Ezrin	5.95	69281
3112	ANXA3	0.017	-1.44	gi|4826643	Annexin A3	5.63	36244
3112	PPA1	0.017	-1.44	gi|11056044	Inorganic pyrophosphatase	5.54	32660

1563	EZR	0.027	-1.45		Ezrin	5.95	69281
1432	EZR	0.012	-1.47	gi|46249758	Ezrin	5.95	69281
1399	LMNA - LMN1	0.026	-1.5	gi|57014047	Lamin A/C	6.44	72224
1073	EEF2	0.042	-1.51	gi|4503483	Elongation factor 2	6.42	95206
1673	HSPA1A	0.0096	-1.53	gi|167466173	Heat shock protein 70 (Hsp70)	5.48	69921
1433	LMNA - LMN1	0.033	-1.62	gi|57014047	Lamin A/C	6.44	72224
1400	LMNA - LMN1	0.0098	-1.68	gi|57014047	Lamin A/C	6.44	72224
1445	EZR	0.018	-1.78	gi|46249758	Ezrin	5.95	69281
1398	LMNA - LMN1	0.045	-1.86	gi|57014047	Lamin A/C	6.44	72224

3.4 Validation of differentially expressed proteins

We searched the literature to try to identify colon CSC markers among our differentially expressed proteins and to identify cellular processes deregulated in CD133+ cells. We included in the group of proteins selected for validation by western blotting, a few species closely related to our protein set. Western blot analyses were performed on freshly prepared protein extracts of CaCo-2 and HCT-116 CD133+ and CD133- cells. GAPDH was used as loading control of each sample. Semi-quantitative analyses of protein expression were performed for each validated protein. We analyzed the expression of the following proteins in the CaCo-2 CD133+ *vs* CD133- cells: SRp20, γ-catenin, β-catenin, CKII, lamin A/C, AnxA1 and AnxA2 (Fig. 23).

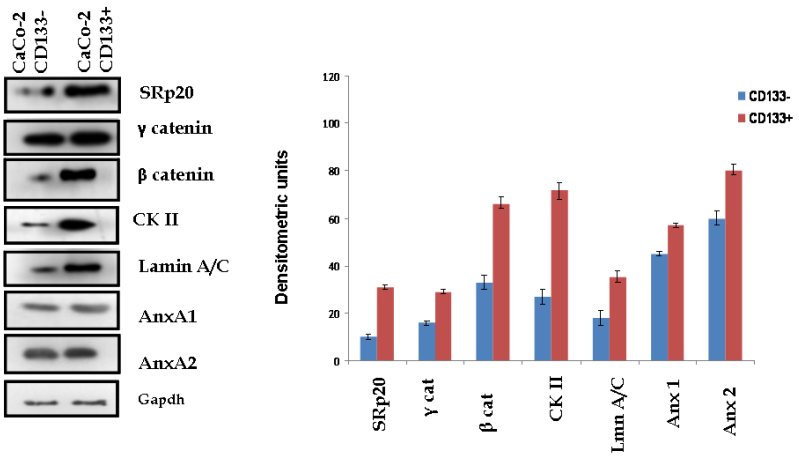


Fig.23: WB analysis and densitometric measurements of protein lysates of CaCo-2 CD133+ and CD133- cells

The western blot experiments confirmed the up-regulation of SRp20, catenin γ, CKII, lamin A/C and AnxA2 in CD133+ cells. In addition, β-catenin and AnxA1 were differentially expressed in the system, all of them being up-regulated.

Hsp27, cytokeratin 8, lamin A/C, AnxA1, AnxA2, SRp20, CKII, β-catenin and γ-catenin were chosen for western blotting and densitometric analysis of HCT-116 CD133+ *vs* CD133- cells (Fig. 24). The analysis confirmed the up-regulation of Hsp27 and AnxA1 and the down-regulation of KRT8 and lamin A/C in CD133+ cells. Interestingly, as demonstrated in the CaCo-2 system, SRp20 and β-catenin were up-regulated in the HCT-116 CD133+ cells, whereas no relevant differences were detected as regards CKII, γ-catenin and AnxA2.

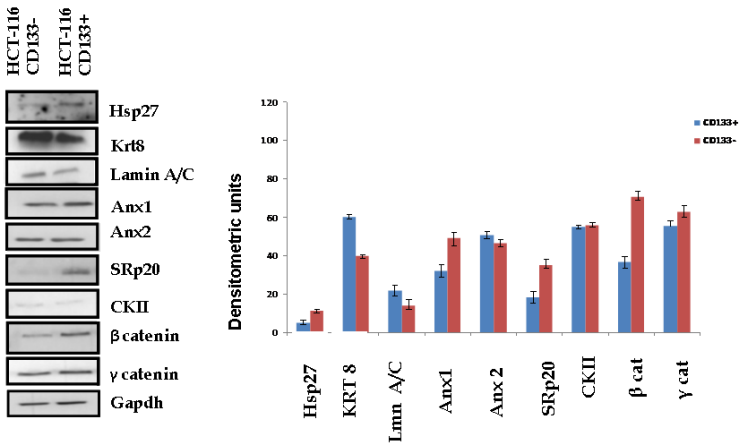


Fig. 24: WB analysis and densitometric measurements of protein lysates of HCT-116 CD133+ and CD133- cells

Finally, to determine whether SRp20, CKII, γ-catenin and β-catenin belong to distinctive molecular pathways of cancer stem-like cells (CSLCs), the level of expression of these proteins was investigated in CD133+ *vs* CD133- cells isolated from another colorectal carcinoma cell line, namely HT-29 (Fig. 25).

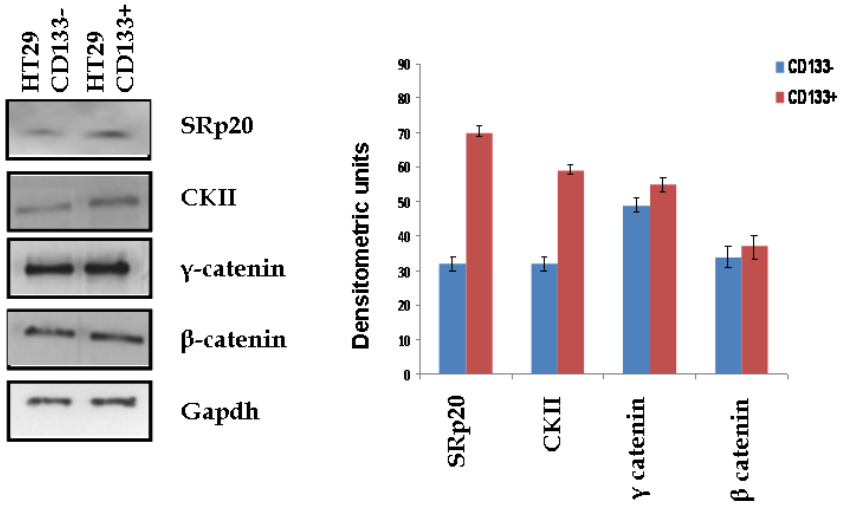


Fig. 25: WB analysis and densitometric measurements of protein lysates of HT-29 CD133+ and CD133- cells

The analysis confirmed the up-regulation of SRp20 and CKII in CD133+ cells, whereas no significant change was observed for γ-catenin and β-catenin.

In conclusion, SRp20 was over-expressed in CD133+ *vs* CD133-cells isolated from all three CRC cell lines: CaCo-2, HCT-116 and HT-29. This overexpression amounts to about two-fold in all the three CRC cell lines.

3.4 Biological network analysis

To identify processes connected with CD133+ cells, we analyzed separately the groups of differentially expressed proteins listed in Table 2 and Table 3 for each cell line using the IPA software. The system created three high-score multidirectional interaction networks for CaCo-2 cells. One was related to "cellular movement" (score=51), one to "energy production, small molecule

biochemistry" (score=33) and one to "cell signaling" (score= 27). To have a complete graphical vision of all the proteins identified in our analysis, we merged the three CaCo-2 networks into a single network (Fig. 26). The network revealed several nodes of proteins, genes and/or molecules (NFkB, Caspase, MCC) directly or indirectly correlated with the proteins we identified. In the case of HCT-116, the IPA output generated three networks: "cellular movement" (score= 44), "energy production, small molecule biochemistry" (score= 24) and "cellular growth and proliferation" (score=9). Figure 27 shows the merged image of all three HCT-116 networks.

We also performed an ontological analysis on our data set and, in agreement with IPA output, the "cellular catabolic process, energy production" GO term was statistically significant for both cell lines (GO: 0044265, CaCo-2 p-value 3.0E-05; HCT-116, GO:0009109 p-value 7.5E-04).

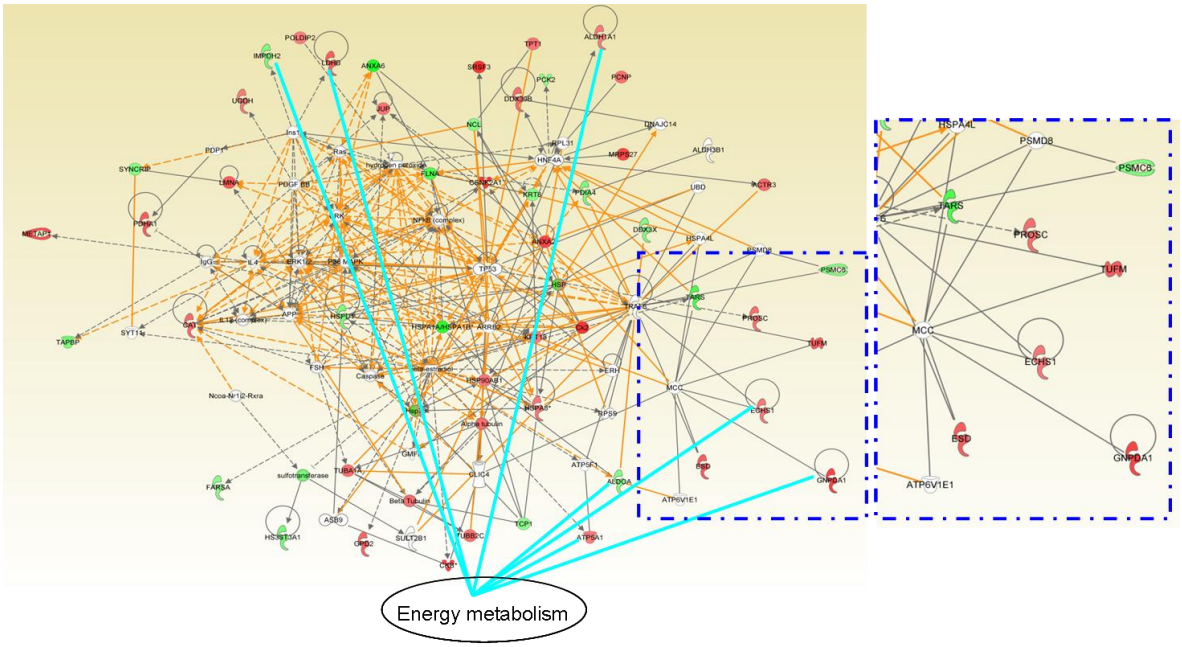


Fig. 26: IPA output for differentially expressed proteins in CaCo-2 CD133+ cells. Merged network. Green indicates a decrease, red indicates an increase in protein expression and white indicates proteins not identified in our study, which allow in connecting all the proteins together in a network. Blue lines indicate single identified proteins that are associated with energy metabolism. Network enlargement showed that MCC is directly correlated with the proteins identified in our analysis

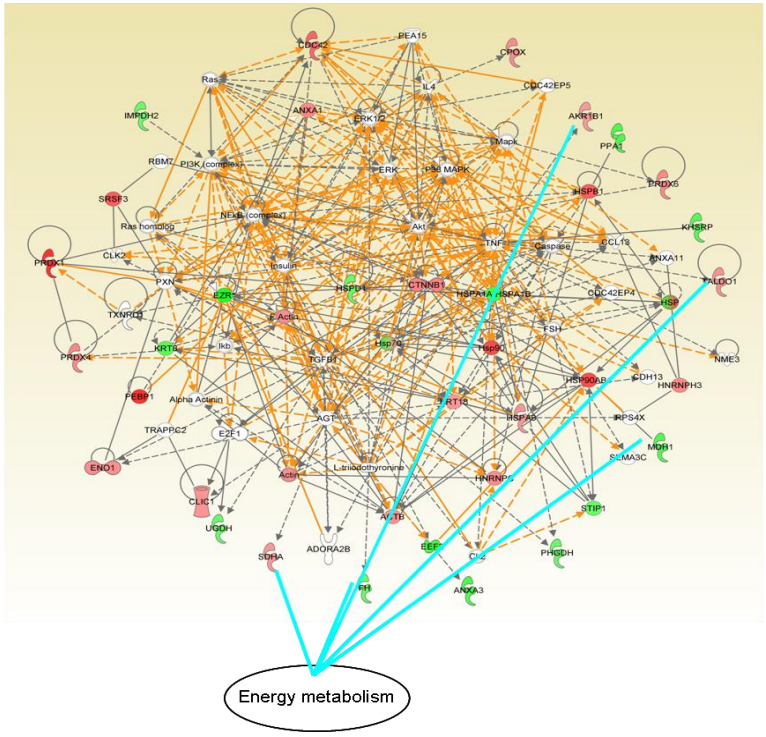


Fig. 27: IPA output for differentially expressed proteins in HCT-116 CD133+ cells. Merged network. Green represents a decrease, red indicates an increase in protein expression and white represents proteins not identified in our study, which allow in connecting all the proteins together in a network. Blue lines underline single identified proteins connected to energy metabolism.

