UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Scienze Chirurgiche, Oncologiche e Gastroenterologiche

SCUOLA DI DOTTORATO DI RICERCA IN

ONCOLOGIA E ONCOLOGIA CHIRURGICA

XXV CICLO

Analysis of blood-based markers as predicting tools of pathologic tumour response in rectal cancer patients receiving neo-adjuvant chemoradiotherapy

Direttore della Scuola: Ch.ma Prof.ssa Paola Zanovello

Supervisore: Ch.mo Prof. Marco Agostini

Dottoranda Maria Vittoria Enzo

SUMMARY

ABSTRACT

Neo-adjuvant chemo-radio therapy (pCRT) has been accepted as a standard care in the treatment of patients with locally advanced rectal cancer. The multimodality treatment has been established to improve tumour downstaging, pathological complete response, and local disease control. However, the response of individual tumors to the treament is not uniform and ranges from complete response to complete resistance. The discovery of new molecular markers that predict the tumour response is surely of wide interest for personalizing the therapy and reducing time, costs and side effects in the patients with resistant tumours. Many potential biomarkers have been evaluated in order to predict the response to pCRT, and to implement targeted therapeutics. However, single-marker or multi-markers analyses, based on pre-treatment tissue biopsies, often obtained conflicting results demonstrating the heterogeneity of the individual tumour response. Moreover, the prediction of histopathological response to neo-adjuvant treatment is complicated by the interaction and the involvement of the microenvironment in modulating the sensitivity to pCRT. In this study, we developed blood-based methods of biomarker analysis in order to evaluate the histopathological response to treatment in a broad contest that can take into account not only the signalling pathway of tumour cells but also the microenvironment as a part of a unique system. Indeed, the non-invasive nature and the dynamism for which different molecules could be detected according with physiological and pathological states has given us the possibility to monitor the response along the administration of the treatment. In particular, we focused on two different kind of circulating molecules: the cell-free DNA (cfDNA) and the circulating low molecular weight (LMW) peptides. In particular, we investigated the presence, the quantity of cfDNA and its integrity (cfDNA integrity = non apototic cfDNA / total cfDNA) along the chemo-radio treatment. For this purpose we measured the cfDNA concentration and cfDNA integrity in a prospective study of locally advanced rectal cancer patients plasma collected before the pCRT, after two weeks from initiation of the pCRT and after the pCRT. We evaluated the association of these markers with the histological response to the chemo-radio therapy, demonstrating a different kinetic of cfDNA integrity in association with the tumour response.

Then we studied the LMW peptidome circulating in plasma, in order to find evidences of possible differences in the peptide profile that could reflect the tumour response. To overcome the technical difficulties in harvesting LMW species, we have employed the mesoporous chip-based technology, developed by the

Nanomedicine Department of The Methodist Hospital Research Institute in Houston, Texas. This mesoporous device, in combination with matrix-assisted laser desorption/ionization - time of flight mass spectrometry (MALDI-TOF MS), allows the isolation and the detection of small peptides from the large proteins. We analyzed plasma of rectal cancer patients, with positive or negative response to the therapy, at the same time points as the cfDNA: before, during, after the chemo-radio therapy. Multivariate analyses of the LMW peptide profile at different time points identified combinations of peptides that revealed high discriminating capacity of the different tumour responses. In particular, before the pCRT, a pattern of five ionic species showed a sensitivity and a specificity of 80% and after the pCRT, a pattern of other five specific ionic species showed a sensitivity of 80% and a specificity of 85% to cluster patients on the basis of histopathologic response to pCRT. Moreover the identification of the amino acids sequences of the response-specific ionic species revealed the presence of protein fragments that could be directly or indirectly valuable for further investigation on the resistance mechanisms of the rectal tumour to the neo-adjuvant chemo-radio therapy.

RIASSUNTO

La radiochemioterapia neoadiuvante (pCRT) è un protocollo standard accettato per il trattamento di pazienti con cancro rettale localmente avanzato. Il trattamento preoperatorio multimodale è stato introdotto per la riduzione dello stadio del tumore, per l'aumento della risposta completa patologica e per il controllo locale della malattia. Tuttavia la risposta patologica al trattamento non è uniforme e varia da una risposta completa alla resistenza totale. La scoperta di nuovi marcatori molecolari in grado di predire la risposta del tumore è sicuramente di grande interesse al fine di personalizzare la terapia, riducendo così i tempi, i costi e gli effetti collaterali nei pazienti con tumori resistenti. Molti potenziali biomarcatori sono stati valutati con l'obiettivo di prevedere la risposta alla pCRT, e di attuare terapie mirate. Finora molti studi su singolo o multi-marcatore sono stati eseguiti prevalentemente su biopsie di tessuto pre-trattamento. I risultati ottenuti, tuttavia, erano spesso contrastanti dimostrando l'eterogeneità individuale della risposta tumorale al trattamento. Inoltre, la predizione della risposta istopatologica alla pCRT è complicata dall'interazione e dal coinvolgimento del microambiente che può modulare la sensibilità del tumore al trattamento. In questo studio, abbiamo sviluppato metodi di analisi di biomarcatori su sangue, al fine di valutare la risposta del tumore al trattamento in un contesto più ampio, che rende conto non solo dell'ambiente strettamente tumorale, ma che prende in considerazione anche il microambiente come parte di un sistema unico. Infatti, la natura non invasiva del materiale biologico analizzato e il dinamismo per cui molecole differenti possono essere rilevate secondo lo stato fisiologico e patologico dell'organismo, ci hanno permesso di monitorare la risposta lungo il tempo di somministrazione del trattamento. In particolare, ci siamo concentrati su due diversi tipi di molecole circolanti: il DNA libero da cellule (cfDNA) e i peptidi a basso peso molecolare (Low Molecular Weight, LMW). In particolare, abbiamo studiato la presenza, la quantità e l'integrità del cfDNA durante il trattamento radio-chemioterapico. A questo scopo abbiamo misurato la concentrazione e l'integrità del cfDNA (cfDNA integrity=cfDNA apoptotico/cfDNA totale) in uno studio prospettico di plasma di pazienti con carcinoma rettale localmente avanzato, raccolto prima della pCRT, dopo due settimane dall'inizio del trattamento e dopo la pCRT. Abbiamo valutato l'associazione di questi marcatori con la risposta istologica alla chemio-radio terapia, dimostrando la presenza di diversa cinetica nell'integrità del cfDNA, in associazione con la risposta tumorale.

Nel plasma, abbiamo quindi studiato il peptidoma circolante a basso peso molecolare, al fine di trovare potenziali differenze nel profilo peptidico che potessero riflettere la risposta tumorale. Per superare le difficoltà tecniche nella rilevazione dei peptidi circolanti a basso peso molecolare, abbiamo utilizzato una strategia basata sull'esclusione dimensionale di un chip di silice mesoporosa (MSC), sviluppato dal Dipartimento Nanomedicina del The Methodist Hospital Research Institute di Houston, Texas. Questo dispositivo mesoporoso, in combinazione con l'utilizzo dello spettrometro di massa MALDI-TOF MS (Matrix-Assisted Laser Desorption/Ionization-Time Of Flight Mass Spectrometry), consente l'efficiente rimozione di grandi proteine e l'isolamento del peptidoma circolante da campioni di fluidi corporei. Abbiamo analizzato il plasma di pazienti con cancro rettale, prelevato in diversi tempi (prima, durante, dopo la chemio-radio terapia) e stratificati secondo la risposta positiva o negativa alla pCRT. L'analisi multivariata del profilo peptidico nei diversi tempi di analisi ha identificato combinazioni di peptidi che evidenziavano un'elevata capacità discriminante della risposta tumorale. In particolare, il modello di regressione logistica ha evidenziato, prima della pCRT, una combinazione di cinque specie ioniche capace di identificare i pazienti che non rispondono al trattamento, con una sensibilità e una specificità del 80%; mentre la stessa analisi con i campioni raccolti dopo la pCRT, ha identificato un'altra combinazione di cinque specie ioniche che evidenziano una sensibilità del 80% e una specificità del 85%. Inoltre, l'identificazione delle sequenze amminoacidiche di alcune tra le specie ioniche discriminanti la risposta alla pCRT, hanno rivelato la presenza di frammenti proteici che possono essere direttamente o indirettamente utili per ulteriori indagini sui meccanismi di resistenza del tumore rettale alla radio-chemio terapia neo-adiuvante.

INTRODUCTION

1 RECTAL CANCER

Rectal and colon tumours are commonly referred as one disease named Colon-Rectal Cancer (CRC). This definition originates from the common anatomic localization; indeed rectum is the continuation of the part of the large bowel of colon and it shows similar morphology and configuration. Furthermore colon and rectal tumours share many features; they have similar aetiology and type of precancerous lesions. However, the issue whether colon and rectal tumours should be considered as a single or two distinct entities is still debated because they showed some different features at both clinical management and molecular level. Differences have been found at the age of tumour formation, the gender of the patients, the clinical behaviour, and the therapeutic strategies (*Li M. et al., 2007)*. In this contest rectal cancer will be presented as CRC for the features that share with colon cancer, highlighting on the distinctive characteristics of rectal cancer.

1.1 *EPIDEMIOLOGY AND AETIOLOGY*

CRC incidence and mortality rates vary markedly around the world. Globally, CRC is the third most commonly diagnosed cancer in males and the second in females, with over 1.2 million new cases and 608,700 deaths estimated to have occurred in 2008. CRC presents a serious public healthcare issue for the population of Europe, as it is the most common newly diagnosed cancer and the second most common cause of cancer deaths in Europe. It has been estimated that in 2006, 412.000 people were diagnosed with CRC in Europe, and more than 200.000 of them die of the disease *(Ferlay J. et al., 2006; Zavoral M. et al., 2009).* In Italy, CRC is the second more common tumour in women (12%, 145.027) preceded only by breast cancer, while in men is the third type of neoplasm (15%, 152.660) preceded by prostate cancer and bladder cancer (*Epidemiologia 2010 suppl.2*). Colorectal tumour includes different anatomic part of the digestive tract. In particular, adeno-carcinoma of the rectum contributes to 28% of the total colon and rectal cancer burden, resulting the seventh tumour type most commonly diagnosed in both females and males (*Julien L.A. et al., 2010*). Globally, the incidence of CRC varies over 10-fold. The highest incidence rates are in Australia and New Zealand, Europe and North America, and the lowest rates are found in Africa and South-Central Asia (*Jemal A.*

et al., 2011). Incidence rates in most western countries have been stable or increased slightly during the last 15 years. In contrast, CRC incidence rates have rapidly increased in several areas historically at low risk, including Spain, and a number of countries within Eastern Asia and Eastern Europe.

Although the incidence of CRC is generally higher in both the gender, rectal tumours show a higher prevalence in males. The risk of the rectal disease is around 17,2% in males and 9,3 % in females. CRC mortality is about 25% higher in men than in women, and both colonic adenomas and CRCs appear to have a more proximal distribution in women, particularly in postmenopausal women. Incidence of CRC increases with the age, as the majority of cases are diagnosed in patients more than 50 years of age. Large bowel cancer is uncommon before the age of 40; the incidence begins to increase significantly between the ages of 40 and 50, and age-specific incidence rates increase in each succeeding decade thereafter. Lifetime incidence of CRC in patients at average risk is about 5%, with 90% of cases occurring after age 50. Interestingly, rectal cancer is rare under the age of 45. About two-third of patients with rectal cancer are over 65 years of age.

Death rates from CRC have declined progressively since the mid 1980s in the United States and in many other western countries. This improvement in outcome can be attributed, at least in part, to detection of CRC at early stages and to use of more effective treatments. The good prognosis of patients with CRC decreased from 90% in locally confined tumours, to 63% in the infiltrating and non-metastatic tumours, until 5% in the case of metastatic tumours (*Li F.Y. et al., 2009*).

1.1.1 Risk factors

Colorectal cancer presents a family's or hereditary derivation in 20% of cases whereas the remaining 80% occurs sporadically. Although cancer of the rectum and colon follow a common scheme of neoplastic progression, they seem to have distinct characteristics and, in particular, the impact of family or hereditary derivation is very low in rectal cancer. Studies on migrants suggest that colorectal cancer risk is determined largely by environmental exposure. Rates in migrants moving from low-risk to high-risk countries tend to increase to the rates of the host countries within the migrating generation itself *(Zampino M.L., 2004).*

Several risk factors have been identified increasing person's chance in developing colorectal cancer. Diet is the most important exogenous factor identified up to now in the aetiology of colorectal cancer. It has been estimated that 70% of colorectal cancers could be prevented by nutritional intervention. Various promoting and

protective factors have been identified. There are many evidences that diets rich in vegetables protect against this type of cancer; vegetables contain a large quantity of micronutrients (carotenoids, folate, ascorbate), bioactive compound (phenols, flavonoids, isothiocyanates) with anti-carcinogenic properties. Diets with high fiber content may be cause lower exposure of the rectal epithelium to carcinogens derived from the environment, as the dietary fibers favour transit through the bowel. Furthermore the role of fibers as a protective factor for CRC was confirmed in European large cohort (*Vainio H. et al., 2006*). On the other hand, several epidemiological studies reported the negative effect of the red meat intake. The mechanism by which the red meat and processed meat increase the colorectal cancer risk include the effect of the fat on bile acid production that increase the possibility of cellular membrane damage. Furthermore, carcinogens can be produced when the meat is processed and cooked. Overcooking the meats or cooking with high temperatures contributes to formation of heterocyclic amines, strong oxidants that might react with DNA by inducing the accumulation of genomic instability (*Burnett-Hartman A.N. et al., 2011*).

The role of other factors, such as the consumption of alcohol and sugar is more controversial. Alcohol intake seems to be associated more with an increased risk in rectal cancer than in colon cancer for which it has a modest effect *(Valentini V. et al., 2008).* Some non-dietary factors like smoking tobacco, radiation, chronic inflammation, increase the risk of colorectal cancer (*Wang J. et al., 2012*). For both incidence and mortality, the association was stronger for rectal cancer than colon cancer. Cigarette smoking, both active and passive, has been identified as risk factors for rectal cancer in case-control and cohort studies. Recent smokers and those that have smoked many packs/years of cigarettes have been estimated to have from a 50% increment to a doubling of risk of rectal cancer compared to non-smokers *(Hooker C.M. et al., 2008).* The association between cigarette smoking and rectal cancer may be impacted by the ability to detoxify polycyclic aromatic hydrocarbons (PAH) generated from cigarette smoking *(Curtin K. et al., 2009).* Increasing evidence suggests that diabetes mellitus is associated with an elevated risk of CRC. One possible explanation linking diabetes to CRC is hyperinsulinemia, because insulin is an important growth factor for colonic mucosal cells and stimulates colonic tumour cells (*Larsson SC et al., 2005).*

1.1.2 Protective factors

A large number of factors have been reported by at least some studies to be associated with a decreased risk of CRC. These include regular physical activity, a

variety of dietary factors, the regular use of aspirin or non-steroidal anti-inflammatory drugs (NSAIDs), and hormone replacement therapy in postmenopausal women. Evidences suggest that aspirin and other non-steroidal anti-inflammatory drugs (NSAIDs) protect against the development of colonic adenomas and cancer. Regular use of aspirin and other NSAIDs are associated with a 20 to 40 percent reduction in the risk of colonic adenomas and colorectal cancer in individuals at average risk. How the aspirin produce this protective factor is not well understood yet. Proposed explanations include the increased apoptosis and impairment of tumour cell growth by inhibition of cyclooxygenase-2 (*Wang J. et al., 2012*).

1.2 *RECTAL ANATOMY*

The large bowel is divided into two anatomical regions: proximal "right" region and distal "left" region. The rectum is part of the distal region that includes: splenic flexure, descending colon, sigmoid colon and rectum. This division is supported by the distinct embryonic ontogenesis. The proximal large intestine develops from the embryonic midgut and it is supplied by the superior mesenteric artery; the distal large intestine forms from the embryonic hindgut and it is supplied by the inferior mesenteric artery. Interestingly, rectum originates from the cloaca. The rectum presents three lateral curvatures: the upper and the lower are convex to the right, the middle convex to the left (Figure 1). The adult rectum is 15-20 cm in length and is divided into equal thirds: the upper third, which is mobile and has a peritoneal coat; the middle third which is the widest part of the rectum and where the peritoneum covers only the anterior and part of the lateral surfaces; and the lowest third, which lies within the muscular floor of the pelvis and has important relations to fascial layers. These fascial layers are surgically important, as they are a barrier to malignant penetration and valuable guides to operation. The artery supplies are divided in superior, middle and inferior rectal arteries. The arteries and their accompanying lymphatic are kept to the back of the rectum by dense connective tissue (mesorectum) that decreases up to disappear in the lowest third. The venous drainage includes the superior haemorrhoidal veins draining the upper half of the anal canal above the dentate line pass upwards to become the rectal veins. It forms part of the portal venous system and ultimately drains into the splenic vein. The rectum presents more abundant diffuse lymphatic tissue and lymph nodules. Rectum lymphatic drainage occurs in the mesorectum, to the inferior mesenteric artery, lateral to the internal iliac artery and down to the inguinal area.

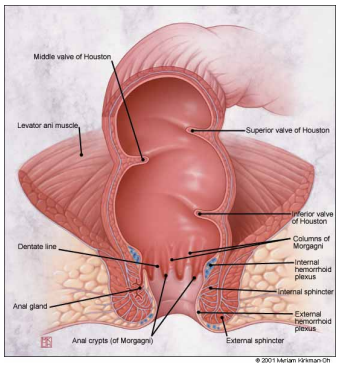


Figure 1. Rectum anatomy

The rectum is composed of an innermost layer of mucosa that lies over the submucosa, and two continuous sheaths of muscle, the inner circular and outer longitudinal layer. Rectum histology resembles that of the colon, but it is distinguished by transverse rectal folds in its submucosa and the absence of tenia coli in its muscularis externa. The mucosa of the rectum largely resembles the colon, with a simple columnar epithelium shaped into straight tubular crypts. Rectum presents crypts that are longer than the crypts in small intestine. The crypts are more widely spaced and a higher proportion of goblet cells are interspersed among the adsorptive cells. Neutral mucopolysaccharide is predominant in the ascending colon, whereas in the rectum acidic mucin is predominant or exclusive. The mucus facilitates the passage of the increasingly solid colonic contents, and covers bacteria and particulate matter. Rectum shows an unusually high concentration of endocrine cells, compared with the other segments *(Bailey & Love's Short Practice of Surgery, 25th edition, 2012).*

1.3 *COLORECTAL CANCER PROGRESSION*

The development of rectal cancer is common, for many aspects, to that of the colon cancer. Morphological data, consolidated with clinical, epidemiological and molecular observations, suggested the hypothesis that its occurrence is the result of a long series of events leading to the definition of the “adenoma-carcinoma sequence” as model of cancer progression (Figure 2). This transformation occurs slowly and gradually in about 10-15 years and progresses through both histological changes and genetic alterations, highlighting a multistep progression at both molecular and pathophysiological level (*Vogelstein B. et al., 2000; Fodde R, 2001*).

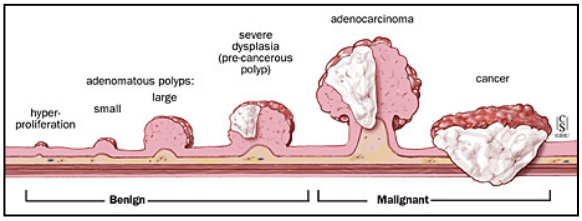


Figure 2. Colorectal cancer morphological progression

Intestinal epithelium shows numerous invaginations called crypts, each of which contains several thousands of differentiated cells and a small number of active stem cells located at the bottom of the crypts (*Kauh J. et al., 2004*). Cells constantly migrate from the basal level to the intestinal lumen in a process 3-6 days long. During the migration the cells grow up and differentiate, losing their ability to divide. In the intestinal lumen they are eliminated by exfoliation that involves apoptotic mechanism of death. In this way, there is a complete and continuous renewal of the mucosa epithelium, while death and growth cell rates are equivalent (*Fodde R., 2001*). It has been shown that in tumours, there exist a subset of stem cell or cancer stem cells that drive tumour growth and possess similar characteristics of proliferation, self-renewal, clonogenicity and multi-potentiality as stem cells show in normal organs *(Buczacki S. et al., 2011*). One hypothesis is that colorectal cancer stem cells could originate from gastrointestinal stem cells, which represent the natural target of tumourigenic mutations due to their long life and their capacity for self-renewal *(Ricci-Vitiani L. et al., 2008).* Alternatively, differentiated cells could re-acquire stem-like characteristics. These two possibilities are not mutually exclusive, and it is not certain in which tumours or specific circumstances either occur *(Buczacki S. et al., 2011*).

The first recognizable histological neoplastic lesion consists in the so-called "aberrant crypt foci" (ACF) that is caused by hyperproliferation of epithelial cells. ACFs are characterized by impaired luminary opening and thickened epithelium. The clonal expansion could lead to the formation of a benign tumour, known as "adenoma" or "polyp" that are limited and enclosed in a fibrotic membrane. Adenoma epithelium is characterized by abnormal morphology including alteration of cell differentiation, presence of hyper-chromatic nuclei and loss of cell polarity. The rectum, along with the sigmoid colon, is the most frequent site of polyps in the gastrointestinal tracts. The rectum shares substantially the same spectrum of polyps as the colon. They can be either single or multiple. Adenoma could be classified according to different histopathological features like the dimension, the histological

type, and the degree of dysplasia. Colon adenomas can be subdivided in a large group of non-progressive lesions (~95%), and a small group of progressive lesions (~5%). It is not well understood which and how many adenomas might evolve in adeno-carcinoma but its features could favour the progression. The progression is marked by a gradual increase in dimension and in aggressiveness. In the first phase adeno-carcinoma is limited to the epithelium (carcinoma in situ); as it evolves, it acquires the ability to invade the tissue underneath the basement membrane (invasive carcinoma). The cancer may eventually acquire the ability to spread in the blood and lymph vessels and lymph node regions and reach distant organs (metastasis) *(Kinzler K.W. et al., 1996; Fearon E.R. et al., 2011*).

For colon cancers, the progression pattern is typically characterized by distal metastasis, usually at liver site whereas rectal cancers usually have a pattern of local recurrence. This is potentially explained by the fact that superior mesenteric vein drains the right colon whereas neither left colon nor rectal vasculature directly drains to liver. Local spread occurs circumferentially rather than in a longitudinal direction. Around 2 years is required for complete encirclement. After the muscular coat has been penetrated, the growth spreads into the surrounding mesorectum. If penetration occurs anteriorly the sex organs are invaded (prostate, seminal vesicles, bladder in male, uterus and vagina in female). Spread via the venous system occurs late. The principal sites for blood-borne metastases are liver, lungs, adrenals *(Bailey & Love's Short Practice of Surgery, 25th edition, 2012)*.

1.3.1 Molecular pathways involved in cancer progression

Cancer progression requires the accumulation of genetic and epigenetic alterations that affect signalling pathways that regulate hallmark behaviours of cancer and create clonal growth advantage. CRC phenotypically appear to be a homogeneous disease while it is a heterogeneous disease at the molecular level, suggesting the existence of multiple CRC subtypes with distinct biological and clinical properties.

Three are the most pathways involved in tumorigenesis *(Worthley D.L. et al., 2007; Beggs A.D, et al., 2008)*. In more than 50% of the cases, CRCs are consequences of events that cause chromosome instability (CIN), associated with amplifications and deletions of chromosomal regions. Chromosome instability occurs mostly through the accumulation of numerical or structural chromosomal abnormalities, aneuploidy, and allelic deletions. In about 15% of sporadic colon cancers and tumours from patients with hereditary non-polyposis colorectal cancer (HNPCC), has been

observed nucleotide instability (MIN). MIN occurs by inactivation of genes involved in Mismatch Repair System (MMR) that has the function of repairing DNA replication errors. These events lead to insertions, deletions or substitutions of a few nucleotides in DNA strands. CpG island methylator phenotype (CIMP) contributes to up to 20% of sporadic CRCs. CpG island methylation phenotype, is characterized by the hypermethylation of genes that play a role in carcinogenesis. Aberrant methylation of CpG-rich regions (CpG islands) at the promotor region or at the 5' region of genes, leads to transcriptional silencing and therefore, inactivation of the gene *(Jass J.R., 2007)*. Most sporadic microsatellite unstable colon tumours are CIMP positive, whereas CIMP is uncommon in HNPCC syndrome-associated cancer that exhibits MIN. CIMP is significantly more frequent in tumours of the proximal colon (30-40 %) than in distal colon or rectal cancers (3-12%).

Rectal cancer is mainly associated with chromosomal instability that frequently exhibits a temporal genetic alteration sequence correlates with the histological cancer progression. Development of the normal gut and subsequently maintenance of colon tissue homeostasis is regulated by the balance between several key signal transduction pathways, in particular the Wnt, Notch, Hedgehog, and BMP pathways. Of these, the Wnt signal transduction pathway is most relevant to maintain the stem cell population. Constitutive activation of Wnt signalling stimulates proliferation and is a common and early event in colon tumour development, occurring in about 80% of colon adenomas *(Jimenez C. et al., 2010)*. The initial event, which is first observed in dysplastic cells present in the foci of aberrant crypts, is more frequently represented by the biallelic inactivation of the *APC* gene (*Adenomatous Polyposis Coli)*. It has been considered that APC inactivation shows the highest malignant potential in cancer progression, as it is altered in 70-80% of cases of sporadic adenomas and carcinomas. *APC* is a tumour suppressor gene on chromosome 5q and it has a broad spectrum of functions, ranging from control of the Wnt signal transduction pathway, to cell adhesion, migration, signal transduction, apoptosis, microtubule assembly and chromosomal segregation at mitosis. *APC* usually lead to the degradation of citoplasmatic β-catenin, inhibiting the activation of Wnt pathway that is usually activated during embryogenesis and for the maintenance of tissue specific compartments of stem cells. *APC* inactivation interferes with the inhibition of Wnt pathway, increasing β-catenin accumulation and translocation to the nucleus where it promotes proliferation through expression of oncogenes like Myc and Ciclin D1 *(Fearon E.R., 2011)*.

The loss of function of *APC* may also interfere with the normal regulation of mitosis, contributing to chromosomal instability *(Worthley D.L. et Al., 2007)*.

Alterations in *APC* and Wnt pathway may represent a convenient mechanism in the progression of early lesions in the later stages as adenomas. Mutations or chromosomal deletions that determine the loss of other genes such as *DCC, SMAD2* and especially *KRAS* and *SMAD4* contribute to the adenoma growth and its progression to neoplasia. *KRAS* (*V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog*) encodes for a GTP-binding protein. Its alteration determines the loss of its GTPase activity and thus constitutive signalling through the downstream, (MAPK) pathway. At the molecular level, *KRAS* alteration induces genomic instability and/or by affecting G1 and G2/M cell-cycle transit times and apoptosis. Activating *KRAS* mutations are found in 35-42% of CRCs and in a similar percentage of advanced adenomas. *DCC* (*Deleted in Colorectal Carcinoma), SMAD2* and *SMAD4 (mothers against DPP homolog 2 and 4)* are all located at 18q21.1 and allelic loss at this site is found in up to 60% of CRCs. *DCC* encode a netrin-1 receptor it can induce apoptosis of epithelial cells when netrin-1 ligand is not bound. Furthermore, DCC is a caspase substrate, and mutation of the site at which caspase-3 cleaves DCC, completely suppresses the pro-apoptotic effect of DCC (*Ren X.R. et al. , 2004*). SMAD proteins are signal transducers and transcriptional modulators that mediate multiple signalling pathways. These proteins mediate the signal of the *Transforming growth factor (TGF-β),* and thus regulate multiple cellular processes, such as cell proliferation, apoptosis, and differentiation.

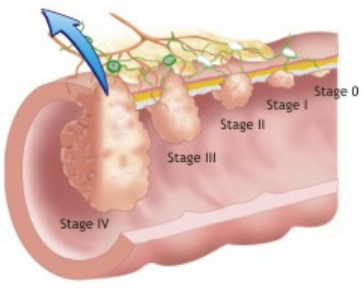
As described above, although colon and rectal cancer are classified as a single entity of study, morphological and clinical-pathological differences exist between these two types of tumour. On the basis of these evidences, studies have been made to highlight if there were molecular differences that could explain the different behaviour. So far the multistep model of rectal progression was confirmed but there are evidences that there may be alternative genetic events during tumour progression, events that occur at different frequency or at different temporal time. In a study of Kapitejan et al *(Kapitejan E. et al., 2001)* a greater presence of β-catenin in the nucleus of cells of rectal tumours than those of the colon has been highlighted; Frattini et al *(Frattini, M. et al., 2004)* showed a broader range of mutation types in APC in the colon than in the rectum. Mutations in APC have been found in 60% of colon tumours, compared to 80% for rectal cancer *(Worthley D.L. et al., 2007)*. On the contrary, there is a higher prevalence of KRAS-dependent pathway alteration in colon cancer compared to the rectum tumour. Pamplona et al *(Pamplona R.S. et al., 2011)* revealed only small gene expression differences between tumours arising in the colon or rectum, that are largely driven by the HOX family of genes. However other studies *(Li J.N. et al., 2012)* found different gene expression profiling and different protein expression, indicating that the situation could be more complex.

1.4 *TNM CLASSIFICATION*

Prognosis of rectal tumours is closely correlated to the degree of tumour extension and the accurate preoperative staging is therefore essential to determine the most appropriate treatment protocol. Current protocols use the *Tumor-Node-Metastasis* (TNM) classification system that allows the physician to establish the grade of tumour progression by studying three parameters:

- T: the size of the primary tumour
- N: involvement of regional lymph nodes; this parameter depends on the number of affected lymph nodes and on the distance of lymph nodes affected from the primary site of the tumour
- M: presence of distant metastases or disseminated tumour cells in organs other than the site of tumour origin.

TNM classification is a dual system that includes a clinical classification (pre-treatment) and a pathological (post-operative). The distinction between the two is crucial, as they are based on different types of examinations and they fulfil different purposes. In general, the classification of the clinical stage cTNM is used to evaluate the extent of the disease prior to the treatment, and it provides information for choosing the most appropriate treatment protocol. Modern imaging techniques (trans-rectal and endoscopic ultrasound and pelvic-rectal MRI), in addition to standard investigations (digital exploration, endoscopy, proctoscopy, trans-rectal scan and surgical exploration), allow physicians to more precisely determine tumour characteristics and prognostic factors in the preoperative setting. This knowledge has been used to improve cancer stage specific treatments. After the surgery, post-surgical pTNM staging help for prognostic evaluation and it is based on the integration of the data collected prior to treatment with those highlighted with pathological examination and surgical (*Rodel C. et al., 2012*) (Figure 3).



AJCC/TNM Staging		Tumor
Stage 0	Tis (carcinoma in-situ)	Superficially involves the mucosa. Has not grown beyond the mucosa
Stage I	T1N0 T2N0	Invades through mucosa Invades through submucosa
Stage II	T3N0 T4N0	Invades through muscle layers Invades nearby tissues or organs
Stage III	Any T, N1-N3	Lymph nodes involved
Stage IV	Any T, Any N, M1	Distant Spread

Figure 3. The colon cancer progression staging and the correspondent TNM classification.

1.5 *RECTAL CANCER THERAPY: SURGERY AND CHEMO-RADIO THERAPY*

Treatment decisions are mainly based on TNM classification and on the *Circumferential Resection Margin* (CRM) at time of diagnosis. Based on their TNM stage, rectal cancers can be treated with a different therapeutic approach *(Zampino M.G. et al., 2009)*. Standardized surgical techniques, specifically *Total Mesorectal Excision* (TME), introduced in the mid-1900s, have reduced local recurrence rates in rectal cancer from 39% to 10% and increased 5-year-survival rates to 71%. TME is one of the most influential factors in rectal cancer outcomes and is now considered the standard of care for clinical practice. T1 or T2, lymph node–negative rectal cancers have very low local recurrence rates and high cure rates after TME surgery *(Julien L.A. et al., 2010).* Although the surgical resection is, for the majority of the cases, curative the recurrence risk and the mortality increase when the tumour locally spread the tissue (TNM II) or the local lymph nodes (TNM III). The prognosis of rectal cancer is clearly related to the degree of penetration of the tumour through the bowel wall, the presence of nodal involvement. Additional relevant parameters are the grading, the angio- or venous invasion and the perineural invasion. In these cases, it has been observed that neo-adjuvant therapy for rectal cancer, implemented as a combination of chemotherapy and radiation therapy, has a beneficial effect on local control.

1.5.1 Neo-adjuvant chemo-radio therapy

Neo-adjuvant chemo-radio therapy has been accepted as a standard care in the treatment of patients with locally advanced rectal cancer. The advantages of neo-adjuvant therapy utilizing radiation are due to improve responsiveness of tissue that is not hypoxic by previous surgery. Theoretically, ionizing radiation is more effective in irradiation of virgin tissue due to the increased oxygen tension in this tissue. Accordingly, preoperative radiation and chemotherapy are more effective in producing tumour cell kill in the non-disturbed pre-surgical tumour bed compared to the hypoxic postsurgical bed. Several other advantages with neo-adjuvant therapy include less radiation-induced small bowel injury in the pelvis, which has not been fixed by previous surgery, and the ability to excise the irradiated rectal segment and perform an anastomosis to healthy, non-irradiated colon, resulting in improved post-operative function. In addition, studies have shown that chemo-radiation therapy, in the preoperative setting, decreased of acute toxic effects and long-term toxic effects compared to giving it post-operatively *(Sauer R. et al., 2004).* Modern regimens, including improved imaging, better chemotherapy, and more accurate and focused

radiation, have resulted in an increased frequency of tumour down-staging (refers to the change in stage between pre-therapy clinical evaluation and post-therapy clinical or pathologic evaluation), a higher likelihood of complete clinical and pathologic responses, and a decreased local recurrence rates in stage II and III rectal cancer. In addition, the utilization of neo-adjuvant therapy in the management of stage IV disease has shown potential for prolonged survival. The combination of the treatments has been established to improve tumour down-staging tumour down-sizing, survival and in some cases, sphincter preservation (*Julien L.A. et al., 2010).*

The major local effects of chemoradiotherapy, that cause cell damage, are elicited by irradiation, whereas the chemotherapy enhances these effect as a radiosensitizer, most often with a small direct effect on tumour cells damage. Ionizing radiation can have multiple effects on cells by disrupting chemical bonds in all the basic components of cells such as membranes, lipids, proteins and most importantly, in the genetic compartment of the cell, the DNA. The most well-studied target of radiation is DNA. Although all components can theoretically damaged by radiation, studies demonstrated that the main target is the nucleus. Radiation damage can cause base damage, DNA-protein crosslink, single-strand or double strand breaks, which occurs directly through ionization or indirectly through the action of chemical radicals that are produced during irradiation. Classic radiobiology classifies radiation cell interaction into 5 categories: repair, redistribution, reoxygenation, repopulation, and molecular regulation (*Harrington K et al., 2007*). Multiple pathways are involved mediating the cellular response to ionizing radiation and they culminate in either cell checkpoint arrest, which allow DNA damages to repair, or the activation of cell death pathways through necrosis or apoptosis. These cells accumulate in G2/M to be repaired. DNA damage starts to be repair within minutes by error-free processes (base excision repair, nucleotide excision repair, homologous recombination) and error-prone processes (nonhomologous end joining). Single strand breaks are generally repaired by nucleotide excision repair whereas double strand breaks are repaired by 2 mechanisms: nonhomologous end joining and homologous recombination. Cells in G2/M are more sensitive to radiation as double strand break will likely lead to aberrant chromosome formation during mitosis causing cell death. This sensitivity is used to improve the cell death through delivery of radiation in multiple fractions. The interval from one radiation fraction to another allows the resistant cells, not killed in response to the first fraction, to progress to G2/M before the next fraction. This process is called redistribution of cells into the cell-cycle phases and it is due to the synchronization and accumulation of cells in the G2/M phase that have to be repaired (*Selzer E. et al., 2012*). Sensitivity in this cell-cycle phase is due to the condensed chromatin that enhances the probability for

the photon to passes through more DNA and causes irreparable double-strand breaks. Fractionation of treatment also improve tumour cell killing through reoxygenation of cell that were hypoxic and therefore resistant before the radiation. Reoxygenation of cells improves the sensitivity to radiation because oxygen makes permanent the damage caused by free radicals *(Woodward W.A. et al., 2008)*. In addition to the DNA damage some studies emphasizes the role of multisystem tissue interaction in radiation response *(Barcellos-Hoff M.H. et al., 2006)*.

Chemotherapeutic agents are used in combination with the radiotherapy because it has been noticed an increment of the cell sensitivity to the radiotherapy. The chemotherapeutic drug commonly used in rectal cancer treatment is *5-fluorouracil (5-FU*), which induces the arrest of DNA synthesis with a consequent interruption of the duplication of the cell. The 5-FU action is directed to the transcriptional inhibition of the gene coding for *Thymidylate Synthetase* (*TS*): an enzyme that catalyzes the transfer of a methyl group to deoxyuridine monophosphate (dUMP) with the formation of thymidine monophosphate, a key metabolite in DNA synthesis. 5-FU is an analogue of uracil with a fluorine atom at the C-5 position in place of hydrogen. It rapidly enters the cell using the facilitated transport mechanism of uracil. Into the cell, 5-FU is converted to several active metabolites: fluorodeoxyuridine monophosphate (FdUMP), fluorodeoxyuridine triphosphate (FdUTP) and fluorouridine triphosphate (FUTP) (Figure 4). These active metabolites disrupt RNA synthesis and the action of *TS*. 5-FU is inactivated by an enzyme to dihydrofluorouracil (DHFU). More than 80% of administered 5-FU is normally catabolized primarily in the liver *(Longley D.B. et al., 2003*).

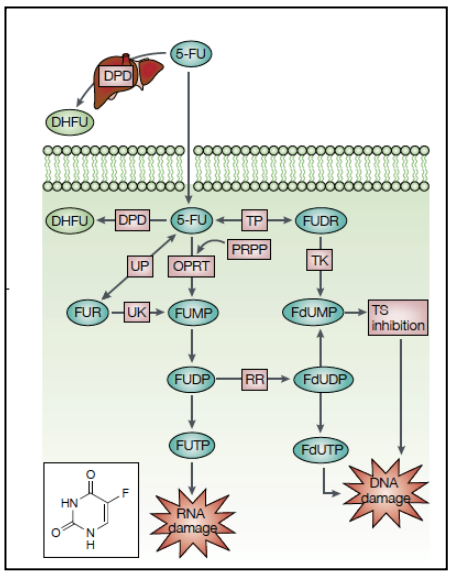


Figure 4. Fluorouracil metabolism. 5-Fluorouracil (5-FU is converted to three main active metabolites: fluorodeoxyuridine monophosphate (FdUMP), fluorodeoxyuridine triphosphate (FdUTP) and fluoro uridine triphosphate (FUTP) by a series of enzyme conversions and phosphorilations. The active metabolites can damage RNA molecules or can inhibit the Thymidylate synthetase (TS), inhibiting the DNA synthesis. Dihydropyrimidine dehydrogenase (DPD)-mediated conversion of 5-FU to dihydrofluorouracil (DHFU) is the rate-limiting step of 5-FU catabolism in normal and tumour cells. Up to 80% of administered 5-FU is broken down by DPD in the liver (Figure from *Longley D.B. et al., 2003*).

Current standard of treatment includes administration of ionizing radiation for 45-50.4 Gy in 25-28 fractions associated with chemotherapy (5-FU) administrated during radiation treatment.

