



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

Università degli Studi di Padova  
Dipartimento di Scienze del Farmaco

SCUOLA DI DOTTORATO DI RICERCA IN:  
BIOLOGIA E MEDICINA DELLA RIGENERAZIONE  
Ciclo XXVI

**UTILIZZO DI *NANOPOROUS SILICA CHIP* NELLO STUDIO  
DEL PROFILO PEPTIDICO PLASMATICO:  
APPLICAZIONE NELLO SVILUPPO E PROGRESSIONE  
DEL CANCRO COLORETTALE**

**Direttore della Scuola:** Ch.mo Prof. Maria Teresa Conconi

**Supervisore:** Ch.mo Prof. Cosimo De Filippis

**Co-supervisore:** Dott. Marco Agostini

**Dottoranda:** Chiara Bedin



# INDICE

RIASSUNTO

ABSTRACT

<b>1. INTRODUZIONE</b>	1
<b>1.1. Cancro coloretale</b>	
<b>1.1.1. Anatomia e funzione dell'intestino crasso</b>	1
<b>1.1.2. Epidemiologia e fattori di rischio</b>	1
<b>1.1.3. Cancerogenesi</b>	2
1.1.3.1. Omeostasi cellulare	3
1.1.3.2. Lesioni pre-neoplastiche	4
1.1.3.3. Sequenza adenoma-carcinoma	6
<b>1.1.4. Screening, diagnosi e terapia</b>	8
1.1.4.1. Screening del CRC	8
1.1.4.2. Diagnosi di CRC	8
1.1.4.3. Trattamento del CRC	9
1.1.4.4. Follow-up	10
<b>1.1.5. Marcatori tumorali</b>	10
1.1.5.1. Fattori prognostici e di predizione da risposta al trattamento	11
<b>1.2. Proteomica</b>	13
<b>1.2.1. Proteomica clinica</b>	13
1.2.1.1. Ricerca di biomarcatori	13
1.2.1.2. Identificazione del profilo proteico	15
1.2.1.3. Frazione peptidica plasmatica e peptidi a basso peso molecolare	16
<b>1.2.2. Nanoproteomica</b>	18
<b>2. SCOPO</b>	21
<b>3. MATERIALI E METODI</b>	23
<b>3.1. Casistica dei pazienti</b>	23
<b>3.2. Nanoporous Silica Chip (NSC): fabbricazione e         caratteristiche</b>	23
<b>3.3. Frazionamento dal plasma con NSC</b>	26
<b>3.4. Analisi del profilo peptidico con MALDI-TOF</b>	26
3.4.1. Processamento dei dati ottenuti dall'analisi MALDI-TOF	27
<b>3.5. Identificazione peptidica con MALDI-TOF/TOF</b>	27
<b>3.6. Analisi dai dati e metodi statistici</b>	30

<b>4. RISULTATI</b>	31
<b>4.1. Ottimizzazione del processo di frazionamento con NSC</b>	31
4.1.1 Valutazione di NSC con diversa dimensione di pori	31
4.1.2. Valutazione della rimozione di proteine dalla superficie di NSC	32
4.1.3. Valutazione del tempo di incubazione del campione	33
4.1.4. Valutazione della temperatura di incubazione del campione	34
<b>4.2. Valutazione del profilo peptidico plasmatico nel CRC</b>	35
4.2.1. Spettri di massa	36
<b>4.3. Fase esplorativa</b>	37
4.3.1. Analisi dei 4 gruppi di soggetti in studio	37
4.3.2. Analisi a 3 gruppi di soggetti	38
<b>4.4. Confronto diretto</b>	42
4.4.1. Controlli vs Adenomi	42
4.4.2. Controlli vs CRC	45
<b>4.5. Identificazione delle specie ioniche</b>	47
<b>5. DISCUSSIONE</b>	55
<b>6. CONCLUSIONI</b>	65
<b>7. BIBLIOGRAFIA</b>	67
<b>8. PUBBLICAZIONI</b>	73
<b>8.1. Sottomissioni a revisione editoriale</b>	82