3.5 Effects of Wnt/β catenin pathway activation on SRp20 expression

There is evidence that SRp20 expression can be regulated by the Wnt/β-catenin pathway [69]. To evaluate whether this occurs in our system, we stimulated pathway activation by adding Wnt ligand to the CaCo-2 culture medium and we estimated SRp20, β-catenin and CKII expression by western blotting (Fig. 28). Cells were withdrawn at 48 h and 72 h. Total β-catenin was increased in the presence of Wnt after 48 h and 72 h, which confirms activation of the Wnt pathway. The expression profile of SRp20 and CKII was the same as that of β-catenin, i.e., it increased 48 h and 72 h after treatment with Wnt ligand.

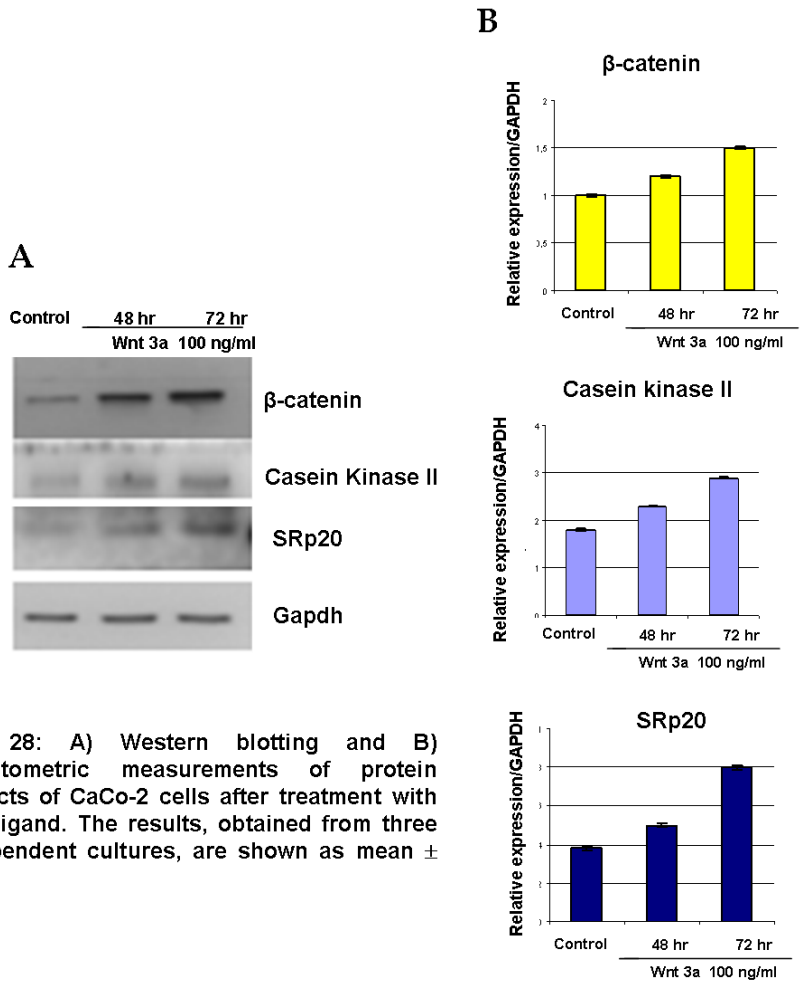


Fig. 28: A) Western blotting and B) densitometric measurements of protein extracts of CaCo-2 cells after treatment with Wnt ligand. The results, obtained from three independent cultures, are shown as mean ± SD.

3.6 Effects of silencing of SRp20

We next investigated the role of SRp20 on cell proliferation and found that CaCo-2 cells proliferation decreased after SRp20 expression was knocked down by a siRNA targeting SRp20 mRNA (Fig. 29 A). This result indicates that increased SRp20 expression is necessary for the indefinite growth of cancer cells. We also found that SRp20 was up-regulated in CD133+ cells, thus indicating that SRp20 could play a crucial role in the higher tumorigenicity of this subpopulation of cells.

We also evaluated if SRp20 silencing affected the expression of MCC, β-catenin and γ-catenin. Interestingly, immunoblotting showed that SRp20 silencing increased MCC expression and decreased the expression of β-catenin and γ-catenin (Fig. 29 B).

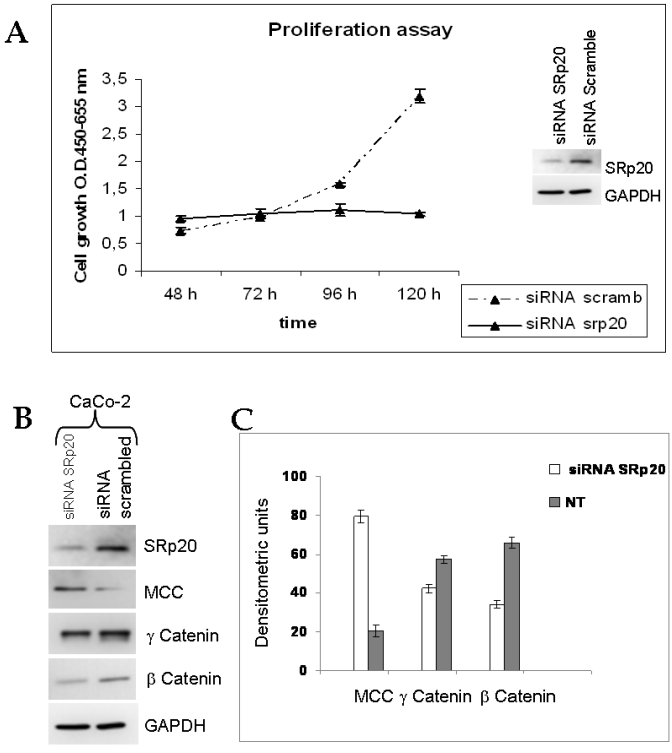


Fig. 29: Effects of SRp20 silencing A) WST-based proliferation assay. SRp20 was silenced in CaCo-2 cells by Ambion siRNA (20nM), while CaCo-2 cells treated with siRNA scramble were used as control. Cellular proliferation was measured after 48 h, 72 h, 96 h and 120 h siRNA transfection. The results of three independent experiments (mean ± SD) are shown. B) Western blot and densitometric measurements showing up-regulation of MCC and down-regulation of β-catenin and γ-catenin after 48 h of transfection of SRp20-siRNA compared to the scrambled control, in CaCo-2 cells. Results are shown as mean ± SD of three independent experiments.

4. DISCUSSION

A tumor is constituted by heterogeneous cells. The long-held view of tumor development, according to the stochastic model, is that each tumor cell is equally capable of initiating neoplastic growth. This theory has recently been challenged by the CSC theory, which suggests that only a small proportion of cells within a tumor possess cancer-initiating potential and that these so-called CSCs sustain tumor growth and may be associated with metastasis, treatment resistance and recurrence [70]. During this PhD program, a differential proteomic analysis on two human colon carcinoma cell lines, CaCo-2 and HCT-116, was conducted. Both cell lines were sorted according to the expression of the stemness marker CD133 in the attempt to identify proteins and/or cellular processes distinctive of colon CSCs. CD133+ cells were assumed to be CSLCs. By counting only once the protein spots that contained the same protein, the comparative experimental approach revealed 49 differentially expressed proteins (31 up-regulated and 18 down-regulated) in CaCo-2 CSLCs and 36 differentially expressed proteins (21 up-regulated and 15 down-regulated) in HCT-116 CSLCs. Some of the identified species were further validated in a third colon carcinoma cell line, namely HT-29. The results of this PhD thesis show that two relevant interconnected processes are activated in the CSLCs subpopulation: energy metabolism and the Wnt pathway.

Both the ontological and the IPA analyses showed that a group of proteins identified are involved in cellular energy metabolism. Among

the proteins differentially expressed in CaCo-2 CD133+ cells (Table 2), glucosamine-6-phosphate isomerase 1 (Gnpda1), lactate dehydrogenase (Ldhb), enoyl-CoA hydratase (Echs1), the well-established CSC marker aldehyde dehydrogenase 1A1 (Aldh1a1) [71-73]. ATP synthase subunit alpha (Atp5a1), aldolase A (Aldoa) and inosine monophosphate dehydrogenase 2 (Impdh2) are involved in energy metabolism. Collectively, these proteins account for 22% (7/32) of the singly identified proteins. A similar result was obtained for proteins differentially expressed in HCT-116 CD133+ cells (Table 3). In fact the species involved in energy metabolism represent 23% (5/22) of the singly identified proteins; these are alcohol dehydrogenase [NADP+] (Akr1a1), the tumor suppressor fumarate hydratase (Fh), transaldolase I (Taldo1), succinate dehydrogenase (Sdha) and cytosolic malate dehydrogenase (Mdh1). Metabolic reprogramming, called the “Warburg effect” [74], is a hallmark of cancer cells that undergo “aerobic glycolysis” thereby giving rise to enhanced lactate production [75]. The expression of most of the identified proteins suggests that metabolic reprogramming is activated in CaCo-2 and HCT-116 CSLCs. Furthermore, examination of the CaCo-2 IPA network (Fig. 26) shows that several differentially expressed proteins are connected with protein MCC (mutated in colorectal cancer), a negative regulator of Wnt/β-catenin signal transduction [76]; in this study there is evidence of Wnt/β-catenin signaling activation.

The Wnt signaling cascade is conserved throughout the animal kingdom and, depending on the context, it plays various roles in stem cell

maintenance, cell proliferation, differentiation, and apoptosis [77]. Given its fundamental role in homeostasis in human adult tissue, it is not surprising that deregulation of the Wnt pathway is associated with various pathologic states, including various types of cancer [78, 79]. The differential proteomic study performed shows that some proteins identified are closely related to the signal transduction pathway of the Wnt protoncogene. SRp20 and CKII were up-regulated in CaCo-2 CD133+ cells, whereas KSRP/KH type-splicing regulatory protein (Khsrp) was down-regulated in HCT-116 CD133+ cells. SRp20 is a serine/arginine-rich splicing factor recently characterized as a novel target of β-catenin/TCF4 signalling through the Wnt canonical pathway [69]. CKII, when over-expressed, induces neoplastic growth, thereby acting as an oncogene [80]. CKII was recently implicated in the regulation of β-catenin stability [81]. In fact, it positively regulates β-catenin phosphorylation at the level of Thr393, and thus inhibits proteasome-mediated degradation of β-catenin [81]. Its kinase activity promotes survival by increasing residual gene expression via β-catenin-TCF/LEF-mediated transcription [82]. Khsrp is a KH-type splicing regulatory protein able to negatively regulate Wnt/ β-catenin signaling at the level of post-transcriptional β-catenin-mRNA stability [83]. Its down-regulation indicates the activation of specific molecular mechanisms aimed at stabilization of β-catenin mRNA. β-catenin was up-regulated in CaCo-2 and HCT-116 CD133+ cells (Figs. 23 and 24). These findings suggested that the Wnt pathway was activated in the CSLCs. The western blot analyses performed in the Caco-2, HCT-116 and HT-29 cell

lines sorted by CD133 confirmed this hypothesis and revealed the up-regulation of the splicing factor SRp20 in all the three CSLC systems under investigation (Figs. 23, 24 and 25). Moreover, by stimulating CaCo-2 cells with Wnt ligand an activation of the Wnt pathway was reproduced and, as a consequence, an up-regulation of CKII and SRp20 was found (Fig. 28).