Although there are numerous positive evidence, the application of the same protocol to tumours with similar characteristics in terms of TNM classification, has shown a broad histopathological response from one patient to another, ranging from complete regression to complete resistance highlighting the individual variability between the patients: in literature, the pathological complete response (pCR) are described with a variability ranging from 5 to 35% of cases (*Reerink O. et al., 2004).*

1.5.2 Tumour regression grade (TRG)

Tumour response to neo-adjuvant therapy appears to be an important prognostic factor and it is evaluated by determining the degree of tumour regression: Tumour Regression Grade (TRG). Tumour regression stratifies the response on the basis of the biological effect of radiation and chemotherapy to the tumour piece. The classification is based on the observation, made by the pathologist, of the presence of tumour cells directly on the piece of surgical resection. Tumour response to CRT is usually scored following the criteria proposed by Mandard AM et al.: TRG-1, pathological complete response (pCR); TRG-2, presence of residual cancer cells; TRG-3, fibrosis outgrowing residual cancer cells; TRG-4, residual cancer cells outgrowing fibrosis; and TRG-5, absence of response. The classification is represented graphically in Figure 5:

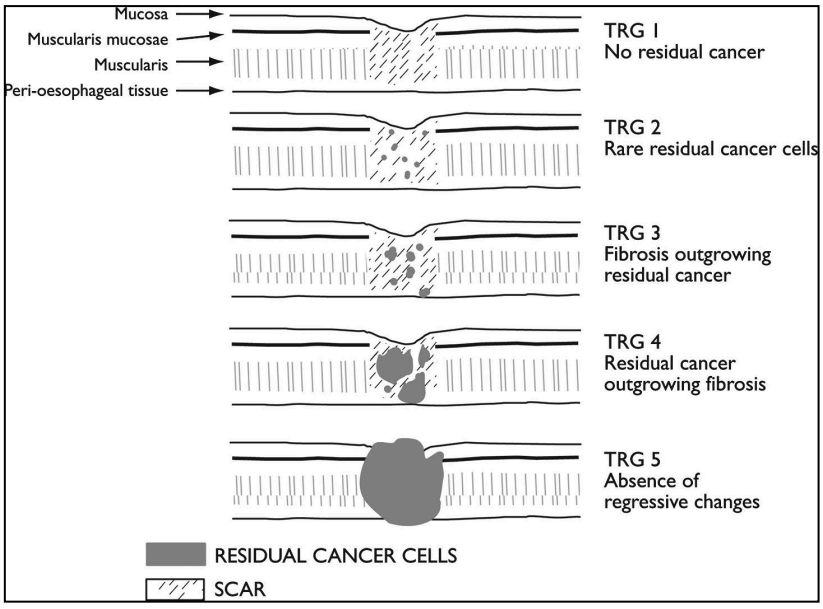


Figure 5. Mandard Tumour Regression Grade used to classify the histopathological response in this study *(took from Smith F.M. et al, 2007).*

1.6 *TUMOUR RESPONSE VARIABILITY TO THE NEO-ADJUVANT CHEMO-RADIO THERAPY*

Different TRG values show a high variability in the biological response that is unrelated to the change of the tumour stage (T) after therapy.

In order to increase the response rates to preoperative chemo-radio therapy, agents such as oxaliplatin and antibody-based regimens have been incorporated into multimodal treatment concepts, and these strategies are currently under extensive evaluation. However, at the same time, the inclusion of more toxic agents has increased the acute toxicity and long-term side effects of these treatments. In some cases, acute organ toxicity necessitates a dose reduction or even termination of therapy *(Cunningham D., 2010)*. Furthermore, any higher grade of acute toxicity impairs the quality of life of the individual patient. The variability treatment-associated tumour regression and toxicity from one patient to another are the major clinical problems, thus attention to the possibility to individualize rectal cancer treatment has increased (*Grade M. et al., 2012*). The ability to predict tumour response before or in the early stage of the treatment may significantly impact the selection of patients for pre-operative combined therapy, establishing a more effective and personalized therapy. The benefits also include the avoiding of unnecessary toxic treatments and the reduction of time, costs and side effects in the patients with resistant tumours *(Smith F.M. et al., 2007)*. The main interest of the scientific community is to customize the treatment so that the majority of individuals might have a targeted therapeutic benefit depending on the characteristics of the tumour. There are many factors that can contribute to the variability in the biological response to the treatment: the size of the tumour, the type and dose of radiation therapy and chemotherapy administered, the time elapsed between the end of treatment and surgery, histomorphological parameters (differentiation, inflammatory response, etc). However, it is a well-recognized fact that individual sensitivity to chemo-radiation therapy is influenced by the genetic background and contributes to the toxicity of treatment effects *(Lindsay KJ. et al., 2007)*. Molecular information, which includes genomic, trascriptomic and proteomic effectors, modulate tumour and stromal cells response causing different sensitivity to the treatment. This involves complex signalling processes that differ by cell type, cell contest, cell differentiation. Understanding the genetic background and the different molecular events related to the different tumour response could provide relevant clinical advantage on the ability to predict the chemo-radio resistance. Furthermore, information on the molecular pathways and the mechanisms underlying the tumour transformation might be used for the development of new

therapeutic agents more effective and less toxic to the body. To this end, extensive investigations have been performed to identify molecular biomarkers differentiating responsive and resistant tumours.

1.7 *MOLECULAR TUMOUR BIOMARKERS*

Molecular tumour biomarkers are molecules produced by tumour itself or by other tissue in response to the presence of tumour or other associated conditions. While clinical exams show the presence of the tumour, molecular markers indirectly detect the tumour. Ideally a molecular marker should be peculiar only of the tumour and it should be easily and consistently identified. However, tumour often shows characteristics similar to the normal tissue and none of the known tumour markers is specific only for tumour cells. A number of different kinds of molecules can be biomarkers: DNA, RNA, proteins, metabolites.

- These molecules could have different functions: they can be part of apoptosis or proliferation process, they can be hormones, enzyme, oncofetal antigens or receptors;
- These molecules could have different location: they can be nuclear, cytoplasmatic or extracellular molecules.

In the last few years active researches has been carried out to find biomarkers with the aim of potentially stratify rectal cancer patients. The majority of the studies have focused on tumour-tissue differences. Many groups investigated single or multi-biomarkers, focused on specific pathways involved in tumour progression, DNA repair systems, cell cycle progression pathway, pathways involved in death mechanisms (*Kuremsky J.G. et al., 2009; Bertolini F. et al., 2007; Reerink O. et al., 2004).* DNA alterations, gene expression patterns and protein expressions were analyzed for detecting reliable biomarkers. However, for most studies, the results were conflicting and still remain inconclusive, highlighting the complexity and heterogeneity of the molecular mechanism involved in chemo-radio resistance.

Since complex phenotypes like tumour responsiveness to chemo-radio therapy, do not depend on the alteration of a single molecule, high-throughput technologies have emerged for investigating multiple genetic events. Gene expression microarray have been applied to interrogate complex genetic pathways and networks simultaneously. Many groups analyzed the gene expression profile of tumour biopsies collected before the chemo-radio therapy. *(Kim I.J. et al., 2007; Watanabe T. et al., 2006;*

Ghadimi B.M. et al., 2005; Kuremsky J.G. et al., 2009; Brettigam-Moore K.H. et al., 2011). However, gene signatures found by different authors had very limited overlap of genes and the results of testing published gene signature on different tumour cohort were useless. Several reasons could lead to this result and they range from technical to clinical differences. Collection of tumour biopsy before the treatment is usually carried out during colonoscopy or rectum exploration and it is often difficult to get enough tumour material. Lack of standardized clinical management rules and definitions, differences on the collection, storage, processing and quality control of biological material, differences in the microarray procedures and in the evaluation of results make difficult the comparisons between different works and make difficult the definition of reliable biomarkers. Furthermore, several recent publications have provided evidence that tumour microenviroment is essential in modulating the tumour response to chemo-radio therapy since neighbour cells can express molecular factors that could be involved in different signalling pathways associated to tumour resistance (chemotactic molecules, growth factors, death factors) and they can regulate the recruitment of immune system cells. Irradiated cells can induce mutagenic response in neighbouring cells, which are not directly traversed by particle radiation, via gap junctions (bystander effect) *(Pajonk F et al., 2010; Mothersill C. et al., 2004)*.

All these features highlight the complexity and heterogeneity of the individual tumour response and the importance of considering response to treatment in a broad contest that can take into account not only the signalling pathway of tumour cells but also the microenvironment as a part of a unique system. A potential alternative to response prediction tissue-based methods is the development of blood-based methods. Blood obtained by venopuncture is the most accessible human specimen, the least invasive, and feasible to monitor over long periods of time. The blood is like a liquid biopsy that captures different molecules from all organs and tissue in health and disease thus reflecting the physio-pathological state of a body. The dynamic nature of the circulatory system and its constituents reflects diverse physiological or pathological states, and the ease with which the blood can be sampled makes it a logical choice for biomarker applications. Blood components that provide an indication of cancer status include various cellular elements such as circulating tumour cells, cell-free DNA and RNA, proteins, peptides and metabolites. These tests are more desirable because they are less dangerous and they don't present the side effects that could have the invasive type of analysis, which also must be performed by health care professionals with high level of specialization. Furthermore, it is possible to study blood-based markers in patients over time, making possible the disease monitoring. This would give many

advantages in early diagnosis as well as in monitoring the recurrences or the molecular alteration during a treatment, for determining the effectiveness.

PROJECT 1. Circulating cell-free DNA (cfDNA) as a promising diagnostic, prognostic and predictive marker of tumour malignancy

2 cfDNA

In the course of the search for new biomarkers, circulating cell-free DNA (cfDNA) has become a popular target of interest.

The existence of nucleic acids (DNA and RNA) in human blood has been reported for the first time in 1948 by Mandel and Metais (*Tsang J.C.H. et al., 2007*). This attracted little attention in the scientific community until the observation, several years later, of a direct correlation between the amount of cfDNA in plasma and serum and the tumour disease. Although circulating cfDNA is present in healthy individuals, elevated concentrations have been found in cancer patients, reflecting the physiological and pathological states. In support of this evidence many studies detected tumour specific alterations, suggesting the application of cfDNA as a potential candidate biomarker of diagnosis (*Schwarzenbach H. et al., 2011*). In the 1994 mutated RAS gene fragments was detected in blood of cancer patient (*Sorenson G.D. et al., 1994*) and then, during the past decade increasing number of scientific groups found different tumour-specific alterations such as microsatellite instability, loss of heterozygosity, mutations, polymorphisms, methylation. Different groups tried to exploit the detection of genetic alteration in cfDNA for several clinical purposes such as early diagnosis, prognosis, and prediction of recurrences. However, cfDNA assays, which target specific mutations, require that the mutations in the tumour occur frequently at a specific genomic site. A major drawback of cfDNA assays is the low frequency of some mutations that occur in tumours resulting in observation of a lower percentage of patients that present tumour specific cfDNA compare with the real situation. The mutational background of cancer is enormously complex and this makes it quite difficult to find cancer specific mutations. Furthermore detection of a specific genetic alteration is complicated by the presence of cfDNA originated from normal cells that increase the background noise making the analysis difficult (*Kohler C. et al., 2011*). Tumours are usually a mixture of different cancer cell clones (which account for different the genomic and epigenomic alterations causing by the tumour heterogeneity) and other

normal cell types, such as stromal and haematopoietic cells (*Schwarzenbach H. et al., 2011)*. Thus, during tumour progression and turnover, both tumour-derived and wild-type (normal) cfDNA can be released into the blood. The proportion of cfDNA that originates from tumour cells varies owing to the state and the size of the tumour. Furthermore, cfDNA is variably fragmented and the usual molecular tools of investigation that exploit the DNA replication strategy such as PCR, qPCR, sequencing, could fail the mutation detection. These problems are relevant in the prospective of use specific cfDNA biomarker for early detection of cancer or prognosis of the disease. Quantitative alteration of cfDNA has been observed to correlate with the tumour disease because elevated levels of cfDNA in serum and plasma have been detected in many cancer types. However the percentage of cfDNA that originates from the tumour are variable from patient to patient and sometimes different studies found discordant results. Concentrations of overall cfDNA vary considerably in plasma or serum of both patient and healthy groups. The causes may be biological (histology, the size, the stadium, the origin of the tumour), or they can be technical such as differences in the processing, in the recovery of cfDNA or in methods of genomic alteration detection. Furthermore, the origin and the mechanism of cfDNA release could be different in relation to different physiological condition and in relation with the presence of tumour.

2.2 *ORIGIN OF cfDNA*

To date, the physiological event that lead to the release of cfDNA during cancer development and progression are still not well understood. Among the various hypotheses, it is thought the release of nucleic acids into the blood could be related to the apoptosis and necrosis of cancer cells in the tumour microenvironment. Secretion has also been suggested as a potential source of cfDNA as well as mitotic catastrophe and autophagy (*Gormally E. et al., 2007*).

In a physiological state, fate of cells are constantly regulated by a fine balance between survival and death signals that are activated in response to internal or external environmental stimuli, regulating the cell homeostasis. In physiological situations the usually exploited process of death is apoptosis, a regulated mechanism in which transcriptional and translational machines are intensively employed to ensure the controlled elimination of dying cells, preventing the involvement of inflammatory response. During apoptosis, genomic DNA is cleaved into DNA fragments of around 180bp in length. This is the result of cutting internucleosome fragments by endonucleases. This mechanism represents a real barrier to tumour

formation, because it is involved in the elimination of cells that undergo alterations to the genetic material exceeding the capacity of cellular repair. Indeed, inactivation of programmed death pathway is one of the important events in the development, progression and aggressiveness in cancer. In this situation the main events of cell death are non-apoptotic, such as tumour necrosis and mitotic catastrophy. In particular, necrosis is a pathological cell death that involves the enzymatic digestion of the cell, which fails to maintain the cell membrane integrity, causing the release of its content outside. This mechanism generates a spectrum of DNA fragments with variable and longer dimensions, made by incomplete and non-organized enzymatic cleavage (*Kumar V. et al., 2009*). Therefore, the increased length (*integrity*) of circulating cfDNA in cancer patients has been postulated to be a consequence of pathologic processes other than apoptosis that could be peculiar of tumour cells. Indeed, the degree of alterations in circulating cfDNA integrity has been shown to reflect the extent of cancer involvement. Different cell death mechanisms seem to be also involved in the response to neo-adjuvant treatment. Chemo-Radio therapy (CRT) exploits cell damages and in particular DNA damage to eliminate cancer cells (*Lindsay KJ et al., 2007; Selzer E. et al., 2012; Woodward W.A. et al., 2008*). Some studies found a correlation between observations of a high intrinsic apoptotic cell tumour index and the positive tumour regression (*Smith F.M. et al., 2006*). Therefore, the analysis of the integrity of cfDNA could provide an indication of the contribution of the various processes of death in the release of cfDNA. (*Wang B.G. et al., 2003; Umetani N. et al., 2006*).

2.3 *THE ROLE OF CFDNA AS PREDICTING TOOL OF HISTOPATHOLOGICAL RESPONSE: THE FIRST APPROACH*

A wider project started before my PhD program, has involved the evaluation of the cfDNA quantity and cfDNA integrity as markers in different cancer diseases, with diagnostic and prognostic purposes. The project has also included analysis of cfDNA concentration and cfDNA integrity fluctuation in rectal cancer, in relation to the tumour response to CRT. The obtained results, which have been recently published in a paper cited at the end of this thesis, are briefly described below.

In this study we measured the cfDNA concentration and the cfDNA integrity in a retrospective study of 67 patients with locally advanced rectal cancer and we evaluated their association with tumour response to preoperative chemo-radio therapy. To maximize the sensitivity of the cfDNA quantification, we performed

quantitative real time PCR (qPCR) assays on the Alu repeat sequences. These sequence elements are part of SINE family (Short Interdispersed nuclear Elements) and they are non-coding, and highly repeated in the genome. The Alu sequence is a genomic region that extends for about 280 bp and it is composed by two tandem repeats of approximately 120 bp.
cfDNA integrity was assessed as a ratio of the concentration of a fragment longer than 180 bp over the concentration of fragment smaller than 180bp. In brief we first demonstrated the significant increment of plasma cfDNA concentration and cfDNA integrity (AUC=0.90) in rectal cancer patients in comparison to healthy subjects with negative colonoscopy. This finding suggested the differential origin of cfDNA according with the pathological status. Moreover, after the CRT, cfDNA integrity was found to be associated with tumour response. During the treatment we observed a significant decrement of cfDNA integrity in Good responders (p-value=0,0048) caused by a decreasing of the longer cfDNA fragments that was not observed in Poor responders (p-value=0,0005).

2.4 *CRITICAL POINTS*

In this study cfDNA concentration and cfDNA integrity were evaluated with two sets of primers, Alu115 and Alu247, as reported by Umetani (*Umetani N. et al., 2006*) where Alu115 amplified both the shorter and longer fragments and Alu247 amplified only the longer DNA fragments. cfDNA quantification was obtained using the Alu115 primers represents the total plasma cfDNA, while quantification of the Alu247 primers reflected the amount of cfDNA released mainly from non-apoptotic cells. qPCR were performed by using SYBR Green non-specific chemistry. SYBR Green is an asymmetrical cyanine dye used as a nucleic acid stain. It binds in non-specific way to DNA molecules. Therefore, we performed a careful standardization of the method for gaining confidence that the correct fragment lengths were amplified. However, even if SYBR Green preferentially binds to double-stranded DNA, it will also stain single-stranded DNA and RNA with low performance. Moreover, SYBR Green exhibits a very strong fluorescent signal, but it has been shown to inhibit the PCR reaction if it is used at concentration sufficient to saturate the double stranded molecules generated during the amplification. Therefore, it must be used at low non-saturating concentrations to prevent reaction inhibition, but in this condition the signal could be less reproducible, because of the variability of the dye molecules number that can bind DNA amplicons. In addition, melt curve analysis using SYBR Green is complicated by its preferential binding to specific DNA fragments containing higher GC%, and by the dye redistribution due to the

non-saturated condition that occurs during the melting analysis altering the detection of low quantity of aspecific products or primer pairing with the melting curve (*Gudnason H. et al., 2007; Eischeid A.C. et al., 2011*). Limits of the technique used to detect the cfDNA and the limited control of the variables that influence the measurement, affect the quality and the reliability of the data. This problem can become important in a study where the quantitative information of fragments, with different lengths, have to be correlated in a ratio index.

In association with Euroclone S.p.A. and in agreement with the aim of my Ph.D. project, we designed and standardized more specific and easy to use qPCR assays in view of a potential development of a scientific kit. We exploited the Taqman chemistry involving the use of a fluorogenic probe that emits fluorescence only if DNA polymerase amplifies the right amplicon, increasing the specificity. A comparison between SYBR green non-specific chemistry and Taqman specific chemistry was performed for evaluating the possible differences in quantification and in sensitivity. We measured cfDNA of the same cases with the qPCR chemistry in parallel.

Then we investigated the cfDNA concentration and cfDNA integrity in a prospective study of locally advanced rectal cancer patients to evaluate the association of these markers with the histological response to the chemo-radio therapy. Accordingly with the previous study, we mainly focused our attention on the kinetics of the size distribution of cfDNA along the chemo-radio therapy. For this purpose we introduced a time point analysis after two weeks from the initiation of the CRT.

3 MATERIAL AND METHODS

3.1 *PATIENTS AND TREATMENT CHARACTERISTICS*

The prospective study included 49 rectal cancer patients who received neo-adjuvant chemo-radio therapy and underwent a surgical resection at the 2nd surgical section of Department of surgical, oncological and gastroenterological sciences (University of Padova). The pre-treatment evaluation of a rectal cancer patients that will undergo to a neo-adjuvant treatment, included a complete clinical history and physical examination, colonoscopy, complete blood cell count, transrectal ultrasound, pelvic computed tomography scan or magnetic resonance imaging, abdominal/chest computed tomography and carcino-embryonic antigen test. The inclusion criteria for CRT were as follows: a) biopsy-proven adenocarcinoma of the mid-low rectum (< 11 cm from the anal verge); b) clinical stage T3-4 and/or node-positive; and c) Eastern Cooperative Oncology Group performance status 0-2. These patients were treated with external beam radiotherapy using high-energy photons (> 6 MV), with conventional fractionation (≥ 50 Gy in 28 fractions, 1.8 Gy/day, 5 sessions per week), and 5-fluorouracil (5-FU)-based chemotherapy administered by bolus or continuous venous infusion. A standard total mesorectal excision was performed 4 to 8 weeks after the completion of CRT. Blood samples were collected before the CRT and before the surgical resection. Another draw blood was collected after 2 weeks from initiation of the CRT. After the surgery, rectal resections were analyzed by a pathologist that observed the presence of tumour cells. The tumour response to CRT was defined using the Tumour Regression Grade (TRG) and was scored following the criteria proposed by Mandard et al. We defined patients with TRG 1 and 2 as Good responders and patients with TRG 3, 4 or 5 as Poor responders. The study protocol was reviewed and approved by the local ethics committee (protocol number 740 P), and each patient provided written informed consent.

3.2 *cfDNA EXTRACTION*

Ten mL of peripheral blood was collected into a purple-top blood collection tube (containing EDTA additive) before physical examination or biopsy. Plasma samples were obtained by centrifugation of peripheral blood at 3,000g for 10 minutes. Plasma samples were carefully collected from the upper portion of the supernatant and stored in aliquots at −80°C. The DNA was extracted from 500µl of plasma with a QIAamp DNA Mini Kit (Qiagen, Hilden, Germany). The procedure involves cellular degradation and protein denaturation by adding Lysis buffer solution (AL

buffer) and proteinase K to the plasma sample in proportion 1:1 and 1:10 for buffer and proteinase K respectively. The extraction method exploits the tendency of DNA molecules to adsorb to a silica column in the presence of high concentration of salts, ethanol and guanidine isothiocyanate. We introduced a carrier RNA poly-dA (Qiagen, Hilden, Germany) in the initial sample solution for increasing the binding efficiency of DNA fragments to the column. Purification of the DNA was performed by using Washing Buffer 1 and 2 (AW1 and AW2), which ensure the complete removal of any residual contaminants without affecting DNA binding. We increased the number of washing step with Washing Buffer 2 to eliminate all the organic residues. Elution of the nucleic acid from the columns was performed with the Eluition Buffer (AE), an ethanol-free solution with low concentration of salts.

3.3 *QUANTITATIVE PCR OF PLASMA cfDNA FRAGMENTS: PRIMER-PROBE SPECIFIC Alu ASSAYS*

On the basis of the Alu repeat consensus sequence, we designed two Taqman PCR assays. We used Primer Express V 3.0 (Applied Biosystem, Milan) to design several primer-probe sets. Primer-probe sets were evaluated in silico to minimize primer and probe complementarities; we used OligoAnalyzer 3.1 (Integrated DNA technologies, IDT) software to evaluate the different primer-probe for hairpins, self-dimers and hetero-dimers. We selected two primer-probe assays of 83 bp (Alu83) and 244bp (Alu244) in length. Alu83 assay was designed within the longer amplicon and thus it amplified both the short and the longer fragments, whereas the second primer set (Alu244) amplified only the long DNA fragments. cfDNA quantification obtained using the Alu83 assay represents the total cfDNA, while quantification of the Alu244 primers reflects the amount of cfDNA released mainly from non-apoptotic cells. We estimated the *cfDNA integrity* for each sample as the ratio of the long over the short fragments (*Alu244/Alu83*). The two assays shared the same probe: 5'-6FAM-CCTGGCCAACATGGTGAAACCCC-TAMRA-3'. Primers for the Alu83 assay were: primer forward 5'-CTGAGGTCAGGAGTTCGAGACC-3'; primer reverse 5'-CCACGCCCGGCTAATTTT-3' and primers for the Alu244 were: primer forward 5'-GCGGTGGCTCACGCCTGTAA-3'; primer reverse 5'-GGAGTGCAGTGGCGCGATCT-3'. Amplicon sizes and specificity were confirmed in gel electrophoresis. Real-time PCR amplification was performed with 10ul of FluoCycle probe master mix (Euroclone, Milan) in a 20ul total. We performed the optimization assays by varying the primer and probe concentrations. We evaluated primer concentration ranging from 50nM to 900nM. Alu244 was

amplified using 900nM and 900nM of primer forward and reverse respectively, while Alu83 was amplified using 300nM and 900nM of primer forward and reverse respectively. Alu probe concentration was set to 250nM. The reaction was performed using 7300 Real-Time PCR System (Applied Biosystem, Milan) with hold 95°C for 5 minutes, followed by 35 cycles of 95°C for 15s and 62°C for 60s. The absolute amount of cfDNA in each sample was determined by a standard curve using 10-fold dilutions (from 10 ng to 0,1 pg) of genomic DNA derived from peripheral lymphocytes of clean-colon healthy subjects. A negative control (water template) and a positive control (cfDNA from a volunteer) were performed in each plate.

3.3.1 Primer-probe specific Alu assays: a comparison with previous SYBR green based assays

A comparison between SYBR green non-specific chemistry and Taqman specific chemistry was performed for evaluating if the two chemistry methods showed differences in quantification and in sensitivity. The setting of the two SYBR green assays are described as follow: Alu115 primers were as follows: forward, 5'-CCTGAGGTCAGGAGTTCGAG-3' and reverse, 5'-CCCGAGTAGCTGGGATTACA- 3'; Alu247 primers were as follows: forward, 5'-GTGGCTCACGCCTGTTAATC- 3' and reverse, 5'-CAGGCTGGAGTGCAGTGG-3' (PRIMM, Milan). Real-time PCR amplification was performed with 10ul of SYBR Green master mix (Applied Biosystem, Milan) in a 20ul total. Alu247 was amplified using 200nM and 600nM of primer forward and reverse respectively, while Alu115 was amplified using 500nM and 800nM of primer forward and reverse respectively. The reaction was performed using 7300 Real-Time PCR System (Applied Biosystem, Milan) with hold 95°C for 10 minutes, followed by 35 cycles of 95°C for 30s and 64°C for 30s, followed by melt curve analysis. We quantified the cfDNA concentration and we assessed the cfDNA integrity using both the Taqman and SYBR green assays. For this purpose we analyzed a cohort of 90 patients with colon cancer and 46 healthy subjects with the two sets of assays in parallel.

3.4 *STATISTICAL ANALYSIS*

Descriptive results are expressed as median and 25%-75% percentiles (Q1-Q3). We performed Mann-Whitney U test to compare plasma concentrations of short, long cfDNA and median cfDNA integrity ratio between histopathologic Good responders and Poor responders for each defined time between groups: before the CRT (T0), after the CRT (T2) and after two weeks from initiation of CRT (T1).
A P-value of < 0.05 was considered statistically significant.
The variation in the levels of cfDNA was studied comparing cfDNA levels of Good responder and of Poor responder patients at different time point draw-blood. The Wilcoxon signed-rank test was performed to assess the significance of changes within each histopathological response group. Results were considered significant at $p < 0.05$ (two-tailed). The statistical analyses were performed using the SAS statistical package (SA, release 9.1.3, Cary, NC, USA) and with GraphPad Prism version 5.0

4 RESULTS

4.1 *PRIMER-PROBE SPECIFIC ALU ASSAYS: A COMPARISON WITH PREVIOUS SYBR GREEN BASED ASSAYS*

We compared the cfDNA quantification results using two different sets of qPCR assays that exploited SYBR green and Taqman chemistry. For this purpose we standardized and optimized the qPCR procedure of the Taqman primer sets that we designed. Figure 1 shows the optimal efficiency and comparability of the two Taqman assays. It is noticed that although we used low range of DNA quantity, the curves of amplification crossed the threshold line soon in the PCR reaction.

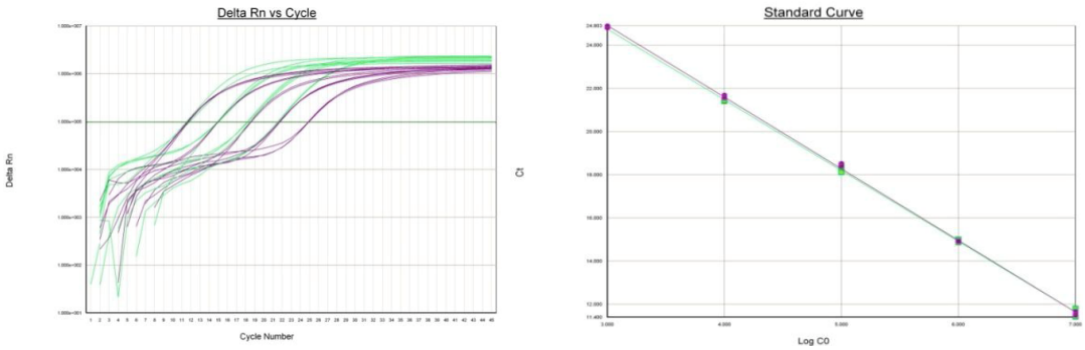


Figure 1. Standard curve of Alu83 (green) and Alu244 (purple) made with serial dilution 1:10 from 10 ng/ul to 1pg/ul. A) Amplification curves; B) standard curve graph that shows the Ct number in relation to the log of the DNA concentration.

We analyzed cfDNA from plasma of a same study, including 136 subjects between patients and controls, with Taqman and SYBR chemistry assays as described in the material and methods. Briefly, we carried out a standard curve for each qPCR assay and we quantified the concentration of cfDNA on plasma (ng/ml). Observing the SYBR green assay results we noticed that the cases that showed cfDNA concentration up to 3ng/ml, measured by Alu115, usually showed high cfDNA integrity value regardless to the physiopathology status of the subject. Therefore we performed two separately comparison analyses on the basis of the total cfDNA quantified by Alu115 using a cut-off of 3ng/ml. Figure 2A shows box plot of cfDNA concentration measured with the two different chemistry assay (Taqman assays: Alu83 and Alu244; SYBR green assays: Alu115, Alu247) for those samples that presented Alu115 cfDNA concentrations lower than 3ng/ml. As said before, samples that presented low values of total cfDNA quantified by Alu115, showed also similar quantities for the longer fragments (Alu247 SYBR green assay), producing a high range of cfDNA integrity that exceed the unit (Figure 2B). However, on the basis of the assays design, it is unlikely that the cfDNA integrity is often above the unit. The

analysis of the same plasma samples performed by Taqman assays revealed an increment of the cfDNA concentration detected with both the Alu244 and Alu83 Taqman assay (Figure 2A). The total cfDNA concentration (short fragment), quantified by Alu83 Taqman assay, was eight times higher compared with that quantified by Alu115 SYBR green assay and the long fragment concentration, evaluated by Alu244 Taqman assay, is two times higher compare to the Alu247 SYBR green assay concentration. According to the increment of the two fragments concentrations, the Taqman cfDNA integrity was considerably reduced in the analyzed cases (Figure 2B). These results indicated a better ability of Taqman assays to detect low quantities of cfDNA fragments, improving the sensitivity of the analysis. However, SYBR green and Taqman assays detected similar cfDNA fragments quantities for the cases where total cfDNA, quantified by Alu115 SYBR green assay, ranged from 3ng/ml to 100ng/ml (Figure 2C). The cfDNA integrity ratios assessed by SYBR green assays and by Taqman assays are comparable in this range of cfDNA concentrations indicating the reliability of both the two chemistry assays.

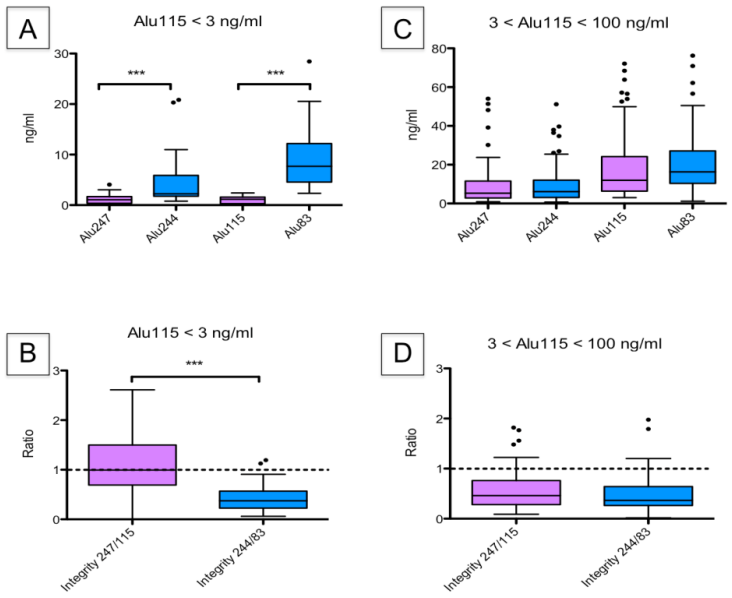


Figure 2. Quantification of cfDNA fragments made by Taqman assays (Alu83 and Alu244) and by SYBR green assays (Alu115 and Alu247) in a same cases study. A) Selection of samples that showed total cfDNA concentrations, detected by SYBR Alu115, up to 3ng/ml of plasma. B) cfDNA integrity related to those samples with total cfDNA up to 3ng/ml of plasma. C) Selection of samples that showed total cfDNA concentrations, detected by SYBR Alu115, ranged from 3ng/ml to 100ng/ml of plasma. D) cfDNA integrity related to those samples with total cfDNA concentration range from 3ng/ml to 100ng/ml of plasma.

4.2 *EVALUATION OF cfDNA CONCENTRATION AND SIZE DISTRIBUTION IN A PROSPECTIVE STUDY OF LOCALLY ADVANCED RECTAL CANCER PATIENTS*

We tested the cfDNA primer-probe specific Taqman assays on a set of 49 plasma samples from prospective patients with locally advance rectal cancer treated with neo-adjuvant CRT**.**

On the basis of TRG, 23 rectal patients were considered Good responders (TRG 1 n=10, TRG 2 n=13) and 26 were considered Poor responders (TRG 3 n=18, TRG 4 n=7, and TRG 5 n=1). Complete details of the patients, tumour, and treatment characteristics are summarized in Table 1.

Table 1. Patients, tumour and treatment characteristics of the 49 patients included in the study.

Patient characteristics		**n°**	**%**
Age	Median (range) yrs	67 (48-80)	
Sex	Male	36	73,5
	female	13	26,5
Tumor distance from the anal verge	≤ 7 cm	18	36,7
	> 7 cm	31	63,3
5-Fluorouracil administration	Continuos infusion	7	14,3
	Bolus	38	77,6
	Not available	4	8,2
Other drugs	5-Fluorouracile alone	20	40,8
	Oxaliplatin	20	40,8
	Not available	9	18,4
pTNM	0	11	22,4
	I	14	28,6
	II	12	24,5
	III	7	14,3
	IV	1	2,0
	not available	4	8,2
Radical surgery	yes	36	73,5
	no	2	4,1
	not available	11	22,4
Acute toxicity	0	2	4,1
	1	16	32,7
	2	18	36,7
	3	6	12,2
	4	1	2,0
	not available	6	12,2
TRG	1	10	20,4
	2	13	26,5
	3	18	36,7
	4	7	14,3
	5	1	2,0
	Good responders (TRG 1-2)	23	46,9
	Poor responders (TRG 3-4-5)	26	53,1

4.2.1 Relation of cfDNA levels and tumour response to CRT: cfDNA levels in plasma pre-CRT and post-CRT

We first evaluated the cfDNA quantity and cfDNA integrity in plasma of Good and Poor responders collected before the CRT.
As we noticed in the published paper, before the CRT, cfDNA levels and cfDNA integrity were not able to classify rectal patients on the basis of tumour response to treatment (Table 2).
In particular, the cfDNA integrity range, showed by Good responders, overlapped the range of cfDNA integrity found in Poor responders, indicating that the mechanism of cfDNA release from apoptotic and necrotic processes might not predict the response of the tumour to the treatment before it is applied.

Table 2. Pre-CRT plasma levels of cfDNA in patients divided by the tumour regression response to CRT. (Good= Good responders; Poor= Poor responders; Q1-Q3= Range from first to third quartile) (*=ng/ml)

Marker	Response	N	Median	Q1-Q3	p-value
Alu244 pre-CRT	**Good**	23	17.36*	*9.44 ;82.52*	*0.3464*
	Poor	26	19.23*	*7.44 ;57.79*	
Alu83 pre-CRT	**Good**	23	44.59*	*19.83 ;74.20*	*0.2793*
	Poor	26	22.49*	*9.5;74.18*	
cfDNA integrity pre-CRT	**Good**	23	0.88	*0.32;1.19*	*0.9202*
	Poor	26	0.73	*0.47;1.18*	

After the CRT we noticed that, in Good responders, the median value of cfDNA integrity decreased comparing with the median value before the CRT while in Poor responders it remained stable. The median value of cfDNA integrity in Good responders decreased about 50% and the same behaviour was noticed in the previously published paper. However, cfDNA integrity between Good and Poor responders did not reach the statistical significance as in the pilot study. Although the median values difference between Poor and Good responders was considerable, the range of cfDNA integrity values of Good responders was wide and it included the Poor responder integrity value range (Table 3).

Table 3. Post-CRT plasma levels of cfDNA and cfDNA integrity in Good responders and in Poor responders. (Good= Good responders; Poor= Poor responders; Q1-Q3= Range from first to third quartile) (*=ng/ml)

Marker	Response	N	Median	Q1-Q3	p-value
Alu244 post-CRT	**Good**	23	20.12*	7.43;36.62	*0.7035*
	Poor	26	18.81*	6.38 ;46.47	
Alu83 post-CRT	**Good**	23	44.95*	16.06;76.55	*0.6165*
	Poor	26	31.16*	19.39;49.66	
cfDNA integrity post-CRT	**Good**	23	0.44	0.19;1.19	*0.3165*
	Poor	26	0.71	0.42;0.89	

4.2.2 Monitoring the tumour response to CRT: variation of cfDNA in plasma pre- and post-CRT

Although the post-CRT cfDNA concentrations of short and long fragments were not showed a statistical difference according with the response to CRT, the variations of cfDNA concentration and cfDNA integrity before and after the CRT were similar to those found in the published paper. Differences of cfDNA concentrations and of cfDNA integrity between pre-CRT and post-CRT did not show any significant tendency in Poor responders. Conversely, in Good responders group the concentration of the longer fragment (Alu244) significantly decreased after the CRT (p-value=0.0024) while shorter fragments did not show a precise variation tendency. These different tendencies of shorter and longer cfDNA caused a significant decrement in cfDNA integrity (p-value=0.012) (Table 4).

Table 4. Variation of the cfDNA levels and cfDNA integrity between pre- and post-CRT in Good responders and Poor responders. (Good= Good responders; Poor= Poor responders; Q1-Q3= Range from first to third quartile; S=signed rank) (*=ng/ml)

Response	Marker variation	N	Median	Q1-Q3	S	p-value
Good	Alu244 post-pre	23	-5.65*	-39.2;1.4	-93	**0.0024**
	Alu83 post-pre	23	-1.09*	-36.5;8.7	-25	0.4593
	cfDNA integrity post-pre	23	-0.2	-0.5;0.0	-79	**0.0126**
Poor	Alu244 post-pre	26	-3.6*	-29.1;8.2	-32.5	0.4198
	Alu83 post-pre	26	-0.53*	-26.4;20.4	-0.5	0.9902
	cfDNA integrity post-pre	26	0.01	-0.57;0.25	-26.5	0.5116

These data confirmed the published paper results in which we showed a different kinetic of distribution of the fragments potentially associated with non-apoptotic death mechanism of release, only in the Good responders group. Poor responders did not show evident change tendencies in the cfDNA integrity between pre- and post- treatment confirming the different behaviour between these two groups. However, differences between the two response groups are still not well-defined and marked; patients within the groups showed high variability in the quantity of shorter and longer cfDNA making the assessment of a cut-off for clinical purposes difficult.

4.2.3 Monitoring the tumour response to CRT: variation of cfDNA in plasma after two weeks from the initiation of the therapy

T1 time draw blood was carried out after two weeks from starting the CRT, meaning in the middle of the neo-adjuvant treatment. We performed the analysis of cfDNA concentration and cfDNA integrity to evaluate if the initial perturbation of the biologic system could give meaningful evidences according to the different tumour response.

The draw blood at this time point is difficult to obtain because multiple variables could influence: patient availability, patient physiologic state, and the different medical hospitals where the patient does the therapy. Therefore we were able to collect only 12/23 T1 time point plasma of Good responders and 13/26 T1 time point plasma of Poor responders.

Table 5. Variation of the cfDNA levels and cfDNA integrity between pre-CRT (T0) and T1 time point in Good responders and Poor responders (Good= Good responders; Poor= Poor responders; Q1-Q3= Range from first to third quartile; S=signed rank) (*=ng/ml)

Response	Marker variation	N	Median	Q1-Q3	S	p-value
Good	ALU 244 (T1-T0)	12	-2.8*	-43.1;9.5	-10	0.4697
	ALU 83 (T1-T0)	12	-5.85*	-19.3;22.9	-1	0.9697
	cfDNA integrity (T1-T0)	12	-0.14	-0.4;-0.01	-26	**0.0425**
Poor	ALU 244 (T1-T0)	13	-14.4*	-19.3;2.2	-27.5	0.0574
	ALU 83 (T1-T0)	13	1.32*	-18.3;17.4	-0.5	10.000
	cfDNA integrity (T1-T0)	13	-0.16	-0.37;0.16	-15.5	0.3054

Analyzing the variation of cfDNA markers before the CRT and after two weeks from the initiation of the therapy we noticed that both the concentration of short and long cfDNA fragment were very variable and this situation is evident in the

evaluation of the variation of the cfDNA concentration compared with the basal level of cfDNA before the CRT (T0). Observing the variation of cfDNA integrity a significant decrease of this marker is showed in Good responder patients whereas Poor responder did not show a defined variation tendency (Table 5). This data gave somehow proof of the difference in size distribution of the cfDNA in association with the histopathology response. However, the number of patients is too small to give confidence of the results. The variation of cfDNA between T1 and T2 time point for this subset of patients did not give any significant tendency in both Good and Poor responders

4.3 *PORTABILITY OF THE CFDNA ANALYSIS STRATEGY*

In parallel with this study we evaluated the cfDNA markers in other tumour types and for different clinical purposes.