## RIASSUNTO

**Background:** Per la ricerca di biomarcatori proteici, l'approccio basato sulla identificazione di un singolo marcatore hanno, finora, dimostrato l'incapacità di individuare inequivocabilmente il cancro, in parte perché gli attuali biomarcatori tumorali sono anche espressi nelle cellule normali. Il profilo proteico si basa, invece, sulla rilevazione di pattern identificativo di centinaia di proteine in un alto numero di campioni. Il contenuto informativo più alto sembra risiedere nelle proteine/peptidi a basso peso molecolare (LMW), la frazione meno abbondante nei fluidi biologici, che sembrano rispecchiare meglio gli stati fisiopatologici dei tessuti. Per l'analisi in spettrometria di massa è necessario selezionare e arricchire questa frazione del plasma.

**Scopo:** Si è focalizzato sull'utilizzo di un dispositivo nanoporoso (*Nanoporous Silica Chip*, NSC) per il recupero della frazione a basso peso molecolare da plasma, in campioni di pazienti con cancro coloretale (CRC) a vari stadi di progressione tumorale e sullo studio del relativo profilo peptidico mediante tecnica MALDI-TOF.

**Materiale e metodi:** NSC è un prototipo creato e brevettato dal Laboratorio del Prof. M. Ferrari (Dip. di Nanomedicina presso The Methodist Hospital Research Institute, Houston, TX, USA) costituito da un supporto in silicone, di circa 10cm, rivestito da un sottile strato di silice con una struttura a nano-pori. È stato necessario lo sviluppo di un protocollo semplice e veloce di frazionamento dei peptidi plasmatici. Il protocollo standardizzato è stato applicato per il frazionamento di campioni di plasma di 34 soggetti sani (controlli), 27 con lesione pre-cancerosa (adenoma) e 33 con CRC a stadio precoce (stadio I-II) e 34 con CRC a stadio tardivo (stadio III-IV). La frazione ottenuta è stata analizzata con spettrometria di massa MALDI-TOF e i dati calibrati, allineati e normalizzati sono stati sottoposti ad attenta e accurata analisi statistica univariata e multivariata con lo scopo di identificare differenze nel profilo peptidico plasmatico nei diversi gruppi di soggetti.

**Risultati:** Si è ottenuta una buona classificazione dei controlli rispetto ai pazienti, ma una scarsa distinzione tra i gruppi di soggetti con adenoma, CRC con stadio precoce e tardivo. Da tale analisi, si sono individuate alcune specie ioniche

rappresentate con diversa intensità nei vari gruppi, che sono state sottoposte a identificazione della sequenza amminoacidica mediante MALDI-TOF/TOF. Sono risultate essere tutte frammenti peptidici di proteine plasmatiche appartenenti alla risposta infiammatoria e al sistema del complemento. In particolare i frammenti peptidici del C3f e C4-A/B sembrano originare dal proprio precursore a seguito di tagli enzimatici a carico di endoproteasi ed esoproteasi. La presenza di alcuni di questi frammenti è risultata essere variabile nei gruppi in esame. Inoltre si è identificato un interessante frammento peptidico che deriva dal propeptide di ITIH4 (Inter- $\alpha$ -trypsin inhibitor heavy chain H4), una proteina di fase acuta secreta dal fegato e coinvolta nello sviluppo e rigenerazione epatica. Una ricerca bibliografica ha identificato che frammenti di ITIH4, possono essere possibili biomarcatori per vari stati patologici come infarto e cancro alla mammella e prostata.

**Conclusioni:** La metodica ottimizzata è semplice e veloce e NSC è un dispositivo con ampie potenzialità di utilizzo. È necessario approfondire lo studio sul particolare pattern proteolitico di peptidi ottenuti, che suggerisce il coinvolgimento di un'attività esoproteasica ben distinta e la presenza di proteasi specifiche del tumore del colon-retto.

## ABSTRACT

**Background:** To date single tumour molecule identification approach for the protein biomarkers discovery is unable to unequivocally recognize the cancer, because current tumor biomarkers are also expressed in normal cells. The protein profiling identify a specific pattern of hundred of protein in numerous specimens. Low molecular weight (LMW) proteins/peptides, the fraction less abundant of the biological fluids, seem to contain disease-specific information and correlate to the tissue pathological status. The detection by MS-analysis of low abundant fraction and LMW peptides remains a critical challenge.