SRp20 is the smallest member of the SR protein family [84]. It modulates alternative splicing of CD44 cell adhesion protein in two colorectal cell lines (SW480, DLD-1) [69]. Its knockdown causes apoptosis in ovarian cancer cells whereas its expression is associated with malignancy of epithelial ovarian cancer [85]. SRp20 is over-expressed in many cancer types and its increase is critical for cell proliferation, tumor induction and maintenance [86].

When SRp20 was silenced in the CaCo-2 cell line, cell proliferation slowed down (Fig. 29A), which suggests that SRp20 plays a role in the tumorigenicity of CSLCs. In addition, the results of this thesis show that SRp20 exerts a powerful effect on MCC protein expression (Fig. 29B). Fukuyama and colleagues reported that MCC expression is dramatically decreased in many CRC cell lines; importantly, they found that MCC interacts with β-catenin and finally that re-expression of MCC in CRC cells inhibits Wnt signaling [76]. The data of this thesis demonstrate that when SRp20 is silenced, MCC expression is increased, while β-catenin and γ-catenin expression is decreased (Fig. 29B), thereby suggesting a slowing-down of the Wnt pathway. These data are in accordance with the hypothesis that SRp20 expression is closely correlated with Wnt

pathway activation and also suggest that SRp20 could act as a regulator of the pathway by modulating MCC expression.

In this study, a positive correlation among CD133 expression, Wnt pathway activation and increased SRp20 expression was found. Based on these findings it is possible to propose a model of putative sequential molecular events that characterize colon CSLCs in which the β-catenin/TCF4 pathway would stimulate gene transcription and thus the production of transcript variants through alternative splicing mediated by SRp20.

In summary, the results of this PhD thesis showed i) activation of metabolic reprogramming in CSLCs that is potentially connected to Wnt pathway activation ii) over-expression of SRp20 in the CSLCs of three different colon cancer cell lines; iii) a direct correlation between Wnt pathway activation and SRp20 expression; iv) the possibility that SRp20 plays a role in the tumorigenicity of colon CSLCs and v) the possibility that SRp20 modulates the Wnt pathway, where also the expression of MCC is involved.

5. REFERENCES

[1] Lapidot, T. et al. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* 1994, **367**, 645–648

[2] Pan CX, Zhu W, Cheng L. Implications of cancer stem cells in the treatment of cancer. *Future Oncol* 2006, **2(6)**:723-31.

[3] Al-Hajj M, Clarke MF. Self-renewal and solid tumor stem cells. *Oncogene* 2004, **23(43)**:7274-82.

[4] Woodward WA, Chen MS, Behbod F, Rosen JM. On mammary stem cells. *J Cell Sci* 2005, **118**(Pt 16):3585-94.

[5] Dalerba P, Dylla SJ, Park IK, Liu R, Wang X, Cho RW, Hoey T, Gurney A, Huang EH, Simeone DM, Shelton AA, Parmiani G, Castelli C, Clarke MF Phenotypic characterization of human colorectal cancer stem cells. *Proc Natl Acad Sci USA*. 2007; **104**:10158–10163.

[6] O'Brien CA, Pollett A, Gallinger S, Dick JE. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature.* 2007; **445**:106–110.

[7] Ricci-Vitiani L, Lombardi DG, Pilozzi E, Biffoni M, Todaro M, Peschle C, De Maria R. Identification and expansion of human colon-cancer-initiating cells. *Nature.* 2007; **445:**111–115

[8] Minal Vaish. Mismatch repair deficiencies transforming stem cells into cancer stem cells and therapeutic implications. Molecular Cancer

[9] Wang JC, Dick JE. cancer stem cells: lessons from leukemia. *Trends Cell Biol*. 2005; **15:**494–501.

[10] Miller SJ, Lavker RM, Sun TT. Interpreting epithelial cancer biology in the context of stem cells: tumor properties and therapeutic implications. *Biochim. Biophys. Acta*. 2005; **1756:** 25–52.

[11] Jemal A, Siegel R, Ward E, et al. Cancer statistics. *CA Cancer J Clin.* 2006; **56:**106–130.

[12] Porter E.M., Bevins C. L., Ghosh D. & Ganz T. The multifaceted Paneth cell. *Cell. Mol*. *Life Sci.* 2002; **59:**156–170

[13] Stierum R, Gaspari M, Dommels Y, Ouatas T, Pluk H, Jespersen S, Vogels J, Verhoeckx K, Groten J, van Ommen B: Proteome analysis reveals novel proteins associated with proliferation and differentiation of the colorectal cancer cell line Caco-2. *Biochim Biophys Acta.* 2003; **1650(1-2)**:73-91.

[14] Radtke F, Clevers H. Self-renewal and cancer of the gut: two sides of a coin*. Science*. 2005; **307:**1904–9.

[15] McDonald SA, Preston SL, Lovell MJ, Wright NA, Jankowski JA. Mechanisms of disease: from stem cells to colorectal cancer. *Nat. Clin. Pract. Gastroenterol. Hepatol*. 2006; **3**:267–74.

[16] Powell DW, Mifflin RC, Valentich JD, Crowe SE, Saada JI, West AB. Myofibroblasts. II. Intestinal subepithelial myofibroblasts. *Am. J.Physiol. Cell Physiol*. 1999; **277:** C183–C201.

[17] Schmidt G. H., Winton D. J., Ponder B. A. Development of the pattern of cell renewal in the cryptvillus unit of chimaeric mouse small intestine. *Development.*1988; **103:** 785–790.

[18] Roth K. A., Hermiston M. L., Gordon J. I. Use of transgenic mice to infer the biological properties of small intestinal stem cells and to examine the lineage relationships of their descendants. *Proc. Natl Acad. Sci*. 1991; **88**:9407–9411.

[19] BM, Wicha MS, Fields JZ, Runquist OA. Symmetric division of cancer stem cells – a key mechanism in tumor growth that should be targeted in future therapeutic approaches. *Clin. Pharmacol. Ther.* 2007; **81**:893–8.

[20] Kim SJ, Cheung S, Hellerstein MK. Isolation of nuclei from label-retaining cells and measurement of their turnover rates in rat colon. *Am J Physiol Cell Physiol*. 2004; **286:**C1464–C1473.

[21] Brabletz T, Jung A, Spaderna S, Hlubek F, Kirchner T. Opinion: migrating cancer stem cells—an integrated concept of malignant tumour progression. *Nat Rev Cancer*. 2005; **5:**744–9.

[22] Cotsarelis G, Sun TT, Lavker RM. Label-retaining cells reside in the bulge area of pilosebaceous unit: implications for follicular stem cells, hair cycle, and skin carcinogenesis. *Cell.* 1990; **61:**1329–133.

[23] Potten CS, Owen G, Booth D (2002). Intestinal stem cells protect their genome by selective segregation of template DNA strands. *J Cell Sci.* 2002; **115:**2381–2388.

[24] Kim SJ, Cheung S, Hellerstein MK. Isolation of nuclei from label-retaining cells and measurement of their turnover rates in rat colon. *Am J Physiol Cell Physiol*. 2004; **286**:C1464–C1473.

[25] Gostjeva EV, Zukerberg L, Chung D, Thilly WG. Bell-shaped nuclei dividing by symmetrical and asymmetrical nuclear fission have qualities of stem cells in human colonic embryogenesis and carcinogenesis. *Cancer Genet. Cytogenet*. 2006; **164**:16–24.

[26] Singh SK, Hawkins C, Clarke ID, Squire JA, Bayani J, Hide T, et al. Identification of human brain tumour initiating cells. *Nature.* 2004; **432:**396–401.

[27] Collins AT, Berry PA, Hyde C, Stower MJ, Maitland NJ. Prospective identification of tumorigenic prostate cancer stem cells. *Cancer Res*. 2005; **65:**10946–10951.

[28] Ma S, Chan KW, Hu L, Lee TK,Wo JY,Ng IO, Zheng BJ, Guan XY. Identification and characterization of tumorigenic liver cancer stem/progenitor cells. *Gastroenterology.* 2007; **132:**2542–2556.

[29] Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci USA*. 2003; **100**:3983–3988.

[30] Li C, Heidt DG, Dalerba P, et al. (2007) Identification of pancreatic cancer stem cells. Cancer Res. 2007; 67: 1030–1037

[31] Prince ME, Sivanandan R, Kaczorowski A, et al. Identification of a subpopulation of cells with cancer stem cell properties in head and neck squamous cell carcinoma. *Proc Natl Acad Sci USA*. 2007; **104:** 973–978.

[32] Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med*. 1997; **3**:730–737.

[33] Miraglia S, Godfrey W, Yin AH, et al. A novel five-transmembrane hematopoietic stem cell antigen: isolation, characterization and molecular cloning. *Blood* 1997; **90:**5013–5021.

[34] Maw MA, Corbeil D, Koch J, et al. A frameshift mutation in prominin (mouse)-like 1 causes human retinal degeneration. *Hum Mol Genet* 2000; **9**:27–34.

[35] Singh SK, Clarke ID, Terasaki M, et al. Identification of a cancer stem cell in human brain tumors. Cancer Res 2003;63:5821–5828.

[36] Bussolati B, Bruno S, Grange C, et al. Isolation of renal progenitor cells from adult human kidney. *Am J Pathol* 2005; **166:**545–555.

[37] Yin S, Li J, Hu C, et al. CD133 positive hepatocellular carcinoma cells possess high capacity for tumorigenicity. *Int J Cancer* 2007; **120:**1444–1450.

[38] Shmelkov SV, Butler JM, Hooper AT, et al. CD133 expression is not restricted to stem cells, and both CD133+ and CD133- metastatic colon cancer cells initiate tumors. *J Clin Invest* 2008; **118:**2111–2120.

[39] Bauer N, Fonseca AV, Florek M, et al. New insights into the cell biology of hematopoietic progenitors by studying prominin-1 (CD133). *Cells Tissues Organs* 2008; **188**:127–138.

[40] Bao S, Wu Q, Sathornsumetee S, et al. Stem cell-like glioma cells promote tumor angiogenesis through vascular endothelial growth factor. *Cancer Res* 2006; **66**:7843–7848.

[41] Bruno S, Bussolati B, Grange C, et al. CD133 renal progenitor cells contribute to tumor angiogenesis. *Am J Pathol* 2006;**169:** 2223–2235.

[42] Zhu L, Gibson P, Currle DS, et al. Prominin 1 marks intestinal stem cells that are susceptible to neoplastic transformation. *Nature* 2009; **457:**603–607.

[43] International Human Genome Sequencing Consortium, *Nature* 2004, 431 (7011): 931 945

[44] Krishna RG, Wold F. "Post-translational modifications of proteins". Adv. Enzymol. Mol. Biol. 1993, **67**: 265-298

[45] Cohen AA, Geva-Zatorsky N, Eden E, Frenkel-Morgenstern M, et al. Dynamic proteomics of individual cancer cells in response to a drug. Science 2008, **322**:1511-1516

[46] Anderson N.L., Anderson N.G. Proteome and proteomics: New technologies, new concepts and new words. *Electrophoresis*. 1998, **19(11**): 1853-1861

[47] Zhou M.,Veenstra T.D., Proteomic analysis of protein complexes. *Proteomics* 2007, **16:** 2688–2697.

[48] Josic D., Clifton J. G., Mammalian plasma membrane proteomics. *Proteomics* 2007, **16**: 3010–3029

[49] Hanash, S. Disease proteomics. *Nature* 2003, **422**: 226–232.