4.3.1 Colon cancer

We investigated cfDNA markers in colon cancer in order to evaluate their diagnostic and prognostic values. In particular we analyzed a cohort of healthy subjects, adenoma patients, and early and late colon tumour patients. We evaluated cfDNA level (Alu83) and cfDNA integrity (Alu244/alu83) in association with hypermethylation analysis of promoter region of two genes involved in colon tumour progression (*SFRP1* and *OSMR*). A significant increment of cfDNA integrity in colon cancer patients compared to healthy subjects was revealed. Interestingly, early stage of colon cancer revealed a significant difference in cfDNA integrity compared to the controls indicating the potential role in early diagnosis. The frequency of cfDNA hypermethylation events and the quantity of hypermethylated genes found in cfDNA increased in association with the tumour progression. A correlation between the cfDNA integrity and the frequency of methylation event was observed, indicating the relation between the tumour cfDNA and the portion of longer cfDNA fragment. cfDNA methylation was associated specifically with the tumour disease but we detected cfDNA methylation in only 65% of the patient with tumour-tissue methylation. These results demonstrated that cfDNA markers showed potential benefit in the colon tumour diagnosis and prognosis in association with the other clinical diagnostic screening procedures. (Paper in preparation)

4.3.2 Thyroids malignancies

cfDNA levels and cfDNA integrity were analyzed in thyroids malignancies with the aim of developing a non-invasive tool for the diagnosis and prognosis. In this study we coupled the analysis of these markers with the cfDNA hypermethylation of specific genes associated with thyroid cancer (*SLC5A8* and *SLC26A4)* and with the evaluation of an approved clinical marker ($BRAF^{V600E}$). cfDNA quantification showed a high ability to discriminate healthy individuals from tumour patients. Furthermore an increment of total cfDNA was observed with the progression of papillary thyroid tumours and correlations between long cfDNA fragment levels and histological types of thyroid tumours were found. Correlations between positive methylation events and the increase concentration of total cfDNA and of long cfDNA fragment were observed as well as the correlation between BRAF mutation with the increment of cfDNA integrity. These results highlighted the amount of information that cfDNA could providefor identification, diagnosis and prognosis of thyroids cancer, in association with marker already in clinical diagnostic. (Publication under submission)

5 DISCUSSION

In recent years, blood-based tests and in particular circulating cfDNA has become a popular target of interest as potential biomarker candidate for tumour detection or prognosis because it can easily be isolated from human plasma, serum and other body fluids. Besides its non-invasive nature that allows the clinician for an easy access to the specimen, at the same time, its use reduces the physical and psychological stress to the patient. Numerous studies reported both qualitative and quantitative alterations in tumour patients indicating its correlation with the proliferative disease (*Schwarzenbach H. et al., 2011*). In this study, we focused our attention on the evaluation of the cfDNA as a potential marker to predict tumour response in patients with rectal cancer who received neo-adjuvant CRT. Many clinical, pathological and molecular tissue-based markers have been investigated according to the response after neo-adjuvant CRT, but, to date, the results has been contradicting and highly variable. Considering the lack of solid molecular information associated with the resistance to the chemo-radio therapy and the lack of differences regarding the clinical characteristics of tumour of patients included in the cases study, we focused our attention on the different pattern of cfDNA fragments length produced by cell death mechanisms. Although many different mechanisms can contribute to the cfDNA release, the nature of cell death, and in particular apoptotic and non-apoptotic mechanisms, has appeared to be of relevance. These different death strategies influence the pattern and size distribution of circulating fragments: while apoptotic death mechanism releases mainly DNA fragments shorter than 200bp into circulation, non-apoptotic death mechanism, and in particular tumour necrosis, is characterized by the presence of different size fragments that are more likely greater than 200bp. Macrophages or other scavenger cells might phagocyte necrotic and apoptotic cells and macrophages that engulf necrotic cells can release digested DNA into the tissue environment (*Choi J. J. et al., 2005*). Physiological alteration of the cellular environment due to the presence of tumour could be caused alteration on the cell death mechanisms and different sizes of fragment could be found in different blood samples. Recently, several studies have showed the diagnostic potential of cfDNA integrity index in comparison to healthy controls (*Holdenrieder S. et al., 2008; Jahr S. et al., 2001; Giacona M.B. et al., 1998*). cfDNA integrity was found higher in breast tumours (*Umetani N. et al., 2006; Deligezer U. et al., 2008; Agostini M. et al., 2012*) and prostate tumours, plasma of head and neck tumours (*Jiang W. W. et al., 2006*), lung malignancies (*Sriram K.B. et al., 2012*), and naso-pharyngeal carcinoma (*Chan K.C.A. et al., 2008).*

In association with Euroclone S.p.A. and in agreement with the aim of my Ph.D. project, we designed and standardized more specific and easy to use qPCR assays in view of providing an easy, reproducible and non-invasive method of investigation. With this improved method we found significant data that highlighted the diagnostic and prognostic significance of cfDNA quantity and cfDNA integrity in different cancer types such as colon cancer and thyroid cancer, as reported in our published and submitted papers.

The aim of neo-adjuvant chemo-radio therapy in locally advanced rectal cancer is the death of tumour cells, for reducing the tumour mass. The different sensitivity of tumour to the chemo-radio therapy causes a different proportion of tumour cell to be able to survive. In addition cell death might occur also for the cells in the tumour microenvironment, for normal cells and for immune system cells, in different proportion and with different death strategies. cfDNA integrity index, indicated as the ratio of fragment longer than 200bp over the total cfDNA, could give information regarding the proportion of different mechanisms involved in cfDNA release.

In the current study we investigated the cfDNA levels and cfDNA integrity in association with therapy monitoring during the treatment of rectal cancer patients undergoing preoperative chemo-radiation. At the time of diagnosis (before CRT; T0), we did not detect significant differences in cfDNA concentration and in cfDNA integrity between Good and Poor responders. Accordingly with our previously published paper these results suggest that, before the therapy, cfDNA mechanism of release is similar between Poor and Good responders, or the cfDNA markers could not differentiate the histopathological response. In particular we observed a high variability of the cfDNA quantities and cfDNA integrity in both Good and Poor responders, leading to a conclusion that the cfDNA release could be affected by numerous different factors. Non-significant differences were also found in the evaluation of the cfDNA concentration and cfDNA integrity after the end of the CRT, although a substantial decrement of cfDNA integrity was observed in Good responders. Despite these results, some scientific groups suggested that the analysis of the kinetics of the plasma cfDNA concentration in cancer patients, could lead to more essential information, in particular during the course of treatment (*Fleischhacker M. et al., 2007*). We therefore focused on the kinetics of short and long cfDNA fragments to assess whether differences along the treatment have the potential to discriminate between Good responders and Poor responders. The kinetics of cell death and cfDNA release into circulation could be altered during the neo-adjuvant therapy period. For this reason we evaluated the variation of the

cfDNA concentration and cfDNA integrity until two weeks from the initiation of the CRT. The analysis, made in a subset of patients, found a statistical significant decrement of cfDNA integrity only for the group of Good responders. However cfDNA concentrations and cfDNA integrity showed high variability in both the response groups. This variability could due to the low number of patients analyzed but it could be due to the latency effect of chemo-radio therapy. Cell death, which is mainly due to radiation therapy, is not showed immediately after the CRT dose. For this reason the treatment is usually fractionated in more doses. Furthermore mechanisms of repopulation, DNA repair, and cellular redistribution could be variably activated in the irradiated tumour cells, increasing the individual disparity. Although comparison of short and long cfDNA fragment concentrations between Good responders and Poor responders revealed comparable plasma cfDNA levels before (T0) and after (T2) the chemo-radio therapy, the variation of cfDNA from T0 to T2 showed a statistically significant difference in long cfDNA fragment concentration and cfDNA integrity, for Good responders. Indeed, Good responder patients showed a significant decrement of long cfDNA fragment concentration (p-value=0.0024) and of cfDNA integrity (p-value= 0.0126) that the Poor responders did not show. The current study confirmed the results of our previous study by detecting a difference in the kinetics of long cfDNA fragments and cfDNA integrity in Good responders, enhancing the idea that the proportion of cfDNA fragment size might be potentially reversible after curative treatment. Decrement of cfDNA integrity in association with the positive response to CRT and therefore with the tumour mass reduction could be due to a differential process of cfDNA release and therefore to a shift of the cell death mechanism. Instead, the persistence of plasma cfDNA integrity level after the CRT may also be associated with many undefined factors including, maybe, the incomplete eradication of the tumour.

Other studies found a correlation between the level of cfDNA in patients with different cancer types and response to a given therapy. Chan K.C. et al. highlighted that the persistence of DNA integrity aberrations after curative radiotherapy in nasopharyngeal carcinoma patients was associated with an increased probability of disease recurrence (*Chan K.C. et al., 2008*).

5.1 Current Limitations

Although we confirmed the significant decrement of cfDNA integrity and the long cfDNA fragment concentration in Good responders, the high variability within the group decrease the accuracy and the overlapping range of data observed in Good and Poor responders make the selection of a cut-off difficult, limiting the clinical potential of these markers. The high variability is subjective to the individual plasma

sample analyzed. Numerous factors could influence the detection of the DNA circulating in blood and they might be of technical nature as well as of biological nature. A major technical issue that hampers consistency in all the cfDNA assays is the efficacy of the extraction procedures, because usually only small amounts of DNA is obtained from plasma and serum. As we seen when we performed the optimization of cfDNA extraction procedures, Fleischhacker et al. found that significantly different amounts of absolute DNA values were obtained from plasma using different DNA isolation methods. Because of the small quantity and highly fragmented nature of cell-free DNA in plasma and serum, a fast, efficient, and reliable isolation method is still a problem and so far there is no agreement on a standardized method. Improvement is ongoing in these aspects for cfDNA analysis to be more robust, consistent, comparative and informative. (*Fleischhacker M. et al., 2011*) To overcome this problem we has begun a collaboration with a group of *INESC Microsistemas e Nanotecnologias* (INESC MN), Lisbon, Portugal, for the development of a hybridization detection system based on magnetoresistive (MR) sensors associated to the use of superparamagnetic micro- and nano-particles as reporter systems to detect cfDNA directly from the body fluids (*Freitas P.P. et al., 2011*). To date this approach is to the first step of fabrication but we hope for implementing the system soon.

On the other hand, the biological factors that could be involved in the variability of cfDNA markers between subjects are numerous and different; the proportion of cfDNA that originates from tumour cells or that might be related to the presence of tumour, could be influenced by clearance, degradation and other physiological filtering events of the blood and lymphatic circulation. The influence of non-tumour cells on DNA fragments is currently not known and will also need to be evaluated. In order to better resolve the complex heterogeneity of tumour progression and histopatological response to chemo-radio therapy, we focused our attention on the circulating proteome because of its composition fluctuates and changes depending on a combination of numerous intracellular and environmental stimuli.

PROJECT 2. Circulating Low Molecular Weight peptidome for cancer biomarkers discovery

6 CANCER BIOMARKERS AND CLINICAL PROTEOMICS

The proteome is the entire set of proteins produced by an organism at certain time or under defined conditions. It comprises much more information than the genome, because it is a high dynamic system which composition fluctuates and changes depending on a combination of numerous environmental factors. In this contest, the proteome could better reflect the complexity of molecular processes involved in cancer since its expression dynamically fluctuate in response to different intracellular and extracellular stimuli. For this reason, one of the most interesting fields of molecular oncology has become the clinical proteomics, the division of proteomics that focused on the analysis of proteins expression changes during the development of a disease or during the progress of a therapy. Identification, characterization of unique patterns of protein expression, or biomarkers, associated with specific diseases is one of the most promising areas of clinical proteomics *(Pietrowska M. et al., 2012*). A lot of effort has been put to the determination of biomarkers whose measurement would be clinically valuable for diagnostic prognostic or predictive purposes and readily obtainable from the patients. In this regard, commonly used blood-based biomarkers include prostate-specific antigen (PSA) for prostate cancer, Carcino-embrionic antingen (CEA) for monitoring colorectal cancer, and estrogen receptor for estrogen-dependent mammary cancer. However, most of these single biomarkers do not reached the level of sensitivity and specificity that it is required for clinical purpose.

Sensitivity problem arise when biomarker is not present in all the patients or if the level is too low to detect it. This is due to the molecular heterogeneity of tumours from patient to patient. Cancer is a complex disease in which alterations of specific physiologic mechanisms that regulate the cell life could occur through modification of signalling pathways at different levels *(Push W. et al., 2003).*

The emergence of new powerful and high-throughput analytical technology has facilitated numerous attempts for cancer biomarker discovery, increasing the sensitivity of detecting trace amount of molecules and increasing the amount of data available. The advances of proteomic technologies have allowed analysis of many proteins simultaneously enabling the systematic interrogation of complex proteomes

and the identification of differentially expressed proteins in cells, tissues and body fluids. Depending on the type of analyte, various techniques of ionization and separation of ions in the analyzer are used in the field. Among proteomic technologies, *Matrix-assisted laser desorption/ionization mass spectrometer* (MALDI-MS) is one of the most widely used proteomic technologies because it has allowed rapid progress in cancer biology investigation. It has become the major tool for proteins, peptides and other biomolecules (oligonucleotides, carbohydrates, lipids) analysis in relatively complex samples because of its valuable and useful features, including high sensitivity, mass accuracy, minimal sample preparation, ease of use and heightened throughput. MALDI-MS is capable of performing parallel rather than serial analyses, and of analyzing changes in global protein expression pattern or specific sub-proteome. These characteristics allow for multiple biomarker screening analyzing, promoting the understanding of pathogenesis of cancer, and might improving early detection, prognosis as well as therapeutic monitoring (*Aebersold R. et al., 2003; Bateson H. et al., 2011*). MALDI-MS was developed in the late 1980s by Hillenkamp and Karas (*Karas M. et al., 1988*) and it exploits the efficient and directed energy transfer during a matrix-assisted laser – induced desorption event that ionizes the peptide residues and allows for the measurement of compounds with high accuracy and sensitivity. MALDI-MS instruments are usually coupled with *time of flight* ion analyzer (MALDI-TOF MS) in which ions separation occurs in a vacuum and the arrival time at the detector is dependent upon the mass to charge ratio, since the kinetic energy of the ions is equal for each ion. Intrinsic properties of the technique include: low sample consumption; high-resolution spectra of peaks by using the reflectron mode and the delay time; tolerance toward moderate amount of salt and buffer allowing proteins to be characterized directly without further purification that can be very difficult and time consuming. Another advantage of this technique is the minimal fragmentation because MALDI-TOF MS is a 'soft' ionization technique, the vaporization and the ionization are mild and it does not break down the analyte. Furthermore it almost exclusively creates singly charged ions, which mean that the m/z value becomes a very real measurement of the actual mass of the particles. Single laser pulse produces a quantity of ions that are detected in a millisecond timescale, allowing the acquisition of data from multiple laser pulses in a short time, allowing for detection of multiple peptides and proteins in the same analysis (protein profiling). MALDI-TOF-MS is used for analysis of proteins and peptides with molecular weight ranging from 500Da to as much as several thousand Da but peptides analysis are more efficient compared with larger proteins, which are difficult to ionize (*Lewis J.K. et al., 2000; Hortin G.L. et al., 2006, Aebersold R. et al., 2003).*

6.1 *RELEVANCE OF LOW MOLECULAR WEIGHT PEPTIDOME PROFILING*

Accordingly with technologies progress, increasing interest has been arisen for the dynamic nature of sub-proteomes presented in blood, as it constantly perfuses tissue and it thus reflects diverse physiological or pathological states (*Taguchi A. et al., 2013*). In particular, circulating *low molecular weight* (LMW) *peptidome* seems to be a source of disease-specific biomarkers. Until recently, the peptidome was considered by the scientific community simply a pool of the protein fragments result by degradation or proteolytical activity of active proteases, with an informational content equivalent to biological 'trash' or 'noise' (*Petricoin E.F et al., 2007*). However, during the past five years, a novel concept has come into sight, suggesting that the peptidome could instead contain useful disease-specific information. While full-length tissue proteins are usually too large to passively diffuse into the circulatory system and only secreted proteins are detectable, peptides and fragments of proteins of different cellular localization can shed from all cell types still representing the parent proteins. Peptides show better permeability between tissue and cell membranes than the corresponding full-lenght proteins. Furthermore, the protein fragment itself can have a powerful diagnostic capability because its origin can be due to a tumour specific event. Proteases are proteins that cleave amino acid sequences at precise locations and they participate in basically every physiological as well as tumour specific process. Tumour specific proteolytic cascades within the tissue can generate fragments that diffuse into the circulatory system. Genomic sequence alteration can cause an atypical protein expression or secretion patterns. Over-expression of proteases and its secretion can increase the circulating level of extracellular domain peptides. Another mechanism of biomarker elevation in biological fluid is the increase of signal peptide secretion and shedding of membrane-bound protein with an extracellular domain (*Kulasingam V. et al., 2008*). However, tumour specific peptides are derived not only from a cancer population of cells that has a small volume compared to the totality of the tissues but they are also derived from the microenvironment tissue surrounding the tumour cells. The interaction with tissue microenvironment is essential for cancer progression and various molecules such as cytokines, growth factors, extracellular matrix molecules mediate this type of communication. Also normal cellular processes can generate peptide fragments. However, the tumour microenvironment, through the aberrant processes of cell growth, cellular invasion, altered immune system function, and angiogenesis can produce an unbalance amount of peptides or it can generate a unique cascade of events that lead to a specific fragmentation. In addition, tissue

invasion by the tumour might permit direct release of molecules into the interstitial fluid and, through the lymphatic circulation, into the blood. The irregular and leaky nature of the new tumour microcirculation and the different hydrostatic pressure increase the peptides shedding. LMW blood peptides could be derived from all cells and tissue types, or they are cleavage products of resident blood proteins. It was also proposed that some LMW information might not come from tissue but might be generated in the blood sample by degradative proteinases that are already present in the blood. (*Petricoin E.F. et al., 2007*)

It has recently indicated that peptides could accumulate on the high-concentration resident proteins such as albumin, and then acquire its longer half-life. Blood resident proteins can thereby protect the bound species from kidney clearance. (*Tirumalai R.S. et a.l, 2003*)

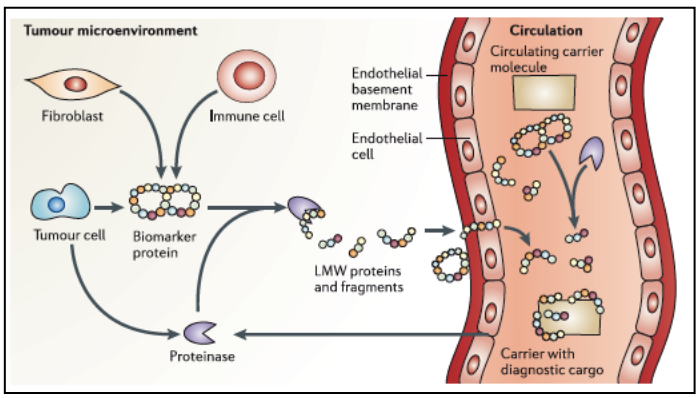


Figure 1. Different ways of peptide generation and elevation into the circulatory system. Circulating peptides could be shed and generate from all cell types present in the tissue microenvironment. *(Petricoin E.F. et al, 2007)*

6.2 *LIMITS OF THE DETECTION OF LOW MOLECULAR WEIGHT PEPTIDES*

One of the major challenges of MS-based serum/plasma proteomics is the high dynamic range of known plasma proteins that spans nine orders of magnitude. This wide range limits the detection of tumour-tissue-derived proteins in the circulation that are present in trace amounts (present at ng/ml to pg/ml) within a large background of abundant and non-relevant proteins. It is known that the 20 most abundant proteins represent over 99% of the total protein amount whereas the low molecular weight low-abundant peptides comprise less than 1% (*Lee H.J. et al., 2006*). Blood protein content is dominated by proteins such as albumin, transferring, haptaglobulins, immunoglobulins, and lipoproteins. This large dynamic

range exceeds the analytical capabilities of traditional proteomic method, making the detection of lower abundance peptides extremely challenging. Components of highest molar abundance, in complex mixtures, generally dominate the MALDI-TOF MS spectrum and tend to suppress the detection of lower abundance proteins *(Hortin G.L. et al, 2006)*. The presence of albumin, representing around 60% of all blood proteins, is the major factor changing the yield of ionization and affecting the ability to register low abundance serum components. In this situation, sample fractionation is a crucial step to generate reproducible and high quality mass spectra enabling the analysis of the peptidome (*Savino R. et al., 2012*).

Several methods have been proposed to reduce the complexity of the biological material on study. Different strategy focused on the depletion of albumin from the serum/plasma specimen prior MS analysis. Among them, several methods have been used such as immunoaffinity depletion, ion exchange chromatography, isoelettrofocusing, and organic solvent extraction. However, even if depletion of the most abundant proteins increased the peptides signal, it is well known that albumin can act as carrier proteins and bind a vast assortment of peptides and protein fragments prolonging their half-life in blood (*Karpova M.A. et al, 2010*). Several methods coupled with denaturating condition have been proposed to reduce the problem (ultrafiltration, continuous elution under denaturation condition). However, these strategies require complex and time-consuming procedures of protein depletion prior to MS analysis, limiting throughput and introducing other concerns about experimental variability, reproducibility, sample handling procedures, and protein stability during sample processing (*Findeisen P. et al., 2005*). The development of nanomaterials, with controllable physicochemical properties, has widely improved the use of nanotechnology in biomedical applications, and among these, in biomarker discovery.

6.3 *NANOMATERIAL-ASSISTED PROFILING STRATEGIES IN CLINICAL PROTEOMICS*

Nanomaterials have been considerably explored in biology due to the matching of their chemical, physical, structural, and topographical properties with those of the molecules and building blocks of most biological systems. The ability to develop objects and structures at the nanoscale and the possibility to precisely control their interactions with cells, organelles, enzymes, proteins, and nucleic acids (DNA or RNAs) permitted the emergence of nanomaterials in the biomedical field with a wide range of applications (*Ferrari M. et al., 2005; Riehemann K. et al., 2009*) such

as pharmaceutical, therapeutics, medical imaging, and diagnostics (*Ferrari M., 2005; Hu y. et al., 2011, Godin B. et al., 2011*). Among those materials, mesoporous silica has received growing interest because of its unique physicochemical properties. The discovery of mesostructured materials was made in the 1990s; the scientists of Mobil Oil Corporation discovered MCM-41 material in 1992. Due to the very ordered and adjustable pore diameters, uniform pore size distributions, large surface areas, confined pore volumes, and well-defined porous structures, mesoporous silica has been extensively applied to the adsorption, separation, immobilization, storage and identification of biomolecules. Mesoporous silica can be synthesized and assembled in different structures, shapes and dimensions (nanoparticles, nanowires, nanotubes, nanorods and thin films) (*Sakamoto J.H. et al., 2010*). Recently, properties of nanomaterials have been increasingly exploited for reducing the complexity of the complex sample (such as blood) and for isolating the low-abundance peptides without losing its clinical value. One of the first studies reported peptides and protein fractionation using zeolite nanocrystals establishing the proof-of-concept of the MS profiling assisted by nanomaterial. As a consequence, different mesoporous silica materials were tested for this purpose showing its suitable features and opening the possibility to extend the detection of LMW species beyond the limit of conventional diagnostic methods (*Zhang, Y. et al., 2005; Terracciano R. et al., 2006*; *Savino R. et al., 2012*).

In this scientific contest the Nanomedicine Department of the Methodist Hospital Research Institute (Houston, Texas) developed a size-exclusion strategy based on a Mesoporous Silica Chip (MSC).

This tool has variable pore dimensions, structures and surface chemistries that allow the efficient depletion of large proteins and the selective isolation of the circulating LMW proteome from body fluids samples (*Hu Y. et al, 2010*). This device is structured as a platform in which a small amount of sample is spotted on the mesoporous silica surface. In this way, many samples can be processed at the same time allowing high throughput processing. In combination with the consistency in the properties, the ease of manufacturing and the fast sample handling procedures it resulted in the development of a reliable, quick and powerful exploratory screening approach for biomarker discovery (Figure 2). In previous studies the Nanomedicine group optimized the synthesis of the MSC presenting various pore structures, chemico-physical characteristics of the porous surfaces and demonstrated the significant enrichment of the peptides and small proteins after using the MSC, confirming the depletion of the majority of the large molecules (*Bouamrani A. et al, 2010; Hu Y. et al, 2010*).

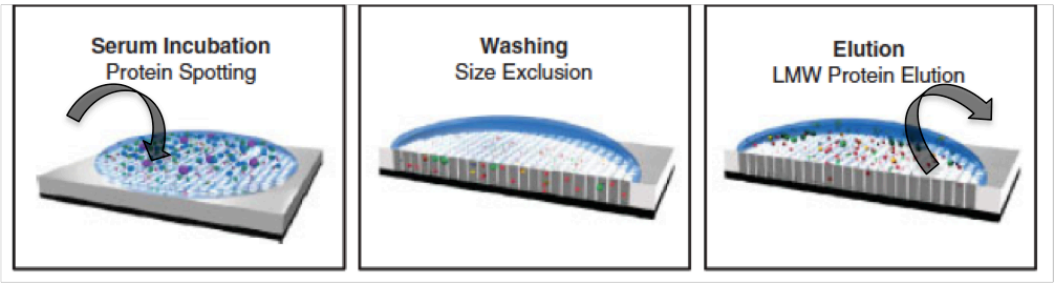


Figure 2. Principle of peptide fractionation using MSC. The rapid 3-step fractionation procedure consist of an incubation step, in which sample was incubated on the surface of the chip LMW proteins and peptides are trapped into the pores; a washing step in which the large proteins, excluded from the pores, were eliminated and an elution step for recovering the peptides subsequently analyzed by MALDI-TOF-MS (*modified by Bouamrani A. et al, 2010).*

In this study we performed a characterization of the mechanism of interaction between specific peptides and MSCs by varying the critical parameters involved in the process. We studied the influence of the peptide features, such as mass and charge, in relation with the physical and chemical properties of MSC, in order to evaluate the specific role of each of these features in the overall process of protein absorption, retention, and fractionation.

We then applied the optimized peptide fractionation procedure, in combination with MALDI-TOF MS, to investigate the circulating low molecular weight peptidome of locally advanced rectal cancer patients treated with neo-adjuvant chemo-radio therapy. Peptidome exploration was carried out at different time points along the neo-adjuvant treatment with the aim of identifying peptide expression patterns that could be able to predict the histopathological response to therapy.

7 MATERIAL AND METHODS

7.1 *MSC FABRICATION AND CHARACTERIZATION*

Mesoporous silica chip has been produced by the Nanomedicine Department at The Methodist Hospital Research Institute, Houston (Texas, USA).

The principal methodology for the synthesis of mesoporous silica is focused on the evaporation induced self-assembly procedure. Hydro-alcoholic solution of a soluble silicate and a structure-directing polymer are mixed and after dip or spin coating, the evaporation of the solvent drives silica/copolymer self-assembly into a uniform thin film. In particular, tetraethyl orthosilicate (TEOS; Sigma-Aldrich, St. Louis, MO, USA) is dissolved in a mixture of ethanol, distilled water and HCl. Separately Pluronic surfactant copolymer (BASF Co.) is dissolved in ethanol by stirring at room temperature. The silicate is mixed with the triblock copolymer solution followed by stirring of the resulting solution for 2h at room temperature. The pH of precursor solution is controlled at 1.5 to prevent the precipitation of mesoporous silicate and balance the procedure between the silicate hydrolysis and condensation of hydrolyzed silicate to polymer micelle. The coating solution is deposited on a Si (100) wafer by spin-coating at the spin rate of 2000 r.p.m. for 20 s. To increase the degree of polymerization of the silica framework in the films and to further improve their thermal stability, the films are heated at 80◦C for 12h. They are then calcinated at 425°C to remove the organic surfactant. The temperature is raised at a rate of 1◦C min^{-1}, and the furnace is heated at 425°C for 5h. The thickness of the thin film could be controlled by adjusting the concentration of polymer in the precursor solution, while the porosity is related to the ratio between copolymer and tetraethyl orthosilicate in the starting material. Oxygen plasma treatment is applied to ensure the hydrophilicity.

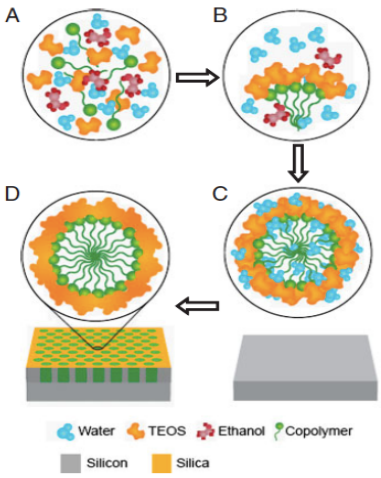


Figure 3. Scheme of chemical formation of the mesoporous silica film during the production of MSC for proteomic applications: A) Fresh coating solution; B) formation of micelles; C) evaporation-induced self-assembly during spin-coating process; D) view of a pore after aging at elevated temperature (*Figure modified from Bouamrani A. et al, 2010).*

The characterization techniques used to study the mesoporous silica include variable-angle spectroscopic ellipsometer (J. A. Woollam Co., M-2000DI) offering the thickness of the thin films and their porosities, transmission electron microscopy (TEM) used to acquire micrographs of the plane view of mesoporous silica thin films with an FEI Technai (FEI Co.) at a high tension of 200kV, N_2 adsorption/desorption applied to measure the adsorption/desorption isotherm and the pore size distribution. The hydrophilicity of the films is evaluated through contact angle measurement.

7.2 *MSC EXPERIMENTAL PROCEDURE*

For each experiment 5ul of standard protein solution or plasma was spotted onto the porous layer of the chip. Plasma was diluted in solution with acetonitrile (final v/v concentration of 5%) because, affecting the protein/protein interactions, it helped to release the low-molecular weight peptides associated with albumin or other carrier proteins that are removed using the MSCs. Then the solutions were removed and the wells were washed several times with 10ul of deionized water to remove surface bound material. Desorption of peptides was carried out by pipetting up and down for 20 times with 5ul of elution buffer made by a 1:1 (v/v) mixture of Acetonitrile (ACN, Sigma) and 0,1% of Trifluoroacetic acid (TFA, Sigma).

7.3 *EVALUATION OF PROTEIN REMOVAL FROM THE SURFACE OF THE MSC*

7.3.1 Protein concentration measurement

Protein concentration was measured using a microBCA assay (Pierce, Rockford, IL) and read with the absorbance at 562nm on a UV/Vis Spectrophotometer (Spectra Max M2, Molecular Devices). In order to determine protein recovery from the MCS, absorbance values were converted into protein amounts using the formula calculated through a standardization curve. For every experiment performed, results were internally standardized using known amounts of Albumin (0.5µg/ml to 30µg/ml) according to the instruction of the microBCA kit.

7.3.2 Surface characterization

7.3.2.1 Attenuated Total Reflectance Fourier Transform Infrared (ATR-FTIR)

Fourier Transform InfraRed (FT-IR) is a method of infrared spectroscopy in which IR radiation is passed through a sample. Some of the infrared radiation is absorbed by the sample and some of it is passed through (transmitted). The result is a spectrum of peaks that correspond to the frequencies of vibrations between the bonds of the atoms making up the material. Because each different material is a unique combination of atoms, no two unique molecular structures produce the same infrared spectrum. Therefore, infrared spectroscopy can result in a positive identification (qualitative analysis) of every different kind of material. In addition, the size of the peaks in the spectrum is a direct indication of the amount of material present.

Surface adsorption was investigated through Attenuated Total Reflectance Fourier Transform Infrared (ATR-FTIR) spectroscopy. ATR-FTIR is a sampling technique that enables samples to be examined directly in the solid or liquid state without further preparation. A beam of infrared light is passed through the ATR crystal in such way that it reflects at least once off the internal surface in contact with the sample. This reflection forms the evanescent wave that extends into the sample. Because the evanescent wave into the solid sample is improved with a more intimate contact, solid samples are usually firmly clamped against the ATR crystal, so that trapped air is not the medium through which the evanescent wave travels, as that would distort the results. The measured signal is digitized and sent to the computer where the Fourier transformation takes place. Before the sample is analyzed, it is necessary to measure a background spectrum that is normally a measurement with no sample in the beam. Thus, all spectral features, which are present, are strictly due to the sample.

In this study pieces of MSCs with different pore size were incubated with plasma and they were analyzed after removal of the flow through, after 2 washes and after 5 washes. Spectra were collected by placing the crystal on the surface of chips. Spectra were recorded using a Nicolet 6700 spectrometer (Thermo Scientific, Waltham, MA) with a Smart iTR diamond ATR attachment. Typically, 64 scans were performed in the wave number range of 4000–600/cm with a spectral resolution of 4 cm^{-1}. Clean chips were used as negative control. *(Introduction to Fourier Transform Infrared Spectrometry, 2001)*

7.3.2.2 X-ray photoelectron spectroscopy (XPS)

X-ray photoelectron spectroscopy is a quantitative spectroscopic technique that measures the elemental composition, the chemical state, and electronic state of the elements that exist within a material.XPS is a surface chemical analysis technique that can be used to analyse the surface chemistry of a material in its state and after some treatments. XPS spectra are obtained by irradiating a material with a beam of X-rays while simultaneously measuring the kinetic energy and number of electrons that escape from the top 1 to 10 nm of the material being analyzed. XPS requires ultra-high vacuum conditions.

A typical XPS spectrum is a plot of the number of electron detected versus the binding energy of the electron detected. Each element produces a characteristic set of XPS peaks at characteristic binding energy values that directly identify each element existing in or on the surface of the material being analyzed. These characteristic peaks correspond to the electron configuration of the electrons within the atoms. The number of detected electrons in each of the characteristic peaks is directly related to the amount of element within the area irradiated.

7.3.2.2.1 XPS surface spectra:

The surface adsorption of the chips was monitored by XPS surface spectra using a PHI Quantera XPS with monochromatic Al Kα X-ray (1486.7 eV) source. The 40W, 15kV and 200μm diameter X-rays were shot on the sample. The vacuum pressure during the analysis was always lower than 5.0E-8 Tor. The XPS survey scan spectra in the 1100.0 eV binding energy range were recorded in 0.5eV steps with a pass energy of 140eV. High-resolution scan spectra of C1s were recorded in 0.1eV steps with pass energy of 26 eV. Low energy electrons and Ar+ ions were conducted for specimen neutralization in each measurement. The spectral lines were referenced against the C1s signal at 284.50 eV. Like to the ATR-FT-IR study, we incubated pieces of different MSCs with plasma and protein solution and we analyzed them by XPS after removal of the flow through, after 2 washes and after 5 washes. The photoelectron peaks considered were C1s, Ni2, Si 2p. Clean chips were used as negative control.

7.3.2.2.2 XPS depth profiling:

In order to investigate the protein enrichment at different depths along the chip thickness, XPS depth profiling was carried out after Ar ion etching of the chips using a rapid etching ion gun (XP-HSIG, JEOL). Ar sputtering was performed over a 2 mm square area. Ar ion bombardment angle was 90°, while current and voltage

was 19 mA and 800 V. XPS spectra were measured after Ar bombardment each time. The sputter rate was determined at different sputtering times by measuring the pit depth after the experiment and dividing by the etching time. The pit depth was measured using a Veeco Dektak Series 3 Stylus Profiler System. To minimize effects of charging at the insulating films, a low-energy electron gun was used for charge neutralization. The photoelectron peaks considered were C1s, Ni2, Si 2p and the depth profiling was carried out in a range from 0 to 630 nm. The depth profiling was performed after each protocol steps: after incubation of the plasma, after the washing procedure and after elution step. The depth profiling of a clean chip was used as a negative control.

7.4 *FACTORS THAT INFLUENCE PEPTIDE ADSORPTION PROCESS*

7.4.1 Evaluation of the effect of time of incubation and of temperature on the adsorption kinetic equilibrium

Human serum and plasma from a healthy donor were bought from Sigma. For each experiment 5ul of protein solution or plasma/serum, added with ACN (final v/v concentration of 5%), were spotted onto the porous layer of the chip. Adsorption kinetic was investigated by incubating samples on a mesoporous silica chips (MSC) for 5, 15, 30 and 60 minutes. The effect of temperature on the amount of peptides retained in the pore was studied by processing the samples at 4°C, 25°C, and 37°C. The protein content of the elution was assessed by microBCA protein quantification method as described above and by MALDI-TOF-MS.

7.4.2 Study of adsorption of specific standard peptides on different pore size MSCs

Adsorption of Angiotensin II, ACTH (CLIP18-37), and Aprotinin on MSC was carried out following the procedure described in “MSC experimental procedure”. Effect of pH on peptide adsorption was evaluated after adjusting the single peptide solutions to pH 4, 7, or 10 by using different concentrations of TFA and ammonium hydroxide (NH_3OH). Adsorption kinetic of peptides on MSC was performed with 4 serial 1:5 dilutions to obtain concentrations ranging from 1.6 to 400pmol/ul.

To study the effect of increased protein complexity on the specific adsorption of peptides in solutions we investigated the recovery of each of the three peptides *I)* alone in solution, *II)* in a simple mixture of peptides (equal combination of Angiotensin II, ACTH (CLIP18-37), and Aprotinin), *III)* in a complex mixture of proteins with a higher range of molecular weight, and *IV)* spiked in a biological sample (plasma) that was checked for the absence of the specific peptide peak to avoid misinterpretation of the results. Each peptide elution and each peptide concentration was analyzed directly on MALDI-TOF-MS. One well was spotted only with the matrix as negative control, for observing matrix-specific peaks. The peptide adsorption study was performed using MSCs with different pore size and structure as described in the result section.

7.5 *ANALYSIS OF CIRCULATING LMW PEPTIDE PROFILE OF HUMAN PLASMA SAMPLES*

We performed the LMW fractionation of plasma from colorectal cancer patients collected during operatively or endoscopic treatment in the Second Surgery Clinic of The Department of Surgical, Oncological and Gastroenterological Sciences (University of Padova). Informed consent was obtained from all patients and the study was approved by the Medical Ethics Committee.

7.5.1 Peptide adsorption study of human plasma

We first investigate plasma samples of three controls (subjects healthy with clean colon by endoscopically check), and three colon-rectal cancer patients (histologically confirmed). We evaluated the peptide profiling of samples by changing the pH of sample incubation and by using different pore size MSC. In particular each sample was incubated at native pH (usually around 7-7.5) and at pH 5 by adding 1ul of 0.5% TFA (Thermo Fisher Scientific, Waltham MA, USA). Then, each sample was fractionated by using 4nm and 6nm pore size MSCs. We incubated the samples one hour onto the MSC and we washed the surface five times. We performed the analyses in triplicate, for each sample. Peptides samples were eluted from the pores by using 5ul of a 1:1 mixture of ACN:TFA (50%:0,1%). The elution samples were stored at -80°C. Elution analysis was performed by using MALDI-TOF-MS.

7.5.2 Analysis of LMW peptide profile of locally advanced rectal cancer patients treated with neo-adjuvant chemo-radio therapy (pCRT)

7.5.2.1 Patient characteristics

Between 2008 and 2010 a total of 69 patients with primary adenocarcinoma of the rectum underwent pCRT followed by surgery. The pre-treatment evaluation of the patients included a complete clinical history and physical examination, colonoscopy, complete blood cell count, transrectal ultrasound, pelvic computed tomography scan or magnetic resonance imaging, abdominal/ chest computed tomography, and carcinoembryonic antigen test. The inclusion criteria for pCRT were as follows: biopsy-proven adenocarcinoma of the mid–low rectum (<11 cm from the anal verge); clinical stage T3–4 and/or node-positive disease; and Eastern Cooperative Oncology Group performance status score of 0–2. 40 patients were included in the study group because they underwent surgery at our institution and they had at least pre- and post CRT plasma.

These patients were provided preoperative external-beam radiotherapy with high-energy photons (>6 MV) with conventional fractionation (>50 Gy in 28 fractions, 1.8 Gy per day, 5 sessions per week), and 5-fluorouracil (5-FU)-based chemotherapy administered by bolus or continuous venous infusion. A standard total mesorectal excision was performed 4 to 8 weeks after the completion of preoperative CRT. We also analyzed the peptidome profile spectra from plasma of 43 subjects with negative colonoscopy.

7.5.2.2 LMW peptidome profiling

We performed the LMW peptide profile fractionation from plasma of rectal cancer patients that was collected at different timing: before the pCRT (T0), two weeks after the initiation of pCRT (T1) and after the conclusion of pCRT, before the surgery (T2). For this study 5ul of plasma with 5%ACN and pH set around 5 was spotted onto the surface of 6nm pore size MSC. We incubated triplicates of each sample for one hour and we washed the MSC surface five times. Peptides samples were eluted from the pores by using 5ul of a 1:1 mixture of ACN:TFA (50%:0,1%). On each MSC we spotted and fractionated commercial plasma bought from Sigma-Aldrich, in triplicate. The elution samples were stored at -80°C.