**Aim:** A *Nanoporous Silica Chip* (NSC) was used to select and purify LMW plasma peptides in samples of colorectal cancer (CRC) patients to study the peptide profiling in association to development and progression of tumour.

**Materials and Methods:** NSC is a patented prototype of Prof. M Ferrari Labs (Dept. of Nanomedicine of The Methodist Hospital Research Institute, Houston TX). It is a disc of 10 cm of diameter with a superficial nanoporous silica thin film. A standardized, fast and simple protocol was validated to perform the selection of plasma LMW peptides. 34 health subjects, 27 patients with pre-cancer lesion (adenoma), 33 patients with early stage of CRC (stage I-II) and 34 patients with late stage of CRC (stage III-IV) was enrolled for this study. MS-analysis by MALDI-TOF instrument was performed on fractionated LMW plasma samples. Data were calibrated, aligned and normalized, and then they were undergone to accurate univariate and multivariate statistical analysis to highlight and identify difference of plasma peptide profile intensity comparing the 4 study group.

**Results:** Good classification of control group was obtained regard patient group, but poor discrimination was observed between adenoma, early stage CRC and late stage CRC. Consequently, 29 ionic species was differentially expressed in study groups. MALDI-TOF/TOF analysis was identified the aminoacid sequence of several ionic species. All of them were fragment peptide of plasma protein arising to inflammatory response and system of complement. Moreover, peptide fragments of C3f and C4-A/B could be generated from precursor peptide by endoprotease and exoprotease cleavage. A peptide fragment was originated from propeptide of ITIH4 (Inter- $\alpha$ -trypsin inhibitor heavy chain H4). ITIH4 was

secreted to liver and it belongs to phase acute proteins involved in inflammatory responses. It may also play a role in liver development and regeneration. Some studies were identified some ITIH4 peptide fragments involved in several pathological status as ischemic stroke, breast, ovarian and prostate cancer.

**Conclusions:** A fast and simple method was set to perform study with NSC. NSC is a new tool with wide potential application. A deep study of obtained peptide proteolytic pattern was necessary suggesting a colorectal cancer specific proteases and exoprotease activity.



# 1. INTRODUZIONE

## 1.1. Cancro Coloretale

### 1.1.1. Anatomia e funzione dell'intestino crasso

L'intestino crasso è l'ultima porzione del canale digerente, è la continuazione dell'intestino tenue attraverso la valvola ileocecale e si apre all'esterno mediante l'orifizio anale. È lungo 1,5-2 m e ha un diametro di 7-8 cm nella parte prossimale e di 2,5-3 cm nella parte distale (Figura 1A).

La parete dell'intestino crasso è costituita da quattro tuniche: la *sierosa*, che è una dipendenza del peritoneo; la *muscolare*, che ha uno strato interno circolare e uno esterno longitudinale non continuo (tenie); la *sottomucosa*, che è costituita da connettivo lasso ed infine la *mucosa*, che è rivestita da un epitelio prismatico semplice con numerose ghiandole intestinali (*cripte di Lieberkuhn*). Queste ultime sono rivestite in prevalenza da cellule mucipare, che aumentano sempre di più in senso distale, da cellule di Paneth, che secernono enzimi e scarsamente presenti o assenti nel colon destro, e da cellule assorbenti con orletto a spazzola (enterociti). Sono numerosi e voluminosi i noduli linfatici solitari che talora raggiungono la sottomucosa (Figura 1B).