[50] Petricoin, E. F., Liotta, L. A. Clinical applications of proteomics. *J. Nutr*. 2003, **133:** 2476S–2484S.

[51] Moseley, F. L., Bicknell, K. A., Marber, M. S., Brooks, G., The use of proteomics to identify novel therapeutic targets for the treatment of disease. J. Pharm. Pharmacol. 2007, **59:** 609–628.

[52] Zhang, C. C., Kast, J. Applications of current proteomics techniques in modern drug design. *Curr. Comput. Aided Drug Des.* 2010, **6:** 147–164.

[53] Rosenblatt, K. P., Bryant-Greenwood, P., Killian, J. K., Mehta, A. et al. Serum proteomics in cancer diagnosis and management. *Annu. Rev. Med*. 2004, **55:** 97–112.

[54] Petricoin, E. F., Belluco, C., Araujo, R. P., Liotta, L. A. The blood peptidome: a higher dimension of information content for cancer biomarker discovery. *Nat. Rev. Cancer* 2006, **6:** 961–967.

[55] Mirza S.P., Olivier M. Methods and approaches for the comprehensive characterization and quantification of cellular proteomes using mass spectrometry. *Physiol Genomics*. 2008, **33(1):**3-11

[56] Köcher T., Superti-Furga G., Mass spectrometry–based functional proteomics: from molecular machines to protein networks. *Nat. Methods*. 2007, **4(10):** 807-815.

[57] Ratushny V., Golemis E. Resolving the network of cell signaling pathways using the evolving yeast two-hybrid system. *Biotechniques* 2008, **44(5):** 655-662

[58] Issaq H.J., Conrads T.P., Janini G.M. and Veenstra T.D., Methods for fractionation, separation and profiling of proteins and peptides. *Electrophoresis* 2002, **23(17):** 3048-61

[59]. Jensen O.N., Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr. Opin. Chem. Biol*. 2004, **8:** 33-41

[60] O'Farell PH. High resolution two-dimentional electrophoresis of proteins. *J. Biol. Chem*. 1975, **250:** 4007-4021

[61] Unlu M, Morgan M E, Minden J S. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* 1997, **18**: 2071-2077

[62] . Fenn J. B., Mann M., Meng C. K., Wong S. F., and Whitehouse C. M., *Science* 1989, **246**: 64–71.

[63] Karas M. and Hillenkamp F., *Anal. Chem*. 1988, **60:** 2299–2301.

[64] Emmett M.R., Caprioli R.M., *J. Am. Soc. Mass Spectrom* 1994, **5:** 605–613

[65] McLuckey SA, Goeringer DE, Glish GL. Collisional activation with random noise in ion trap mass spectrometry. *Anal. Chem*. 1992, **64:** 1455-1460

[66] Mann M., Højrup P., Roepstorff P., *Biol. Mass Spectrom.* 1993, 22(6): 338–45.

[67] Willis, N. D., Cox, T.R., Rahman-Casañs, S. F., Smits, K. et al., Lamin A/C is a Risk Biomarker in colorectal cancer. *Plos One* 2008, **3(8):**e2988

[68] Belt, EJ, Fijneman, RJ, van den Berg, EG, Bril H et al., Loss of lamin A/C expression in stage II and III colon cancer is associated with disease recurrence. *European Journal of Cancer* 2011, **47:** 1837-1845

[69] Gonçalves, V., Matos, P., Jordan, P., The beta-catenin/TCF4 pathway modifies alternative splicing through modulation of SRp20 expression. *RNA* 2008, **14:** 2538-2549

[70] Soltanian, S., Matin, M.M., Cancer stem cells and cancer therapy. *Tumour Biol*. 2011, **32(3):** 425-440.

[71] Ginestier, C., Hur, M.H., Charafe-Jauffret, E., Monville, F. et al., ALDH1 Is a Marker of Normal and Malignant Human Mammary Stem

Cells and a Predictor of Poor Clinical Outcome. Cell Stem Cell 2007, 15: 555-567.

[72] Balicki, D., Moving Forward in Human Mammary Stem Cell Biology and Breast Cancer Prognostication Using ALDH1. *Cell Stem Cell* 2007, **15:** 485-487.

[73] Jiang, F., Qiu, Q., Khanna, A., Todd, N.W. et al., Aldehyde dehydrogenase 1 is a tumor stem cell-associated marker in lung cancer. *Mol Cancer Res*. 2009, **7(3):** 330-338.

[74] Kroemer G., Pouyssegur, J., Tumor Cell Metabolism: Cancer's Achilles' Heel Cancer *Cell* 2008, **13:** 472

[75] Gatenby, R.A., Gillies, R.J., Why do cancers have high aerobic glycolysis? *Nat. Rev. Cancer* 2004, **4:** 891–899.

[76] Fukuyama, R., Niculaita, R., Ng, K.P., Obusez, E. et al., Mutated in colorectal cancer, a putative tumor suppressor for serrated colorectal cancer, selectively represses β-catenin-dependent transcription. *Oncogene* 2008, **27:** 6044-6055

[77] Clevers, H., Wnt/β-Catenin Signaling in Development and Disease. *Cell* 2006; **127:** 469–80

[78] Malanchi, I., Peinado, H., Kassen, D., Hussenet, T. et al., Cutaneous cancer stem cell maintenance is dependent on β-catenin signaling. *Nature* 2008; **452:** 650–653

[79] Wong, S.C., Lo, S.F., Lee, K.C., Yam, J.W. et al., Expression of frizzled-related protein and Wnt-signalling molecules in invasive human breast tumours. *J Pathol* 2002; **196:** 145–153

[80] Tawfic, S., Yu, S., Wang, H., Faust, R. et al., Protein kinase CK2 signal in neoplasia. *Histol Histopathol*. 2001, **16:** 573–575.

[81] Song, D.H., Sussmani, D.J., Seldin, D.C., Endogenous protein kinase CK2 participates in Wnt signaling in mammary epithelial cells. *J Biol Chem*. 2000, **275:** 23790–23797.

[82] Song, Z., Yao, X., Wu, M., Direct interaction between survivin and Smac/DIABLO is essential for the anti-apoptotic activity of survivin during taxol-induced apoptosis. *J Biol Chem*. 2003, **278:** 23130–23140.

[83] Bikkavilli, R.K., Malbon C.C., Dishevelled-KSRP complex regulates Wnt signaling through post-transcriptional stabilization of β-catenin mRNA. *J Cell Sci* 2010, **123:** 1352-1362.

[84] Càceres, J.F., Misteli, T., Screaton, G.R., Spector, D.L. et al., Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *J. Cell Biol.* 1997, **138:** 225–238.

[85] He, X., Arslan, A.D., Pool, M.D., Ho, T.T. et al., Knockdown of splicing factor SRp20 causes apoptosis in ovarian cancer cells and its expression is associated with malignancy of epithelial ovarian cancer. *Oncogene* 2010, **30(3):** 356-365.

[86] Jia, R., Li, C., McCoy, J.P., Deng, C.X. et al., SRp20 is a proto-oncogene critical for cell proliferation and tumor induction and maintenance. *Int J Biol Sci*. 2010, **6(7):** 806-826

6. APPENDIX I

Table 4: MS details of CaCo-2 identified proteins

Spot No.	Id Method	Protein name	Mascot Protein Score	Sequence coverage (%)	Identified peptide sequence	Ion Score (MS/MS)	Precursor mass (m/z)	Mass error (ppm)
3560	MSMS	Splicing factor, arginine/serine-rich 3 (SRp20)	141	10	VYVGNLGNNGNK	74	625,03	240
					VYVGNLGNNGNKTELER	67	626,70	79
2635	MS	Casein Kinase II (CKII)	159	51	VYTDVNTHRPR		1357,55	110
					EYWDYESHVVEWGNQDDYQLVR		2830,00	84
					GKYSEVFEAINITNNEK		1955,75	112
					GGPNIITLADIVKDPVSR		1864,80	134
					TPALVFEHVNNTDFK		1731,68	109
					QLYQTLTDYDIR		1528,56	130
					FYMYEILK		1122,44	98
					LIDWGLAEFYHPGQEYNVR		2306,85	117
					KEPFFHGHDNYDQLVR		2001,72	119
					EPFFHGHDNYDQLVR		1873,64	117
					VLGTEDLYDYIDKYNIELDPR		2544,43	71
					FNDILGR		834,36	108
					FVHSENQHLVSPEALDFLDK		2324,87	120
					YDHQSR		805,28	99
					EAMEHPYFYTVVK		1629,59	104
3076	MSMS	Glucosamine-6-phosphate deaminase 1	411	30	LVDPLYSIK	47	524,20	210
					IIQFNPGPEK	49	571,68	235
					EVMILITGAHK	44	614,20	225
					TLAMDTILANAR	74	653,18	253
					TFNMDEYVGLPR	68	729,18	219
					TVFVCDEDATLELK	72	820,17	280
					VPTMALTVGVGTVMDAR	57	875,24	228
2747	MSMS	Annexin A2	365	16	TPAQYDASELK	58	612,12	65
					TNQELQEINR	67	623,02	233
					GLGTDEDSLIETICSR	48	889,80	90
					SLYYYIQQDTKGDYQK	63	672,10	54
					RAEDGSVIDYELIDQDAR	76	689,20	184
					AYTNFDAERDALNIETAIK	53	719,31	190
2626	MSMS	Creatine Kinase B type	590	33	LLIEMEQR+ Oxidation (M)	44	524,29	28
					DLFDPIIEDR	46	616,90	146
					VLTPELYAELR	45	652,51	222
					LAVEALSSLDGDLAGR.	99	793,90	19
					LEQGQAIDDLMPAQK	69	836,91	6
					LGFSEVELVQMVVDGVK.L + Oxidation (M)	85	933,14	160
					GTGGVDTAAVGGVFDVSNADR	55	655,80	229
					AIEKLAVEALSSLDGDLAGR	89	676,82	172

					TDLNPDNLQGGDDLDPNYVLSSR	58	840,10	47
2625	MSMS	Creatine Kinase B type	360	22	VLTPELYAELR	48	625,50	153
					LAVEALSSLDGDLAGR.	102	794,04	145
					LEQGQAIDDLMPAQK + Oxidation (M)	56	558,41	215
					LGFSEVELVQMVVDGVK + Oxidation (M)	58	933,09	107
					GTGGVDTAAVGGVFDVSNADR	51	655,77	183
					RGTGGVDTAAVGGVFDVSNADR	45	707,62	89
	MSMS	28S ribosomal protein S27, mitochondrial	219	11	EALDVLGAVLK	50	564,44	186
					LVEQLDIEETEQSK	123	831,04	150
					ALTSADGASEEQSQNDEDNQGSEK	46	837,20	159
2555	MSMS	Methionine aminopeptidase 1	93	5	YRELGNIIQK	43	412,13	40
					NGYHGDLNETR	50	771,80	18
3125	MSMS	Lactamase, beta 2	139	9	SINNDTTYCIK	61	665,10	195
					EEIIGNGEQQYVYLK	78	892,16	300
2667	MSMS	Creatine Kinase B type	323		LLIEMEQR	49	524,32	76
					VLTPELYAELR	44	652,51	222
					LAVEALSSLDGDLAGR.	93	794,06	170
					LEQGQAIDDLMPAQK + Oxidation (M)	74	837,06	173
					GTGGVDTAAVGGVFDVSNADR	63	983,05	81
1701	MS	Lamin A/C	162	33	LQEKEDLQELNDR		1629,81	0
					SLETENAGLR		1089,52	27
					ITESEEVVSR		1148,56	17
					AAYEAELGDAR		1165,53	16
					EGDLIAAQAR		1043,37	95
					TLEGELHDLR		1182,62	8
					QLQDEMLRR		1187,53	32
					NIYSEELR		1023,44	68
					LADALQELR		1028,49	77
					NSNLVGAAHEELQQSR		1752,81	28
					IRIDSLSAQLSQLQK		1699,91	35
					LALDMEIHAYR.		1347,62	37
					LSPSPTSQR		972,46	51
					SSFSQHAR		919,39	54
					VAVEEVDEEGKFVR		1605,77	25
					ASASGSGAQVGGPISSGSSASSVTVTR		2365,1	25
					SVGGSGGGSFGDNLVTR		1566,75	0
					TQSPQNCSIM		1165,5	0
2532	MSMS	Elongation factor Tu	577	26	GTVVTGTLER	53	516,71	249
					VEAQVYILSK	64	575,21	208
					YEEIDNAPEER	58	682,66	252
					QIGVEHVVVYVNH	65	733,83	238