7.6 *MATRIX-ASSISTED LASER DESORPTION/IONIZATION TIME OF FLIGHT MASS SPECTROMETRY (MALDI-TOF-MS)*

MALDI-TOF-MS provides for sensitive detection of low molecular weight proteins, whereas other traditional methods of analysis are employed for detecting larger proteins. This characteristic yields the MALDI-TOF-MS ideal for providing new insights into small peptides and protein components.

Typically a small amount of analyte is co-crystallized on a specific plate with a large molar excess of matrix molecule, a UV-absorbing weak organic acid. Matrix is a low molecular weight organic compound comprising of an aromatic ring and carboxylic acid that provides the chromaphore for absorbing laser energy at the correct wavelength and creating an acidic environment for protonating the analyte, respectively. The matrix is usually present in 1000- to 10000-fold excess of the analyte to protect it from laser ablation decay. A nanosecond-duration laser radiation pointed on this mixture causes both the sample and the matrix to be vaporized and released (desorbed) from the surface. The matrix enhances indirectly the analyte vaporization by strongly absorbing laser light energy. Furthermore the matrix is used as a proton donor and receptor allowing the ionization of the analyte in both positive and negative ionization modes, respectively. The ionization depends on the matrix-analyte combination, but it is not dependent only on the number of acidic or basic group of analyte. The mass analyzer typically used with the MALDI ionization source is a linear-time-of-flight (TOF) or a TOF reflectron. The linear TOF is the simplest devices and the analysis is based on accelerating a set of ions to a detector where the ions have the same energy. The separation occurs in a vacuum and the arrival time at the detector is dependent upon the mass to charge ratio, since the kinetic energy of the ions is equal for each ion. Ion species with low mass/charge (m/z) ratios are accelerated to higher velocities and reach the detector before the ions with higher ratio. When the ion reaches the detector, this produces a single electric impulse that is converted to digital signal. Single laser pulse produces a quantity of ions that are detected in a millisecond timescale, allowing the acquisition of data from multiple laser pulses in a short time. The resolution of this instrument is enhanced by the introduction of an electrostatic analyzer, the reflectron that extend the ion way, increasing the amount of time that the ions need to reach the detector. The accuracy and the resolving power is increased by the delay extraction function that “cools” the ions immediately after the MALDI ionization. This cooling period reduce the temporal spread of ions once they enter to the TOF analyzer. This technique detect a relative low mass range, typically <10000 m/z and it provides optimal performance for m/z up to 4000. Furthermore fast ions, with low m/ z,

produce stronger detector responses than slow ions (*Hortin G.L. et al, 2006; Lewis J.K. et al, 2000*).

7.6.1 Analysis of proteins and peptides by MALDI-TOF MS

During desorption/ionization of proteins by laser pulses, both positive and negative molecular ions are formed, usually single charged. The yield of positive ions is greater for most proteins and peptides and for this reason they are usually analyzed in the positive ion mode. The ionization efficiency of the peptides is highly variable, with preferential ionization of peptides containing Arginine. Multiple ionic forms, doubly charged ions or ions coupled with sodium, matrix or other ions, complicate the profile and they depend on variables of sample composition, sample preparation and type of matrix used. There is competition among different proteins and salts in a specimen for ionization and this causes the phenomenon termed ion suppression, which reduces the sensitivity of detection and the yield of specific ions.

Analysis of low molecular weight peptides was performed using α-cyano-4-hydroxycinnamic acid (CHCA,Sigma-Aldrich, St. Louis, MO, USA). A matrix saturated solution of CHCA in 100% ACN and 0.1% TFA in ratio 1:1 (v/v) was used for peptide detection. We spotted four or three elution replicates of each sample (of the different study analyses performed) onto the MALDI plate. We used 0,5ul of sample elution that was let it dry and then we spotted 0,5ul of matrix solution onto the sample. MALDI-TOF MS experiments were performed on a 4700 MALDI-TOF/TOF Analyzer (Applied Biosystems Inc., Framingham, MA, USA) operating in linear and reflectron positive ion mode using a 337 nm nitrogen laser. In linear mode samples were evaluated at the range of 800-10000 m/z with focus mass at 4000 m/z. Settings were optimized at a laser intensity of 4000. In reflectron mode, samples were evaluated in the range of 600-4000 m/z with focus mass at 2000 m/z Before any analysis the plate's calibration settings were update through external mass-calibration (m/z) based on the monoisotopic values of [M + H]+ of Bradykinin (1-7), Angiotensin II, P14R, ACTH clip (18–39), Insulin oxidized B chain at m/z of 757.3997, 1046.5423, 1533.8582, 2465.1989, 3494.6513 respectively (ProteoMass peptide MALDI-MS calibration kit, Sigma-Aldrich). The single spotted sample was analyzed automatically. Each mass spectrum was averaged with three thousand laser shots across the spot. In each MALDI plate we spotted a negative control, made by spotting CHCA alone and triplicates of a commercial plasma sample elution (Sigma-Aldrich).

7.6.2 Peptide identification: MALDI-TOF/TOF analysis of plasma samples

MSC elutions of plasma rectal cancer patients collecting at different timing were used for identifying different peptides. Spectra have been obtained in the range 500-3500Da with a Bruker Ultraflex II TOF/TOF instrument equipped with a nitrogen laser (λ=337 nm) and operating in reflectron positive ion mode. A peptide calibration standard mixture (calibration range 700-3200 Da) has been used for the instrumental calibration before analysis. For each spectrum, at least 2500 laser shots were collected. Once the most interesting peptides were selected, their fragmentation spectra were acquired in the laser-induced dissociations (LID) mode (LIFT® technology, Bruker Daltonics).

- *Compounds and sample preparation:*

We prepared the matrix by dissolving the α-cyano hydroxyl cinnamic acid (CHCA) in a 0.1% TFA/acetonitrile 50/50 solution. Peptide identification of elution sample was performed without using any enzymatic digestion.

We mixed 5µL of MSC elution sample with 5µL of a saturated solution of CHCA. 1 µL of mixed solution was placed near the calibrants onto the MALDI plate and left to crystallize before analysis.

- *Proteins and peptides identification procedure:*

Before the identification procedure, all spectra have been manually processed in order to eliminate peaks originating from the HCCA matrix. A direct comparison of samples with a blank solution (HCCA alone) was used. As shown in Figure 4, we identified several matrix adducts in the mass range 500-900Da (e.g., at m/z 568, 656, 666, 672 and 877).

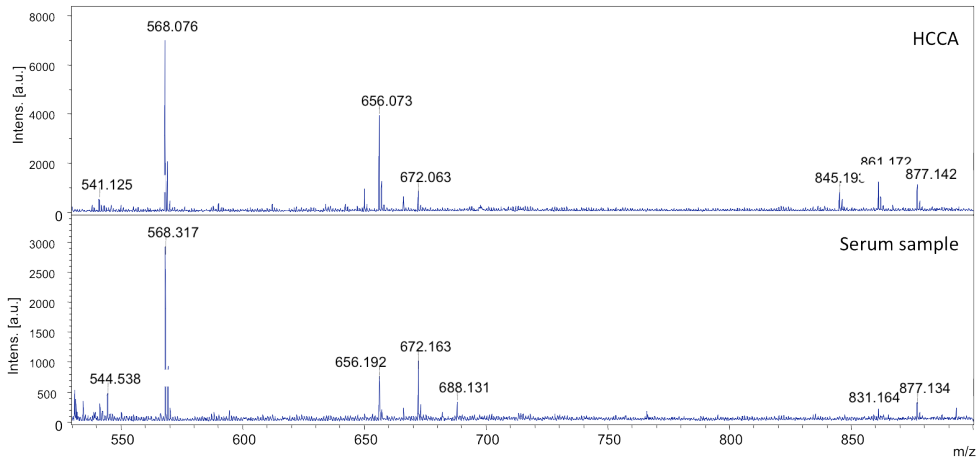


Figure 4: HCCA alone spectrum (upper panel) compared to a sample spectrum (lower panel).

After the matrix peaks removal, MS/MS analysis has been performed in order to obtain the fragmentation spectra. The fragmentation was performed directly on the peptides presented on the total MS spectra; each m/z of interest (precursor ion) was selected and its fragments (product ions) were simultaneously detected. Fragment ions cannot be observed in linear mode because they have the same velocity as the precursor ion. However in a reflectron TOF mass analyzer these fragments penetrate the reflectron at different depths and they arrive at different flight times at the reflector detector. These fragments are represented in the mass spectrum and provide structural information. The obtained peptide mass fingerprints, as well as the precursor ion's molecular weight are used to search a database.

Fragmentation spectra have been processed using *flexAnalysis* (version 3.3) software and, after baseline subtraction, a peaks list has been generated using *Snap* as peak detecting algorithm.

The peptide identification was performed by correlating the peaks list produced through the experimental MS/MS spectra with a peaks list from theoretical spectra predicted for each peptide and contained in a protein sequence database.

The peaks list has been used for peptides/proteins identification via on-line search engines such as: Mascot (http://www.matrixscience.com) and MS-Tag (http://prospector.ucsf.edu). The search has been performed using large and comprehensive protein databases (such as NCBInr) when possible.

A summary of the most important parameters used it for the on-line search is reported:

Database	UniProtKB
Enzyme	No enzyme
Taxonomy	Homo Sapiens
modification	None
modifications	Oxidation (HMW)
tolerance	1.2 Da
tolerance	0.6 Da
Instrument	MALDI-TOF/TOF

In both the search engines that we used, MS/MS data were matched against sequence tags predicted for all proteins in the selected database. This means that all searches are based on comparisons between the experimentally observed fragmentation pattern and the predicted fragments for all hypothetical peptides having the same molecular mass. A series of hits is obtained, ranked by a MOWSE (Molecular Weight Search) based scoring.

In the case of Mascot search engine, a typical results report consists of a protein score graphically represented (Figure 5) and of a peptide view window in which it is possible to check the fragment matching. This ions score is calculated as -10xLog (P), where P is the probability that the peptide match is due to a random event.

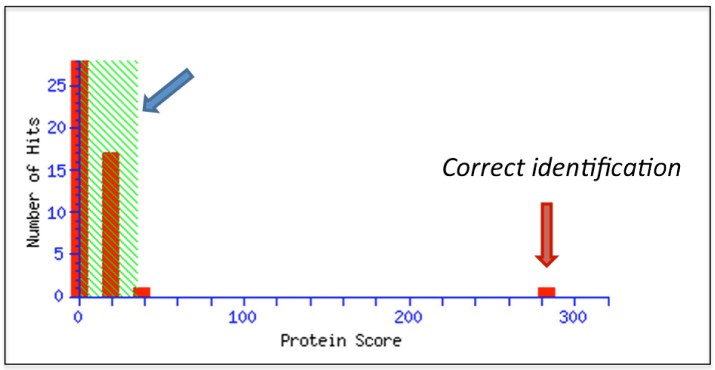


Figure 5. Graphical protein score in which proteins are represented by the red bars. The proteins labelled with the blue arrow are under the significant threshold.

For several peptides, for which a correspondence in a search engines were not find, the peptide sequence was obtained directly from the spectra, without referring to sequence database help. The primary sequences were identified using protein BLAST (Basic Local Alignment Search Tool), a tool that calculate sequence similarity that aligns the peptide sequence with a BLAST databases of protein sequences available from NCBI. The program compares protein sequences to sequence database and calculates the statistical significance of matches.

7.7 *DATA PROCESSING AND STATISTICAL ANALYSIS*

Before conducting the statistical analysis, the raw t_2d MALDI-TOF-MS files were converted to text files using ConvertPeakList software. Then they were imported into SpecAlign software version 1.22 for the pre-processing procedure: all spectra were aligned using the PAFFT correlation method and intensity was normalized to TIC. The baseline was corrected and the negative values were removed prior to the analysis.

7.7.1 Study of adsorption of specific standard peptides on different pore size MSCs

Absolute intensity of four replicates of each peptide was evaluated by calculating the mean, standard deviation and the standard error. Then the mean values were used for subsequent analysis. We performed kinetic analysis of MALDI-TOF MS

signal intensity in relation to the different peptide concentrations and the different factors that potential affect the fractionation procedure: time of incubation, temperature of incubation. Kinetic analysis was performed for each peptide, incubated in solution at different pH and fractionated with different pore size MSCs. Statistical analysis among the different conditions tested were performed by using Mann-Whitney non-parametric test or ANOVA test followed by student's t-test for unpaired data, and a significant difference was assumed for p-value<0.05. Bars plot, kinetic curves and statistical analysis were performed with GraphPad Prism version 5.0

7.7.2 Analysis of circulating LMW peptide profile of human plasma samples

Absolute intensity of triplicates of each peptide was evaluated by calculating the mean, standard deviation and the standard error. Then the mean values were used for subsequent analysis.
We used classification techniques, such as cluster analysis (CA) and principal component analysis (PCA), as useful statistical tools able to discover structures in the experimental data. These techniques have the advantage to reduce the dimensionality of the data, and to allow independent classification of cases.

7.7.2.1 Principal component analysis (PCA)

PCA is a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables. The data are represented in a dimensional space of *n* variables, which are reduced to a few principal components that are linear combinations of the original variables, indicating the maximum variation within the data. PCA is a method of data visualization useful in identifying new meaningful underlying variables, and it can detect the presence of clusters within multivariate data.
Sample analysis was performed using MarkerView™ Software v1.1 (AB Sciex, Concord, Canada) on the matrix of m/z values for all patients and controls. The software provides for supervised (PCA-DA) and unsupervised (PCA) analysis. Unsupervised methods allow the structure within the data to be determined and visualized. PCA processing was carried out unsupervised mode with a Pareto scaling method. The loading plot provides valuable insight into variables that lead to sample clustering illustrating which variables are up- or down-regulated. Supervised methods use prior knowledge of the sample groups to determine the variables that distinguish the groups. For this purpose, MarkerView software

performs a combination between PCA and the supervised technique, the Discriminant Analysis (DA). This software also allows t-test analysis to be performed, useful for determining which variables lead to the significant differences between groups, through the calculation of the probability for each m/z variable to distinguish two different groups. We also analyzed each variable with Mann-Whitney non-parametric test to enhance differences of median values. We performed multivariate analysis between control subjects and rectal cancer patient plasma collected before the pCRT, for identifying a peptide profiling peculiar of the rectal cancer disease. Then we examined the peptide profiling of rectal cancer patient in relation to the histopathological response to the treatment. In this contest, we compared plasma peptides relative abundances of Good and Poor responders collected in different time point: before the pCRT (T0), during the pCRT (T1), and after the pCRT (T2). We also performed analysis in relation to the time-points of the treatment, to monitoring the potential fluctuation of peptide abundances over the time of treatment in a individual group of response.

7.7.2.2 Cluster analysis

Another classification method is Cluster analysis. Hierarchical clustering was performed using Cluster software and visualized with TreeView Software. The MALDI-data were log-transformed, normalized and median centered. Pearson correlation was used to calculate the distance between the samples and complete linkage clustering was performed.

For each analysis that was performed, a pattern of several significant peptides were identified. Each peptide was individually analyzed to observe its discriminative power according with the specific analysis. We carried out Mann-Whitney non-parametric test and *receiver-operating characteristic curves* (ROC-Curve) to compare the ability of the different ionic species to discriminate between the different groups of patients. Sensitivity and specificity, statistical measures of the performance of a diagnostic test, were calculated on the basis of ROC-Curves. Sensitivity is defined as the probability of a positive test among patients with disease (true positive) whereas specificity is the probability of a negative test among patients without disease (true negative). The area (AUC) under the ROC curve indicates the ability of the test to discriminate between patients.
Box plots and Roc curves, kinetic curves were performed with GraphPad Prism version 5.0

7.7.2.3 Multiple Logistic regression analysis

The specific panels of m/z ionic species found at T0 and T2 time point, in relation with the histopathological response, were also analyzed by multiple logistic regression analysis in order to investigate the adjusted association of the selected m/z ionic species. The m/z ionic species associated with the probability of histopathological non-responsive event (Poor responders with TRG: 3-5) were determined by univariate analyses using the logistic regression model. Variables with p-value<0.3, in univariate logistic regression analysis, were included in the multiple regression. The best subset of predictor variables was determined using the stepwise selection method according to a selection entry and stay criterion of 0.1. The goodness of fit for the final logistic regression model in interpolating the observed data was evaluated using the Hosmer-Lemeshow test. The test assesses whether or not the observed event rates match the expected event rates in subgroups of the model population. It is similar to a Chi Square test, and indicates the extent to which the model provides better fit than a null model with no predictors, or, in a different interpretation, how well the model fits the data, as in log-linear modeling. The Hosmer and Lemeshow's goodness of fit test divides subjects into deciles based on predicted probabilities, and then it computes a chi-square from observed and expected frequencies. If chi-square goodness of fit is not significant, then the model has adequate fit the data. The area under the ROC-Curve was used to quantify the predictive performance of the fitted model. An AUC-ROC equal to 1 denotes perfect discrimination between distinct groups of patients, while a value equal to 0.5 denotes the lack of discrimination. It was computed with 95% confidence intervals.

8 RESULTS

8.1 *MESOPOROUS SILICA CHIP (MSC) FABRICATION AND CHARACTERIZATION*

The MSC were produced by the evaporation-induced self-assembly procedure under acidic conditions using Pluronic triblock copolymers as structural templates.
Different nanostructures and pore configurations can be produced by varying the molecular weight of the block copolymer, the ratio of the hydrophilic (PEO) to hydrophobic (PPO) block volume in the amphiphilic polymer, and the proportion of the polymer in the starting material. We used two typical triblock copolymers with different molecular weight and composition:
PEO_5 -PPO_{70} -PEO_5 (Pluronic L121), PEO_{106} -PPO_{70} –PEO_{10} (Pluronic F127).
Furthermore, by adding a swelling agent to the precursor solution, it is possible to enlarge the MSC pore size. In table 1 and in figure 6 are shown the characteristics of MSC that we used in this study.

Table 1. Characteristics of Mesoporous Silica Chips used for protein adsorption study

Sample label	Surfactant	Pore structure	Average pore size (nm)	Porosity (%)	Contact angle
4nm	F127	Cage-like	3.7	56.0	18.0
6nm	L121	Random	6.0	55.0	7.4
7nm	L11+PPG	Radom	6.7	57.0	10.4

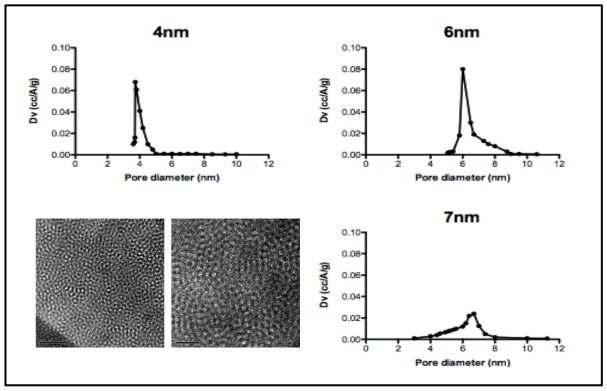


Figure 6. Chip characterization: Pore distribution graphs of 4nm, 6nm and 7nm MSC, TEM images of the MSC surface structure.

8.2 *EVALUATION OF PROTEIN REMOVAL FROM THE SURFACE OF THE MSC*

In order to investigate the adsorption of specific peptides into the MSC pores, a verification of complete removal of proteins stuck on the chip surface was needed. We spotted commercial serum and plasma on the MSC and we performed different numbers of serial washing steps. After each washing step, we measured the protein content with *microBCA* kit and we performed a MALDI analysis on both the washed and eluted samples. The amount of proteins removed during the washing procedure decreased increasing the number of the washing steps (Figure 7A). After five washes, proteins were almost absent on the MSC surface and the amount of proteins recovered from the chip with the elution step was significantly higher than the amount of proteins measured in the last 2 washing steps.

Moreover, we characterized the surface of the chip before and after the washing steps by using FTIR and XPS (Figure 7B). As shown in the FTIR spectra, the clean chip (control chip) displayed peaks at 1059.29 cm^{-1} , which originated from the Si-O bond of the silica surface (*Roy S. et al, 2010*). In contrast, the spectra produced with the chip after protein incubation showed weaker peak for Si-O moieties, which is due to the immobilization of proteins on the chip shielding the silica surface. The spectra also presents bands at 3000 and at 2800 cm^{-1} assigned to C-H stretching modes of CH_2 and CH_3 groups of protein and bands at 1474, 1460 and 1398 cm^{-1} attributed to C-H asymmetric deformation modes of the alkyl groups. Moreover, two strong peaks at 1645 and 1525 cm^{-1} on the spectra corresponded to amide I and amide II modes that link amino acids of the protein adsorbed on the silica surface. The amide I band primarily represents the C=O stretching vibrations of the peptide bond groups while the amide II band is caused by C-N stretch coupled with N-H bending mode (*Kulikova G.A. et al, 2010*). However, as shown in the IR spectra, these peaks decreased significantly after two washes and almost disappeared after five washes. Conversely, increasing the washing steps, the peak originating from the Si-O bond of silica surface became significantly sharpened, as well as the peak shown in the spectra of control chip.

Figure 7B (on the right) displays XPS survey peaks attributing to relative amount of C, N and Si atoms of protein adsorbed on the chip surface before and after the washes. The chip before washing procedure exhibited a significant amount of N and C atoms which originated from the coated protein molecules while the chip after the washing steps showed diminishing peaks of these atoms and increasing peak of Si atoms. On the basis of these considerations, we concluded that a number of five

washes were optimal to remove proteins from the chip surface and we performed all the experiments described in the paper accordingly.

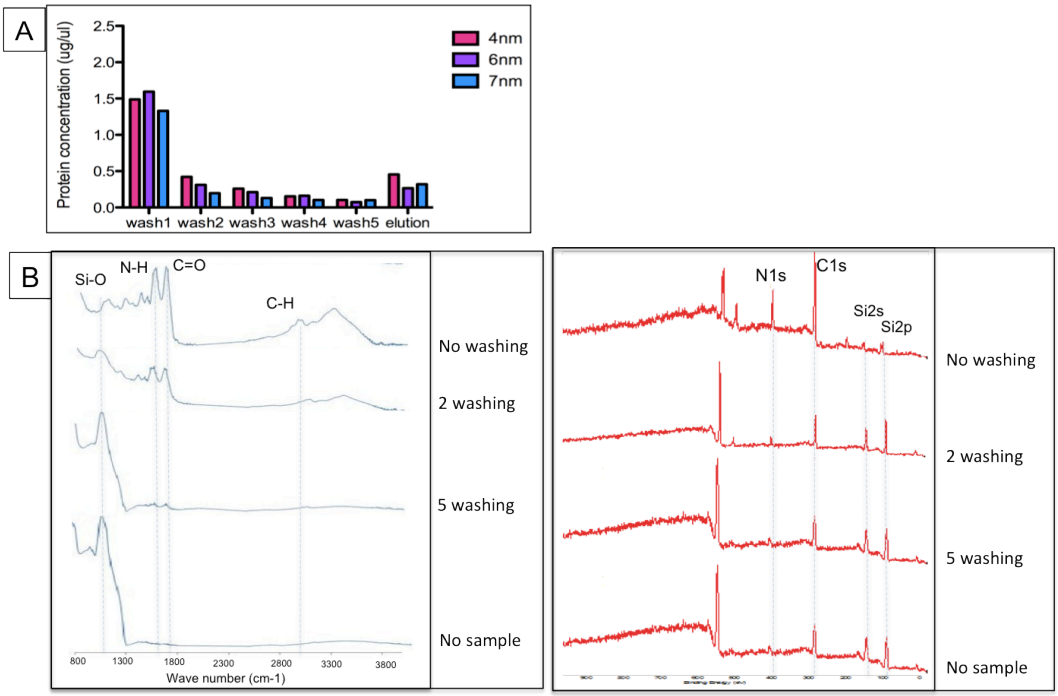


Figure 7. A) Protein concentration recovery from each washing step and from the elution step using 4nm (pink) 6nm (purple) and 7nm (blue) MSC . B) ATR-FTIR spectra (on the left) and XPS characterization (on the right) of MSC surface after sample loading and sequential washing compared with a clean MSC. The graphs show the presence of protein stuck on the surface when the sample is loaded and in the first two washing steps. After 5 washing the presence of proteins are almost absent.

8.3 *XPS DEPTH PROFILING*

In order to quantify the amount of peptides penetrated into the MSC and to determine the depth at which the peptides can reach within the chips, XPS depth profile analyses were performed. This study allowed the possibility to quantify the capture and reclaim of peptides from the MSC as a function of depth, with nanometric resolution. By means of surface XRD analysis, coupled with sequential removal of superficial MSC layers through *Ar* sputtering, we performed a depth profile of the nitrogen and carbon, typical protein components, and of Si, the component of MSC. The intensity of the scanned elements was recorded after each etching and plotted versus the depth (nm) from the surface. At first, we carried out a series of depth profiles on a clean MSC film. We then spotted plasma sample and we carried out the profiling of nitrogen, carbon, and silicon atoms in depth both before and after the washing procedure. A clean chip showed the presence of silica component along the depth of the chip (data not shown). Carbon environmental

contamination is found mainly in the proximity of the surface, but the presence is still tracked deeper into the film, while environmental nitrogen was absent. On the basis of these preliminary results we exploited the presence of nitrogen in amino acids to study the penetration of peptides into the porous MSC. After the sample spotting, a thick protein layer was present on the MSC (Figure 8A) that prevent the depth profiling analysis. Figure 8B shows the XPS depth profile after the washing procedure. The signals of silica and nitrogen showed an opposite behaviour. Initial Si 2p signal increased with increasing the depth of etching and after 150 nm the signal was stabilized. The presence of a low Si 2p signal is due to the fact that XPS does not only read the top surface, but actually penetrated the surface for some nm in depth. On the opposite side, N signal decreased with the depth: after 50nm the signal decreased by about 75% compare to the signal found near the surface, while after 100nm the signal stabilized. This indicated that the higher amount of peptides is recovered in the first 50nm in depth from the surface. After the elution (Figure 8C), nitrogen signal decreased indicating that the elution was effective along the depth in which there is the higher quantity of peptides.

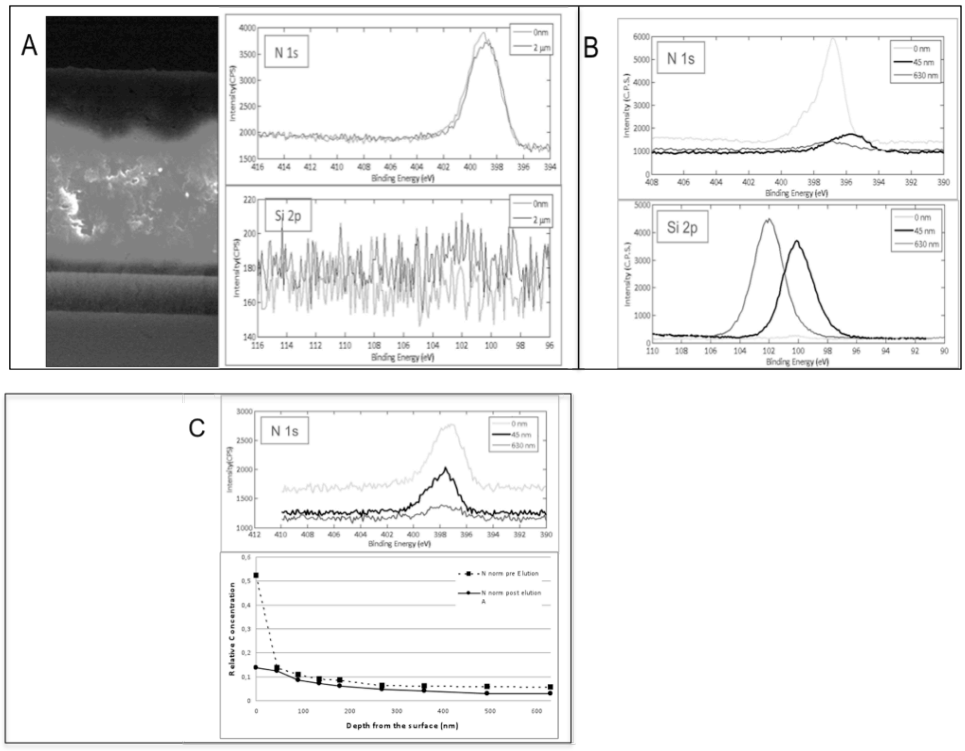


Figure 8. A) TEM of a MSC cross section after plasma spotting and XPS spectra of N1s and of Si 2p both at the plasma surface and at 2um in depth from the surface. B) XPS spectra of a MSC at 0 nm, 45nm, and 630nm depth of nitrogen and silica after 5 washing. The protein density within the chip varies in depth. C) XPS depth profile spectra of a MSC at 0nm, 45nm, and 630nm of Nitrogen after elution with the buffer and graph of the protein recovery rate after elution. The protein recovery within the chip is effective in the first 45nm in depth from the surface.

8.4 *FACTORS THAT INFLUENCE PEPTIDE ADSORPTION PROCESS*

8.4.1 Effect of time of incubation

Commercial plasma was incubated for different time (5, 15, 30, and 60 min) on the MSCs and, after five washings, elutions were performed and measurement of protein concentration was made with *microBCA* kit. The adsorption kinetic of the peptides on the MSC pores as a function of contact time is shown in Figure 9. The Langmuir isotherm model, the most widely used protein adsorption model, adequately reproduces our experimental data and the parameters are summarized in Figure 9A. Langmuir adsorption isotherm is characterized by binding site saturation over the time. A possible explanation refers to the time to reach the equilibrium in which the available pore surface is completely filled with adsorbed molecules. As predicted by the model, the adsorption rate is fast, with half of binding sites occupied in less than 6 min and the equilibrium capacity is reached after 30 min. No major differences were found in the adsorption kinetic for the three different chips in the present study. The fast equilibrium coverage of the adsorbent constitutes an essential parameter for a rapid and efficient enrichment procedure.

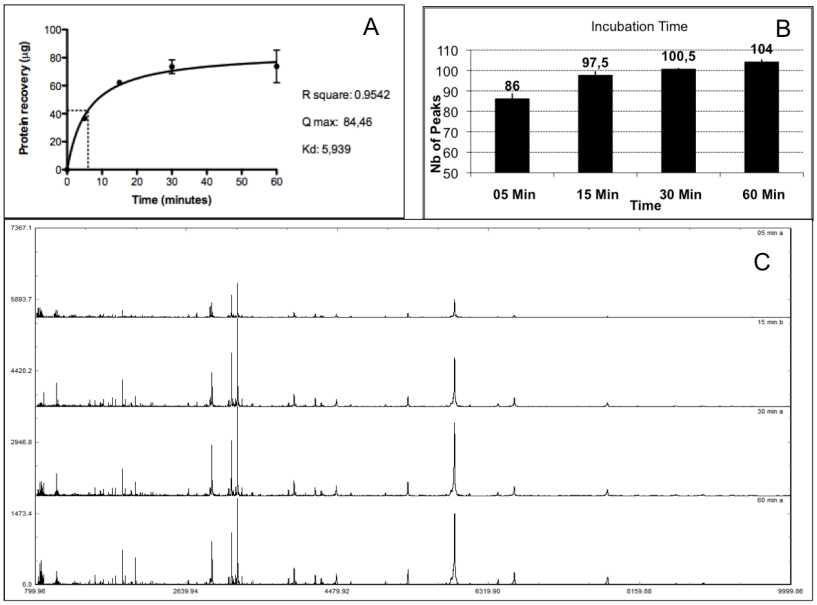


Figure 9. Kinetic of proteins adsorption as function of incubation time. A) Kinetic of the recovery after fractionation of protein sample, at different incubation time. The kinetic fit with Langmuir isotherm model. B) Histogram of the number of peaks recovered at different protein sample incubation time. C) MALDI-TOF MS spectra of proteins recovered after elution at different incubation time. Mean values and standard deviations of 3 independent experiments are shown.

On the basis of these results, we performed the subsequent experiments with a sample incubation of 60 minutes to reach the equilibrium capacity.

8.4.2 Effect of temperature

Adsorption protein behavior was studied under three different temperature conditions 4°C, 25°C, and 37 °C, respectively (Figure 10). In Figure 10A we reported the protein content for both the washing steps and for the elution. The amount of proteins removed through the washing steps was low at 4°C, compared to 25 °C and 37°C which gave comparable results. Similar results were observed for protein recovery in the elution, indicating a lower protein harvesting at 4°C. The difference in protein recovery could be explained by the mechanism of protein diffusion into the pores. Increased adsorption rates can be expected due to an accelerated diffusivity of proteins towards the adsorbent surface.

At low temperatures, molecules have a lower kinetic energy to enter the pores. Higher temperatures increase the kinetic energy of molecules enhancing the rate of diffusion as well as the probability of interaction between the proteins and the silica surface chip. Although the amount of proteins was analogous at 37°C and 25°C, the MALDI-TOF MS spectra showed better results in terms of number of detected peaks (Figure 10B) and peaks intensity (Figure 10C) when the procedure has been done at 25°C. Protein recovery at 37°C can be affected by degradation processes (proteolytic and enzymatic process accelerated the degradation of peptides).

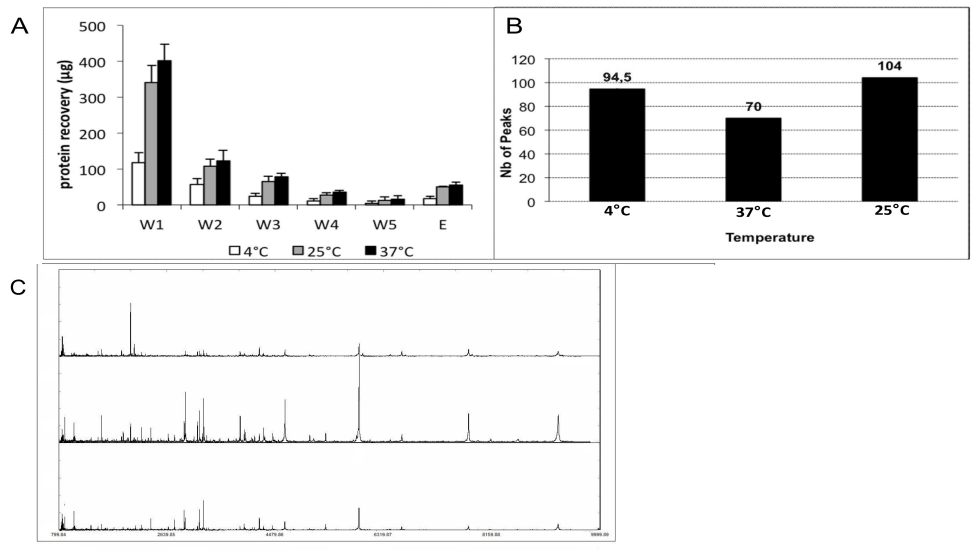


Figure 10. Effect of temperature on the amount of protein retained in the MSC. A) Histogram of the amount of proteins washed out from the surface for each washing step in relation to the amount of proteins recovered from the pore after elution, at different temperature: 4°C, 25°C, and 37°C respectively. B) Histogram with the number of peaks obtained by a MALDI analysis of elution at different temperature. C) MALDI spectra profile of the elutions at different temperature. Mean values and standard deviation of three independent experiments are shown.

8.4.3 Study of adsorption of specific standard peptides in different pore size MSC

Adsorption behavior of three peptides with molecular weight ranging from 1000Da to 6500Da and with different isoelectric points was specifically studied at different pH, with increasing environmental complexity, on both the three different MSCs. The physicochemical properties of the peptides are described in Table 2 and the structures are shown in Figure 11.

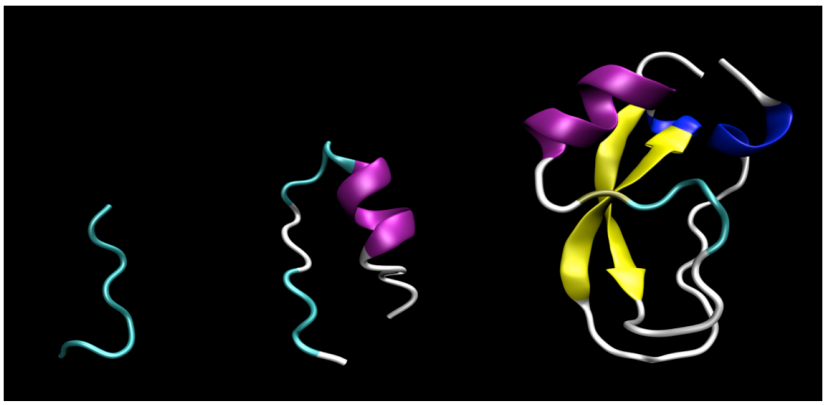


Figure 11. Conformation of the three peptides used in the adsorption study: Angiotensin II, ACTH(CLIP), and Aprotinin.

Table 2. Physicochemical characteristics of the three peptides analyzed

	Number residues	pI	Molecular weight (Da)	Max diameter (nm)	Charge at pH7 (e)
Angiotensin II	8	7.7	1046	1.7	0.1e
ACTH(CLIP17-38)	22	4.3	2466	2.0	-3e
Aprotinin	58	10.5	6511	3.1	+6e

In Figure 12, adsorption isotherms of the three proteins are shown for 4, 6, and 7nm MSC pore size. Concentrations ranging from 1,6pmol/ul to 400pmol/ul for each peptide were incubated on the different MSCs and were processed according with the optimized procedure. We run the elution products on MALDI-TOF and then we plotted the absolute intensities of each peptide as a function of the initial concentrations. Each graph presents the specific peptide adsorption curve for the different MSC.

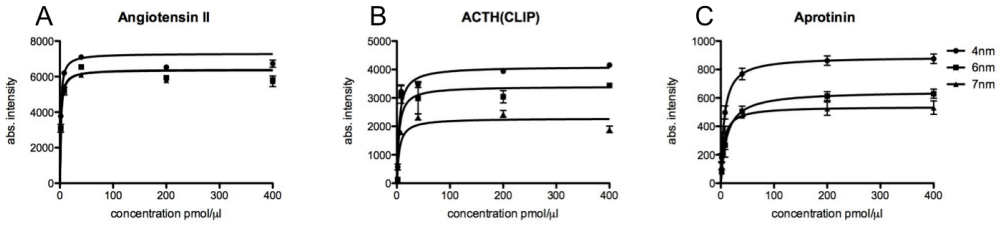


Figure 12. Adsorption kinetic isotherm of a specific peptide incubated at a pH of 7 on a MSC synthesized with pore average of 4nm, 6nm, and 7nm. The x-axis indicated the peptide concentration, the y-axis indicate the absolute intensity of peptide signal at MALDI-TOF. (A) Angiotensin II, (B) ACTH(CLIP 18-37), (C) Aprotinin. Error bars correspond to standard deviation of 4 replicates.

The Langmuir isotherm model seems to reproduce adequately the experimental data for all the peptides tested. The sharp initial rise of the curves suggested a high affinity between the protein and the adsorbent surface. Increasing the concentration of the peptide the isotherms reached a plateau because of the saturation of the pores. As in the kinetic equilibrium, the Langmuir equation is expressed as:

$$Y = \frac{Q_{max} * C}{K_d + C}$$

where C is the peptide concentration, Q max represents the maximum adsorption amount (the pore saturation uptake) and Kd is the adsorption constant. The uptake rate of the peptide on MSC is expected to depend strongly on the pore size structure of the sieve material as well as on the molecule dimensions. However, the graphs showed a higher uptake capacity with 4nm pore size MSC irrespectively to the dimension of the peptide. The data can be explained considering that each peptide can fit in all the tested pore sizes but the 4nm MSC possesses a higher number of pore entrances compared to the MSC with larger pores and that increases the available surface improving the capacity rate.

8.4.3.1 Effect of pH on protein adsorption capacity

The protein adsorption efficiency and stability on MSC is strongly affected by the pH of the solution. The driving forces for the peptide uptake on the silica surface include different kind of interactions between proteins and between silanol groups and proteins. pH of the solution determines the electrostatic state of proteins causing differences in surface charge and surface charge density.

Electrical charges of the proteins results from the ionization of carboxylic, tyrosyl, amine, and imidazole groups of amino acid side chains that can drive intermolecular attraction and repulsion between amino acids residues, while hydrophobic and electrostatic interaction can be performed between amino acid residues and silanol groups on the surface of the silica pore.

We investigated the absorption recovery of the three peptides at three different pH: 4, 7, and 10 respectively. The silica surface is mainly negatively charged at pH>2, so the important factor is the isoeletric point (pI) of the peptide and the amino acid residues that can be electrically charged.

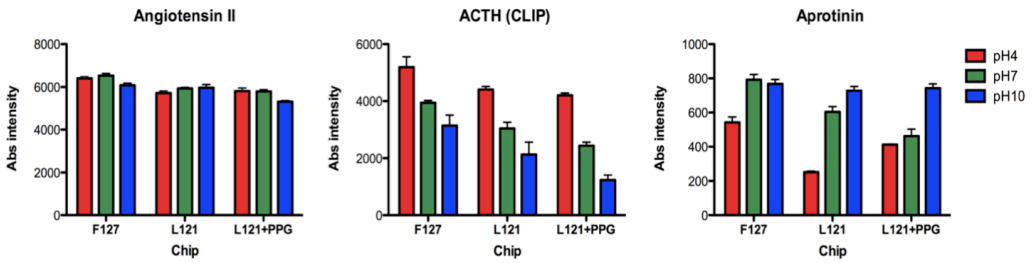


Figure 13. Effect of pH on peptide adsorption capacity. Histograms of absolute intensity recovered with MALDI-TOF MS (x-axis) after peptide incubation on MSC with different pore size (y-axis). Peptides were incubated in solution with pH4, 7, and 10. A) Angiotensin II; B) ACTH; C) Aprotinin. Error bars correspond to standard deviation of 4 replicates.

In Figure 13 three histograms related to the three specific peptides are shown. MALDI-TOF signal intensity that correlates with the recovery of peptide is plotted in relation to the pH and the different MSC. The graph 13A shows the adsorption recovery of the smaller peptide, Angiotensin II. In contrast with what we expected, the uptake rate is not significantly affected by the pH. The isoelectric point of Angiotensin II is around 7, hence, the peptide is positively charged at pH below its pI and negatively charged when the pH is higher. However, Angiotensin II is composed by eight amino acids with only two amino acids chargeable residues. This information, according with the experimental data, suggested that the net charge variation of the protein at pH4 and pH10 might not be enough to significantly affect the interactions involved in the peptide adsorption. Moreover, as described above, the small dimension of the peptide improved its entry in the pores. For ACTH(CLIP) and Aprotinin the uptake is clearly a function of pH solution (Figure 13B and C). A higher amount of protein adsorbed and eluted from all different pore size MSCs were obtained when ACTH fragment and Aprotinin were incubated at pH close to the isolectric point (pI) of the peptide, at pH4 and pH10 respectively.

When the pH equals the pI of a protein, the number of positive and negative charges is in balance resulting in a neutral molecule. In this condition, the lateral Coulombic repulsion between the protein molecules due to the patches of amino acids sequences with a similar local charge is minimized. At isoelectric point, the ionic forces between the peptide and the silica are small and the molecules and silica adsorbent surface can be close enough to promote stabilization through Van deer Waals attraction that acts on all molecules irrespectively to the electrical charge.
Consequently higher packing density on the surface is possible and a smaller surface area is required for protein to adsorb, increasing the monolayer capacity. The adsorption at pI may greatly depend on the mesoporous surface structure and on its relation to the hydrophobic interactions. At a pH higher than the pI the adsorption of ACTH fragment is reduced.
Under these conditions the peptide is negatively charged and the electrostatic repulsion between the peptide molecules and silanol groups of silica film decreases the amount of adsorption. This is quite evident in the Figure 14B where the Langmuir curves of ACTH incubated at different pH solution was spotted and eluted from the 3 different chips. The maximum capacity was reached when the peptide was incubated in a solution with a pH close to the peptide pI. Moreover at pH=pI the experimental data fitted better with the Langmuir model. Increasing the pH, ACTH fragment showed a significant and constant uptake reduction, for every kind of chip. In particular both the 4nm and the 6nm MSCs showed a decrease of the maximum uptake by 27% and by 58% for pH7 and pH10 respectively and the uptake decreased by 55% and by 78% for 7nm MSC (Table 3).

Table 3. Langmuir adsorption isotherm parameters (B_{max}: maximum uptake recovered in percentage, K_d: adsorption constant, R^2: coefficient of correlation) for each specific peptide (Angiotensin II, Aprotinin, ACTH(CLIP 17-38) and for each pore size MSC (4nm, 6nm, 7nm).
Percentage of B_{max} was calculated for each peptide and each MSC pore size by setting the highest value as the 100%

	Angiotensin II		
	pH4	pH7	pH10
4nm			
B_{max} (%)	88,5	100,0	89,2
Kd (pmol/ul)	1,05	1,48	1,00
R^2	0,94	0,99	0,88
6nm			
B_{max} (%)	96,8	94,9	100,0
Kd (pmol/ul)	1,25	1,73	1,87
R^2	0,94	0,90	0,98
7nm			
B_{max} (%)	100,0	99,0	95,7
Kd (pmol/ul)	1,25	1,51	1,59
R^2	0,96	0,95	0,84

	ACTH(CLIP 17-38)		
	pH4	pH7	pH10
4nm			
B_{max} (%)	100,0	72,8	42,1
Kd (pmol/ul)	10,97	4,31	12,70
R^2	0,96	0,90	0,70
6nm			
B_{max} (%)	100,0	71,0	44,2
Kd (pmol/ul)	15,84	3,95	17,73
R^2	0,97	0,73	0,88
7nm			
B_{max} (%)	100,0	45,7	22,3
Kd (pmol/ul)	29,35	4,55	11,20
R^2	0,98	0,79	0,82

	Aprotinin		
	pH4	pH7	pH10
4nm			
B_{max} (%)	82,0	100,0	99,8
Kd (pmol/ul)	69,12	6,30	6,47
R^2	0,98	0,93	0,90
6nm			
B_{max} (%)	57,0	84,7	100,0
Kd (pmol/ul)	158,40	22,49	12,02
R^2	0,99	0,99	0,97
7nm			
B_{max} (%)	54,6	66,3	100,0
Kd (pmol/ul)	4,19	5,53	9,07
R^2	0,92	0,88	0,96

Adsorption performance decreased also when Aprotinin was incubated in solutions with pH lower than the pI, although the protein net positive surface charge should favor the interaction with silica surface. The involved driving forces could be the electrostatic force between the positive charges of ionized proteins and the negative charges of the silica film itself, and the repulsive positive force between peptides.
With decreasing pH, Aprotinin surface charge increases, and repulsive coulombic forces between amino acid residues (lateral interaction) become more pronounced, reducing the protein adsorption on the silica (Figure 14C). Table 3 clearly showed that uptake (B_{max}) diminished by decreasing the pH.

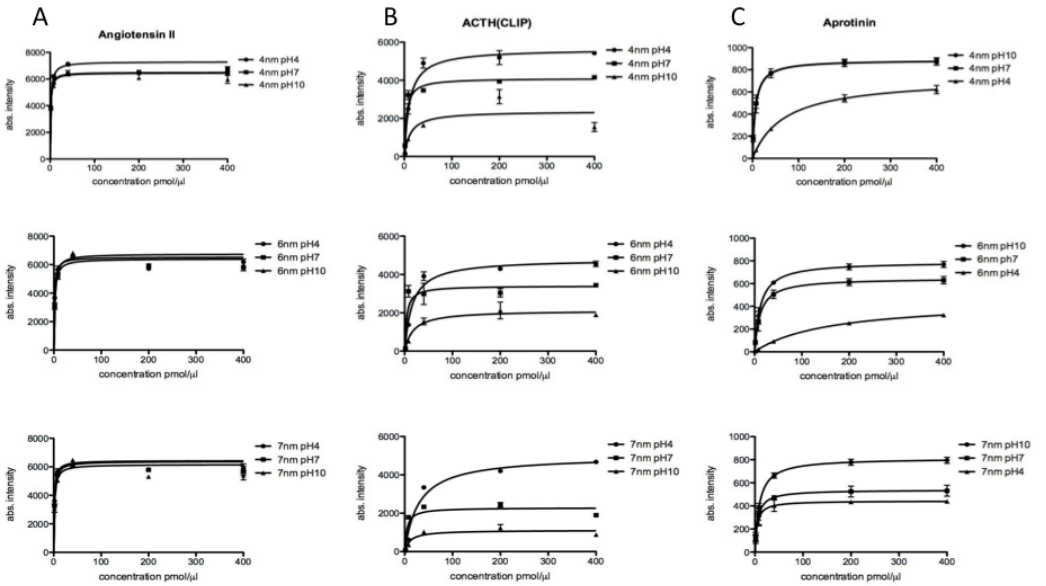


Figure 14. Adsorbed Angiotensin II (panel A), ACTH (CLIP18-37) (panel B), Aprotinin (panel C) as function of protein concentration (adsorption isotherm) for different pH solution and different pore size MSCs. Data are fitted with Langmuir isotherm. Error bars correspond to standard deviation of 4 replicates.

However, looking at the adsorption curves of Aprotinin (Figure 14C) we noticed that at the initial protein concentrations the uptake curve of protein at pH equal to the pI is similar to the curve for incubation at pH 7. In solution, proteins rotate freely whereas on a surface each protein will adapt a certain orientation, which determines the part of the molecules interacting with the surface. It can be assumed that adsorbed proteins and peptides initially adapt an orientation to maximize favourable surface-peptide interaction. At low concentrations the peptide-peptide repulsion is negligible because of the large distance between the molecules. Increasing the molecule density on the surface results in a smaller peptide-peptide distance.

This interaction becoming increasingly dominating and it can trigger orientational changes that lead to a decreased peptide-surface interactions. These considerations are proper for pH solutions that are not too far from the peptide pI.

In a situation in which the charge of the proteins is more pronounced, the lateral interaction between molecules is higher and this could reduce the effect of attraction with the silica surface also at very low peptide concentrations. However, the pH is not the only variable involved in the adsorption process. The pore size also modulated the peptide uptake.

Figure 14, supported by Table 3, shows that at the iso-electric point the peptide uptake is similar, irrespective to the different pore size chip whereas, when the specific peptide acquires a net charge, the uptake decreases even if the pore size of the MSC increased. A possible simplistic explanation can be provided by thinking at the peptides dimension according to the pore size. The ratio between the pore size diameter and the rough peptide diameter can give us a simplify idea of the number of molecules that can fit into the pore section area. In the Figure 15 is shown a schematic model of the pores and the peptides that fit into them. Different number of ACTH molecules can fit in the different MSC pore area: roughly 2, 6, or 7 molecules can fit in the 4nm, 6nm, or 7nm pore size MSC, respectively. For Aprotinin, 1 molecule can fit in the 4nm pore while around 2 and 3 molecules can fit in the 6nm and 7nm pore size MSC, respectively. As we described before, considering the Langmuir kinetic model of monolayer adsorption, when the pH of the solutions is close to the pI of the peptide, the repulsive forces are minimal and the molecules can be packed into the pore. Therefore, similar recovery uptake of the different pore size chips can be due to the higher number of entrance pores that increases the available surface of the smaller pore size chip that compensates, on the other hand, the higher uptake from the larger pores.

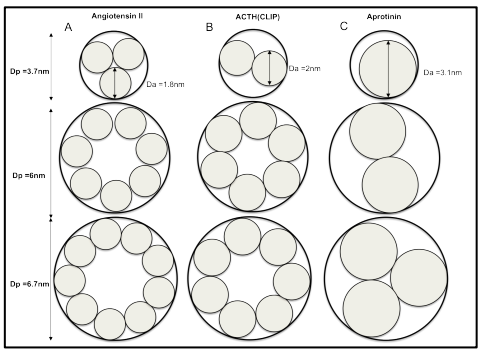


Figure 15. Schematic models of the specific peptides for the different pore size MSC. A) Angiotensin II; B) ACTH(CLIP);C) Aprotinin.

When the pH differs from the pI of the peptide, the repulsion between molecules is more evident with larger pore size that cannot accommodate the same number of molecules in a pore section. In support of this consideration, although the maximum uptake recovery of Aprotinin at pH7 was the same than pH10 using the 4nm MSC, it decreased when increasing the pore size. On Table 3, at pH7, B max of Aprotinin was attested to 85% and 66% for 6nm and 7nm pore sizes respectively. At pH4 the uptake of Aprotinin decreased from 82% with 4nm to 57% and 54% for 6nm and 7nm respectively. Compared to the 4nm pore size MSC, ACTH maximum uptake is similar in the 6nm MSC but in the 7nm MSC it was attested around to 60% and to 50% at pH7 and at pH10, respectively.

8.4.3.2 Effect of increasing complexity of sample solution in peptide recovery

The effects of electrolytes and of increased protein complexity, on the adsorption of specific peptides, were studied for each MSC pore size.

Four experiments in which the specific peptide was incubated in solutions with increasing complexity were reported in order to show the selective harvesting capability of MSC:

I. Interaction with one component in solution;
II. Interaction of a simple mixture of 3 peptides;
III. Interaction of a complex mixture of peptides and proteins with a wide range of MWs mimicking a biological fluids
IV. Interaction with crude human plasma;

The recovery rate of the MSC was calculated using the same amount of peptide used in the experiment and directly analyzed with MALDI-TOF MS. We performed the recovery rate for each peptide in each condition and the relative histograms are shown in the Figure 16.

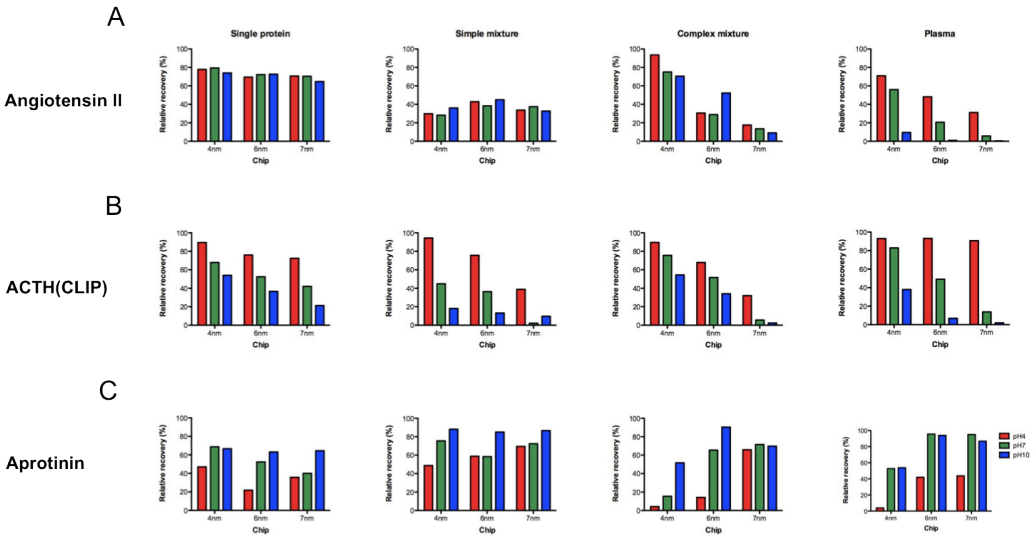


Figure 16. Relative recovery of specific peptide incubated in solution with increasing complexity. Panel A) Angiotensin II, Panel B) ACTH(CLIP 18-37), panel C) Aprotinin. The relative recovery varies increasing protein complexity and introducing the effect of electrolytes.

Figure 16 panel A presents the graphs related to Angiotensin II. As we discussed before, the peptide incubated alone on the MSC showed the same relative recovery independently from the different chips and the different pH solutions.

The result was maintained when the Angiotensin II was incubated with a simple mixture of 3 proteins, while when it was incubated in a more complex mixture, both the pore dimension and the pH affected the peptide recovery.

In particular, the chip with smaller pore size (4nm) presented a better recovery of the specific peptide, compared to the other pore size chips. When increasing the pore dimension, Angiotensin II uptake decreased constantly. The pH conditions of crude plasma under which the peptide presented a slightly positive charge gave the higher uptake irrespectively to the pore size of the MSC. In Figure 16 panel B, ACTH fragment relative uptake is presented. The relationship between ACTH fragment adsorption and the pH of the solution was maintained in every different conditions tested, with a higher recovery at pH solution close to the pI. However, increasing the complexity of the solution, there presented a decrease of the overall uptake according to the pore dimension, particularly at the conditions in which the molecule was incubated in solution with pH different from the pI. Compared to the relative recovery obtained with a simple mixture, ACTH uptake on the 7nm MSC is clearly reduced when the mixture complexity is increased. Spiking the peptide in the plasma sample showed a reduction of the adsorption even on the 6nm MSC. Aprotinin showed the opposite behavior compared to the other 2 peptides.

In a complex mixture or in the crude plasma, the 6nm and 7nm pore size MSC revealed a higher uptake, while with the 4nm pore size MSC it was reduced by about 40%. Considering the number of molecules that can simply fit in a pore section area we observed that in a complex mixture the uptake is improved when the pore is large enough to accept two or three molecules. With an increased complexity of the solution, the pH condition that facilitates the uptake remains the one close to the pI of the specific peptide. In addition, it has been observed that the recovery is also promoted when the peptide has a positive charge.

8.5 *PEPTIDE ADSORPTION STUDY OF HUMAN PLASMA PATIENTS*

We evaluated the ability of the MSC to fractionate and capture LMW peptides and proteins from human plasma according to different pore sizes and different pH. Four different procedures were compared. Plasma samples from three CRC patients and three healthy control cases were incubated either at pH 5 or at native pH (around pH 7-7,5) on both 4nm and on 6nm MSC pore size. Before the sample spotting on the MSC, we added ACN (5% v/v) to the sample in order to partially denature abundant protein allowing the release of the peptides that were bound to albumin or to other proteins (*Tirumalai R.S. et al 2003*). We analyzed each sample in triplicate.

8.5.1 Data pre-processing

Identification of disease specific proteins could provide biological information as well as potential diagnostic markers or drug targets. However there is little to no agreement in determining which ions are the most significant for classifying the samples. Protein mass spectra analysis raises a number of technical challenges highlighting the limitations of existing analysis method. Mass spectra mining involve a high risk of finding pattern in spectra noise. Elimination of confounding factors that are introduced by the analytical method applied is very important. For this reason each MALDI-TOF spectrum was manually inspected for the presence of peaks and for the presence of repetitive noise and matrix peaks. Matrix can generate peaks in the low molecular range of MALDI-TOF spectrum. Matrix peak intensities fluctuate in relation to many variables such as sample quality, matrix deposition, concentration of sample peptides, distribution of the analytes in the MALDI-TOF well and formation of good crystal of sample and matrix. Blank control was used to

generate a spectrum of potential matrix peaks. Spectrum of matrix was compared with sample spectra and peaks with the same m/z as the matrix peaks were not used for subsequent analysis. Moreover, raw MALDI spectra are not amenable to any type of quantitative analysis by direct comparison. For this reason, prior to data analysis, mass spectra data require more meticulous and customized quality control, cleaning, and transformation than most other type of data. Multiple factors govern data production such as sample collection, handling and instrumentation. Therefore we carried out a pre-processing analysis of the raw spectra to reduce experimental variance within the data set for obtaining subsequent reliable statistical analysis of data. Pre-processing involves a complex analysis of digital signal processing, data exploration, and data engineering techniques. (*Norris J.L. et al, 2007; Hilario M. et al, 2006*) No standard method has been established so far regarding the pre-processing steps, including the order in which the steps might be performed. However, pre-processing analysis includes several processes:

- *Filtering or denoising.* This function helps to remove high-frequency interfering signal caused by sources unrelated to the bio-chemical nature of the sample (electrical interference, random ion motions, statistical fluctuation in the detector gain or chemical impurities).
- *Baseline subtraction.* Profile spectra typically exhibit an intense and a variable chemical noise background that must be quantified and eliminate before accurate measures of ion intensities can be determined. There are a number of algorithms that have been reported for estimating the baseline, but the difficult in developing such algorithm is the lack of any quantitative evaluation criteria.
- *Normalization* is the process that corrects for systematic variation between spectra minimizing the spectrum-to-spectrum differences in ion intensity to facilitate direct comparison of spectra. These differences are derived from a number of source including instrument variation, sample variability, differences in sample preparation. Several normalization processes are described in the literature and they can be divided in collective and individual normalization process as well as global and local normalization. In this study we used the Total Ion Current (TIC) that is one of the global normalization procedures and makes the total amount of detected ions (the sum of all intensity values) equal in all spectra.
- *Alignment of the spectra* is often required when the spectra are generated over a long period of time. Misalignment must be corrected to ensure that the same protein intensities are correctly identified in a sample (*Wong J.W. et al., 2005*). There are different alignment methods but usually there must be

references across the entire spectra: an approach consists to identify peaks common to most or all the spectra and spanning the entire mass range of interest. The alignment can be performed before the peak identification, using different algorithms. in this study we used the *Peak Alignment by Fast Fourier Transform algorithm (PAFFT)* that is not dependent on picking peaks but it works by dividing a spectrum into segments before the evaluation of the best shift FFT cross-correlation (*Wong J.W.H et al., 2005*) .

- *Format of data analysis*. The desorption/ionization step in MALDI-TOF MS is a complex process involving optical and mechanical phenomena, as well as thermodynamic and physicochemical processes of phase transition and ionization, which are not well understood. Also the matrix (co)crystallization has been derived empirically. Different matrix molecules crystallize in different shapes and dimensions, proteins tend to accumulate in a non-homogeneous way in the MALDI well, and the composition of the matrix solution and the rate of crystal growth influence the spectral output. These phenomena produce shot-to-shot variation, which is related to sampling different parts of the target surface and progressive sample ablation with repeated sampling (*Albrethsen J., 2007*). Repeated measurements may, therefore, result in largely different absolute intensities. Furthermore, spectrum peaks absolute intensities can vary significantly between measurements especially if the abundances of the peptides are low. For these reasons we decided to primarily analyzed the relative abundances of the peaks instead to use the absolute intensities. However, analyses using absolute intensities were also performed to verify that results between the different approaches did not vary significantly.

Pre-processing analysis were performed using Specalign software that allows for a good interaction from the user. We smoothed the signal; we normalized the ion intensities to minimize spectrum-to-spectrum differences in peak intensity. We used the Total Ion Current (TIC) that makes the total intensity amount of detected ions equal in all spectra. The baseline was subtracted and spectra were binned to a size of a unit of m/z to reduce the complexity of the spectra and to obviate any amplification of differences from species that might be very close to one another. We realigned spectra using PAFFT method to ensure the correct comparison of the same ions across a number of samples. We then picked peaks on the basis of the average spectrum and we exported the relative intensities of each spectrum.

8.5.2 Reproducibility

The semiquantitative nature of the peak intensity and the high-throughput capability of MALDI-TOF MS permit its use in preclinical explorative research of biomarkers.

However, the reproducibility of peak intensities is of high importance when we use MALDI-TOF MS protein profiling for biomarker research because it is based on the comparison of peak intensities between two different conditions (*Albrethsen J., 2007*). If the system is not reproducible, the analysis could lead to significant alterations of the proteomic profiling and the generation of artifacts. Bouamrani et al, at the Nanomedicine department, already performed an intra-experiment variability reproducing the same experiment in six replicates highlighted the stability and reproducibility of the methods. Furthermore we evaluated the inter-experiment reproducibility incubating a commercial plasma, in triplicates, on the five MSC that we used for the biomarker analysis. We assessed the consistency of the system evaluating the presence of the peaks in every spectrum. The variability of the peaks signal, measured by the average CV, was estimated at 13%.

8.5.3 Evaluation of selective peptide fractionation patterns

To evaluate the influence of different sample pH in combination with different pore size MSC we used exploratory techniques such as hierarchical clustering and PCA analyses. Unsupervised hierarchical clustering was performed on the complete MALDI-TOF MS spectra profiles using the Euclidean distance metrics. The distribution of ionic species in all 4 procedures is shown in Figure 17. Irrespective to the group of subjects, cluster analysis showed a clear division of the samples according to the pH, while, within the pH groups, the separation of the different pore sizes is not defined.

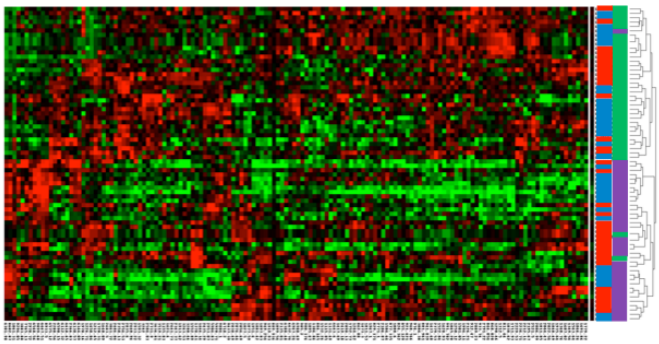


Figure 17. LMW proteins fractionated with the 4 different procedures: A) Clustering Analysis of MALDI-TOF dataset of the four procedures tested: samples incubated at pH 5 (green) and pH7,5 (purple) and fractionate with 4nm MSC (red) and with 6nm MSC (blue).

Data exploration was further carried out with unsupervised and supervised PCA. Figure 18A showed the unsupervised 2D plot of the two first principal components (PC) for the different pH of incubation. The score plot is shown on the left of the graph and it presents the spatial separation of samples in relation to the selected m/z variables that are showed in the loading plot (on the right of the graph). We selected the subset of m/z variables that showed a higher discriminatory power among the groups. The m/z variables were selected after we performed a t-test analysis of each individual m/z variable. T-test determines if the mean of an individual m/z variable was significantly different between the two groups, given the standard deviation and the number of samples. A clear separation between the two different conditions is spatially showed along the Principal Component 1 (PC1) confirming the result presented in the hierarchical cluster. Moreover, looking at the loading graph we noticed that there were more variables expressed in the group of samples incubated at pH5 than those peculiar of the groups of samples incubated at pH7.5. We could assume that at pH5 more peptides have slightly positive charge, confirming the results presented in the section above, with standard peptides incubated in complex samples. In addition variables that contributed more to the group of samples incubated at pH7.5 showed higher m/z compared to the group of sample incubated at pH5. We analyzed the samples separation in relation to the MSC pore size. The unsupervised PCA score plot in figure 18B showed a less clear separation between samples fractionated with 4nm or 6nm pore size MSC. We noticed, as well as the pH, a major contribution of variables with higher m/z, for samples fractionated on the 6nm pore size MSC. However, looking at supervised PCA of all the four experimental procedures (Figure 18C) we observed a less separation between MSC of 4nm and 6nm when the samples were incubated at pH5. m/z variables that contributed to those groups were spatially close and they were more numerous than those peculiar of the samples incubated at pH7.5. Furthermore the classification of control and CRC samples was spatially close for those two groups. On the basis of these experiments we observed that some harvested species are peculiar of the different operative procedures. However the major contribution to the different fractionation is determined by the pH. All the procedures allowed the healthy and the pathological samples to be differentiated, although more consistent results were obtained with pH 5, with both pore sizes. These observations demonstrate that the MSC selective harvesting and fractionation of peptides and proteins in complex biological samples like plasma, is principally due to their charge more than the pore sizes.

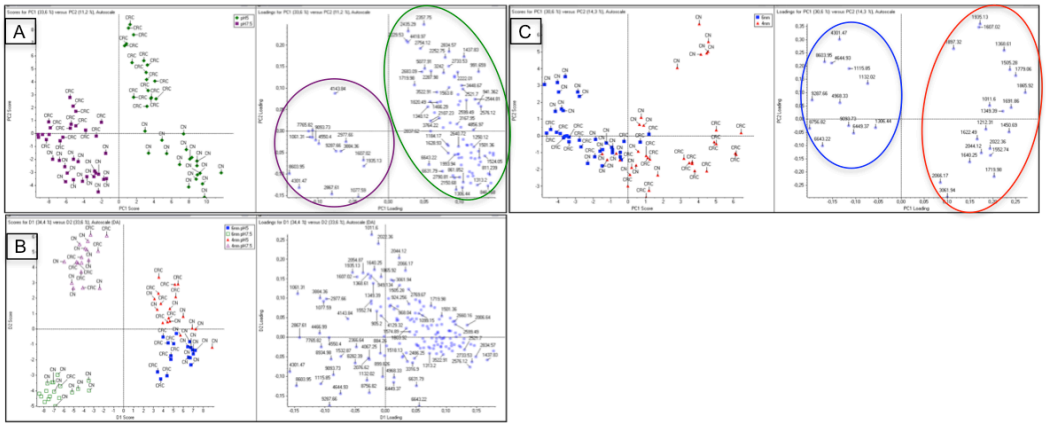


Figure 18. Unsupervised and supervised PCA-DA of LMW proteins fractionated with different condition of pH (pH 5 and pH7.5) and pore size MSC (4nm and 6nm pore size). On the left it is showed the score plot with the spatial disposition of the sample patients while on the right the loading plot showed the different contribution of the single m/z to the separation between the groups. A) Sample fractionation with pH5 (green) or pH7.5 (purple) B) PCA of sample fractionation with 4nm MSC (red) and 6nm MSC (blue). C) Supervised PCA of sample solutions with pH 5 incubated on 4nm MSC (full red triangle), of sample solutions with pH 7.5 incubated on 4nm MSC (empty purple triangle), of sample solutions with pH 5 incubated on 6nm MSC (full blue square) and of sample solutions with pH 7.5 incubated on 4nm MSC (empty green square).

8.6 *ANALYSIS OF CIRCULATING LOW MOLECULAR WEIGHT PROFILE OF RECTAL CANCER PATIENT TREATED WITH NEO-ADJUVANT CHEMO-RADIO THERAPY (pCRT)*

8.6.1 Patient characteristics

We performed MALDI-based comparative analysis of the low-molecular-weight fraction of plasma proteome of 40 patients with locally advance rectal cancer treated with neoadjuvant chemoradio-therapy.

Rectal cancer TRG was distributed as follow: TRG 1: n=9; TRG 2: n=11; TRG 3:n=13; TRG 4: n=6; and TRG 5: n=1. On the basis of the TRG distribution, 20 (50%) patients were considered Good responders (TRG 1 to 2), and 20 (50%) were considered Poor responders (TRG 3 to 5). Complete details are shown in Table 4

Table 4. Patients, tumour and treatment characteristics of the 40 patients included in the study

Patient characteristics		**n°**	**%**
Age	Median (range) yrs	66 (48-80)	
Sex	Male	29	72,5
	female	11	27,5
Tumor distance from the anal verge	≤ 7 cm	25	62,5
	> 7 cm	15	37,5
5-Fluorouracil administration	Continuos infusion	3	7,5
	Bolus	33	82,5
	Not available	4	10,0
Other drugs	5-Fluorouracile alone	17	42,5
	Oxaliplatin	14	35,0
	Not available	9	22,5
pTNM	0	10	25,0
	I	11	27,5
	II	8	20,0
	III	6	15,0
	IV	1	2,5
	Not evaluable	4	10,0
Radical surgery	yes	28	70,0
	no	2	5,0
	Not available	10	25,0
Acute toxicity	0	2	5,0
	1	13	32,5
	2	14	35,0
	3	5	12,5
	4	1	2,5
	Not available	5	12,5
TRG	1	9	22,5
	2	11	27,5
	3	13	32,5
	4	6	15,0
	5	1	2,5
	Good responders (TRG 1-2)	20	50,0
	Poor responders (TRG 3-4-5)	20	50,0

Plasma samples were collected in different time points related to the chemo-radio therapy procedure. For each patient, draw samples were collected before the start of therapy, after two weeks from the start of the therapy, before the surgery (after 2-4 weeks from the end of the therapy). Plasma from group of 43 healthy control subjects, with negative colonoscopy were analyzed

8.6.2 Analysis of Low Molecular Weight peptide profile

We performed the LMW peptide analysis and we collected the MALDI-TOF spectra for each plasma draw. The MALDI-TOF dataset was aligned and normalized using Specalign software and the values were binned to 1 m/z as describe before. We applied the same protocol to all the spectra acquired from different draw blood timing and from different patients. We picked peaks on the basis of the average spectrum and we exported the peaks absolute and relative intensities of each spectrum.

Looking at the sample MALDI spectra individually, we noticed the presence of few peaks with high intensity compare with the other peaks presented; in particular ionic species at 1076.6 and 2021.9 showed high intensities and 1060.6 usually showed the highest intensities in the majority of the spectra. Therefore, relative abundances were calculated on the intensity of m/z at 1060.6. Data from MALDI spectra were analyzed in relation to the pathological status, to the response to CRT, and to the time point draw samples. The explorative research of biomarkers was performed using Principal Component Analysis (PCA) that is considered a useful statistical tool able to discover structures in the experimental data.

8.6.3 Control subjects versus not treated rectal cancer patients

We first analyzed plasma of a group of control subject versus plasma of the rectal cancer patients collected before the CRT (T0), irrespectively to the class of response, for highlighting the peptidomic differences related to the pathological condition. We carried out the pre-processing data of the raw MALDI spectra for all control subjects and rectal patients. By a first visual inspection of the MALDI-TOF spectra, differences seem to be present between the overall peak pattern of the patients and the healthy controls. Interestingly, spectra of control cases showed a lower number of peaks compare to the patient spectra, and of the peaks that the two

groups shared, several species showed lower abundances in control spectra (Figure 19).

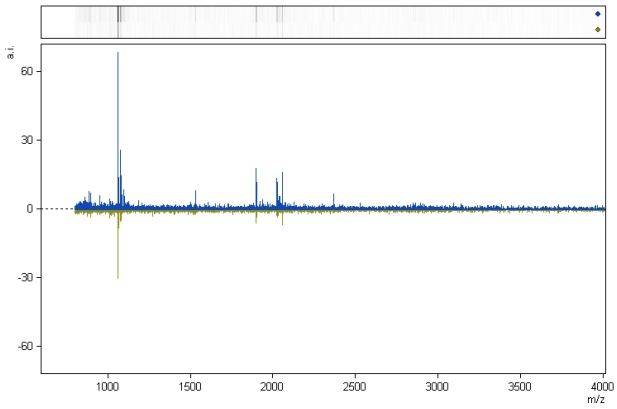


Figure 19. A graphic visualization of two MALDI-TOF spectra; the upper spectrum is related to a rectal cancer patient (Blue) while the lower spectrum is related to a healthy control spectrum (green).

We performed a PCA multivariate analysis including all the peaks found in rectal cancer patient spectra and in healthy control spectra and we confirmed that a good discrimination between the two groups was mainly due to the differential presence of several peaks, listed below (Figure 20 and Table 5). In figure 20A we used all the peaks presented in healthy controls and in patients spectra while in figure 20B we performed the multivariate analysis with the subset of peaks that resulted preferentially present in one of study groups. The picture shows, on the left, a spatial separation of the sample of the two groups and, on the right, the spatial distribution of the m/z ionic species that highlight its differential contribution to the separation.

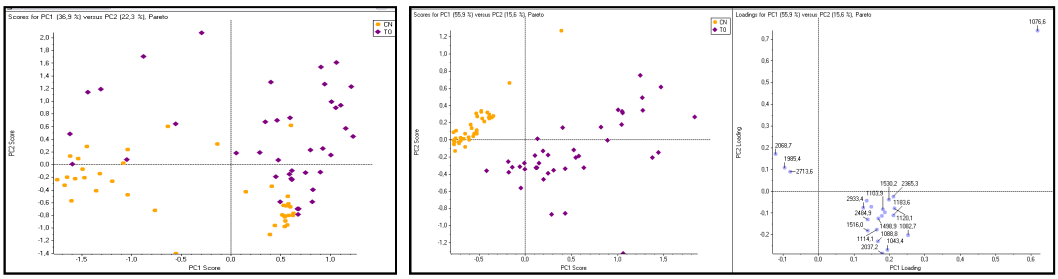


Figure 20. A) Unsupervised PCA made with all the peaks found in healthy subjects spectra and in rectal cancer patient spectra. B) Unsupervised PCA between group of healthy control and rectal cancer patients using the m/z variables differentially expressed in the two groups. Samples spatial separation is showed in the score plot on the left, while loading plot is showed on the right.

Table 5. m/z variables that were present either in rectal cancer patients or in healthy control subjects.

m/z present only in rectal cancer patient spectra		**m/z present only in healthy control spectra**
1043.4	1934.8	2068.7
1082.7	2037.2	1985.4
1103.9	2287.7	2713.6
1114.1	2484.9	
1120.1	2861.7	
1183.6	2878.1	
1498.9	2933.4	
1516.0		

Then we analyzed only the peaks that were in common with both the groups to look for potential differences in the intensity levels. We found 20 peaks in common with both the groups. We exported the relative intensities and for each triplicate we calculated the mean and the standard deviation. We performed an unsupervised PCA between the two groups (Figure 21A) but the multivariate analysis was not able to separate the groups, maybe because the variance of the m/z variables within the groups was large. In order to improve the separation between the two groups we tested one m/z variable at the time for its ability to discriminate the pathology using t-test analysis. The subsequent unsupervised PCA suggested that although the healthy subjects group and the patients group are scattered within the space, a separation between the two groups is present. Interestingly, the majority of the significant m/z variables are down-regulated in control subjects compare to the rectal cancer patients. m/z at 1865.2 is slightly more abundant while m/z at 1098.0, 1530.2, 1896.4, and 2365.3 are down-regulated in the control cases (Figure 18B).

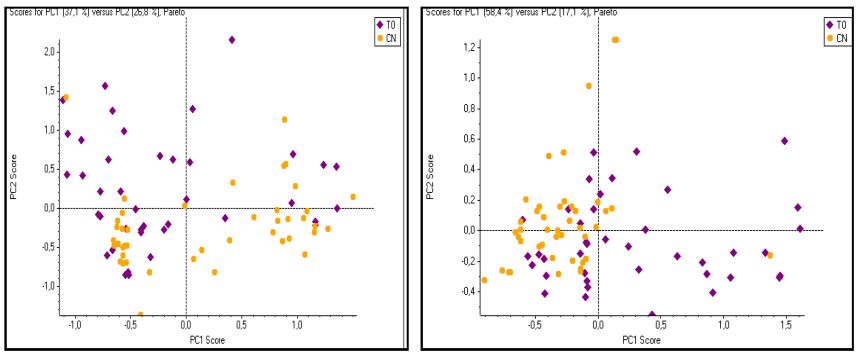


Figure 21. Unsupervised PCA of the group of healthy subjects (yellow) and the patients with locally advanced rectal cancer (purple). A) Unsupervised PCA with all the m/z variables in common between rectal cancer patients and healthy subjects. The picture shows a large overlap of the two groups, indicating the inability of the m/z species to classify the groups. B) Unsupervised made with the m/z variables that showed higher discriminatory power among the groups, at the t-test analysis.

To deepen the analysis we then performed a PCA combined with a Discriminant analysis (supervised PCA-DA) dividing the rectal cancer patients on the basis of the histopathological response. The software first performs a PCA as usual, which reduces the dimensionality of the data by generating the PC's; then the data are used in the Discriminat Analysis (DA) that combined the PC's with the group information to find m/z combinations that maximize the variance between groups while minimizing the variance within groups (supervised approach) (Figure 22). The separation of the pathological condition (healthy versus cancer) is spatially observed along the D1 score plot where the control group shows negative score. The rectal cancer groups that defined the tumour response to CRT are separated along the D2 score plot but the division is less defined. The m/z variables seem to have a lower capacity to separate the control group from the group of rectal cancer that respond to the therapy.

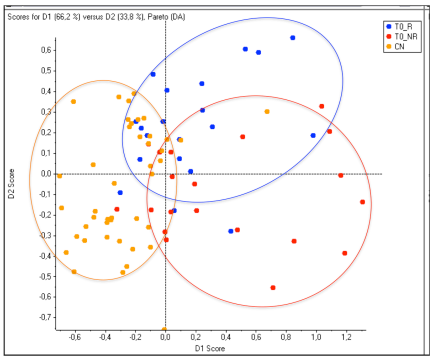


Figure 22. Supervised PCA-DA of healthy subjects and rectal cancer patients divided on the basis of the tumor response. Healthy subjects (yellow) are separate from rectal cancer patients along the D1. Good responder group (blue) highlighted a larger spatial overlap with healthy cases compare with spatial area of Poor responder group (red).

The comparison of a single class of response with the control group confirmed the different capacity for the selected m/z variables to distinguish control cases from rectal cases stratified on the basis of the histopathological response. The peptide expression trends of individual variables between the healthy subjects, the Good and the Poor responders are shown in the box plot in Figure 23. Interestingly box plots of some of the selected m/z variables (in particular m/z at 1530.2 and 1880.9) show a progressive increased trend from the healthy controls to the Good responders and to the Poor responders. This might mean that some peptides increased their expression in blood according with the pathology but also according with the tumour response to CRT, even before applying the therapy. Anova statistical analyses highlighted significant differences between Poor responders group and healthy group for m/z at 2365.3, 1880.9 (p-value < 0.05) and 1530 (p-value < 0.0001). Statistical analysis of single m/z variable between two groups at the time was performed with non-parametric Mann-Whitney test for unpaired data, and a significant difference was assumed for $p < 0.05$ (two-tailed test).