					KYEEIDNAPEER	69	746,65	229
					LLDAVDTYIPVPAR	51	771,90	146
					GEETPVIVGSALCALEGR	50	929,26	234
					TIGTGLVTNTLAMTEEEK	65	962,22	275
					ADAVQDSEMVELVELEIR	102	1031,27	225
	MSMS	Pyruvate dehydrogenase E1-alpha precursor	240	11	EILAELTGR	56	501,15	269
					YGMGTSVER	60	508,08	295
					VDGMDILCVR	50	597,28	244
					LEEGPPVTTVLTR	74	706,20	276
3047	MSMS	Esterase D	169	29	DDQFLLDGQLLPDNFIAACT EK	68	842,15	126
					GEDESWDFGTGAGFYVDATE DPWK	48	893,97	280
					SYPGSQLDILIDQGKDDQFL LDGQLLPDNFIAACTEK	53	1380,20	240
2945	MSMS	L-lactate dehydrogenase B	245	13	IVVVTAGVR	62	457,12	128
					GLTSVINQK	51	480,20	104
					MVVESAYEVIK	42	642,14	230
					IVADKDYSVTANSK	90	755,63	280
2791	MSMS	Annexin A2	722	48	AYTNFDAER	43	543,58	300
					DALNIETAIK	51	544,17	248
					TPAQIDASELK	61	611,63	278
					DIISDTSGDFR	56	613,27	16
					TNQELQEINR	67	622,80	24
					SYSPYDMLESIR	49	738,75	218
					GVDEVTIVNILTNR	82	711,82	280
					GLGTDEDSLIETICSR	63	889,81	165
					TDLEKDIISDTSGDFR	43	604,77	220
					AEDGSVIDYELIDQDAR	78	954,64	300
					RAEDGSVIDYELIDQDAR	75	688,80	290
					AYTNFDAERDALNIETAIK	54	718,89	280
		Glyceraldeide-3-phosphate dehydrogenase	121	5	VPTANVSVVDLTCR	59	765,70	260
					VIHDNFGIVEGLMTTVHAITATQK	62	870,90	300
2418	MSMS	ARP3 actin-related protein 3 homolog variant	328	18	LSEELSGGR	59	474,37	260
					FMEQVIFK.Y + Oxidation (M)	48	502,36	236
					LKLSEELSGGR	49	594,95	193
					QYTGINAISKK	48	603,46	215
					NIVLSGGSTMFR.D + Oxidation (M)	68	649,47	215
					DREVGIPPEQSLETAK	56	590,43	209
	MSMS	Spliceosome RNA helicase BAT1	125	6	ILVATNLFGR	54	552,30	45
					GLAITFVSDENDAK	71	740,48	148

2128	MS	Catalase	135	30	ADVLTTGAGNPVGDKLNVITVGPR		2363,99	126
					GAGAFGYFEVTHDITK		1712,82	111
					VFEHIGKK		957,54	104
					KTPIAVR		784,50	89
					TPIAVR		656,40	91
					FSTVAGESGSADTVRDPR		1851,67	113
					NLSVEDAAR		974,39	102
					LSQEDPDYGIR		1292,47	108
					LFAYPDTHR		1119,43	116
					LGPNYLHIPVNCPYR		1812,74	99
					VANYQR		750,32	120
					FNTANDDNVTQVR		1493,54	107
					AFYVNVLNEEQR		1481,57	114
					NAIHTFVQSGSHLAAR		1708,70	111
3180	MSMS	Proline synthetase co-transcribed homolog	293	32	DLPAIQPR	46	455,17	110
					VMVQINTSGEESK	61	719,14	277
					TFGENYVQELLEK	82	785,16	286
					HGLPPSETIAIVEHINAK	46	642,53	233
					LMAVPNLFMLETVDSVK	58	969,76	247
3496	MSMS	PEST proteolytic signal-containing nuclear protein	138	15	DTPTSAGPNSFNK	43	668,58	172
					SAEEEAADLPTKPTK	95	529,78	239
2188	MSMS	Tubulin beta-2C chain	218	14	LAVNMVPFPR + Oxidation (M)	68	580,45	224
					EVDEQMLNVQNK + Oxidation (M)	58	731,85	62
					AVLVDLEPGTMDSVR + Oxidation (M)	44	809,54	160
					MSMKEVDEQMLNVQNK.N + 3 Oxidation (M)	48	658,03	96
	MSMS	Alpha-tubulin	176	10	DVNAAIATIK	61	508,40	197
					AVFVDLEPTVIDEVR	45	851,55	111
					TIGGGDDSFNTFFSETGAGK	70	1004,56	109
3434	MSMS	Catenin gamma	164	5	LLNDEDPVVVTK	48	671,24	186
					TMQNTSDLDTAR	61	684,70	143
					ALMGSPQLVAAVVR	55	714,35	70
		Fortilin (p23)	87	16	VKPFMTGAAEQIK	44	479,73	248
					EDGVTPYMIFFK	43	731,66	260
1823	MSMS	Heat shock cognate 71 kDa protein isoform 1	265	8	FEELNADLFR	60	627,41	159
					TTPSYVAFTDTER	90	744,49	181
					SFYPEEVSSMVLTK Oxidation (M)	50	817,01	141
					NQVAMNPTNTVFDAK.R + Oxidation (M)	65	833,48	96
	MSMS	Heat shock protein 90kDa alpha (cytosolic)	138	3	YIDQEELNK	56	576,39	182

					GVVDSEDLPLNISR.	82	757,50	138
3352	MSMS	Enoyl-CoA hydratase, mitochondrial	302	27	AFAAGADIK	44	432,51	34
					GKNNTVGLIQLNRPK	43	551,78	248
					ESVNAAFEMTLTEGSK	68	865,80	293
					TFEEDPAVGAIVLTGGDK	88	916,37	291
					AQFAQPEILIGTIPGAGGTQ R	59	1063,44	227
1826	MSMS	Heat shock cognate 71 kDa protein isoform 1	259	8	FEELNADLFR	55	627,39	159
					TTPSYVAFTDTER	80	744,48	181
					SFYPEEVSSMVLTK Oxidation (M)	70	817,01	140
					NQVAMNPTNTVFDAK.R + Oxidation (M)	54	833,46	95
2179	MS	Aldheyde dehydrogenase 1A1	134	37	LADLIERDR		1100,66	45
					TIPIDGNFFTYTR		1544,85	51
					LIKEAAGK		829,52	12
					RVTLELGGK		972,63	51
					IFVEESIYDEFVR		1645,89	48
					IFVEESIYDEFVRR		1802,04	72
					RSVER		646,40	62
					YILGNPLTPGVTQGPQIDKEQYDK		2674,50	48
					ILDLIESGKK		1115,73	53
					GYFVQPTVFSNVTDEMR		2006,09	79
					IAKEEIFGPVQQIMK		1747,02	40
					SLDDVIKR		945,58	42
					ANNTFYGLSAGVFTK		1589,87	44
					ELGEYGFHEYTEVK		1700,88	58
2178	MS	UDP glucose dehydrogenase	100	26	VTVVDVNESR		1117,69	98
					YIEACAR		882,50	102
					VLIGGDETPEGQR		1370,82	94
					LAANAFLAQR		1074,71	93
					YWQQVIDMNDYQR		1774,95	90
					IIDSLFNTVTDKK		1493,98	107
					YLMDEGAHLHIYDPK		1818,02	93
					EQIVVDLSHPGVSEDDQVSR		2209,29	99
					MLKPAFIFDGR		1310,83	106
					RIPYAPSGEIPK		1327,85	82
					FSLQDPPNK		1045,66	120
2409	MSMS	Cytokeratin 18	357	16	VIDDTNITR	65	523,79	28
					IVLQIDNAR	58	521,38	144
					IIEDLRAQIFANTVDNAR	54	687,18	208
					TVQSLEIDLDSMR + Oxidation (M)	60	761,94	85
					QAQEYEALLNIK	72	710,58	282

					QSVENDIHGLR + Pyro-glu (N-term Q)	48	625,84	48
2742	MS	DNA polymerase delta interacting protein 2	217	46	GVVLFPWQAR		1172,73	59
					LYDRDVASAAPEK		1434,83	76
					GKTHTYYQVLIDAR		1664,98	60
					THTYYQVLIDAR		1479,85	60
					DCPHISQR		1012,54	79
					SQTEAVTFLANHDDSR		1790,97	78
					FLLYDQTK		1027,62	68
					APPFVAR		757,49	66
					NHPWLELSDVHR		1502,85	66
					ETTENIR		862,50	81
					VTVIPFYMGMR		1345,78	89
					EAQNSHVYWWR		1475,80	81
					LENLDSDVVQLR		1400,84	71
					ERHWR		783,46	76
					IFSLSGTLETVR		1322,82	68
					GRGVVGR		700,43	19
					GVVGREPVLSK		1140,75	70
					FERPDGSHFDVR		1461,79	68
2227	MSMS	ATP synthase, H+ transporting, mitochondrial	177	8	VVDALGNAIDGK	82	586.61	85
					NVQAEEMVEFSSGLK	50	843,04	136
					GMSLNLEPDNVGVVVFGNDK	45	707,44	280
1808	MSMS	t-complex polypeptide 1	90	4	EQLAIAEFAR	45	574,42	191
					FATEAAITILR	45	603,47	199
2130	MS	Phenylalanyl-tRNA synthetase	84	21	STKHWELTAEGEEIAR		1856,67	129
					HWELTAEGEEIAR		1540,61	84
					EGSHEAR		785,30	63
					SIPPEGLAQSELMR		1527,67	72
					LQLVR		628,37	63
					LGEKER		731,35	68
					THSQGGYGSQGYK		1369,50	80
					THTTSASAR		931,38	85
					KPFTPVK		816,43	85
					YFSIDR		800,34	62
					YGINNIR		849,39	82
					LDAEPRPPPTQEAA		1491,61	87
3053	MSMS	Cytokeratin 8	179	9	SLDMDSIIAEVK.A + Oxidation (M)	60	668,91	112
					TEMENEFVLIK.K + Oxidation (M)	59	684,98	175
					R.LEGLTDEINFLR.Q	60	710,51	183
2487	MSMS	Aldolase A	84	14	GVVPLAGTNGETTTQGLDGL SER	42	1137,08	127
					VDKGVVPLAGTNGETTTQGL DGLSER	42	872,42	220

1917	MSMS	Heterogeneous nuclear ribonucleoprotein Q (hnRNP Q) isoform3	229	9	LMMDPLTGLNR + 2 Oxidation (M)	49	646,94	185
					TGYTLDVTTGQR	54	656,48	229
					NLANTVTEEILEK	82	737,49	129
					DLFEDELVPLFEK	44	797,52	144
		Phosphoenolpyruvate carboxykinase (GTP)	124		EVLAELEALER	63	636,41	102
					LGTPVLQALGDGDFVK	61	815,52	92
1778	MSMS	Dead box, X isoform	308	14	DLLDLLVEAK	50	564,90	124
					SPILVATAVAAR	60	584,92	111
					DSLTLVFVETK.	48	626,40	88
					ELAVQIYEEAR	54	660,93	128
					MLDMGFEPQIR.R + 2 Oxidation (M)	48	684,97	219
					VGNLGLATSFFNER	48	763,01	151
		Nucleolin	154	6	NDLAVVDVR	55	500,87	190
					GFGFVDFNSEEDAK	45	781,46	147
					GLSEDTTEETLKESFDGSVR	54	734,11	131
2647	MS	Proteasome 26S ATPase subunit 6	139	38	DKALQDYR		1008,46	49
					LLEHKEIDGR		1209,61	41
					ELREQLK		915,48	43
					YVVGCR		753,35	26
					LKPGTR		671,40	29
					VALDMTTLTIMR		1396,65	71
					GCLLYGPPGTGK		1219,57	32
					VVSSSIVDKYIGESAR		1709,81	58
					EMFNYAR		946,37	42
					FSEGTSADREIQR		1495,65	40
					KIHIDLPNEQAR		1433,71	55
					IHIDLPNEQAR		1305,65	45
					HGEIDYEAIVK		1273,59	39
					LSDGFNGADLR		1164,51	42
					NVCTEAGMFAIR		1384,59	28
					ADHDFVVQEDFMK		1596,63	44
1722	MSMS	Chaperonin (HSP60)	206	11	AAVEEGIVLGGGCALLR	92	843,00	53
					ALMLQGVDLLADAVAVTMGPK	57	710,47	122
					TALLDAAGVASLLTTAEVVVTEIPKEEK	57	957,03	170
		Protein disulfide-isomerase A4 precursor	119		IDATSASVLASR	76	595,92	168
					MDATANDVPSDR + Oxidation (M)	43	654,32	61
2226	MS	Inosine monophosphate dehydrogenase 2	227	39	KYEQGFITDPVVLSPK		1820,97	5
					VRDVFEAK		963,53	0

					HGFCGIPITDTGR		1430,7	13
					EANEILQR		972,5	0
					GKLPIVNEDDELVAIIAR		1965,15	25
					LPIVNEDDELVAIIAR		1779,98	7
					NRDYPLASK		1063,64	84
					DKYPNLQVIGGNVVTAAQAK		2086,18	20
					NLIDAGVDALR		1156,63	0
					VSEYAR		724,37	13
					RFGVPVIADGGIQNVGHIAK		2048,19	22
					FGVPVIADGGIQNVGHIAK		1892,03	4
					YRGMGSLDAMDK		1376,62	14
					HLSSQNRYFSEADK		1681,85	35
					YFSEADKIK		1100,55	0
					IKVAQGVSGAVQDK		1399,64	107
					SLTQVR		703,42	17
					AMMYSGELK		1045,56	8
					TSSAQVEGGVHSLHSYEK		1915,89	10
2098	MSMS	Tapasin	302	14	LAPEYEAAATR	65	596,44	226
					DGEEAGAYDGPR	57	618,88	194
					ELSDFISYLQR	51	685,93	116
					DLLIAYYDVDYEK	59	810,48	105
					SDVLELTDDNFESR	70	820,48	128
2888	MS	Heparan sulfate glucosamine 3-O-sulfotransferase 3A1	69	52	QLPQAIIIGVKK		1307,67	130
					GGTRALLEFLR		1232,6	89
					AVGAEPHFFDR		1245,59	8
					DLMPRTLDGQITMEK		1763,79	4
					TLDGQITMEKTPSYFVTR		2102,73	142
					ISAMSKDTK		996,55	50
					AISDYTQTLSK		1226,78	81
					AISDYTQTLSKRPDIPTFESLTFK		2758,24	68
					TAGLIDTSWSAIQIGIYAK		2008,44	149
					HLEHWLRHFPIR		1640,86	18
					QMLFVSGER		1082,17	220
					RIITDK		745,41	67
					HFYFNK		855,15	17
1906	MSMS	Heat-Shock 70kd Protein	130	9	TTPSYVAFTDTER	76	744,46	141
					NQVALNPQNTVFDAK	54	830,04	132
1600	MSMS	Filamin A	155	1,5	AGNNMLLVGVHGPR	91	484,43	275
					NGQHVASSPIPVVISQSEIG DASR	64	816,87	65
		Threonyl-tRNA synthetase	110	5	NELSGALTGLTR	61	616,61	120
					QLENSLNEFGEKWELNSGDG AFYGPK	49	971,56	65
1893	MSMS	Heat shock 70 kDa protein	282	10	TTPSYVAFTDTER	77	744,46	141
					NQVAMNPTNTVFDAK	69	830,04	132
					IINEPTAAAIAYGLDR	81	844,52	77

					DAGVIAGLNVLR	55	599,46	183
1845	MSMS	Annexin A6	231	9	ALIEILATR	58	500,41	190
					SELDMLDIR + Oxidation (M)	49	554,4	235
					DAFVAIVQSVK	70	588,92	144
					EDAQEIADTPSGDKTSLETR	54	721,75	106

Table 5: MS details of HCT-116 identified proteins

Spot No.	Id Method	Protein name	Mascot Protein Score	Sequence coverage (%)	Identified peptide sequence	Single ion Score (MS/MS)	Precursor mass (m/z)	Mass error (ppm)
4928	MSMS	Peroxiredoxin-1	266	37	TIAQDYGVLK	49	554,42	207
					QITVNDLPVGR+ Pyro-glu (N-term Q)	49	597,95	209
					GLFIIDDKGILR	45	454,03	206
					QGGLGPMNIPLVSDPK	69	820,01	97
					TIAQDYGVLKADEGISFR	54	661,84	247
		Neuropolypeptide h3	194	18	LYTLVLTDPDAPSR	54	781,02	134
					GNDISSGTVLSDYVGSGPPK	140	975,56	87
2988	MSMS	Heat shock protein 90kDa alpha (cytosolic)	112	3	EQVANSAFVER	49	625,18	208
					GVVDSEDLPLNISR	63	757,53	178
4323	MSMS	Cell division control protein 42 homolog isoform 1	96	19	NVFDEAILAALEPPEPK	48	618,31	86
					TPFLLVGTQIDLRDDPSTIEK	48	786,9	190
3826	MSMS	Heat shock protein 27	167	18	QLSSGVSEIR	43	538,64	83
					LFDQAFGLPR	75	582,64	51
					VSLDVNHFAPDELTVK	49	595,7	72
4018	MSMS	Phosphoribosyltransferase domain-containing protein 1	157	16	NVLIVEDVVGTGR	87	685,96	109
					NDQSMGEMQIIGGDDLSTLAGK + 2 Oxidation (M)	70	771,49	86
2989	MSMS	Annexin A1	397	29	TPAQFDADELR	87	632,09	120
					GTDVNVFNTILTTR	71	776,3	71
					SEDFGVNEDLADSDAR	106	870,73	143
					MYGISLCQAILDETK	73	879,79	250
					GGPGSAVSPYPTFNPSSDVAALHK	60	786,6	106
		Heterogeneous nuclear ribonucleoprotein H3	180	10	STGEAFVQFASK	53	636,58	180
					DGMDNQGGYGSVGR	83	715,07	220
					ATENDIANFFSPLNPIR	44	960,17	280
4000	MSMS	Peroxiredoxin-4	166	9	QITLNDLPVGR + Pyro-glu (N-term Q)	54	604,97	223
					LVQAFQYTDK	56	606,9	140
					DYGVYLEDSGHTLR	56	813,01	154
		Peroxiredoxin-6	101	12	LPFPIIDDR	50	543,4	92
					DGDSVMVLPTIPEEEAK+ Oxidation (M)	51	923,24	216
2941	MSMS	Coproporphyrinogen oxidase	375	32	FGLFTPGSR	50	491,39	265
					YFEVEEADGNK	47	650,91	184

					YVEFNLLYDR	54	666,44	157
					IESILMSLPLTAR + Oxidation (M)	46	730,5	130
					AGVSISVVHGNLSEEAAK	59	590,13	254
					ATSLGRPEEEEDELAHR	49	646.98	51
					GIGGIFFDDLDSPSKEEVFR	70	1114,55	44
		Heterogeneous nuclear ribonucleoprotein C (C1/C2)	234	16	SDVEAIFSK	45	498,37	231
					KSDVEAIFSK	43	562,4	169
					GFAFVQYVNER	54	665,43	143
					MIAGQVLDINLAAEPK+ Oxidation (M	92	850,07	129
3934	MSMS	Actin beta	111	10	GYSFTTTAER	44	566,86	105
					SYELPDGQVITIGNER	67	859,92	33
		Enolase	109	9	VVIGMDVAASEFFR + Oxidation (M)	68	779,03	180
					DATNVGDEGGFAPNILENK	41	981,04	76
2903	MSMS	Transaldolase	91	6	TIVMGASFR + Oxidation (M)	45	499,32	120
					LLGELLQDNAK	46	607,42	123
3578	MSMS	Chloride intracellular channel protein 1	251	33	GVTFNVTTVDTK	101	641,43	140
					NSNPALNDNLEK	55	664,92	143
					VLDNYLTSPLPEEVDETSAEDEGVSQR	50	998,2	63
					VLDNYLTSPLPEEVDETSAEDEGVSQRK	45	1040,9	64
1627	MSMS	Heat shock cognate 71 kDa protein isoform 1	632	27	FEELNADLFR	48	627,41	159
					TTPSYVAFTDTER	77	744,51	208
					SFYPEEVSSMVLTK Oxidation (M)	74	816,98	104
					NQVAMNPTNTVFDAK + Oxidation (M)	82	833,52	96
					LSKEDIER	48	495,38	232
					VQVEYKGETK	50	590,94	212
					DAGTIAGLNVLR	52	600,46	200
					VEIIANDQGNR	66	614,94	195
					NSLESYAFNMK	87	660,41	166
					STAGDTHLGGEDFDNR	48	564,71	230
2684	MSMS	Septin 2	206	12	TIISYIDEQFER	67	757,18	264
					TVQIEASTVEIEER	77	802,18	287
					ASIPFSVVGSNQLIEAK	62	880,24	270
1692	MSMS	Succinate dehydrogenase complex, subunit A, flavoprotein (Fp)	109	7	GEGGILINSQGER	43	665,41	105
					IDEYDYSKPIQGQQK	66	906,56	127
2351	MSMS	keratin 18	166	7	IVLQIDNAR	49	521,6	19

					QSVENDIHGLR	43	626,08	130
					AQIFANTVDNAR	74	660,59	190
2761	MSMS	Alcohol dehydrogenase [NADP+]	399	22	ALEALVAK	68	407,68	184
					SPAQILLR	43	449,2	178
					YALSVGYR	45	464,66	194
					GLVQALGLSNFNSR	75	738,12	232
					VFDFTFSPEEMK	58	746,61	268
					GLEVTAYSPLGSSDR	56	776,16	241
					DPDEPVLLEEPVVLALAEK	54	1038,19	260
1759	MS	hsp70/hsp90 organizing protein(HOP)	93	23	WVNELKEK		1045,44	120
					LDPHNHVLYSNR		1464,57	109
					KAAALEFLNR		1132,51	123
					TYEEGLKHEANNPQLK		1870,83	48
					TLLSDPTYR		1065,44	112
					ELIEQLR		900,49	22
					ETKPEPMEEDLPENKK		1929,85	26
					YKDAIHFYNK		1298,55	75
					SLAEHR		712,3	98
					SLAEHRTPDVLK		1365,3	51
					LAYINPDLALEEK		1488,62	111
					HYTEAIKR		1017,48	59
					LILEQMQK		1002,49	70
					LMDVGLIAIR		116,55	80
2053	MSMS	Phosphoglycerate dehydrogenase	286	11	VTADVINAAEK	62	566,15	35
					ILQDGGLQVVEK	77	650,15	154
					GTIQVITQGTSLK	85	673,4	300
					TQTSDPAMLPTMIGLLAEAGVR	62	768,84	290
		Inosine monophosphate dehydrogenase	128	6	NLIDAGVDALR	69	579,24	69
					YEQGFITDPVVLSPK	59	847,17	59
2319	MSMS	Fumarate hydratase precursor	329	15	EYDTFGELK	61	607,85	90
					AIEMLGGELGSK	73	610,67	237
					VAALTGLPFVTAPNK	51	749,85	113
					IYELAAGGTAVGTGLNTR	84	882,5	45
					SGLGELILPENEPGSSIMPGK	60	1070,9	126
1388	MSMS	Chaperonin (HSP60)	87	8	AAVEEGIVLGGGCALLR	43	843,09	160
					TALLDAAGVASLLTTAEVVVTEIPKEEK	44	957,10	254
2063	MS	UDP-glucose dehydrogenase	92	29	VTVVDVNESR		1117,58	0
					EADLVFISVNTPTK		1533,71	71
					VLIGGDETPEGQR		1370,7	8
					LAANAFLAQR		1074,6	10
					DVLNLVYLCEALNLPEVAR		2201,18	9