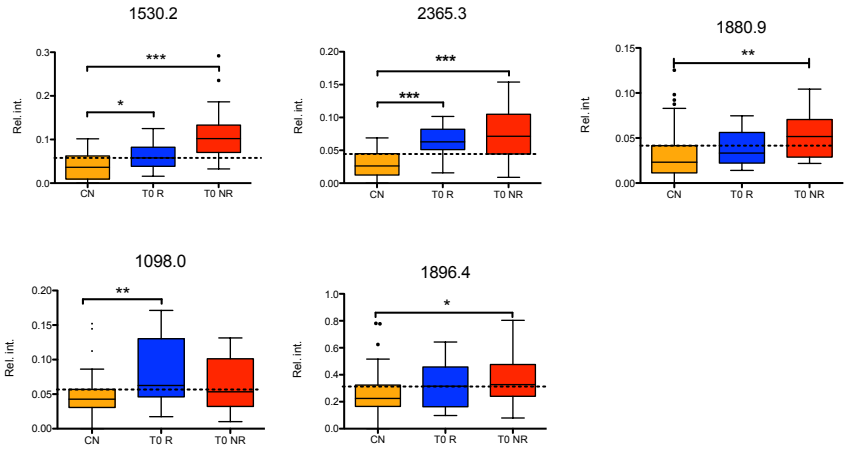


Figure 23. Box plot of the m/z variables selected with the multivariate analyses for the healthy (yellow), Good responders (blue) and Poor responders (red) groups. Mann Whitney tests were performed for two groups at the time. (*= p-value <0.05; **= p-value < 0.01; ***=p-value<0.001 (two-tails p-value)).

8.6.4 Prediction of the histopathological response to the CRT

Observing MALDI-TOF spectra of the patients we noticed a slightly different number of peaks from one patient to another. Usually spectra of T0 and T2 time showed more peaks than spectra of the same patient at T1 time. However the additional peaks commonly showed low intensity and they were present in few patients. For this reason, we choose to study the peaks that were in common with the majority of spectra and that are present through the time points for monitoring its variation tendency.

8.6.4.1 Before the CRT

Data exploration was carried out on the complete MALDI-TOF spectra profiles of rectal cancer patients before the chemo-radio therapy with the aim of identifying a differential expression pattern that could be able to predict the histopathological response to therapy.

The unsupervised PCA did not allow a good separation of the patients according to the tumour response and we performed a supervised PCA-DA analysis (Figure 24). The loading plot of the supervised PCA-DA showed that few m/z ionic species were associated to the Good responder spatial classification. These variables mainly had m/z lower than 1500. A large number of m/z gave low contribution for discriminating the two classes. On the other hand, some m/z ionic species are shown to have higher discriminatory power; as already shown in the previous analysis with the healthy subjects, m/z at 1530,3 showed a higher positive contribution for Poor

responder group. Figure 24 shows negative percentage values for m/z at 1082.7 and 1498.9 suggesting a higher presence in Good responder group.

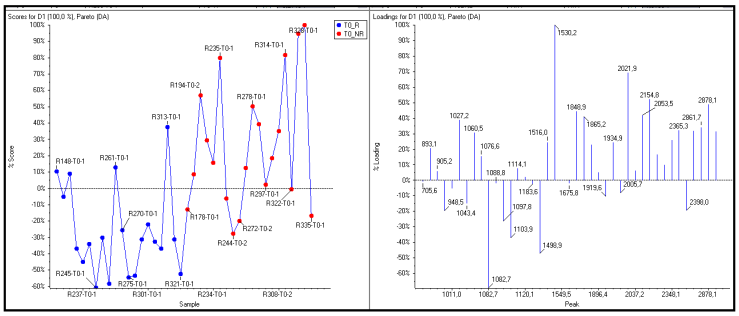


Figure 24. PCA-DA of Good responders (blue) and Poor responder (red) groups at T0 time point.

Therefore we reduced the m/z investigated to enhance the discriminatory capacity of the tumour response classes. We performed a t-test of each individual m/z variable and we selected the m/z variables based on the probability that t-test showed for separating the groups. Ionic species that resulted discriminating the two group at the t-test (with a probability higher than 99%) were: m/z at 1530.3, 1934.9, 2154.8, and 1848.9 that were more abundant in Poor responders and m/z at 1082,7 that was more abundant in Good responders. However, the PCA-DA performed with these m/z variables showed a slight improvement of the discriminatory capacity (Figure 25).

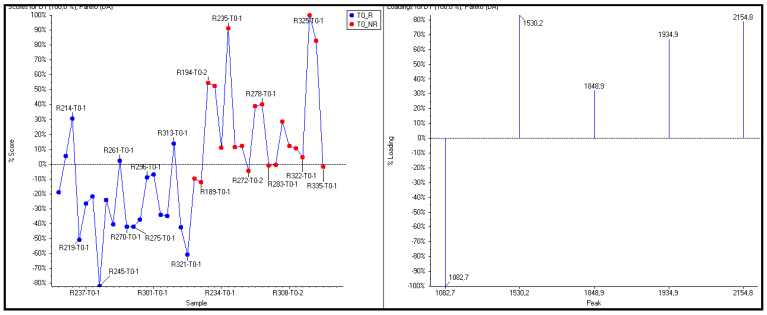


Figure 25. PCA-DA between Poor responder (red) and Good responder (blue) groups using the m/z variables that showed t-test probability >99%.

By combining different information derived from the t-tests, the unsupervised and the supervised multivariate analyses, we found a combination of m/z variables that improved the discrimination of the groups (Figure 26).

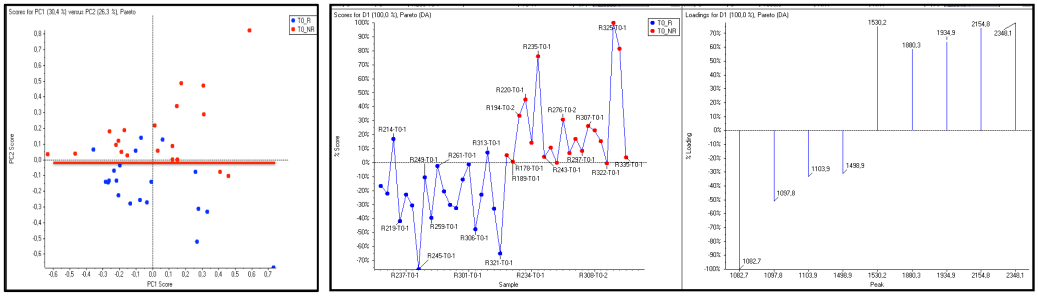


Figure 26. PCA analysis between Good (blue) and Poor (red) responders carried out with m/z variables selected combining PCA analysis with single variable t-test analysis A) Unsupervised PCA shows a discrimination between the 2 groups with a misclassification of four Good responders and two Poor responders. B) Supervised PCA-DA shows the contribution of the selected m/z variables (loading plot) on the classification.

Some Poor responder and Good responder patients are spatially close in the unsupervised score plot PCA, lying in the space in which the two groups overlapped. This means that the combination of the m/z ionic species may not be able to unequivocally classify them.

The t-test analysis is a good way to test each m/z species for its ability to discriminate the condition under investigation and to find the top most significant species for developing the subsequent statistical model. However, unsupervised and supervised multivariate analyses evaluate the synergy among the species, and sometimes variables that could be not significant if they are considered individually, could improve the results when used in combination with other variables. Box plot of each m/z ionic species in Good and Poor responders are showed in Figure 27.

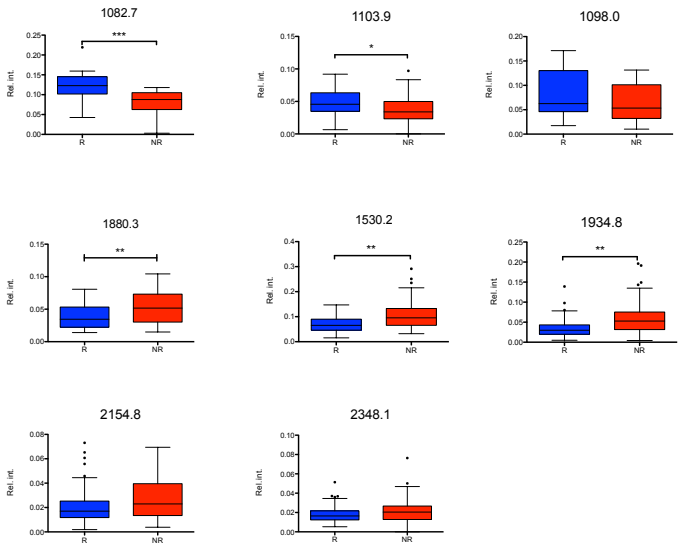


Figure 27. Box plot of the m/z variables selected with the t-test and the multivariate analyses that discriminate between Good responders (blue) and Poor responders (red) groups. m/z at 1498.9 did not show high differences between Poor and Good responders. Mann Whitney tests were performed for each m/z variable. (*= p-value <0.05; **= p-value < 0.01; ***=p-value<0.001 (two-tails p-value))

The ROC-curves for each m/z variable are showed in Table 6.

Table 6. ROC-Curve features for each m/z variables selected. ROC-Curves were carried out for each m/z variable comparing data from Good responder and Poor responder groups .

Good responder, Poor responder	AUC (95%CI)	p-value	Cut-off	Sensitivity (95%CI)	Specificity (95%CI)
V1082.7	0.73 (0.57-0.86)	0.0056	<0.0785	70 (45.7-88.1)	80 (56.3-94.3)
V1098.0	0.60 (0.43-0.75)	0.2796	<0.0382	35 (15.4-59.2)	90 (68.3-98.8)
V1103.9	0.70 (0.53-0.83)	0.0234	<0.0291	55 (31.5-76.9)	85 (62.1-96.8)
V1530.2	0.75 (0.59-0.87)	0.0020	>0.0923	65 (40.8-84.6)	90 (68.3-98.8)
V1880.3	0.67 (0.51-0.81)	0.0422	>0.044	60 (36.1-80.9)	75 (50.9-91.3)
V1934.8	0.73 (0.57-0.86)	0.0048	>0.0488	65 (40.8-84.6)	90 (68.3-98.8)
V2154.8	0.71 (0.54-0.84)	0.0149	>0.0258	55 (31.5-76.9)	90 (68.3-98.8)
V2348.1	0.72 (0.56-0.85)	0.0065	>0.0133	90 (68.3-98.8)	50 (27.2-72.8)

An interesting finding is that among the m/z ionic species, three were not found in healthy control subjects (1082.7 1103.9, 1934.8) whereas three were found down-regulated in healthy controls.

Multivariate logistic regression analysis

We performed a multivariable logistic regression analysis in order to investigate the role of the selected m/z ionic species in predicting the histopathological response. We estimated the probability of the non-responsive event (Poor responders with TRG:3-5) in function of these variables. For this purpose we performed univariate logistic regression analysis for each of the selected m/z ionic species as shown in Table 7.

Table 7. Univariate logistic regression analysis for each of the previous selected m/z variables

		DISTRIBUTION					LOGISTIC MODEL			
		N	Minimo	Massimo	Media	DS	Unit	OR	95% CI	p-value
V1082.7	**Good**	20	0.03	0.21	0.10	0.04				
	Poor	20	0.01	0.13	0.07	0.03				
	Tot	40	0.01	0.21	0.09	0.04	0.01	0.77	0.63-0.95	0.0151
V1098.6	**Good**	20	0.02	0.17	0.08	0.05				
	Poor	20	0.01	0.13	0.06	0.04				
	Tot	40	0.01	0.17	0.07	0.04	0.01	0.91	0.78-1.06	0.2201
V1103.9	**Good**	20	0.01	0.09	0.05	0.02				
	Poor	20	0.00	0.10	0.04	0.02				
	Tot	40	0.00	0.10	0.04	0.02	0.01	0.76	0.57-1.03	0.0733
V1530.2	**Good**	20	0.02	0.13	0.06	0.03				
	Poor	20	0.00	0.29	0.11	0.07				
	Tot	40	0.00	0.29	0.09	0.06	0.01	1.27	1.04-1.56	0.0179
V1880.3	**Good**	20	0.01	0.07	0.04	0.02				
	Poor	20	0.02	0.10	0.05	0.02				
	Tot	40	0.01	0.10	0.05	0.02	0.01	1.35	0.99-1.84	0.0576
V1934.8	**Good**	20	0.00	0.14	0.03	0.03				
	Poor	20	0.01	0.14	0.06	0.04				
	Tot	40	0.00	0.14	0.05	0.04	0.01	1.28	1.01-1.63	0.0440
V2154.8	**Good**	20	0.00	0.04	0.02	0.01				
	Poor	20	0.01	0.06	0.03	0.01				
	Tot	40	0.00	0.06	0.02	0.01	0.01	2.13	1.13-4.02	0.0190
V2348.1	**Good**	20	0.01	0.05	0.02	0.01				
	Poor	20	0.01	0.08	0.03	0.02				
	Tot	40	0.01	0.08	0.02	0.01	0.01	1.7	0.93-3.10	0.0820

All variables entered the multiple logistic regression model and the stepwise selection method identified 5 variables that better predicted the outcome in the final model (Table 8). The estimated probability of having a Poor response event (TRG3-5) increased as the variables m/z 1530.2, 1880.3 and 2348.1 increased, while it decreased as the variables m/z 1082.7 and 1098.6 m/z increased.

Table 8. Logistic regression model carried out with the m/z variables selected through the stepwise selection method

	LOGISTIC MODEL			
	Unit	**OR**	**95% CI**	**p-value**
V1082.7	0.01	0.28	0.08-0.92	0.0356
V1098.6	0.01	0.37	0.13-1.07	0.0667
V1530.2	0.01	2.94	1.03-8.37	0.0432
V1880.3	0.01	5.32	0.90-31.4	0.0649
V2348.1	0.01	16.2	1.01-260	0.0494

The Hosmer-lemeshow goodness of fit test is not significant (p=0.8654), suggesting that the model adequately fit the data. The ROC-Curve of the logistic model, plotted in Figure 28, shows a very high AUC (98%). For the probability of histopathological non-response=0.5, the false positive rate is 20% and the false negative rate is 20%. The sensitivity and the specificity of the model are both 80%.

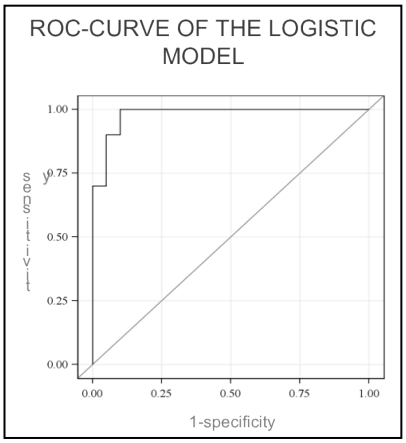


Figure 28. Roc-Curve of the logistic regression model, with the five selected m/z variables. The Area under the curve is 0.98

8.6.4.2 *After the CRT*

We performed the peptide profiling study on plasma of the rectal cancer patients after the CRT. It is noticed that waiting a several of weeks from the end of CRT before doing the surgery improves the prognosis. Hence blood samples were usually done just before the surgery.

We followed the same analytical procedure as previously described. We first performed an unsupervised PCA and a PCA-DA with m/z variables from the complete spectra (Figure 29).

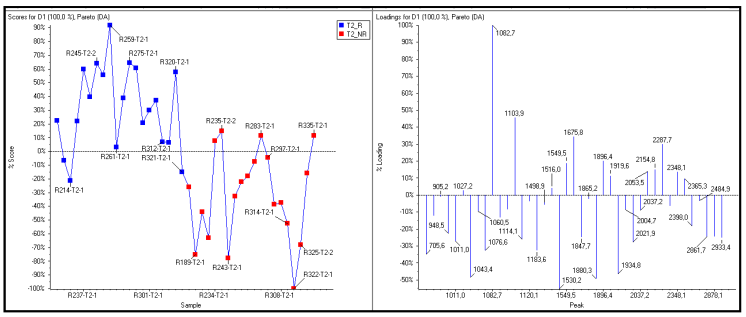


Figure 29. PCA-DA of Good responders (blue) and Poor responder (red) groups at T2 time point.

Looking at the loading plot of the m/z variables, in Figure 29, we noticed that a higher number of m/z ionic species have negative loadings that are associated to the Poor responder spatial classification, as already seen in T0 analysis. However, ionic species that are associated to the Good responder group cover the all range of

spectra. We performed student t-test analyses to determine the ability of each m/z variable to discriminate the response classes. We found that variables had less discriminating power compared to the T0 time point. m/z at 1082,7 showed a t-test probability of 99% for Good responders, whereas Poor responders showed higher intensity for m/z at 1043.4, 1088.3, and 1934.8 although t-test probability were 96,7%, 96,7% and 95,8% respectively. Using both the m/z variables with the higher t-test scores and the m/z variables that showed higher contributes in PCA-DA we were able to define a set of ionic species that is showed in Figure 30. As the T0 time point analysis, there are m/z variables that show low value at the t-test but they improved the discrimination between groups.

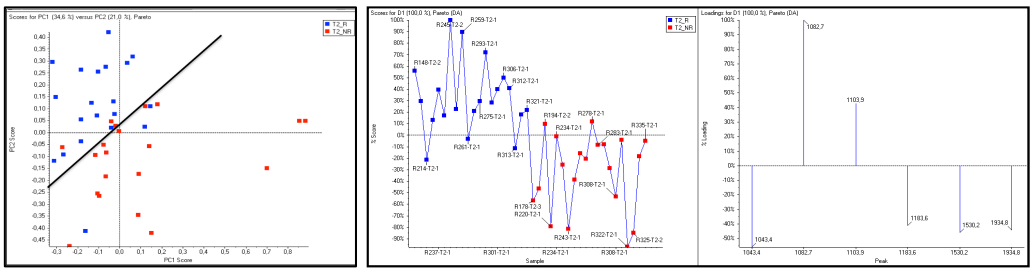


Figure 30. PCA analysis between Good (blue) and Poor (red) responders carried out with m/z variables selected combining PCA analysis with single variable t-test analysis A) Unsupervised PCA shows a discrimination between the 2 groups with a misclassification of three Good responders and three Poor responders. B) Supervised PCA-DA shows the contribution of the selected m/z variables (loading plot) on the classification.

The m/z variables that contributed more in Poor responders classification were: 1043.4, 1183.6, 1530.2, and 1934.8. The m/z which expression was associated to Good responders group were: 1082.7 and 1103.9. In Figure 31, a box plot for each ionic specie is presented. We noticed that, although m/z at 1530.2 was useful to improve the discrimination between Good and Poor responders in multivariate analysis, the individual analysis of this ionic species did not show significant difference between the two groups, indicating that the combination of the ionic species is more powerful than the individual variable.

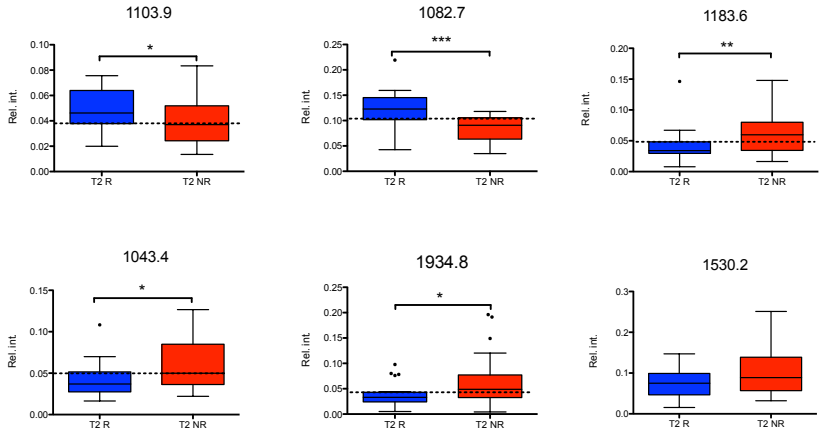


Figure 31. Box plot of the m/z variables selected with the multivariate analyses for Good responders and Poor responders groups at the T2 time point. Mann Whitney tests were performed for each m/z ionic species two groups. (*= p-value <0.05; **= p-value < 0.01; ***=p-value<0.001 (two-tails p-value)).

The ROC-Curve of each m/z variable is showed in Table 9.

Table 9. ROC-Curve features for each m/z variables selected. ROC-Curves were carried out for each m/z variable comparing data from Good responder and Poor responder groups .

Good responder, Poor responder	AUC (95%CI)	p-value	Cut-off	Sensitivity (95%CI)	Specificity (95%CI)
V1043_36	0.67 (0.51-0.81)	0.0489	>0.0603	45 (23.1-68.5)	90 (68.3-98.8)
V1082_72	0.82 (0.67-0.93)	<0.0001	<0.118	100 (83.2-100)	55 (31.5-76.9)
V1103_9	0.63 (0.46-0.77)	0.1650	<0.0373	55 (31.5-76.9)	80 (56.3-94.3)
V1183_62	0.72 (0.55-0.85)	0.0099	>0.0555	55 (31.5-76.9)	90 (68.3-98.8)
V1530_21	0.63 (0.46-0.78)	0.1467	>0.0486	90 (68.3-98.8)	35 (15.4-59.2)
V1934_8	0.66 (0.50-0.81)	0.0664	>0.039	65 (40.8-84.6)	75 (50.9-91.3)

Multivariate logistic regression analysis

We performed a univariate logistic regression analysis for each of the selected m/z ionic (Table 10).

Five m/z variables were included in the subsequent multiple regression analysis. By applying the stepwise method we selected the best subset of m/z variables and we performed a multivariate logistic regression (Table 11). The estimated probability of having a Poor response event (TRG3-5) increases increasing the variables at m/z 1043.4, 1183.8 and 1530.2 while it decreases increasing the variables at m/z 1082.7 and 1103.9 m/z variables.

Table 10. Univariate logistic regression analysis for each of the previous selected m/z variables

		DISTRIBUTION					LOGISTIC MODEL			
		N	**Min**	**MaX**	**Mean**	**SD**	**Unit**	**OR**	**95% CI**	**p-value**
V1043.4	**Good**	20	0.02	0.11	0.04	0.02				
	Poor	20	0.02	0.13	0.06	0.03				
	Tot	40	0.02	0.13	0.05	0.03	0.01	1.32	1.01-1.73	0.0458
V1082.7	**Good**	20	0.04	0.22	0.12	0.04				
	Poor	20	0.00	0.12	0.08	0.03				
	Tot	40	0.00	0.22	0.10	0.04	0.01	0.63	0.45-0.87	0.0049
V1103.9	**Good**	20	0.01	0.08	0.05	0.02				
	Poor	20	0.01	0.08	0.04	0.02				
	Tot	40	0.01	0.08	0.04	0.02	0.01	0.82	0.58-1.15	0.2526
V1183.6	**Good**	20	0.01	0.15	0.04	0.03				
	Poor	20	0.02	0.15	0.06	0.03				
	Tot	40	0.01	0.15	0.05	0.03	0.01	1.28	0.97-1.69	0.0750
V1530.2	**Good**	20	0.02	0.15	0.07	0.04				
	Poor	20	0.03	0.25	0.10	0.06				
	Tot	40	0.02	0.25	0.09	0.05	0.01	1.13	0.97-1.31	0.1103
V1934.8	**Good**	20	0.01	0.10	0.04	0.02				
	Poor	20	0.00	0.20	0.07	0.06				
	Tot	40	0.00	0.20	0.05	0.04	0.01	1.21	0.99-1.48	0.0681

Table 11. Logistic regression model carried out with the m/z variables selected through the stepwise selection method

	LOGISTIC MODEL			
	Unit	**OR**	**95% CI**	**p-value**
V1043.4	0.01	1.93	0.93-3.99	0.0767
V1082.7	0.01	0.30	0.13-0.70	0.0055
V1103.9	0.01	0.28	0.09-0.89	0.0306
V1183.6	0.01	1.51	0.93-2.44	0.0932
V1530.2	0.01	1.66	1.03-2.69	0.0374

The Hosmer-lemeshow goodness of fit is not significant (p=0.9713), suggesting that the model adequately fit the data. The ROC-Curve of the logistic model plotted in figure x shows a very high AUC (97.25%) (Figure 32). For the probability of histopathological non-response=0.5, the false positive rate is 15.8% and the false negative rate is 19%. The sensitivity and the specificity of the model are 80% and 85% respectively.

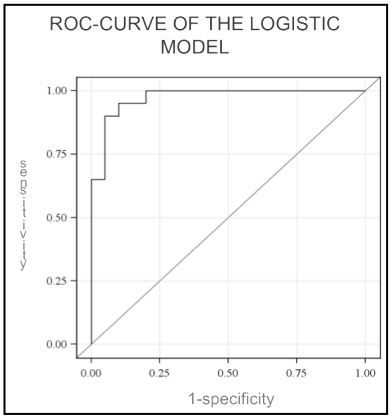


Figure 32. Roc-Curve of the logistic regression model, with the five selected m/z variables. The Area under the curve is 0.975

Interestingly R214 and R313, the good responder patients that were misclassified at T0 time point, are still misclassified at the T2 time point. We checked the clinical parameters available for these patients and we compared the values with those of the other patients, but we did not find any explanation for justifying these results.

8.6.4.3 In the middle of the CRT: T1 time point

T1 time draw blood was carried out after 2 weeks from starting the CRT, meaning in the middle of the neo-adjuvant treatment. We performed the peptide profile analysis to evaluate if the initial perturbation of the biologic system could give meaningful evidences according to the different histopathological response to CRT. The so-called "induced resistance to therapy" is referred to the mechanisms that cells and tissue could activate in response to a stress. These comprise modulation of gene expression and protein expression as well as modulation of a number of pathways and processes such as cell cycle and cell death. The draw blood at this time point is difficult to obtain because multiple variables could influence: patient availability, patient physiologic state, and the different medical hospital where the patient does the therapy. Therefore we were able to analyze only 12/20 T1 plasma of Good responders and 14/20 plasma of Poor responders. We found that peak intensities at this time point were lower than those for T0 and T2 and supervised and unsupervised PCAs showed difficult to discriminate the two groups. Combining individual t-test analyses and principal component analyses data, we found 5 m/z variables that were able to show a kind of separation between Poor and Good responders although it was not clear. In particular, unsupervised PCA shows that Poor responder samples lie in a big spatial area, overlapping the spatial area of Good responders (Figure 33). Box plots in Figure 34 confirmed the difficult of the m/z ionic species to distinguish clearly the two groups.

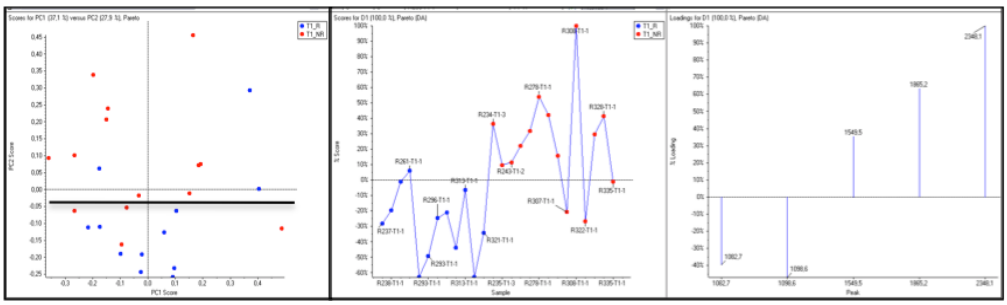


Figure 33. PCA analysis between Good (blue) and Poor (red) responders carried out with m/z variables selected combining PCA analysis with single variable t-test analysis A) Unsupervised PCA shows a high spatial area for Poor responders indicating a variable peptides behaviour of patients within the group. B) Supervised PCA-DA shows the contribution of the selected m/z variables (loading plot) on the classification.

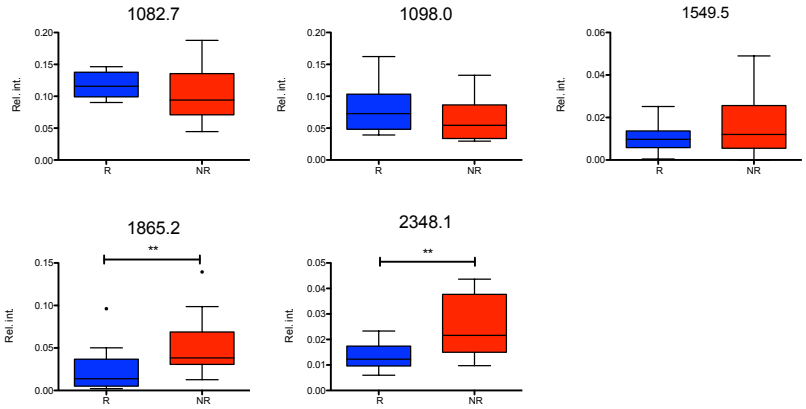


Figure 34. Box plot of the m/z variables selected with the multivariate analyses for Good responders and Poor responders groups at the T2 time point. Mann Whitney tests were performed for each m/z ionic species two groups. (*= p-value <0.05; **= p-value < 0.01; ***=p-value<0.001 (two-tails p-value)).

Table 12 summarizes the m/z variables that were differentially expressed between Good and Poor responders at the individual time points. As it is showed, some m/z ionic species were found differentially expressed only in one time-point whereas some other were found discriminative in 2 different time points. Interestingly there are m/z ionic species that are discriminative before and after the CRT but not during the CRT (1103.9, 1530.2, and 1934.8). Three m/z ionic species were in common between T0 and T1 time point where four ionic species were in common between T0 and T2 time points. Ionic specie at m/ 1082.7 is the only one that showed the same behaviour through the time points and it was found more expressed in Good responders patients. Another interesting finding is that among the m/z ionic species, six were not found in Healthy control subjects (1082.7, 1043.4, 1103.9, 1183.6,

1498.9, 1934.8) whereas three were found down-regulated in healthy controls (1098.0, 1530.2, 1880.3).

Table 12. Summary of the m/z variables found differentially expressed between Good responders and Poor responders at each time point analyzed (T0, T1, T2).

Patients	T0	T1	T2
Good Responders	1082,7	1082,7	1082,7
	1098.0	1098.0	
	1103,9		1103,9
	1498,9		
Poor Responders	1530,2		1530,2
	1934,8		1934,8
	2348,1	2348,1	
	2154,8		
	1880,3		
		1549,5	
		1865,2	
			1043,4
			1183,6

3.4.2.2.4 Monitoring the CRT: variables tendencies through the time points

We investigated the m/z variables intensity through the different time points to determine if there were tendencies that could be correlated to CRT. We first analyzed the relative intensities of all the variables, through the three time points, for single response class. Supervised PCA-DA of Good responder and of Poor responder groups showed an overlap between the three time points. We were not able to find m/z variables which intensities could separate the time-points groups (Figure 35A,B) hence we investigated the differences between 2 time-points at the time (T0 versus T1, T0 versus T2, and T1 versus T2) for each histopathological response class.

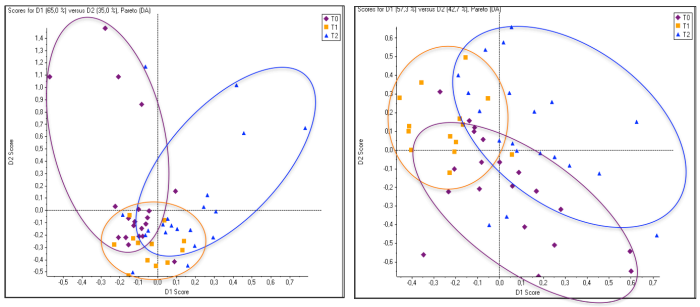


Figure 35. PCA-DA at the three time-points: T0 (purple), T1 (yellow), and T2 (blue) for each histopathological response class. PCA-DA of Good responders (on the left) and PCA-DA of Poor responders (on the right).

We applied t-test analysis in combination with supervised PCA-DA and, although we found specific set of ionic species that showed a better separation between the

group, the unsupervised analysis with the same ionic species did not confirm the classification, suggesting that the PCA-DA forced the separation between groups, enhancing little or random differences of intensity. An example is showed in Figure 36. These differences could be due to the CRT time points but at the current state we could not prove it.

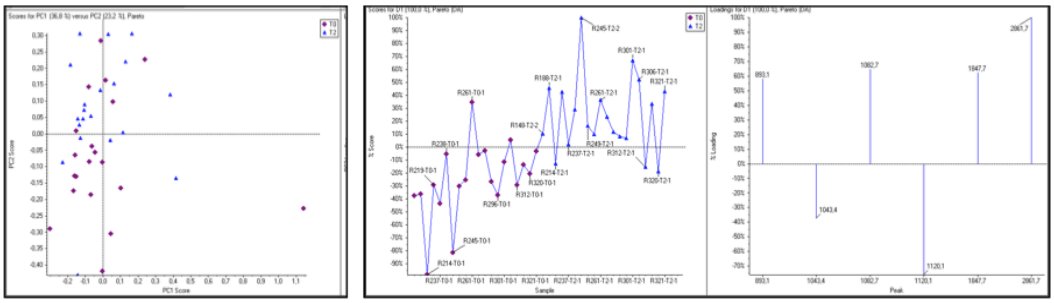


Figure 36. Unsupervised PCA (on the left) and supervised PCA-DA (on the right) of Good responders between T0 (purple) and T2 (blue) time points.

Then we used a different method to better visualize the possible differences between the time points. We used the intensities of each m/z at T0 time as the baseline expression before the treatment and we calculated the ratio of the time T1 and T2 over the T0 time (T1/T0 and T2/T0). Therefore a ratio equal to 1 indicates that the expression was equal to the baseline expression, ratio higher than 1 implied an over-expression and a ratio lower than 1 implied a down-regulation of the specific m/z ionic species. Although the differences were not strong, some m/z variables showed a different expression tendency compare to the baseline expression. Looking at the graphs in Figure 37 we could notice that, globally in good responders, more m/z variables had the median value separate from 1 than those present in Poor responders; 36,5% (15/41) and 26,8% (11/41) for Good and Poor responders respectively. At T1 time-point many m/z variables showed a decrease of intensity compare to T0 time-point, although many of these showed only a slight decrement. Observing the T1/T0 ratio, in Good responders 4/15 of m/z variables show a more accentuated decreasing tendency of intensity and 1/15 of m/z variables presents significant increasing tendency of intensity compared with that observed at the baseline time point (graph A in Figure 37, purple box plots with red border). For Poor responders 4/11 of m/z ionic species showed decreasing tendency of intensity while one m/z variable over 11 showed significant increasing tendency of intensity compare with that observed at the baseline time point (graph B in Figure 37, purple box plots with red border). Observing the T2/T0 ratio we noticed that the majority of m/z ionic species showed an increment after the CRT compare to the baseline levels

or the T1 time point, underlined an increment of intensity for the majority of m/z variables. For Good responders no variables showed a decreasing tendency of intensity whereas 10 m/z had median value of ratio significantly higher than 1(Figure 37 graph A, blue box plots with red border). For Poor responders one m/z showed a decreasing tendency of intensity while 5 m/z ionic species showed median value significantly higher than 1 (Figure 37, graph B, blue box plots with red border). m/z ionic species that presented a T2/T0 ratio higher than the T1/T0 ratio were 8 and 4 for Good and Poor responders respectively. However the T2/T0 ratio of these ionic species usually presents highly variable ranges. m/z that showed a decrement of T2/T0 ratio compared with T1/T0 ratio were: m/z at 1098.6 for Good responders and m/z at 1516.0 and 2037.2 for Poor responders. In particular m/z at 1098.6 and 2037.2 showed an increase of expression in T1 time-point whereas in T2 the expression were similar to T0. On the other hand m/z at 1516.0 showed a decreased expression in T1 that is emphasize in T2 time point.

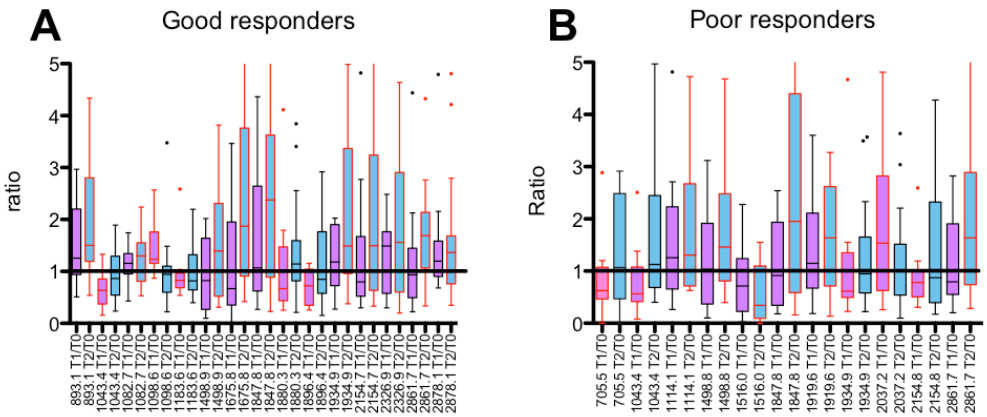


Figure 37. Ratio of relative intensity of each peptide at T1 or T2 time point over the intensity of the baseline time point (T0): T1/T0 (purple) and T2/T0 (blue) ratio were evaluated for each patient. Box plots of T1/T0 ratio (purple) and T2/T0 ratio (blue) of the subset of m/z ionic species that showed higher fluctuation in intensity during the CRT in A) Good Responders and B) in Poor responders. Ratio higher than 1 indicates increased expression compare to the baseline while ratio lower than 1 indicates down-regulation of the peptide expression compare to the baseline. Box plots with red borders indicate a statistically different tendency compare to the baseline level (equal to 1).

When we compared the box plot pictures of Good and Poor responders we noticed that there are some ionic species that are present in both the groups (m/z at 1498.9, 1847.8, 2861.7). Interestingly these ionic species showed the same tendency over the time-points, in both the groups: at the T1 time-point these ionic species the intensity do not show significant differences with the T0, while there is a significant increment of the ionic species at the T2 time-point.

We monitored the tendencies of the peaks that resulted discriminative of the histophatological response before and after the therapy in Good and in Poor responders as shown in the Figure 38. There are no clear rules or evident trend in this subset of ionic species along the therapy. Ionic species did not show big difference in the intensity tendencies between Good and Poor responders. Along the CRT treatment we could observe that in Good responders the ionic species 1498.9, 1934.8, and 2154.8 increased after the end of CRT (T2), 1098.6 increased in T1 while 1043.4, 2348.1 and 1183.6 decrease in T1. Between Good and Poor responders many ionic species showed similar tendencies in T1 and T2. The majority of m/z variables showed intensities similar to those found at the baseline time point for both T1 and T2 time point or they could show a increased tendencies in T2, compared with those found in T1. The m/z variables that showed different tendencies between Good and Poor responders were m/z at 1098.6, 2154.7, and 2348.1. 1098.6 increased in T1 and decreased in T2 in Good responders whereas in Poor responders it showed expression similar to T0 time point. m/z at 2154.7 showed an increase in T2 for Good responders that is poorly observe in Poor responders and 2348.1 showed a inverse tendencies of T1 and T2 ratio between Good and Poor responders. Ionic species that were found differentially expressed in both the T0 and T2 analysis (1082.7, 1103.9 and 1530.2) did not show particular difference between Good and Poor responders.

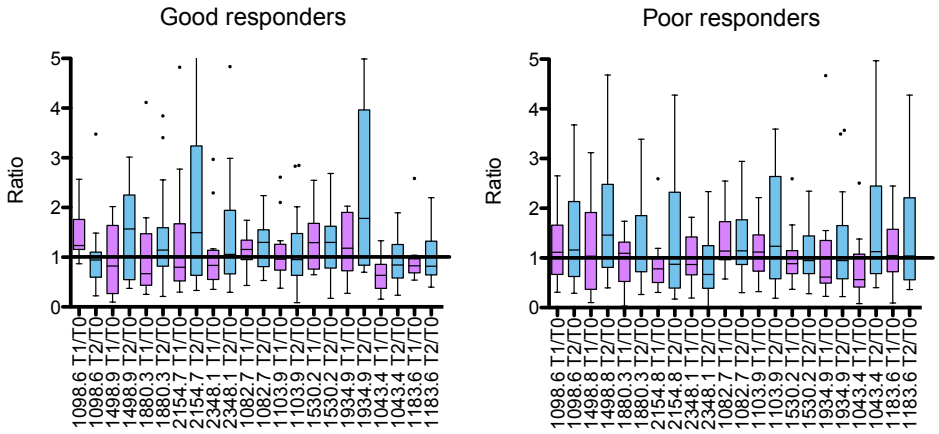


Figure 38. Box plots of T1/T0 ratio (purple) and T2/T0 ratio (blue) of the subset of m/z ionic species that are differentially expressed before and after the CRT between Good Responders (graph on the left) and Poor responders (graph on the right).

8.7 *PEPTIDE SEQUENCE IDENTIFICATION*

Several peptides have been identified in the MSC elution samples. We easily recognized the most intense peaks observed in the MALDI-TOF spectra, using common search engines like Mascot.