					IIDSLFNTVTDKK	1493,8	16
					YLMDEGAHLHIYDPK	1817,88	27
					EQIVVDLSHPGVSEDDQVSR	2209,13	8
					MLKPAFIFDGR	1310,7	8
					RIPYAPSGEIPK	1327,74	8
					FSLQDPPNK	1045,54	0
1421	MS	lamin A/C isoform 3	212	44	SGAQASSTPLSPTR	1359,65	29
					LQEKEDLQELNDR	1629,81	0
					LAVYIDR	849,42	70
					SLETENAGLR	1089,52	27
					ITESEEVVSR	1148,56	17
					AAYEAELGDAR	1165,53	16
					TLEGELHDLR	1182,62	8
					RVDAENR	859,4	46
					LQTMKEELDFQK	1525,68	52
					NIYSEELR	1023,5	9
					LVEIDNGKQR	1171,6	17
					LADALQELR	1028,54	29
					AQHEDQVEQYKK	1502,67	33
					NSNLVGAAHEELQQSR	1752,85	5
					IRIDSLSAQLSQLQK	1699,84	17
					LRDLEDSLAR	1187,63	8
					MQQQLDEYQELLDIKLALDMEIHAYR	3223,2	186
					LALDMEIHAYR	1347,63	29
					LALDMEIHAYRK	1475,63	61
					LLEGEEERLR	1243,67	8
					VAVEEVDEEGKFVR	1605,79	12
					QNGDDPLLTYR	1274,64	31
					ASASGSGAQVGGPISSGSSASSVTVTR	2365,14	8
					SVGGSGGGSFGDNLVTR	1566,74	6
2215	MS	keratin8	237	42	SYTSGPGSR	911,41	11
					ISSSSFSR	870,4	14
					VGSSNFR	766,35	39
					WSLLQQQK	1030,45	98
					WSLLQQQKTAR	1358,57	100
					SNMDNMFESYINNLR	1879,75	21
					NKYEDEINK	1152,4	97
					LEGLTDEINFLR	1419,66	63
					QLYEEEIR	1062,4	94
					ELQSQISDTSVVLSMDNSR	2108,93	37
					SLDMDSIIAEVK	1336,58	20
					AQYEDIANR	1079,42	51
					SRAEAESMYQIK	1428,6	62
					AEAESMYQIK	1185,46	73
					YEELQSLAGK	1137,48	89
					HGDDLR	712,3	87
					TKTEISEMNR	1224,5	42
					ASLEAAIADAEQR	1344,58	70
					GELAIKDANAK	1129,52	74
					EYQELMNVK	1169,46	68
					LALDIEIATYR	1277,58	84

					LLEGEESR		932,4	64
					LESGMQNMSIHTK		1507,61	53
1466	MS	Ezrin	127	28	PKPINVR		823,45	72
					FYPEDVAEELIQDITQK		2038,16	70
					EGILSDEIYCPPETAVLLGSYAVQAK		2823,66	88
					SGYLSSER		898,37	59
					DQWEDR		848,3	67
					IQVWHAEHR		1175,58	80
					IAQDLEMYGINYFEIK		1963,06	30
					IGFPWSEIR		1104,55	29
					KAPDFVFYAPR		1310,68	60
					APDFVFYAPR		1182,58	43
					RKPDTIEVQQMK		1488,8	89
					AKEELER		874,4	90
					SQEQLAAELAEYTAK		1651,85	32
					QLLTLSSELSQAR		1445,77	17
					THNDIIHNENMR		1509,72	39
1207	MSMS	KSRP /KH type-splicing regulatory protein	282	12	MMLDDIVSR	55	556,13	234
					VPDGMVGLIIGR	48	621,72	201
					LASQGDSISSQLGPIHPPPR	57	686,22	194
					MILIQDGSQNTNVDKPLR	43	686,87	243
					TSMTEEYRVPDGMVGLIIGR	79	752,52	240
1401	MSMS	Lamin A/C	668	23	LQEKEDLQELNDR.	44	815,51	122
					MQQQLDEYQELLDIK	79	995,56	99
					ASASGSGAQVGGPISSGSSASSVTVTR	81	1183,06	21
					NSNLVGAAHEELQQSR	60	585,07	188
					VAVEEVDEEGKFVR	57	803,51	124
					SVGGSGGGSFGDNLVTR	87	783,99	140
					TALINSTGEEVAMR + Oxidation (M)	68	754,47	126
					LRDLEDSLAR	53	594,42	160
					AAYEAELGDAR	68	583,35	120
					LADALQELR	71	514,89	194
3206	MSMS	Cytosolic malate dehydrogenase	104	8	LGVTANDVK	49	459,07	98
					VIVVGNPANTNCLTASK	55	879,95	62
1413	MS	Ezrin	98	28	EVWYFGLHYVDNK		1669,69	59
					VSAQEVR		788,32	65
					FGDYNK		743,28	67
					SGYLSSER		898,32	45
					LIPQR		626,3	57
					DQWEDR		848,31	65
					IQVWHAEHR		1175,55	85
					IAQDLEMYGINYFEIK		1963,06	30
					IGFPWSEIR		1104,55	29
					KAPDFVFYAPR		1310,68	15
					APDFVFYAPR		1182,58	43
					RKPDTIEVQQMK		1488,8	13

					ALQLEEER		987,46	23
					AKEELER		874,4	39
					SQEQLAAELAEYTAK		1651,85	18
					IALLEEAR		914,5	21
					QLLTLSSELSQAR		1445,7	25
					THNDIIHNENMR		1509,7	41
3112	MSMS	Annexin A3	198	13	ALLTLADGR	50	465,29	32
					SEIDLLDIR	56	537,3	9
					GIGTDEFTLNR	46	611,95	237
					SDTSGDYEITLLK	46	721,46	145
		Inoraganic pyrpohosphatase	101	11	DKDFAIDIIK	54	589,45	212
					VIAINVDDPDAANYNDINDVKR	47	815,48	90
1563	MS	Ezrin	99	24	PKPINVR		823,54	36
					EVWYFGLHYVDNK		1669,84	23
					SGYLSSER		898,47	18
					DQWEDR		848,4	58
					IQVWHAEHR		1175,63	17
					IAQDLEMYGINYFEIK		1962,97	10
					IGFPWSEIR		1104,61	27
					KAPDFVFYAPR		1310,7	7
					APDFVFYAPR		1182,62	25
					RKPDTIEVQQMK		1488,8	13
					AKEELER		874,49	34
					SQEQLAAELAEYTAK		1651,85	18
					IALLEEAR		914,5	21
					QLLTLSSELSQAR		1445,82	13
					THNDIIHNENMR		1509,74	33
1432	MS	Ezrin	87	18	PKPINVR		823,45	72
					FYPEDVAEELIQDITQK		2038,16	70
					IQVWHAEHR		1175,63	17
					IGFPWSEIR		1104,61	27
					KAPDFVFYAPR		1310,7	7
					APDFVFYAPR		1182,62	25
					RKPDTIEVQQMK		1488,8	13
					QQLETEKK		1003,49	49
					EKEELMLR		1063,52	28
					QLLTLSSELSQAR		1445,82	13
					THNDIIHNENMR		1509,74	33
1399	MS	Lamin A/C isoform 3	133	28	SGAQASSTPLSPTR		1359,65	29
					LQEKEDLQELNDR		1629,81	0
					SLETENAGLR		1089,52	27
					EGDLIAAQAR		1043,37	95
					TLEGELHDLR		1182,62	8
					NIYSEELR		1023,5	9
					LADALQELR		1028,54	29
					NSNLVGAAHEELQQSR		1752,85	5
					DLEDSLAR		918,4	54
					IDSLSAQLSQLQK		1430,7	56
					LALDMEIHAYR		1347,63	29

					QNGDDPLLTYR		1274,64	31
					ASASGSGAQVGGPISSGSSAS SVTVTR		2365,14	8
					SVGGSGGGSFGDNLVTR		1566,74	6
1073	MSMS	EEF2	93	3	VNFTVDQIR	43	546,39	174
					VFSGLVSTGLK	50	554,4	135
1673	MS	Hsp70	171	32	AAAIGIDLGTTYSCVGVFQHGK		2265,18	22
					VEIIANDQGNR		1228,56	81
					TTPSYVAFTDTER		1487,68	13
					IINEPTAAAIAYGLDR		1687,87	17
					ATAGDTHLGGEDFDNR		1675,74	6
					LVNHFVEEFK		1261,6	47
					TLSSSTQASLEIDSLFEGIDFYT SITR		2981,56	33
					ARFEELCSDLFR		1542,71	19
					FEELCSDLFR		1315,54	18
					LDKAQIHDLVLVGGSTR		1822	10
					AQIHDLVLVGGSTR		1465,78	20
					LLQDFFNGR		1109,5	63
					QTQIFTTYSDNQPGVLIQVYEG ERAMTK		3216,37	55
					ITITNDKGR		1017,51	59
					LSKEEIER		1003,45	99
					YKAEDEVQR		1137,47	61
1433	MS	Lamin A/C isoform 3	245	38	SGAQASSTPLSPTR		1359,65	29
					LQEKEDLQELNDR		1629,81	0
					SLETENAGLR		1089,52	27
					ITESEEVVSR		1148,56	17
					AAYEAELGDAR		1165,53	16
					ARLQLELSK		1057,6	37
					EAALSTALSEKR		1275,64	39
					TLEGELHDLR		1182,62	8
					NIYSEELR		1023,5	9
					LVEIDNGKQR		1171,6	17
					LADALQELR		1028,54	29
					NSNLVGAAHEELQQSR		1752,81	28
					IRIDSLSAQLSQLQK		1699,91	35
					LALDMEIHAYR.		1347,62	37
					LLEGEEERLR		1243,67	8
					VAVEEVDEEGKFVR		1605,77	25
					ASASGSGAQVGGPISSGSSAS SVTVTR		2365,1	25
					SVGGSGGGSFGDNLVTR		1566,75	0
1400	MS	Lamin A/C isoform 3	283	49	SGAQASSTPLSPTR		1359,65	29
					LQEKEDLQELNDR		1629,81	0
					EDLQELNDR		1131,44	60
					SLETENAGLR		1089,52	27
					ITESEEVVSR		1148,56	17
					AAYEAELGDAR		1165,53	16
					VREEFK		807,36	99
					EGDLIAAQA		1043,46	86

					TLEGELHDLR	1182,53	67
					QLQDEMLRR	1187,53	32
					RVDAENR	859,37	80
					LQTMKEELDFQK	1525,68	52
					NIYSEELR	1023,43	89
					RHETR	698,3	100
					AQHEDQVEQYKK	1502.77	3
					NSNLVGAAHEELQQSR	1752,81	28
					IRIDSLSAQLSQLQK	1699,91	35
					MQQQLDEYQELLDIK	1909,98	31
					LALDMEIHAYR.	1347,62	37
					LLEGEEERLR	1243,67	8
					ASSHSSQTQGGGSVTK	2365,07	38
					SSFSQHAR	919,36	87
					QNGDDPLLTYR	1274,61	7
					ASASGSGAQVGGPISSGSSAS SVTVTR	2365,1	25
					SVGGSGGGSFGDNLVTR	1566,75	0
					TQSPQNCSIM	1165,5	0
1445	MS	Ezrin	171	33	PKPINVR		
					EVWYFGLHYVDNK	823,44	85
					VSAQEVR	1669,84	23
					FYPEDVAEELIQDITQK	788,32	65
					SGYLSSER	2037,85	73
					DQWEDR	898,47	18
					IQVWHAEHR	848,4	58
					IAQDLEMYGINYFEIK	1175,63	17
					GTDLWLGVDALGLNIYEK	1962,97	10
					IGFPWSEIR	1976,86	86
					KAPDFVFYAPR	1104,61	27
					APDFVFYAPR	1310,7	7
					RKPDTIEVQQMK	1182,62	25
					EKEELMLR	1488,8	13
					ALQLEEER	1063,45	94
					AKEELER	987,43	81
					SQEQLAAELAEYTAK	874,39	80
					QLLTLSSELSQAR	1651,85	18
					THNDIIHNENMR	1445,82	13
						1509,74	33
1398	MS	Lamin A/C isoform 3	238	35	SLETENAGLR		
					AAYEAELGDARK	1089,52	27
					TLEGELHDLRGQVAK	1293,62	15
					LQTMKEELDFQK	1665,82	42
					NIYSEELRETK	1525,76	0
					LADALQELR	1381,64	43
					AQHEDQVEQYKK	1028,53	38
					NSNLVGAAHEELQQSR	1502.77	3
					IRIDSLSAQLSQLQK	1752,81	28
					MQQQLDEYQELLDIK	1699,91	35
					LALDMEIHAYR.	1909,98	31
					SSFSQHAR	1347,62	37
					VAVEEVDEEGKFVR	919,39	54
					QNGDDPLLTYR	1605,77	25