In Figure 39 is reported the case of peptide at m/z 1060.6, for which all ion fragments have been positively attributed to Bradykinin peptide. The calculated threshold score for this peptide was set to 12 by the program. Since the obtained score was 47, this is considered a significant match.

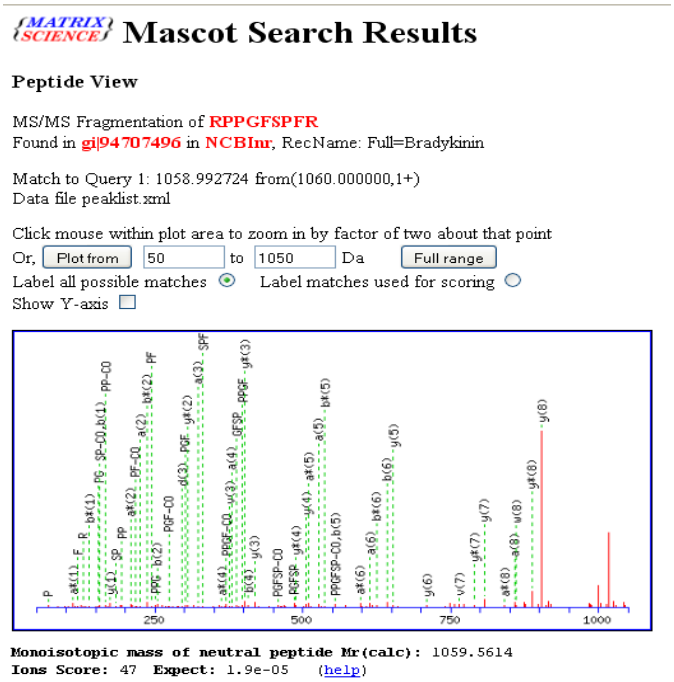


Figure 39. Mascot peptide window result for the ion at m/z 1060. The MS/MS spectrum is shown with the matched fragment ions and the ion score associated.

Similarly, other peptides have been positively recognized by using the search engines. All these peptides are naturally occurring in the plasma and, moreover, they are produced in vivo by specific proteases. The peptides identified are listed in the table below:

Table 13. Peptides identified by using search engines. The table shows the m/z of the spectra, the amino acidic sequence and the putative name of the protein from which derived the peptide.

Family	Peptide	m/z	Sequence
Kininogens	Bradykinin	1060.6	RPPGFSPFR
	Bradykinin oxidized	1076.6	RPP(Ox)GFSPFR
	Kininogen HMW	2365.2, 1183 (2+)	KHNLGHGHKHERDQGHGHQR
Complement	C3f	2021.2, 1011.1 (2+)	SSKITHRIHWESASLLR
	C4a	1896.1, 948 (2+)	NGFKSHALQLNNRQIR
Clusterin	Clusterin B chain	1530.9	RPHFFFPKSRIV

Kininogen, C3f, and C4a fragments were identified by two different ion species where one is half of the m/z value of the other (e.g. C3f were identified by m/z 2021.2 and m/z 1011.1). This is due because of the formation of both single charged and doubly charged (2+) species during the sample ionization process. We observed a Post Translational Modification (PTM) of Bradykinin. The oxidized form of proline (P) residue in position 3 of the Bradykinin fragment has been manually identified and subsequently it has been confirmed by using the oxidation (P) option as the post-translational modification parameter for the search engine.

The m/z that were not matching with any peptide via on-line search, have been identified via *de novo* sequencing. This is a manual peptides sequencing, based on known rules for peptide fragmentation. We used the sequences (or fragment of sequences) obtained for protein identification via BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins). *De novo* sequencing strategy is affected by the quality of MS/MS spectra. For this reason we did not obtain reliable identification for some low abundant peptides which produced low MS/MS spectra (such as the peptide at m/z 2484, Figure 5, picture on the left). On the other hand, sometimes the obtained fragments were not enough to get the complete peptide sequence, although we reached good fragmentations (Figure 40, picture on the right).

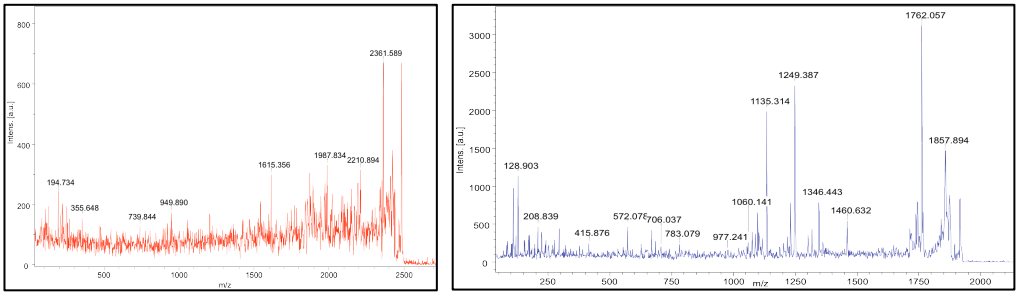


Figure 40: Tandem mass spectra of peptides at m/z 2484 (on the left) and of peptides at m/z 1919 (on the right).

Finally, a complete list of all m/z identified in the samples is reported in Table 14. This list collects peptides identified using MS/MS data, manual attribution and literature references.

Table 14. List of m/z peptides identified and the relative list of attribution name

m/z	Attribution	m/z	Attribution
948,63	bicharge of 1896	1.847,20	Complement C3f fragment
1.011,15	bicharge of 2021	1.865,22	Complement C3f fragment
1.060,66	bradykinin	1.896,20	Complement C4a fragment
1.076,66	bradykinin oxidized	1.934,23	Complement C3f fragment
1.098,64	Complement C3f fragment	2.021,29	Complement C3f
1.114,65	Complement C3f fragment, + oxygen	2.037,27	Complement C3f + oxygen
1.183,19	bicharge of 2365	2.053,27	Complement C3f +2 oxygen
1.498,91	Complement C4a fragment	2.365,37	Kininogen fragment
1.531,00	Clusterin	2.861,58	Fibrinogen α fragment

We noticed that few peptides could be different portions of the same protein; this is the case for Bradykinin (1060.6) and the Kininogen fragment at m/z 2365.4: both peptides derived from High molecular weight-kininogen. However, most of peptides recognized with this approach originated from a same protein and they shared the same sequence core. As Villaneuva et al noticed, few peptide sequences clustered into sets of overlapping fragments lined up within each group at either C- or N-terminal end and with ladder-like truncations at the opposite ends (*Villanueva J.G. et al, 2007*). In our study many of the recognized peptides were clustered in two groups, where the peptide precursors were a fragment of Complement C4a and of Complement C3f (Table 15).

Moreover, most of these peptides derived from proteins or fragment of proteins that usually are resident in blood. Bradykinin and full-lengh founder complement C3f peptide (2021,3 m/z) are usually generated from various plasma proteins through endoproteolytic cleavage: Bradykinin is usually cleaved from high molecular

weight-kininogen by plasma kallikrein while C3f is released by factor I and H after prior conversion of C3 to C3b. The full-length peptides that end with Arg or Lys amino acid preceded by a hydrophobic amino acid (Val, Leu, Phe) are cleaved by trypsin-like protease activity. Then the presence of fragments of peptides with ladder-like truncation at the opposite ends could be mainly due to some exopeptidase activities (Table 15 and Table 16). The exopeptidases require a free N-terminal amino group, C-terminal carboxyl group or both, and hydrolyze a bond not more than three residues from the terminus. The exopeptidases are further divided into aminopeptidases, carboxypeptidases, dipeptidyl-peptidases, peptidyl-dipeptidases, tripeptidyl-peptidases and dipeptidases (MEROPS database, Rawlings N.D. et al, 2008). For example, Arg/Lys amino acid could be removed by a carboxypeptidase except when it is preceded by Pro amino acid. Further exoproteases then could cleave either at the N terminal or C terminal.

Table 15: m/z corresponding to the exopeptidase activity against complement proteins.

	m/z	N-term exopeptidase
Complement C3f	2021.1	SSKITHRIHWESASLLR
	1934.1	SKITHRIHWESASLLR
	1847.1	KITHRIHWESASLLR
	1098.6	HWESASLLR
	m/z	**C-term exopeptidase**
Complement C3f	2021.1	SSKITHRIHWESASLLR
	1864.9	SSKITHRIHWESASLL
Complement C4a	1896	NGFKSHALQLNNRQIR
	1498	NGFKSHALQLNNR

Table 16: Ionic species identified and the C- and N- potential site of cleavage of exopeptidases and endopeptidases.

	m/z	N-cleavage	peptide sequence	C-cleavage
Complement C3f	2021.1	R	SSKITHRIHWESASLLR	S
	1934.1	RS	SKITHRIHWESASLLR	
	1847.1	RSS	KITHRIHWESASLLR	
	1098.6	RSSKITHRI	HWESASLLR	

	m/z	N-cleavage	peptide sequence	C-cleavage
Complement C3f	2021.1	R	SSKITHRIHWESASLLR	S
	1864.9		SSKITHRIHWESASLL	RS

	m/z	N-cleavage	peptide sequence	C-cleavage
Complement C4a	1896	STGR	NGFKSHALQLNNRQIR	GLEE
	1498	STGR	NGFKSHALQLNNR	QIRGLEE

	m/z	N-cleavage	peptide sequence	C-cleavage
Bradykinin	1060	MK	RPPGFSPFR	SS
HMW-Kininogen	2365	EKQR	KHNLGHGHKHERDQGHGH	GHG

	m/z	N-cleavage	peptide sequence	C-cleavage
Clusterin	1530	PHR	RPHFFFPKSRIV	R

Looking at the overall MALDI-TOF spectra that we used in this study and taking into account the sequence identification data, we observed that the ionic specie at 1060,6, which was present in all sample at equally high levels, corresponded to Bradykinin. A study of Yi et al*),* investigating the different spectra between blood samples collected using different anti-coagulant reagent and using protease inhibitor, showed the abundant presence of Bradykinin associated with EDTA sample tube *(Yi J. et al, 2006*. The founder peptide C3f at m/z 2021.1 was also present in all the spectra at high intensities.

Comparing the peptide profiling of the rectal cancer patients at T0 time point and the spectra of healthy subjects, we noticed that among the ionic species that were not present in the healthy subject spectra, the amino acids sequences of two were positively identified. m/z at 1934.1 was identified as a N-terminal truncated fragment of C3f peptide. In particular we observed that this fragment was produced by the removal of an N-terminal Ser from the C3f full-length peptide. m/z at 1498.9 was identified as a fragment of Complement C4a protein. It was a C-terminal fragment of a peptide with three more amino acids at m/z 1896.2 that moreover resulted down-expressed in healthy subjects. Among the 6 peptides that were differentially expressed in both healthy subjects and rectal cancer patients we noticed that the one at m/z 1098.6 represented another fragment of C3f that was

truncated at the N-terminal side. Two fragments were produced from the same precursor protein; m/z at 2365.4 was fragment of high molecular weight kininogen. Another ionic specie that was found increasingly expressed in rectal cancer patient was the one at m/z 1530.9 that was identified as a fragment of Clusterin. Moreover this peptide was found differentially expressed in relation with the histopathological response before and after the CRT and in particular, it was found more expressed in Poor responders. Clusterin is an enigmatic protein that has different/opposed functions. Several studies reported that CLU gene encoded different protein isoforms with distinct subcellular locations and functions: a pro-survival secretory form of 80kDa, a pro-survival cytoplasmatic form of 60kDa and a proapoptotic nuclear form of 50kDa are the most studied isoforms. The twelve amino acid clusterin fragment that we detected in association with the rectal malignancies and the histopathological response to CRT is located at the C-terminus of the β chain of the secretory form. Clusterin β chain is located at N portion of clusterin protein while α chain is the C-terminal part. Chains are separated by a cleavage of Val-Arg-Xaa bond and the peptide that we detected corresponds to C-terminal portion of β chain excluding the Arg at the C-terminal. Beside the clusterin fragment we identified amino acid sequence of other three fragments differentially expressed between Good and Poor responders, before the CRT: the complement C4a fragment (1498.9) and the C3f fragments (1934.1 and 1098.6). Of notice, ionic species at 1498.9 and 1934.1 were not found in the control samples and 1098.6 was found down-regulated in healthy samples. In this analysis the C4a fragment and the smaller C3f fragment were associated with Good responders while 1934.1 C3f fragments was associated with Poor responders.

After the CRT, the differences between the histopatological response classes were maintained for the fragment of C3f (1934.1) and for clusterin fragment (1530.9) even if for the last one there is no statistical significance. m/z at 1183.2 has been revealed as a bicharge ion of HMW kininogen fragment. We were not able to identify amino acid sequences other ionic species.

9 DISCUSSION

In the past decade circulating Low Molecular Weight peptidome has received increasing attentions as a powerful source of potential biomarkers that could improve diagnosis, monitoring and prognosis of various diseases including cancer. This new class of molecules have been recently available for biomarker investigation thanks to the technical progresses in clinical proteomic profiling and in simplification of complex clinical specimens. The ability to screen and discover multiple circulating molecules simultaneously have been made concrete by the recent implementation of Matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF) coupled with methods of high-abundant blood-resident protein depletion. Among different methods of protein depletion, often difficult and time consuming, fractionation methods that exploit the properties of nanomaterial have emerged. Nanodevices show the skill to reduce the complexity of the sample for isolating the low-abundance peptides, improving the ease of sample handling and the high throughput of analysis. The aim of the study has been to evaluate the Low Molecular Weight region of plasma proteome of locally advanced rectal cancer patients in order to find useful biomarkers that could be able to predict histopathologic response to neoadjuvant chemo-radio therapy. For improving the detection of circulating peptides we has developed a collaboration with the Nanomedicine Department of the Methodist Hospital Research Institute (Houston, Texas) for exploiting a size-exclusion strategy based on a Mesoporous Silica Chip (MSC). We first performed a characterization of the mechanisms of adsorption and of interaction between specific peptides and MSCs by varying the critical parameters involved in the process. The ability to precisely control the physical (chip thickness, porosity and texture) and chemical properties of the MSC allowed to isolate the specific role of each of these features in the overall process of peptide absorption, retention, and fractionation. We studied the protein adsorption of MSC with different composition and different pore size (4nm, 6nm, 7nm pore size diameter) and we selected three different peptide molecules—Angiotensin II, ACTH (CLIP) and Aprotinin - for their different characteristics in terms of charge and mass, in order to have a broad view of the adsorption process. We studied the adsorption behaviour and kinetics of the single specific peptide incubated at different pH, in relation to its pI, its molecular weight, and to the different pore size

MSCs. Besides the nanostructure surface properties, the peptide adsorption is influenced by the conditions under which the experiments are conducted and by the nature of the molecular species to be examined. The pH of the solution that is close to the pI of the specific peptide favours its adsorption into the pore irrespectively of the pore size of MSC. However, the adsorption of peptides changes differently using sample solution with different complexity in protein content, where the recovery rate was modulated according to the mass of the peptide and the pore size. In this contest we observed that the peptide uptake was promoted when the peptide was not charged or when it was slightly positive charged. Understanding the influence of these factors was essential to control and optimize the process of biomarker enrichment and its quantification in order to avoid experimental biases that might mask the pathological variations in the sample leading to the erroneous interpretation of the results.

We then focused our attention on the of the peptidomic profiles of patients with rectal cancer who received preoperative chemo-radio therapy. Although several research groups used serum, we choose to analyze plasma samples for MALDI-MS determinations (*Smith F.M. et al, 2007*). Serum is obtained after the blood sample clotting that requires a certain time and this might interfere with analyte stability. Furthermore serum proteins are subjected to proteolytic degradation, enhanced by the activation of the coagulation process and blood cell metabolism that occur ex vivo. Therefore serum collection procedure and its intrinsic characteristic could differentially alter the composition of peptides that could be potential biomarkers introducing uncontrolled clinical and analytical inter-individual variations. On the other hand, plasma samples are obtained more easily and quickly by simply centrifuging blood samples. Yi et al (*Yi J. et al, 2006*) showed a high presence of peaks in serum sample, the majority of these ones ranged from 1000 m/z to 1600 m/z. However, intrinsic plasma proteolysis is not fully inhibited by the anti-coagulant reagent added in the blood tubes. Plasma samples that we investigated were collected and processed and stored in the same range of time and using the same procedure, and we assumed that the partially protease activity could affect the spectra reasonably in the same dimension.

The peptidomic profiling analysis between healthy control subjects and rectal cancer patients plasma collected before the pCRT revealed the presence of peptides that were peculiar of the rectal cancer patients, demonstrating the diagnostic ability of this approach for discriminating the tumour disease. Furthermore, multivariate analyses of the ionic species that were observed in both healthy subjects and rectal cancer patients, identified several peptides that were differentially expressed

according with the pathological status. In particular, the analysis revealed a general over-expression in rectal cancer patients for the majority of the specific ionic species. Interestingly, analyzing several m/z individually, we observed that while Poor responders showed significant higher intensity compare to the controls, the good responders showed less pronounced expression differences (m/z at 1530.2, 1880.9).

Analyses of peptidomic profile according to the histopathologic response to CRT were performed at the three time points collected along the neo-adjuvant treatment (before the pCRT: T0, after two weeks from the initiation of the pCRT: T1, after the CRT: T2) to identify specific peptide patterns with potential predictive value. Unsupervised and supervised principal component analyses, combined with individual t-test analysis, were used to investigate the peptidomic differences between classes of response. This approach has allowed the identification of different subsets of ionic species that could potentially discriminate the histopatologic response, at each time point. In particular, before the chemo-radio therapy, we found a set of eight peptidic markers that differentiated the histopathologic response; three peptides (m/z 1082.7, 1103. 9, and 1098.6) were found increased in the Good responder group whereas five peptides (m/z 1530.2, 1880.3, 1934.8, 2154.8, 2348.1) were found increased in Poor responder groups. The multivariate logistic regression analysis confirmed the correlation of five of those peptides revealing a sensitivity of 80% and a specificity of 80%. Analyses of plasma collected after the CRT revealed a panel of six peptidic markers differentially expressed, in particular ionic species at m/z 1082.7 and 1103.9 were over-expressed in Good responders, and ionic species at m/z 1043.4, 1183.6, 1530.2 were over-expressed in Poor responders. Five of those were confirmed through the multivariate logistic regression, revealing a sensitivity of 80% and a specificity of 85%. Peptidomic profile analysis at T1 time point identified a set of five peptides that showed a better separation between the histopathologic response: ionic species at m/z 1082.7 and 1098.6 were increased in Good responders whereas m/z at 1549.5, 1865.2, and 2348.1 were increased in Poor responders. However the peptide intensity differences between the response classes were slightly significant, revealing misclassification for several patients. This finding might suggest that during the pCRT the peptide intensities could be less powerful in the response discrimination. Modulation of gene expression and protein expression as well as modulation of a number of pathways and processes such as cell cycle and cell death could be occur during the pCRT and this could complicate and alter the mechanisms of peptide formation. Moreover, observing the overall spectra at T1 time point we noticed a decreased of intensity for many ionic species compare to the baseline time

point (T0). Another limitation for this specific analysis could be the low number of plasma patients that we were able to enrol at this time point.

Comparing the specific peptide pattern that discriminated the histopathologic response groups in the three time points, we observed that several peptides that were found before the pCRT were also found after the pCRT or at T1 time point, whereas other peptides did not maintained its discriminative capacity. In particular we observed that the ionic specie at 1082.7 was found significantly expressed in Good responder group along all the three time points, indicating that its discriminative power is maintained along the therapy, whereas the ionic specie at m/z 1098.6 was found at T0 and at T1 time point and m/z at 1103.9 was found before and after the pCRT. This suggests that the presence of peptides in blood could be differentially affected during the chemo-radio therapy. To deepen this matter we monitored the presence and the different fluctuation of the ionic species along the pCRT administration, for each class of response.

Principal component analysis did not showed specific subset of peptides that could be peculiar of one time point during the pCRT, whereas ratio of single peptide intensity at T1 or T2 time point over the baseline intensity (at T0 time point) highlighted fluctuation of several ionic species. As described before, many of these m/z ionic species showed a decreasing intensity at T1 time point compare with the baseline T0 time point, whereas at T2 time point we noticed an increased intensity of many variables compared to T0 and to T1 time point in both Good and Poor responders. This trend might suggest that the pCRT could alter the release or the production of peptides. Only few peptides showed an increased trend in T1 time point and it could be interesting to deepen the analysis in the future. Observing the ionic species that showed higher fluctuation during the pCRT we could noticed that some m/z ionic species differentially changed accordingly with the histopathologic response whereas other ionic species showed tendencies similar for both Good and Poor responders, confirming the idea that at the base of the peptidic fluctuations there might be different and complex mechanisms that we could not prove in this study.

The identification of the amino-acids sequence of the higher peptides and the significant peptides was performed for potentially correlating the specific fragment protein to the pCRT response and for promoting possible hypothetical mechanisms underlying the different response. However, we could identify only few peptides of those observed as potentially involved in the discrimination of the histopathological response. Interestingly, the identified peptides were fragments of protein naturally

occur in blood or secreted protein. Some of these represented internal fragments of the precursor peptide or protein; other represented bioactive molecules or cleaved pro-peptides. In particular we identified Bradykinin, Bradykinin oxidized, the complement C3f fragment, a fragment of complement C4a, of HMW kininogen and of Clusterin. Interestingly, some of these peptides clustered into sets of overlapping fragments with ladder-like truncations at opposite ends, in particular we found 2 overlapped peptides of Complement C4a fragment and many overlapped peptides of Complement C3f.

Among the peptides that were found with high intensities and in the majority of spectra, independently to the pathology or to the response to pCRT, we identified Bradykinin, Bradykinin oxidized and the complement C3f fragment, indicating the absence of their clinical value. In particular, peaks related to Bradykinin were found in the study of Yi et al, who observed this peak as peculiar of spectra of plasma collected with EDTA. Conversely other ionic species seem to be peculiar of the tumour pathology or peculiar of the tumour response to pCRT. In particular, before the pCRT, a fragment of complement C4a and two fragment of the Complement C3f were showed associated with Good responders while Clusterin fragment complement C3f fragment, different from that associated with Good responder, were observed related to Poor responders. After the pCRT, Clusterin and the fragment of Complement C3f were maintained for Poor responders whereas in T1 time point was maintained the C3f fragment associated to Good responders. Several fragments of complement C3f with truncated ends were found associated with the histopathological response to CRT. The different peptides of the same precursor fragment might be generated by exopeptidase activities that potentially confer cancer specific signature. Exoproteases form a heterogeneous group of enzymes that play different role in the regulation of biologically active peptides. The small number of precursor proteins that are source of these fragments are naturally not biomarkers but they could serve as endogenous substrate pool for other kind of marker, the proteases. Proteolytic degradative peptides could hold important information and utility in cancer detection and in tumour response prediction. Fragments that we found differentially expressed in association with the histological response could be the indirect proof of the presence of other cancer specific molecules. Many studies were performed using serum draw blood in which protease activity as well as pathway of complement activation and of coagulation is not inhibited, suggesting that the peptidomic pattern might be produced ex-vivo by numerous protease that include both endo- and exo-protease (*Villanueva J. et al, 2006; Findeisen P. et al, 2012*). Cleavage site of the precursor fragments, from which derive the overlapping fragments with ladder-like truncation at the ends, are

generally consistent with trypsin- or chymotripsin-like activities of serine protease such as kallikreines, plasmin, thrombin, factor I. In this study we could observe similar pattern of peptide ladder even if we used plasma sample as starting biologic material. Plasma is made by adding anti-coagulant reagent like EDTA that inhibits coagulation and other similar protease activities. Although this inhibition is not completely, the ladder-like pattern might be also due to in-vivo tumour specific activity. In any case it is important to point out that we found very different control spectra comparing to the rectal cancer patient spectra before the treatment and the majority of the ionic species found discriminative of the histological response were absent or were found differentially expressed from healthy control spectra, enhancing the association with the cancer-specific event. Proteases and exo-proteases play an important role in different biological processes including cell differentiation, tissue remodelling, immunity, angiogenesis, apoptosis etc. Specifically, in solid tumours, the dysregulation of protease activity is a well-known factor that promotes tumour initiation as well as local progression and distant metastasis. Complex proteolytic networks interact with other important signalling pathways in tumour biology involving the regulation of chemokines, cytokines and active molecules (*Lopez-otin C. et al, 2008*). The tumor-associated proteases are either secreted directly by the tumour or originated from the surrounding connective tissue and from the infiltrating leucocytes as a result of the communication between stroma and tumor. There are mechanisms that concentrate the protease activity in the microenvironment such as expression of membrane-anchored proteins, activation of soluble proteases or presence of cell-surface receptors activators (*Findeisen P. et al, 2012*).

Many research groups focused on the identification of functional tumour-proteases, their putative substrates and their cleavage site, with different techniques (Villanueva J.*,2008 and 2009)*. Up to now several databases are available with information regarding protease-substrate interaction. (Merops and degradome database) (*Rawlings N.D et al, 2008; Quesada V. et al, 2009*). However tumor-associated proteases remain mostly unknown and the disease-related proteolytic activities should further be characterized. Recent reports also demonstrated and characterized some functional proteases activity in serum and plasma (*Findeisen P. et al, 2012*), but again, little is known about the stability of the proteolytic activity in biologic fluid because it is difficult to study specific protease activity in sample in which many factors and enzymes could act in parallel.

Another interesting finding is related to the presence of the proteic fragment at m/z 1530 that showed an increasing expression from control subjects to poor responders,

before the therapy. This corresponded to a fragment of secreted Clusterin β-chain. Clusterin is an enigmatic protein which role is controversial because it could be present in several isoforms that show different localization and different functions: among different isoforms, the secreted isoform is a pro-survival glycoprotein of 80kDa (sCLU). In the past few years it has become increasingly clear that Clusterin isoforms could be involved in numerous physiological processes important for carcinogenesis and tumour growth, including apoptotic cell death, cell cycle regulation, DNA repair, cell adhesion, tissue remodelling, lipid transportation, membrane recycling, and immune system regulation (*Shannan B. et al, 2006*). Furthermore Clusterin has been reported to be a stress inducible gene and it has been involved in resistance to therapy. Several controversial data have been published on colon cancer discussing its role as tumor suppressor or prosurvival factor in colon cancer. In a published paper Andersen et al (*Andersen C.L. et al, 2007*) found a Clusterin expression by neuroendocrine cells in normal colon mucosa whereas CLU gene were found *de novo* synthesized, and often secreted, in 25% of CRCs. Other research groups found a progressively increased expression of secretory CLU correlated to a significant increase of CLU in serum and stool of patients whereas the nuclear pro-apoptotic form was down-regulated (*Mazzarelli P. et al, 2009*). Within the cell, sCLU can bind Bax and prevent it translocation to the mitochondria, thereby blocking apoptosis (*Zhang H. et al, 2005*). Due to its pro-survival functions, sCLU overexpression seems to be partially responsible for increased resistance of cancer cells to various stresses including radiation (*Klokov D. et al, 2005*). Outside the cell, sCLU seems to act as a molecular chaperone, binding to stress-induced unfolded proteins, lipids and other molecules, and works to clear cell debris after tissue trauma. These observations strongly suggest that elevated sCLU levels in cancer, as well as induction of sCLU after cytotoxic agent exposures, may result in the consequent resistance of cancer cells to therapy, including cancer stem. Goetz et al highlighted the complex connection between exogenous or endogenous DNA damage and the induction of sCLU, activated by ATM kinase in response to double strand break, mediated by IGF-1 and regulated by numerous factors such as p53/NF-YA, p21, TGF-B1 cells (*Goetz E.M. et al, 2011)*.

Based on our results, we suggest that low molecular weight peptide analysis is a promising approach complementary to the conventional tumour-tissue based tests and to the clinical current strategy for predicting the response to multimodality preoperative treatment of rectal cancer.

We highlight the need to identify the other significant ionic species that we were not able to identify in the first analysis, such as the m/z at 1082.7 that is found associated with Good responders in all three time point along the pCRT.

Furthermore, the showed data require to be confirmed by analysing other cases and by implementing the study also with patients with inflammatory- non-malignant disease for evaluating the protease pattern that could affect this status. It is noticed that numerous malignant diseases associated with inflammation are influencing the protease pattern in body fluid. On the other hand these data prepare the ground for subsequent investigation on potential peptidases and proteases involved in cancer progression and suggested the potential for improving the personalized neo-adjuvant therapy. Further investigations on the specific roles of Clusterin could be interesting for understanding its potential relation with tumour therapy resistance and eventually thinking about alternative therapy target-specific.

10 CONCLUSION

The Ph.D. project has involved the development and the evaluation of blood-based approaches for discovering potential biomarkers associated with the histopathological response of locally advanced rectal cancer treated with pre-operative chemo-radio therapy (pCRT). The non-invasive nature and the dynamism for which different molecules could be detected in blood, according with physiological and pathological states, has given us the possibility to monitor the response along the administration of the treatment. For this purpose we investigated the circulating cell free DNA and its integrity (cfDNA integrity) along the chemo-radio treatment, demonstrating a different kinetic of cfDNA integrity in association with the histopathological response. The decrement of cfDNA integrity for the positive tumour response, during the CRT, could be potentially related to the tumour mass reduction and to a possible different mechanism of cfDNA release. The positive association between the presence of tumour with cfDNA concentration and with cfDNA integrity, was also confirmed in other studies, performed in different cancer types, in which we observed an interesting indication of their potential diagnostic and prognostic significance. However, current limitations in clinical practice arise by the high inter-individual variability that could be affected by numerous technical and biological factors, decreasing the accuracy of the methods.

More interesting findings have been found by evaluating the low molecular weight region of the proteome circulating in blood, in relation to the histopathological response to pCRT. The development of a controlled method of peptide fractionation and of peptide analysis has allowed the identification of specific patterns of peptides (either before and after the CRT) that correlated with the histopathological response to CRT, highlighting the potential utility of this method as a clinical response prediction tool. Moreover, the results of this study have opened the way to further investigation in different fields that could be involved in modulating the response to chemo-radio therapy. In particular, we pointed out the presence of proteolytic degradative pattern of peptides that suggested the involvement of potential specific functional tumour-protease and exopeptidase activity. In addition, the association of a fragment of Clusterin secretory isoform with the Poor response to pCRT, both before and after the pCRT, stimulate our interest to future deepen its role with the tumour therapy response. Furthermore these data suggested the potential for improving the personalized neo-adjuvant therapy and eventually the design of future clinical trials.

Based on our results, cfDNA and the low molecular weight peptide analysis suggest the potential to become promising approaches complementary to the conventional tumour-tissue based tests and to the clinical current strategy for predicting the response to multimodality preoperative treatment of rectal cancer.

11 REFERENCES

- Aebersold R. et al., *Mass spectrometry-based proteomics*, (2003) Nature 422(13) 198-207.
- Albrethsen J., *Reproducibility in Protein Profiling by MALDI-TOF Mass Spectrometry*, (2007) Clinical Chemistry 53:5 852–858
- Andersen C.L. et al, *Clusterin expression in normal mucosa and colorectal cancer, molecular and cellular proteomics*, (2007) 6.6 1039-1048
- Andrew D. et al., *The Genomics of Colorectal Cancer: State of the Art*, (2008) Current Genomics 9:1-10
- Barcellos-Hoff M.H. et al., *A systems biology approach to multicellular and multi-generational radiation responses,* (2006) Mutat Res 2:32–38
- Bateson H. et al., *Use of matrix-assisted laser desorption/ionisation mass spectrometry in cancer research,* (2011) Journal of Pharmacological and Toxicological Methods, 64:197–206
- Beggs A.D. et al., *The genomics of colorectal cancer: state of the art*, (2008) Curr Genomics 9(1):1-10.
- Bertolini F. et al., *Prognostic and predictive value of baseline and posttreatment molecular marker expression in locally advanced rectal cancer treated with neoadjuvant chemoradiotherapy,* (2007) Int. J. Radiation Oncology Biol. Phys. 68(5): 1455-1461
- Bouamrani A. et al, *Mesoporous silica chips for selective enrichment and stabilization of low molecular weight proteome,* (2010) Proteomics 10(3): p. 496-505.
- Brettingham-Moore K.H. et al, *Using gene expression profiling to predict response and prognosis in gastrointestinal cancers—the promise and the perils,* (2011) Ann Surg Oncol 18(5): 1484-1491
- Buczacki S. et al., *Stem cells, quiescence and rectal carcinoma: an unexplored relationship and potential therapeutic target*, (2011) British Journal of Cancer 105: 1253-1259
- Burnett-Hartman A.N. et al., *Colorectal polyp type and the association with charred meat consumption, smoking, and microsomal epoxide hydrolase polymorphisms,* (2011) Nutr Cancer
- Chan K.C.A. et al., *Persistent Aberrations in Circulating DNA Integrity after Radiotherapy Are Associated with Poor Prognosis in Nasopharyngeal Carcinoma Patients*, (2008) Clin Cancer Res 14(13): 4141–4145.
- Cho K.R. et al., *Genetic Alterations in the Adenoma-Carcinoma Sequence,* (1992) Cancer 70:1727-1731.
- Choi J. J. Et al., *The role of macrophages in the in vitro generation of extracellular DNA from apoptotic and necrotic cells*, (2005) Immunology 115, 55–62.

- Cunningham D. et al., *Colorectal cancer*, (2010) Lancet 375(9719): 1030– 1047
- Curtin K. et al., *Somatic alterations, metabolizing genes, and smoking in rectal Cancer*, (2009) Int J Cancer. 125(1): 158–164
- De Bruin E.C., *Prognostic Value of Apoptosis in Rectal Cancer Patients of the Dutch Tota lMesorectal Excision Trial: Radiotherapy Is Redundant in Intrinsically High-Apoptotic Tumors*, (2006) Clin Cancer Res 12(21)
- Deligezer U. et al., *Effect of Adjuvant Chemotherapy on Integrity of Free Serum DNA in Patients with Breast Cancer,* (2008) Ann NY Acad Sci 1137:175-179
- Eischeid A.C. et al., *SYTO dyes and EvaGreen outperform SYBR Green in real-time PCR*, (2011) BMC Research Notes 4:263.
- Fearon E.R., *Molecular Genetics of Colorectal Cancer*, (2011) Annu. Rev. Pathol. Mech. Dis 6:479-507.
- Ferlay J. et al., *Estimates of the cancer incidence and mortality in Europe in 2006*. (2007) Ann Oncol. 18: 581-592
- Ferrari M., *Cancer nanotechnology: opportunities and challenges*, (2005) Nat Rev Cancer 5(3): p. 161-71.
- Ferrari M., *Nanotechnology-enabled medicine*, (2005) Discov Med. 5(28):363-366.
- Findeisen P. et al., *Functional protease profiling with reporter peptides in serum specimens of colorectal cancer patients: demonstration of its routine diagnostic applicability*, (2012) Journal of Experimental & Clinical Cancer Research 31:56
- Findeisen P. et al., *Functional protease profiling for diagnosis of malignant disease,* (2012) Proteomics Clin. Appl. 6, 60-78
- Findeisen P. et al., *Preanalytical impact of sample handling on proteome profiling experiments with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry*, (2005) Clin. Chem. 51:2409–2411.
- Fleischhacker M. et al, *Methods for isolation of cell-free plasma DNA strongly affect DNA yield*, (2011) Clin Chim Acta 20;412(23-24):2085-2088.
- Fleischhacker M. et al., *Circulating nucleic acids (CNAs) and cancer–a survey*, (2007) Biochim Biophys Acta 1775: 181–232.
- Fodde R. et al., *Mutations in the APC tumor suppressor gene cause chromosomal instability*. (2001) Nature 3:433-438.
- Fodde R. et al., *Wnt/beta-catenin signaling in cancer stemness and malignant behavior*, (2007) Curr Opin Cell Biol 19(2):150-158.
- Frattini M. et al., *Different Genetic Features Associated with Colon and Rectal Carcinogenesis,* (2004) Clinical Cancer Research 10: 4015–4021.
- Freitas P.P. et al., *Optimization and integration of magnetoresistive sensors*, (2011) Journal of SPIN, (1):71-91
- Ghadimi B.M. et al., *Effectiveness of gene expression profiling for response prediction of rectal adenocarcinomas to preoperative chemoradiotherapy*, (2005) J Clin Oncol 23(9): 1826-1838
- Giacona M.B. et al., *Cell-free DNA in human blood plasma: length measurements in patients with pancreatic cancer and healty controls,* (1998) Pancreas 17:89-97.