ASASGSGAQVGGPISSGSSAS SVTVTR	**1274,61**	**7**
SVGGSGGGSFGDNLVTR	**2365,1**	**25**
TQSPQNCSIM	**1566,75**	**0**
	1165,5	**0**

7. APPENDIX II

Publications

1) G. Chiappetta*, **C. Corbo*,** A. Palmese*, F. Galli, M. Piroddi, G. Marino, A. Amoresano. «Quantitative identification of protein nitration sites». Proteomics, 2009 Mar; 9(6):1524-37.
*Co-authors

2) **Claudia Corbo**, Stefania Orrù, Marica Gemei, Rosa Di Noto, Peppino Mirabelli, Esther Imperlini, Margherita Ruoppolo, Luigi Del Vecchio, Francesco Salvatore. "Protein cross-talk in CD133+ colon cancer cells indicates activation of the Wnt pathway and up-regulation of SRp20 that is potentially involved in tumorigenicity" Proteomics, Under Review

Oral communication

Corbo C., Orrù S., Gemei M., Di Noto R., Mirabelli P., Imperlini E., Ruoppolo M., Del Vecchio L., Salvatore F. "The protein expression pattern of CD133+ colon cancer cells indicates activation of the Wnt pathway and potential alteration of splicing mechanisms". 2nd MS-J-DAY, Naples, Italy, 12th December 2011

Proceedings

1) G. Chiappetta, **C. Corbo,** G. Marino, A. Amoresano " A novel method to selectively detect, identify and quantify post-translational modifications by MS/MS/MS fragmentation". ItPA II Congresso Nazionale, Aci Trezza, Catania, Italy, 26-29 June 2007.

2) **C. Corbo**, G. Chiappetta, A. Amoresano, G. Marino «Quantitative analysis of protein nitration sites by using iTRAQ». XX Congresso Nazionale di Chimica Analitica, San Martino al Cimino, Viterbo, Italy, 16-20 September 2007.

3) G. Chiappetta, **C. Corbo**, A. Palmese, G. Marino, A. Amoresano «New applications of iTRAQ in protein chemistry and proteomics». ItPA III Congresso Nazionale, Selva di Fasano, Brindisi, Italy, 11-14 June 2008.

4) **Claudia Corbo**, Luigi Del Vecchio, Rosa Di Noto, Marica Gemei, Esther Imperlini, Peppino Mirabelli, Stefania Orrù, Margherita Ruoppolo, Francesco Salvatore «Characterization of proteins involved in intracellular pathways distinctive for colon cancer stem cells».5th SIBBM Seminar, Frontiers in molecular biology, Napoli, Italy, 4-6 June 2009.

5) A. Palmese G. Chiappetta, **C. Corbo,** G. Marino, A. Amoresano «Expanding the role of iTRAQ chemistry : quantitative analysis of post-translational modifications». ItPA IV Congresso Nazionale, Milano Italy, 22-25 June 2009.

6) **Claudia Corbo,** Luigi Del Vecchio, Rosa Di Noto, Marica Gemei, Esther Imperlini, Peppino Mirabelli, Stefania Orrù, Margherita Ruoppolo, Francesco Salvatore «Characterization of proteins involved in cellular pathways distinctive for colon cancer stem cells» 34th FEBS Congress Prague, Czech Republic, 4-9 July 2009.

7) **Claudia Corbo,** Luigi Del Vecchio, Rosa Di Noto, Marica Gemei, Esther Imperlini, Peppino Mirabelli, Stefania Orrù, Margherita Ruoppolo, Francesco Salvatore «Characterization of proteins involved in cellular pathways distinctive for colon cancer stem cells» 22nd IGB Meeting, Capri, Italy 10-13 October, 2009

8) **Claudia Corbo,** Luigi Del Vecchio, Rosa Di Noto, Marica Gemei, Esther Imperlini, Peppino Mirabelli, Stefania Orrù, Margherita Ruoppolo, Francesco Salvatore «Characterization of proteins involved in cellular

pathways distinctive for colon cancer stem cells» 35th FEBS Congress Gothenburg, Sweden, June 26-July 1, 2010.

9) **Corbo C.,** Orrù S., Gemei M., Di Noto R., Mirabelli P., Ruoppolo M., Del Vecchio L., Salvatore F. “Il pattern di espressione proteica di cellule CD133+ di cancro del colon indica attivazione del pathway di WNT e probabile alterazione del meccanismo di splicing”. Giornate Scientifiche Polo delle Scienze 2010, Facoltà di Medicina e Chirurgia, Napoli, Italy, 23-26 November, 2010.

10) **Corbo C.,** Orrù S., Gemei M., Di Noto R., Mirabelli P., Ruoppolo M., Del Vecchio L., Salvatore F. “The protein expression pattern of CD133+ colon cancer cells indicates activation of the Wnt pathway and potential alteration of splicing mechanisms”. 36th FEBS Congress Turin, Italy, 25-30 June , 2011.

11) **Corbo C.,** Orrù S., Gemei M., Di Noto R., Mirabelli P., Ruoppolo M., Del Vecchio L., Salvatore F. “The protein expression pattern of CD133+ colon cancer cells indicates activation of the Wnt pathway and potential alteration of splicing mechanisms”. 3th EMBO Meeting Wien, Austria, 10-13 September , 2011

12) **Corbo C.,** Orrù S., Gemei M., Di Noto R., Mirabelli P., Ruoppolo M., Del Vecchio L., Salvatore F. “The protein expression pattern of CD133+ colon cancer cells indicates activation of the Wnt pathway and potential alteration of splicing mechanisms”. FEBS Advanced Lecture Course on Translational Cancer Research, Albufeira, Portugal, September 27-October 4, 2011

13) **Corbo C.,** Orrù S., Gemei M., Di Noto R., Mirabelli P., Ruoppolo M., Del Vecchio L., Salvatore F. “The protein expression pattern of CD133+ colon cancer cells indicates activation of the Wnt pathway and potential alteration of splicing mechanisms”. 2nd PhD Symposium “Life in motion” Barcelona, Spain, 17-18 November 2011

Courses

March 2010

2D-DIGE Workshop (GE Healthcare), theoretical and practical course on differential proteomics at National Institute for Research and Treatment of Tumors “G. Pascale”, Naples, Italy

August 2010

Practical EMBO course titled: “Post-translational modifications of proteins: from discovery to functional analysis”. Uppsala, Sweden

October 2011

Awarded with a FEBS Youth Travel Fund (YTF) Fellowship to attend the **FEBS Advanced Lecture Course on Translational Cancer Research**, 27th September- 4th October 2011, Albufeira, Portugal

Visiting appointment

May-August 2010

Short term fellowship during the PhD program under the supervision of Prof. M. Clench and Dr. S. Francese at the BMRC Institute (Biomedical Research Centre), Sheffield Hallam University, United Kingdom. Training on **MALDI IMAGING MASS SPECTROMETRY**

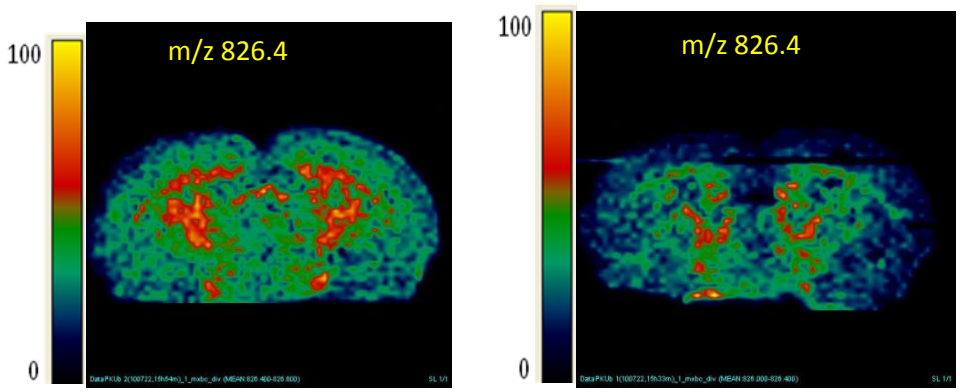
Direct analysis of tissues of biological and clinical interest using MALDI MS has been shown to be successful for the study of the mid- to low molecular weight proteome. Because this technology analyzes intact tissue, avoiding homogenization and separation steps, the spatial distribution of molecules within the tissue is preserved. The process is relatively simple in that a matrix (typically a small aromatic organic molecule dissolved in an organic solvent) is deposited on top of a tissue

section followed by irradiation with a laser (e.g., nitrogen, 337 nm). Molecules are subsequently desorbed and ionized.

MALDI is often coupled with TOF mass analyzers that allows a unlimited mass range is with analytes > 200 kDa capable of being measured.

Image analysis of discrete molecules in tissue can be acquired by using MALDI MS to determine their spatial localization with a lateral resolution of 10–100 µm. A thin (10 µm) tissue section is collected on a target plate, and matrix is applied over the surface of the tissue by a robotic liquid dispensing device followed by desorption, ionization, and separation processes. Spectra are recorded in a systematic fashion over the tissue by moving the sample stage underneath a fixed laser beam. Thus, a spot array over the entire sample then constitutes the image dataset analogous to pixels in a digital photograph. Each laser-irradiated spot (pixel) gives rise to a mass spectrum that is correlated to discrete a X,Y coordinate location on the tissue. Thus each spot or pixel contains a dataset having thousands of channels (m/z values) with each channel having its own brightness (intensity). The intensity of each m/z value can be expressed over the array of pixels as a 2D ion density map. Commercial or custom software can be used to generate images depicting the localization and relative intensities of hundreds of ions in a single acquisition from a tissue section. During the time spent in Dr Francese's laboratory at Sheffield Hallam University, UK, the above described experimental workflow was applied to coronal sections of mice brains. It was acquired good experience in this innovative field of

research. The following images show the distribution of phospholipid at m/z 826.4 in two different slices of a PKU brain as an example of the image obtained with MALDI imaging analysis followed by software elaboration of data.



MALDI MS image of phospholipid at m/z 826.4 in two different slices of a PKU brain

ACKNOWLEDGMENTS

Special thanks to the European School of Molecular Medicine (SEMM) for the four-years Ph.D. fellowship. In particular, I wish to express my gratitude to my Ph.D. tutor, Professor Francesco Salvatore for his continuous encouragement and invaluable suggestions during this work. I would also like to express my gratitude to my Ph.D. Internal Supervisor Prof. Margherita Ruoppolo and to Prof. Stefania Orrù for their scientific guidance and support.

I thank Prof. Del Vecchio who is, together with Prof. Salvatore, the designer of this ambitious project and I thank Dr. Esther Imperlini, Prof. Rosa Di Noto, Dr. Marica Gemei and Dr. Peppino Mirabelli for our scientific collaboration during these years.

I wish to thank my external supervisor, Dr. Simona Francese, for her availability during my stay in her laboratory in Sheffield and for her important suggestions during the editing of this thesis.

A very special thanks is for my lab mates Esther, Irene, Marianna, Sara and Simona for their daily presence in my lab life. They were always available to support and to encourage me when necessary, nevertheless we spent together very funny and challenging moments.

Finally, I want to strongly thank my family, particularly my parents for their loving never-ending support.