- Godin B, et al, *Multistage nanovectors: from concept to novel imaging contrast agents and therapeutics*, (2011) Acc Chem Res, 44(10): 979-89.
- Goetz E.M. et al., *ATM-dependent IGF-1 induction regulates secretory clusterin expression after DNA damage and in genetic instability*, (2011) Oncogene 30: 3745–3754
- Gormally E. et al., *Circulating free DNA in plasma or serum as biomarker of carcinogenesis: Practical aspects and biological significance*, (2007) Mutation Research 635:105–117.
- Grade M. et al., *The molecular basis of chemoradiosensitivity in rectal cancer: implications for personalized therapy*, (2012) Langenbecks Arch Surg 397: 543–555.
- Gudnason H. et al., *Comparison of multiple DNA dyes for real-time PCR: effects of dye concentration and sequence composition on DNA amplification and melting temperature*, (2007) Nucleic Acids Res. 35(19): e127.
- Harrington K. et al., *Molecular Biology for the Radiation Oncologist: the 5Rs of Radiobiology meet the Hallmarks of Cancer*, (2007) Clinical Oncology 19: 561-571
- Hilario M. et al., *Processing and classification of protein mass spectra,* (2006) Mass Spectrometry Reviews, 25: 409– 449
- Holdenrieder S. et al., *DNA Integrity in Plasma and Serum of Patients with Malignant and Benign Diseases,* (2008) Ann N Y Acad Sci 1137(1):162–170.
- Hooker C.M. et al, *A prospective cohort study of rectal cancer risk in relation to active cigarette smoking and passive smoke exposure*, (2008) Ann Epidemiol. 18:28–35.
- Hortin G.L., *The MALDI-TOF Mass Spectrometric View of the Plasma Proteome and Peptidome,* (2006) Clinical Chemistry 52:7 1223–1237
- Hu Y, et al., *Nanodevices in diagnostics*, (2011) Wiley Interdiscip Rev Nanomed Nanobiotechnol. 3(1): 11-32.
- Hu Y, et al., *Nanotexture Optimization by Oxygen Plasma of Mesoporous Silica Thin Film for Enrichment of Low Molecular Weight Peptides Captured from Human Serum*, Sci China Chem, 2010. 53(11): 2257-2264.
- Hu Y, et al., *Tailoring of the nanotexture of mesoporous silica films and their functionalized derivatives for selectively harvesting low molecular weight protein*, ACS Nano, 2010. 4(1): 439-51
- *Introduction to Fourier Transform Infrared Spectrometry*, 2001 Thermo Nicolet Corporation
- Jahr S. et al., *DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells,* (2001) Cancer Res 61:1659-65
- Jass J.R., *Classification of colorectal cancer based on correlation of clinical, morphological and molecular features*, (2007) Histopathology 50(1): 113-130.
- Jemal A. et al, *Global cancer statistics*, (2011) Cancer J. Clin. 61(2):69

- Jiang W. W. et al., *Increased plasma DNA integrity index in head and neck cancer patients,* (2006) Int J Cancer, 119: 2673-2676
- Jimenez C.R. et al., *Proteomics of colorectal cancer: Overview of discovery studies and identification of commonly identified cancer-associated proteins and candidate CRC serum markers*, (2010) Journal of proteomics 73: 1873-1895
- Julien L.A. et al., *Current Neoadjuvant Strategies in Rectal Cancer*, (2010) Journal of Surgical Oncology 101:321–326
- Jung K. et al., *Cell-free DNA in the blood as a solid tumor biomarker—A critical appraisal of the literature,* (2010) Clin. Chim. Acta 411:1611–1624.
- Kapiteijn E. et al., *Mechanisms of oncogenesis in colon versus rectal cancer*, (2001) Journal of Pathology 195: 171-178.
- Karas M. et al., *Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons,* (1988) Analytical Chemistry 60: 2299–2301.
- Karpova M.A. et al., *Cancer-specific MALDI-TOF profiles of blood serum and plasma: Biological meaning and perspectives,* (2010) Journal of proteomics 73: 537-551
- Kauh J. et al., *Colorectal cancer prevention*, (2004) Curr Probl Cancer 28: 240-264.
- Kim I.J. et al., *Microarray gene expression profiling for predicting complete response to preoperative chemoradiotherapy in patients with advanced rectal cancer*, (2007) Dis Colon Rectum 50 (9): 1342-1353
- Kinzler K.W. et al., *Lessons from Hereditary Colorectal Cancer*, (1996) Ceel 87:159-170
- Klokov D, et al, *IR-inducible clusterin gene expression: a protein with potential roles in ionizing radiation-induced adaptive responses, genomic instability, and bystander effects*, (2004) Mutation Research 568: 97-110
- Kohler C. et al., *Cell-free DNA in the Circulation as a Potential Cancer Biomarker, Anticancer Research* (2011) *31*: 2623-2628
- Kulasingam V. et al., *Strategies for discovering novel cancer biomarkers through utilization of emerging technologies,* (2008) Nature clinical practice oncology 5(10)
- Kulikova, GA, et al., *In vitro studies of interaction of modified silica nanoparticles with different types of immunocompetent cells*. J Biomed Mater Res A, 2010. 95(2): p. 434-9.
- Kumar V. et al., *Robbins and Cotran's pathologic basis of disease* 8th edition, 2009
- Kuremsky J.G., *Biomarkers for response to neoadjuvant chemoradiation for rectal cancer*, (2009) Int. J. Radiation Oncology Biol. Phys., 74(3): 673–688
- Larsson S.C. et al, *Diabetes mellitus and risk of colorectal cancer: a meta-analysis*, (2005) J Natl Cancer Inst. 97(22):1679
- Lee H.J. et al., *Biomarker discovery from the plasma proteome using multidimensional fractionation proteomics*, (2006) Current Opinion in Chemical Biology 10:42–49

- Lewis J.K. et al., *Matrix-assisted Laser Desorption/Ionization Mass Spectrometry in Peptide and Protein Analysis,* (2000) Encyclopedia of Analytical Chemistry, R.A. Meyers (Ed.) pp. 5880–5894, John Wiley & Sons Ltd, Chichester
- Li F.Y. et al., *Colorectal cancer, one entity or three*, (2009) J Zhejiang Univ Sci B 10(3): 219-29
- Li J.N. et al., *Differences in gene expression profiles and carcinogenesis pathways between colon and rectal cancer*, (2012) Journal of Digestive Diseases 13: 24–32
- Li M. et al., *Colorectal Cancer or Colon and Rectal Cancer? Clinicopathological Comparison between Colonic and Rectal Carcinomas*, (2007) Oncology 73:52–57
- Longley D.B. et al., *5-Fluorouracil: mechanisms of action and clinical strategies* , (2003) Nature review 3:330-338 (2003)
- López-Otín C. et al., *Proteases: Multifunctional Enzymes in Life and Disease*, (2008) The Journal of Biological Chemistry, 283: 30433-30437.
- Martin K. J. et al., *A need for basic research on fluid-based early detection biomarkers*. Cancer Res, 70: 5203-5206.
- Mazzarelli P. et al., *The dual face of CLU: from normal to malignant phenotype*, (2009) Adv Cancer Res. 105:45-61.
- Menard C. et al., *Exposure with Serum Proteomic Analysis Discovering Clinical Biomarkers of Ionizing Radiation*, (2006) Cancer Res 66:1844-1850
- Mothersill C. et al., *Radiation-induced bystander effects- implications for cancer*, (2004) Nature Review cancer 4: 158-164
- Nannini M. et al., *Gene expression profiling in colorectal cancer using microarray technologies: results and perspectives,* (2009) Cancer Treat Rev 35(3): 201–209
- Norris J.L. et al., *Processing MALDI Mass Spectra to Improve Mass Spectral Direct Tissue Analysis,* (2007) Int J Mass Spectrom. 260(2-3): 212–221.
- Pajonk F. et al., *Radiation Resistance of Cancer Stem Cells: The 4 R's of Radiobiology Revisited*, (2010) Stem Cells 28(4): 639–648.
- Petricoin E.F., *The blood peptidome: a higher dimension of information content for cancer biomarker discovery*, (2006) Nature review 6: 961:966
- Pietrowska M. et al., *MALDI-MS-Based Profiling of Serum Proteome: Detection of Changes Related to Progression of Cancer and Response to Anticancer Treatment*, (2012) International Journal of Proteomics, ID 926427, 10 pages
- Push W. et al., *Mass spectrometry-based clinical proteomics,* (2003) Pharmacogenomics 4(4): 463-476.
- Quesada V. et al., *The Degradome database: mammalian proteases and diseases of proteolysis,* Nucleic Acids Research, (2009) 37, Database issue D239–D243
- Rawlings N.D. et al, *MEROPS: the peptidase database*. (2008) Nucleic Acids Res. 36 Database issue, D320-325.
- Reerink O. et al., *Molecular prognostic factors in locally irresectable rectal cancer treated preoperatively by chemoradiotherapy*, (2004) Anticancer Res. 24:1217–1221.

- Ren X.R. et al., *Focal adhesion kinase in netrin-1 signaling*, (2004) Nat. Neurosci. Nat 7(11): 1204-1212.
- Ricci-Vitiani L. et al., *Colon cancer stem cells,* (2008) Gut 57;538-548.
- Riehemann K, et al*, Nanomedicine--challenge and perspectives*, (2009) Angew Chem Int Ed Engl, 48(5): 872-97.
- Rodel C. et al., *Rectal cancer: state of the art in 2012*, (2012) current opinion oncology 24(4).
- Roy S, et al., *Novel multiparametric approach to elucidate the surface amine-silanization reaction profile on fluorescent silica nanoparticles*, (2010) Langmuir, 26(23): 18125-18134.
- Sakamoto J.H. et al, *Enabling individualized therapy through nanotechnology*, Pharmacol Res, 2010. 62(2): p. 57-89.
- Sanz-Pamplona R. et al., *Gene Expression Differences between Colon and Rectum Tumors*, (2011) Clin Cancer Res 17:7303-7312.
- Sauer R. et al., *German Rectal Cancer Study Group. Preoperative versus postoperative chemoradiotherapy for rectal cancer*, (2004) N Engl J Med 351:1731–1740.
- Savino R. et al., *Mesopore-assisted profiling strategies in clinical proteomics for drug/target discovery*, (2012) Drug Discovery Today 17 (3/4)
- Schmidt B. et al., *Integrity of Cell-Free Plasma DNA in Patients with Lung Cancer and Nonmalignant Lung Disease,* (2008) Ann N Y Acad Sci 1137(1):207–213.
- Schwarzenbach H. et al., *Cell-free nucleic acids as biomarkers in cancer patients*, (2011) Nat Rev Cancer, 11(6): 426-437
- Selzer E. et al., *Basic principles of molecular effects of irradiation*, (2012) Wien Med Wochenschr 162/3–4: 47–54
- Shannan B. et al*., Challenge and promise: roles for clusterin in pathogenesis, progression and therapy of cancer*, (2006) Cell Death and Differentiation 13: 12–19.
- Smith F.M. et al., *Pathological and molecular predictors of the response of rectal cancer to neoadjuvant radiochemotherapy*, (2006) EJSO 32: 55–64
- Smith F.M. et al., *Combination of SELDI-TOF-MS and Data Mining Provides Early-stage Response Prediction for Rectal Tumors Undergoing Multimodal Neoadjuvant Therapy*, (2007) Annals of Surgery 245(2): 259-266
- Sorenson, G. D. *et al. Soluble normal and mutated DNA sequences from single-copy genes in human blood*, Cancer Epidemiol. Biomarkers Prev. **3**, 67–71 (1994).
- Sriram K.B. et al., *Pleural fluid cell-free DNA integrity index to identify cytologically negative malignant pleural effusions including mesotheliomas*, (2012) BMC Cancer, 12:428.
- Stewart B.W., Kleihus P., Editors. *World Cancer Report*. Lyon: IARC Press, (2003).
- Taback B. et al., *Quantification of Circulating DNA in the Plasma and Serum of Cancer Patients,* (2004) Ann N Y Acad Sci 1022(1):17–24.

- Taguchi A. et al., *Unleashing the Power of Proteomics to Develop Blood-Based Cancer Markers,* (2013) Clinical Chemistry 59:1000–000
- Terracciano R. et al., *Selective binding and enrichment for low-molecular weight biomarker molecules in human plasma after exposure to nanoporous silica particles*, (2006) Proteomics, 6(11): 3243-50.
- *The rectum*, chapter 71, Bailey & Love's Short Practice of Surgery, 25th edition, 2012).
- Tirumalai R.S. et al., *Characterization of the Low Molecular Weight Human Serum Proteome,* (2003) Molecular & Cellular Proteomics 2.10 1096:1103
- Tsang J. C. H. et al., *Circulating nucleic acids in plasma/serum*, (2007) Pathology 39(2):197-207
- Umetani N. et al., *Increased integrity of free circulating DNA in sera of patients with colorectal or periampullary cancer: direct quantitative PCR for ALU repeats,* (2006) Clin Chem 52: 1062-1069
- Umetani N. et al., *Prediction of breast tumor progression by integrity of free circulating DNA in serum,* (2006) J Clin Oncol 24: 4270-4276,
- Vainio H. et al., *Fruit and vegetables in cancer prevention*, (2006) Nutr. Cancer 54(1): 111-142.
- Valentini V. et al., *Evidence and research in rectal cancer,* (2008) Radiotherapy and Oncology 87: 449–474.
- Villanueva J et al, *Differential eoprotease activities confer tumor-specific serum peptidome patterns*, the journal of clinical investigation, 2006. 116(1):271-284
- Villanueva J. et al., *Monitoring Peptidase Activities in Complex Proteomes by MALDITOF Mass Spectrometry,* (2009) *Nat Protoc*. 4(8): 1167–1183
- Vogelstein B. et al., *Surfing the p53 network*, (2000) Nature, 408: 307-310.
- Wang B.G. et al., *Increased Plasma DNA Integrity in Cancer Patients*, (2003) Cancer Res 63(14): 3966–3968.
- Wang J. et al*., Carcinogen metabolism genes, red meat and poultry intake, and colorectal cancer risk*, (2012) Int J Cancer. 130(8):1898-907.
- Watanabe T. et al., *Prediction of sensitivity of rectal cancer cells in response to preoperative radiotherapy by DNA microarray analysis of gene expression profiles*, (2006) Cancer Res 66(7): 3370–3374
- Wong J.W.H. et al., *SpecAlign—processing and alignment of mass spectra datasets*, (2005) Bioinformatics application notes, 21(9):2088–2090
- Wong, J.W.H. et al., *Application of Fast Fourier Transform Cross-Correlation for the Alignment of Large Chromatographic and Spectral Datasets.,* (2005) *Anal. Chem*., **77**: 5655-5661
- Woodward W.A. et al., *Molecular basis of radiation therapy*, chapter 50 of the textbook "The molecular basis of cancer" by Mendelsohn J., Howley P.M., Israel M.A., Gray J.W., Thompson C.B., third edition, 2008
- Worthley D. L. et al., *Colorectal carcinogenesis: road maps to cancer,* (2007) World Journal of Gastroenterology, 13(28): 3784-3791

- Yi J. et al., *Inhibition of intrinsic protelytic activities moderates preanalytical variability and instability of human plasma*, (2006) Journal of proteome research, 6:1768-1781.
- Yang Y. ET AL., *Integrative Genomic Data Mining for Discovery of Potential Blood-Borne Biomarkers for Early Diagnosis of Cancer*, (2008) PLosONE 3(11).
- Zampino M.G. et al., *Rectal cancer*, (2004) Critical Reviews in Oncology/Hematology 51:121–143
- Zampino M.G. et al., *Rectal cancer,* (2009) Critical Reviews in Oncology / Hematology 70(2): 160-182
- Zavoral M. et al., *Colorectal cancer screening in Europe***,** (2009) World J Gastroenterol 15(47): 5907-5915
- Zhang, Y. et al., *Enrichment of low-abundance peptides and proteins on zeolite nanocrystals for direct MALDI-TOF MS analysis*, (2005) Angew. Chem. Int. Ed. 44: 615–617
- Zhang H. et al, *Clusterin inhibits apoptosis by interacting with activated Bax*, (2005) Nature Cell Biology 7(9): 909-915
- Zitt M. et al., *Circulating cell-free DNA in plasma of locally advanced rectal cancer patients undergoing preoperative chemoradiation: a potential diagnostic tool for therapy monitoring*, (2008) Dis. Markers, 25: 159-165

- Figure 1: http://www.aafp.org/afp/2001/0615/p2391.htm
- Figure 2: http://www.hopkinscoloncancercenter.org/CMS/CMS_Page.aspx?CurrentUDV=59&CMS_Page_ID=0B34E9BE-5DE6-4CB4-B387-4158CC92408

12 PUBLICATIONS

- Parodi A, Quattrocchi N, van de Ven AL, Chiappini C, Evangelopoulos M, Martinez JO, Brown BS, Khaled SZ, Yazdi IK, Enzo MV, Isenhart L, Ferrari M, Tasciotti E, *Synthetic nanoparticles functionalized with biomimetic leukocyte membranes possess cell-like functions*, Nat Nanotechnol. 2013 Jan;8(1): 61-8.
- Agostini M, Bedin C, Pucciarelli S, Enzo MV, Briarava M, Seraglia R, Ragazzi E, Traldi P, Molin L, Urso ED, Mammi I, Viel A, Lise M, Tasciotti E, Biasiolo A, Pontisso P, Nitti D. *APC I1307K mutations and forkhead box gene (FOXO1A): another piece of an interesting correlation*. Int J Biol Markers. 2012, 27(1):13-19
- Agostini M, Enzo MV, Bedin C, Belardinelli V, Goldin E, Del Bianco P, Maschietto E, D'Angelo E, Izzi L, Saccani A, Zavagno G, Nitti D. *Circulating cell-free DNA: a promising marker of regional lymph node metastasis in breast cancer patients*. Cancer biomarkers, 2012 11(2-3): 89-98
- Agostini M, Pucciarelli S, Enzo MV, Del Bianco P, Briarava M, Bedin C, Maretto I, Friso ML, Lonardi S, Mescoli C, Toppan P, Urso E, Nitti D., *Circulating cell-free DNA: a promising marker of pathologic tumor response in rectal cancer patients receiving preoperative chemoradiotherapy*, Ann Surg Oncol. 2011, 18(9):2461-8.
- Agostini M, Pucciarelli S, Calore F, Bedin C, Enzo MV, Nitti D. *miRNAs in colon and rectal cancer: a consensus for their true clinical value*. Clin Chim Acta. 2010 Sep 6;411(17-18):1181-6
- Agostini M, Enzo MV, Morandi L, Bedin C, Pizzini S, Mason S, Bertorelle R, Urso E, Mescoli C, Lise M, Pucciarelli S, Nitti D, *A ten markers panel provides a more accurate and complete microsatellite instability analysis in mismatch repair-deficient colorectal tumors*, Cancer Biomark, 2010;6(1):49-6

In press

- Agostini M, Bedin C, Enzo MV, Molin L, Traldi P, D'Angelo E, Maschietto E, Serraglia R, Ragazzi E, Prevedello L, Foletto M, Nitti D, *Multivariate analysis approach to the serum peptide profile of morbidly obese patients*, Disease markers

Submitted

- Zane M, Agostini M, Enzo MV, Casal Ide E, Del Bianco P, Torresan F, Merante Boschin I, Pennelli G, Saccani A, Nitti D, Pelizzo MR, Circulating cell-free DNA, *SLC5A8* and *SLC26A4* hypermethylation, $BRAF^{V600E}$: a non-invasive tool panel for detection of Thyroid Tumors

In preparation

- Enzo MV, Ziemyes A, Hu Y, Bouamrani A, Sakamoto J, Chan D, Khaled SM, Chiappini C, Bedin C, Agostini M, Nitti D, Ferrari M, Tasciotti E., *Evaluation of nano scale selective fractionation, chimica-physical determinants of adsorption and retention of peptides on mesoporous silica chip*.

nature nanotechnology

ARTICLES

PUBLISHED ONLINE: 16 DECEMBER 2012 | DOI: 10.1038/NNANO.2012.212

Synthetic nanoparticles functionalized with biomimetic leukocyte membranes possess cell-like functions

Alessandro Parodi[1,2‡], Nicoletta Quattrocchi[1,3‡], Anne L. van de Ven[1], Ciro Chiappini[1†], Michael Evangelopoulos[1], Jonathan O. Martinez[1,4], Brandon S. Brown[1,4], Sm Z. Khaled[1], Iman K. Yazdi[1,5], Maria Vittoria Enzo[1,6], Lucas Isenhart[1], Mauro Ferrari[1] and Ennio Tasciotti[1]*

The therapeutic efficacy of systemic drug-delivery vehicles depends on their ability to evade the immune system, cross the biological barriers of the body and localize at target tissues. White blood cells of the immune system—known as leukocytes—possess all of these properties and exert their targeting ability through cellular membrane interactions. Here, we show that nanoporous silicon particles can successfully perform all these actions when they are coated with cellular membranes purified from leukocytes. These hybrid particles, called leukolike vectors, can avoid being cleared by the immune system. Furthermore, they can communicate with endothelial cells through receptor–ligand interactions, and transport and release a payload across an inflamed reconstructed endothelium. Moreover, leukolike vectors retained their functions when injected *in vivo*, showing enhanced circulation time and improved accumulation in a tumour.

Biophysical barriers protect the body by regulating the trafficking, exchange and clearance of foreign materials[1]. For example, the endothelial cells of blood vessels are responsible for identifying and capturing potential hazards[2]. Upon entering the blood, systemically injected drugs and particles are tagged with proteins (a process called opsonization[3]) and subsequently removed by the mononuclear phagocyte system[4]. To function properly, systemic agents must therefore avoid clearance by the immune system, negotiate their way past the vascular barrier, and localize at the target tissue in sufficient quantities to be effective[5].

Encapsulating free drugs in nanoparticles is advantageous, because it prolongs the drug half-life, improves site-specific targeting, reduces side effects and enhances therapeutic efficacy[6–8]. Moreover, the physical and chemical properties of the nanoparticles themselves confer new capabilities *in vivo*[9,10]. For example, optimization of particle size[11,12], shape[13] and surface charge[14] enhances passive tumour targeting via a mechanism known as enhanced permeation and retention[15]. Furthermore, modification of the surfaces of particles with polyethylene glycol (PEG) improves the biodistribution of chemotherapeutics[16], and bioconjugation of active targeting molecules enhances delivery to specific cells[17,18]. More recently, efforts have focused on the development of multistage vectors that decouple each of these functions *in vivo*[19]. Based around a nanoporous silicon (NPS) platform[18,20], these particles can carry a variety of cargoes[20–22], navigate through blood flow[21,23], recognize and bind specific endothelial targets[23,24] and protect therapeutic cargo so as to achieve enhanced efficacy[22,25].

However, particle-based drug delivery has not yet reached its full therapeutic potential[21,26]. Avoiding opsonization and non-specific clearance remains a challenge[27], and the use of PEG cannot completely prevent clearance[28] and results in activation of the human complement system[29]. Particles with longer circulation times have demonstrated an increased probability of tumoritropic accumulation in tumours with fenestrated endothelia[30,31]. Regrettably, not all tumours are characterized by increased vessel permeability. Under these circumstances, the use of current delivery systems is highly ineffective and the negotiation of vascular barriers is a mandatory step to attain the desired drug biodistribution and acceptable therapeutic indices.

Given the complexity of mass transport in the vascular compartment, it is not surprising that 'biomimetic camouflage' strategies are gaining popularity[32,33]. Virus-based carriers[34], targeted protocells[35] and bio–nano hybrid systems[36,37] have been proposed as potential strategies for overcoming vascular barriers to drug delivery. Here, we describe an approach to transfer bioactive cellular components to the surface of synthetic particles in order to confer unique functions not otherwise attainable through current bioconjugation techniques. This new generation of injectable carriers, termed leukolike vectors (LLVs), are produced by camouflaging NPS particles with cellular membranes isolated from freshly harvested leukocytes. Using a combination of *in vitro* and *in vivo* experiments, we show that LLVs are able to avoid opsonization, delay uptake by the mononuclear phagocyte system, preferentially bind inflamed endothelium, and facilitate chemotherapeutics transport across the endothelium while eluding the lysosomal pathway.

Assembly and characterization

LLVs were produced by cloaking NPS with cellular membranes isolated from freshly harvested leukocytes (Fig. 1a). The membranes

[1]Department of Nanomedicine, The Methodist Hospital System Research Institute, Houston, Texas 77030, USA, [2]Department of Experimental Oncology and Molecular Medicine, Fondazione IRCCS Istituto Nazionale Tumori, Milan 20133, Italy, [3]Department of Experimental Medicine, University of Milano Bicocca, Milan 20052, Italy, [4]The University of Texas Health Science Center at Houston, Houston, Texas 77030, USA, [5]Department of Biomedical Engineering, University of Houston, Houston, Texas 77204, USA, [6]Department of Oncological and Surgical Sciences, 2nd Surgical Clinic, University of Padua, Padua 35128, Italy; [†]Present address: Department of Materials, Imperial College London, South Kensington Campus, London SW7 2AZ, UK; [‡]These authors contributed equally to this work. *e-mail: etasciotti@tmhs.org

Int J Biol Markers 2012;27(1):13-19
DOI: 10.5301/JBM.2011.8908

ORIGINAL ARTICLE

APC I1307K mutations and forkhead box gene (*FOXO1A*): another piece of an interesting correlation

Marco Agostini[1-3#], Chiara Bedin[1,3#], Salvatore Pucciarelli[1], Mariavittoria Enzo[1,3], Marta Briarava[1], Roberta Seraglia[4], Eugenio Ragazzi[5], Pietro Traldi[4], Laura Molin[4], Emanuele Damiano Urso[1], Isabella Mammi[6], Alessandra Viel[7], Mario Lise[1], Ennio Tasciotti[2], Alessandra Biasiolo[8], Patrizia Pontisso[8], Donato Nitti[1]

[1]Department of Surgical, Oncological and Gastroenterological Sciences, 2nd Surgical Clinic, University of Padua, Padua - Italy
[2]The Department of Nanomedicine, The Methodist Hospital Research Institute, Houston, TX - USA
[3]Pediatric Research Institute - Città della Speranza, Padua - Italy
[4]CNR-ISTM, Padua - Italy
[5]Department of Pharmacology and Anesthesiology, University of Padua, Padua - Italy
[6]ULSS 13 Dolo, Venice - Italy
[7]Division of Experimental Oncology, Centro Riferimento Oncologico, IRCCS, Aviano - Italy
[8]Fifth Medical Clinic, Department of Clinical Experimental Medicine, University of Padua, Padua - Italy

[#]These authors contributed equally to this article.

ABSTRACT

Purpose: **Germline nonsense and frameshift mutations in the adenomatous polyposis coli (*APC*) gene are found in approximately 90% of individuals affected by familial adenomatous polyposis (FAP) and a genotype-phenotype relationship has been observed. Missense mutations have also been found in a few cases, even if their role in FAP is still unknown. An association between a missense mutation, *APC* I1307K, and the risk of sporadic colorectal cancer (CRC) has been reported. In order to improve the knowledge about the genetic effect of APC I1307K on the phenotype, we tried a new approach using matrix-assisted laser desorption/ionization mass spectrometry (MALDI/MS).**

Experimental design: **An *APC* mutation (I1307K) was found in an index case of a non-Jewish woman and her son with attenuated familial adenomatous polyposis (A-FAP) and no family history of cancer. In order to evaluate whether the presence and abundance of the ionic species are related to the presence of cancer or the presence of mutation, comparative analyses of 11 healthy clean-colon subjects, 59 patients with CRC (stage II n=19, stage III n=23, stage IV n=17) without polyps, and 9 FAP patients, carriers of a nonsense mutation in the *APC* gene, were evaluated.**

Results: **Comparative analysis of serum protein profiles of the index patient and her healthy son, FAP and sporadic CRC patients, and subjects with preneoplastic lesions showed a characteristic abundance of ionic species at *m/z* 905, which was not present in healthy controls. Two peptides were identified from MALDI/MS/MS spectra of *m/z* 905 belonging to the kininogen-1 precursor and the human forkhead box protein 01A (FOXO1A). FOXO1A was present in only two subjects carrying I1307K, but not in other patients.**

Conclusions: **Our findings seem to suggest a relationship between *m/z* 905, FOXO1A and the development and growth of colorectal cancer. FOXO1A fragment determination in serum with MALDI/MS might be a promising approach for early detection of colon carcinoma or for the development of targeted therapies.**

Key words: FOXO1A, Colorectal cancer, *APC* I1307K

Received: June 10, 2011; Accepted: August 30, 2011

INTRODUCTION

Familial adenomatous polyposis (FAP) has autosomal dominant inheritance caused by mutations in the adenomatous polyposis coli (*APC*) tumor suppressor gene located on chromosome 5q. Pathogenetic germline mutations in the *APC* gene are nonsense and frameshift in about 90% of affected individuals (1) and a genotype-phenotype relationship has been observed (2). Missense mutations have also been found in a few cases, even if their role in FAP is still unknown. In a mouse model, different *APC* mutations confer different degrees of susceptibility to tumorigenesis (3).

APC I1307K is a missense mutation involving a T>A

Cancer Biom arkiv 11 (2011/2012) 89–98
DOI 10.3233/CBM-2012-0265
IOS Press

Circulating cell-free DNA: A promising marker of regional lymphonode metastasis in breast cancer patients

M. Agostini[a,b,c,*], M.V. Enzo[a,c], C. Bedin[a,c], V. Belardinelli[a], E. Goldin[a], P. Del Bianco[d], E. Maschietto[a,c], E. D'Angelo[a,c], Leo Izzi[a], A. Saccani[e], G. Zavagno[a] and D. Nitti[a]

[a]*Department of Surgical, Oncological and Gastroenterological Sciences, 2nd Surgical Clinic, Padova, Italy*
[b]*The Department of Nanomedicine, The Methodist Hospital Research Institute, Houston, TX, USA*
[c]*Istituto di Ricerca Pediatrica- Città della Speranza, Viale della Ricerca, Padova, Italy*
[d]*Clinical Trials and Biostatistics Unit, Veneto Oncological Institute – IRCSS, Padova, Italy*
[e]*Euroclone S.p.a, Pero (MI), Italy*

Abstract. *Purpose:* We undertook the current study with untreated breast cancer to (1) role the variations in the plasma levels of cfDNA and the size distribution in early stage, (2) determine the frequency in plasma of methylation of three candidate genes, *RASSF1A*, *MAL*, and *SFRP1*, and (3) to determine whether detection of cfDNA variations and methylation changes in plasma might have specific clinical utility.
Methods and materials: Thirty-nine patients women patients (median age 64 years; range, 36–90 years) who underwent surgery for primary BR and 49 healthy females' subjects (control group without any breast lesion) were evaluated. The cfDNA levels were analyzed using quantitative real-time polymerase chain reaction of β-globin. Based on the ALU repeats, the cfDNA was considered as either total (fragments of 115 bp, ALU115) or tumoral (fragments of 247 bp, ALU247). The association between the levels of the ALU247, ALU115 repeat, and ALU 247/115and the pathologic tumor characteristics was analyzed.
Used methylight qPCR method, cfDNA from plasma samples of healthy donors and patients with breast cancer were evaluated for the diagnostic value of the methylation status of three genes (*RASSF1A*, *MAL*, *SFRP1*) frequently methylated in breast cancer.
Results: The baseline levels of cfDNA were significantly higher in the patients with cancer, and the level of ALU247 was the most accurate circulating cfDNA marker in discriminating the cancer from non-cancer subjects.
A high statistical significance was found by considering the T stage and patients with regional LN metastasis positive cancers showed significantly higher cfDNA level of ALU247. Moreover, patients with methylation of at least one of the gene under investigate showed a higher quantity of cfDNA ALU115 ($p < 0.0001$) and ALU247 level ($p < 0.0001$).
Conclusions: We observed that necrosis could be a potential source of circulating tumour-specific cfDNA ALU247; and that cfDNA ALU247 and methylated cfDNA (*RASSF1A*, *MAL* and *SFRP1*) are both a phenotypic feature of tumour biology.

Keywords: Breast cancer, regional LN metastasis, cfDNA

1. Introduction

In most developed and many developing countries, breast cancer (BR) is the most frequent cancer and the leading cause of cancer death in women. Current screening methods fail to detect many cancers that present at a later date as a result of symptoms [1,2]. Early detection and a reliable follow-up of breast cancer are crucial for successful treatment. But a definitive diagnosis, a tumour biopsy is required. However, a non-invasive test for early detection of the disease and for monitoring disease progression has been a goal for many researchers.

Although the potential of biomarkers to aid in the early detection, diagnosis, prevention, and treatment of breast cancer is broadly recognized and numerous

*Corresponding author: Marco Agostini, MD, Department of Oncological and Surgical Gastroenterological Sciences, Surgical Clinic II, University of Padova, Via Giustiniani 2, 35128 Padova, Italy. Tel.: +39 049 8214574; Fax: +39 049 670891; E-mail: m.agostini@unipd.it.

ISSN 1574-0153/11/$27.50

Ann Surg Oncol (2011) 18:2461–2468
DOI 10.1245/s10434-011-1638-y

Annals of
SURGICAL ONCOLOGY
OFFICIAL JOURNAL OF THE SOCIETY OF SURGICAL ONCOLOGY

ORIGINAL ARTICLE – COLORECTAL CANCER

Circulating Cell-Free DNA: A Promising Marker of Pathologic Tumor Response in Rectal Cancer Patients Receiving Preoperative Chemoradiotherapy

Marco Agostini, PhD[1], Salvatore Pucciarelli, MD[1], Maria Vittoria Enzo, PhD[1], Paola Del Bianco, PhD[2], Marta Briarava, PhD[1], Chiara Bedin, PhD[1], Isacco Maretto, MD[1], Maria Luisa Friso, MD[3], Sara Lonardi, MD[4], Claudia Mescoli, MD[5], Paola Toppan, MD[1], Emanuele Urso, MD, PhD[1], and Donato Nitti, MD[1]

[1]Department of Oncological and Surgical Sciences, 2nd Surgical Clinic, University of Padova, Padova, Italy; [2]Clinical Trials and Biostatistics Unit, Veneto Institute of Oncology IOV – IRCCS, Padova, Italy; [3]Department of Radiotherapy and Nuclear Medicine, Veneto Institute of Oncology IOV – IRCCS, Padova, Italy; [4]Department of Medical Oncology, Unit 1st, Veneto Institute of Oncology IOV – IRCCS, Padova, Italy; [5]Department of Pathology, University of Padova, Padova, Italy

ABSTRACT

Purpose. The circulating cell-free DNA (cfDNA) in plasma has been reported to be a marker of cancer detection. The aim of this study was to investigate whether the cfDNA has a role as response biomarker in patients receiving preoperative chemoradiotherapy (CRT) for rectal cancer.

Methods. Sixty-seven patients (median age 61 years; male/female 42/25) who underwent CRT for rectal cancer were evaluated. After tumor regression grade (TRG) classification was made, the patients were classified as having disease that responded (TRG 1–2) and that did not respond (TRG 3–5) to therapy. Plasma samples were obtained from patients before and after CRT. The cfDNA levels were analyzed by quantitative real-time polymerase chain reaction of β-globin. On the basis of the Alu repeats, the cfDNA was considered as either total (fragments of 115 bp, Alu 115) or tumoral (fragments of 247 bp, Alu 247). The association between the pre- or post-CRT levels and between variations during CRT of the Alu 247, Alu 115 repeat, and Alu 247/115 ratio (cfDNA integrity index) and the pathologic tumor response was analyzed.

Results. The baseline levels of cfDNA were not associated with tumor response. The post-CRT levels of the cfDNA integrity index were significantly lower in responsive compared to nonresponsive disease ($P = 0.0009$). Both the median value of the Alu 247 repeat and the cfDNA integrity index decreased after CRT in disease that responded to therapy ($P < 0.005$ and $P < 0.005$, respectively) compared to disease that did not respond to therapy ($P = 0.83$ and $P = 0.726$, respectively). The results of the multivariable logistic regression analysis showed that only the cfDNA integrity index was significantly and independently associated with tumor response to treatment.

Conclusions. The plasma levels of the longer fragments (Alu 247) of cfDNA and the cfDNA integrity index are promising markers to predict tumor response after preoperative CRT for rectal cancer.

Preoperative chemoradiotherapy (CRT) is currently considered the standard treatment for locally advanced rectal cancer. The oncological outcomes using this approach are encouraging, with rates of local and distant recurrence at the 5-year mark ranging 6–9 and 33–36%, respectively.[1–3] However, the rates of chemotherapy-, radiotherapy-, and surgery-related toxicity, as well as bowel and sexual dysfunction, are disappointing. The peculiar aspects of this approach are related to clinical overstaging, which may result in an unnecessary neoadjuvant treatment in almost one-fifth—18% in a German trial—of patients.[2] In addition, the disease of approximately 40% of patients shows poor or no response to preoperative CRT, while a

Presented as a podium presentation at the American Association Cancer Research annual meeting, April 2009, Denver, CO.

First Received: 1 October 2010;
Published Online: 17 March 2011

S. Pucciarelli, MD
e-mail: puc@unipd.it

Clinica Chimica Acta 411 (2010) 1181–1186

Contents lists available at ScienceDirect

Clinica Chimica Acta

journal homepage: www.elsevier.com/locate/clinchim

Invited critical review

miRNAs in colon and rectal cancer: A consensus for their true clinical value

Marco Agostini *, Salvatore Pucciarelli, Federica Calore, Chiara Bedin, MariaVittoria Enzo, Donato Nitti

Clinica Chirurgica II, Dipartimento Scienze Oncologiche e Chirurgiche, Università di Padova, Italy

ARTICLE INFO

Article history:
Received 22 February 2010
Received in revised form 3 May 2010
Accepted 3 May 2010
Available online 7 May 2010

Keywords:
microRNAs
Colorectal cancer

ABSTRACT

Numerous miRNAs are deregulated in human cancers and experimental evidence indicates that they can play roles as oncogenes or tumor suppressor genes.
Colorectal cancer represents a wide and exciting area of research for molecular biology, due to the growing need of a molecular classification as well as prognostic and predictive molecular factors that may guide oncologists in the clinical management of patients. The aim of this review is to analyze the state of art of the miRNA expression profiles in colorectal cancer to explore some perspectives in this research field.

Contents

1. Introduction

Colorectal cancer (CRC) is one of the most common causes of cancer-related deaths, and it causes death primarily through liver metastasis. Recent progress in diagnosis and treatment has enabled clinicians to save the lives of many patients at early stages of the disease, but the prognosis for patients with advanced disease or systemic metastasis is still very poor. About 30% of recurrent CRC patients have liver metastasis, and more than 70% of them are not candidates for the curative resection [1]. Therefore, predicting the potential aggressiveness of a primary tumor could help in improving patient survival by identifying those who should receive intensive pre and post-operative adjuvant chemotherapy.

CRC has a single clonal origin at the initial stage of the disease, but a malignant tumor contains multiple cell populations with different properties, such as growth rate, karyotype, immunogenicity, sensitivity to various drugs and the ability to invade and develop metastases [2].

This heterogeneity is the major obstacle to effective treatment of CRC because of the variation in clinical patterns and treatment efficacies. Nowadays, molecular biology represents one of the most interesting topics in medical oncology because it provides a global and detailed view on the molecular changes involved in tumor progression, leading to a better understanding of the carcinogenesis process, to discovering new prognostic markers and novel therapeutic targets. Despite the fact that clinical and pathological parameters are available for the classification and prognostic stratification of cancer, they may be inadequate in everyday practice due to the great biologic and genetic heterogeneity of this multiform disease. Moreover, in the era of patient-tailored therapies, a more comprehensive knowledge of downstream signaling pathways is strictly necessary for the discovery of novel tumor

* Corresponding author. Clinica Chirurgica II, Dipartimento di Scienze Oncologiche e Chirurgiche, Università di Padova Via Giustiniani 2, 35128 Padova, Italy. Tel.: +39 049 8214374 2069; fax: +39 049 651891.
E-mail address: m.agostini@unipd.it (M. Agostini).

0009-8981/$ – see front matter © 2010 Published by Elsevier B.V.
doi:10.1016/j.cca.2010.05.002

Cancer Biomarkers 6 (2009/2010) 49–61
DOI 10.3233/CBM-2009-0118
IOS Press

A ten markers panel provides a more accurate and complete microsatellite instability analysis in mismatch repair-deficient colorectal tumors

Marco Agostini[a,1,*], Maria Vittoria Enzo[a,1], Luca Morandi[b,1], Chiara Bedin[a], Silvia Pizzini[a], Silvia Mason[c], Roberta Bertorelle[d], Emanuele Urso[a], Claudia Mescoli[e], Mario Lise[f], Salvatore Pucciarelli[a] and Donato Nitti[a]

[a]*II Section of Surgery Clinic, Department Oncologic and Surgical Sciences, University of Padova, Italy*
[b]*Department of Ematology and Oncological Science "L. & A. Seragnoli" Section of Pathology, Bellaria Hospital, University of Bologna, Italy*
[c]*AB ANALITICA s.r.l. Padova, Italy*
[d]*Istituto Oncologico Veneto, IRCCS, Padova, Italy*
[e]*Section of Pathology,University of Padova Italy*
[f]*Surgical Oncology, Regional Oncologic Center (CRO), Aviano, Italy*

Abstract. Tumour microsatellite instability (MSI) is useful in identifying patients with hereditary non-polyposis colorectal cancer (HNPCC) with defective *DNA mismatch repair* (*MMR*) genes. A reference Bethesda panel has limitations resulting from the inclusion of dinucleotide markers, which are less sensitive and specific for detection of tumours with MMR deficiencies. We developed a multiplex PCR assay with additional four mononucleotide markers and one dinucleotide marker (NR-21, NR-24, BAT-40, TGF-BetaR and D18S58) for a rapid and proper classification of MSI-H, MSI-L and MSS colorectal cancers. Two tetranucleotide markers were added to identify sample mix-ups and/or contamination. Results: all the 44 cases test cases were in agreement with previous classification except for three cases: one case MSI-H-Bethesda unstable only for dinucleotides markers shifted to MSI-L category and two cases MSI-L-Bethesda unstable for mononucleotide markers shifted to MSI-H category. Immunohistochemistry analysis revealed that these two MSI-H cases did not expressed hMLH1 and they were found to be methylated at the *MLH1* promoter, while the first one that shifted to MSI-L showed MMR protein expression. Conclusion: a complete panel of ten markers including four dinucleotide and six mononucleotide microsatellites allows accurate evaluation of tumor MSI status.

Keywords: Microsatellite instability, colorectal cancer, mononuclotide repeats

1. Introduction

Colorectal carcinoma (CRC) is the second leading cause of cancer-related deaths in the Western world. There are ~110,000 new cases and 50,000 deaths due to the disease per year in the U.S. A small fraction of patients (~5%) have a hereditary colorectal cancer syndrome [1]. One of these syndromes, hereditary non-polyposis colorectal cancer (HNPCC), accounts for 1% to 2% of all colorectal carcinomas [2]. In most HNPCC colorectal tumors, MSI has been shown to result from defects in DNA mismatch repair. Mutations in the *hMLH1* or *hMSH2* genes are the most common defects

[1]The first three authors contributed equally to this article.
*Corresponding author: Marco Agostini PhD, Clinica Chirurgica II, Dipartimento di Scienze Oncologiche e Chirurgiche, Università di Padova, Via Giustiniani 2, 35128 Padova, Italy. Tel.: +39 049 8214374; Fax: +39 049 651891; E-mail: m.agostini@unipd.it.

ISSN 1574-0153/09,10/$27.50

RINGRAZIAMENTI

Voglio esprimere la mia riconoscenza alla Prof.ssa Zanovello, direttrice della Scuola di Dottorato in Scienze Oncologiche e Chirurgiche e al Prof. Donato Nitti, direttore del Dipartimento di Scienze Chirurgiche, Oncologiche e Gastroenterologiche, cui la Scuola di Dottorato afferisce.

Un particolare ringraziamento al Dott. Marco Agostini per aver seguito il mio percorso formativo e al Prof. Mauro Ferrari per avermi dato la possibilità di svolgere un periodo di collaborazione presso il Dipartimento di Nanomedicina dell'istituto che dirige, The Methodist Hospital Research Institute, in Houston, Texas.

Ringrazio il Prof. Ennio Tasciotti per aver seguito il mio lavoro presso il Dipartimento di Nanomedicina e per avermi dato l'opportunità di collaborare con persone che provenivano da diversi rami professionali.

Le esperienze positive come quelle negative hanno contribuito nel rendermi più forte facendomi crescere a livello personale oltre che professionale. Per questo un ringraziamento particolare va a tutte le splendide persone conosciute a Houston, il cui affetto, amicizia e supporto mi accompagneranno sempre.

Ringrazio anche Chiara, Silvia, Loris, Mariangela e Clara e tutte le persone del laboratorio a Padova, per il loro supporto lavorativo e morale.

Un profondo grazie alla mia famiglia, a Stefania e a tutte le persone che mi sono state vicine in questi anni, al loro amore e alla fiducia senza cui nulla di tutto ciò sarebbe potuto accadere.