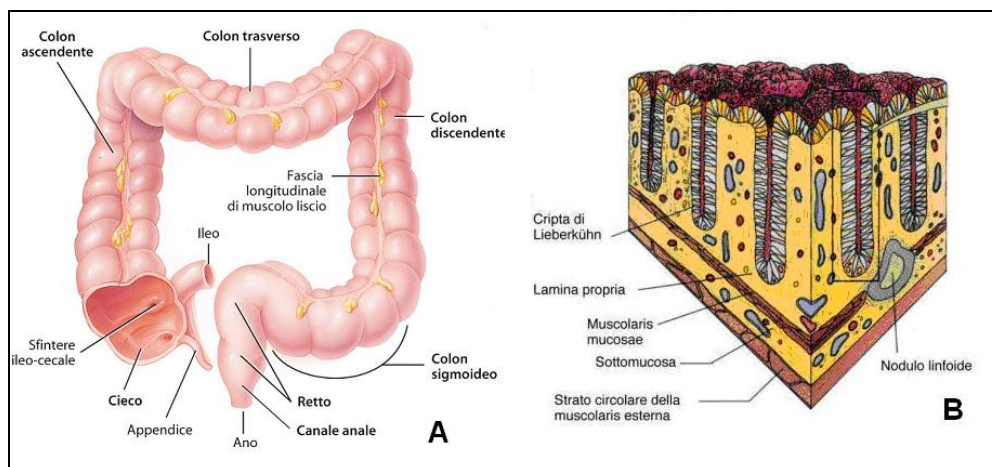


Figura 1. Rappresentazione macroscopica (A) e istologica (B) del colon.

La mucosa dell'intestino crasso produce un secreto alcalino, molto ricco di mucina e privo di enzimi digestivi; inoltre nel colon è presente un'abbondante

flora batterica che determina la decomposizione di molecole del chimo residue. Nell'intestino crasso si verificano complessi processi rappresentati essenzialmente dalla digestione delle sostanze non digerite nell'intestino tenue, dalla secrezione di metalli ed elettroliti, dalla produzione di vitamine K, B<sub>1</sub>, B<sub>2</sub> e B<sub>12</sub> e dalla grande capacità di assorbimento di acqua (circa 300-400 ml) e di elettroliti (60 meq di Na<sup>+</sup>) [1,2].

### 1.1.2. Epidemiologia e fattori di rischio

Il cancro coloretale (CRC) è una delle maggiori cause di morbilità e di mortalità nei Paesi occidentali e la sopravvivenza a 5 anni è inferiore al 60% [3]. In Europa il CRC è la più comune forma di cancro (412 900 casi, 12.9% di tutti i casi di cancro) dopo di quello alla mammella, e la seconda causa di morte dopo il cancro ai polmoni [4].

In Italia, il CRC è in assoluto il tumore a maggiore insorgenza nella popolazione, con quasi 55.000 diagnosi stimate per il 2013. Tra i maschi si trova al terzo posto (14% di tutti i nuovi tumori), preceduto da prostata e polmone, nelle femmine è al secondo posto con 14%, preceduto dalla mammella (Tabella 1). Sulla base della situazione attuale si prevedono per il 2013 quasi 20.000 decessi per CRC (di cui il 54% negli uomini), che è al secondo posto nella mortalità per tumore con 10% nei maschi e 12% nelle femmine (Tabella 2) [5].

**Tabella 1. Primi cinque tumori più frequentemente diagnosticati e proporzione sul totale dei tumori (esclusi i carcinomi della cute) per sesso. Pool AIRTUM 2006-2009.**

Posizione	Maschi	Femmine	Tutta la popolazione
1°	Prostata (20%)	Mammella (29%)	Colon-retto (14%)
2°	Polmone (15%)	Colon-retto (14%)	Mammella (13%)
3°	Colon-retto (14%)	Polmone (6%)	Prostata (11%)
4°	Vescica (10%)	Utero corpo (5%)	Polmone (11%)
5°	Stomaco (5%)	Tiroide (5%)	Vescica (7%)

**Tabella 2. Prime cinque cause di morte tumorali più frequenti e proporzione sul totale dei decessi oncologici per sesso. Pool AIRTUM 2006-2009.**

Posizione	Maschi	Femmine	Tutta la popolazione
1°	Polmone (26%)	Mammella (16%)	Polmone (20%)
2°	Colon-retto (10%)	Colon-retto (12%)	Colon-retto (11%)
3°	Prostata (8%)	Polmone (11%)	Mammella (7%)
4°	Fegato (7%)	Pancreas (7%)	Stomaco (6%)
5°	Stomaco (7%)	Stomaco (6%)	Pancreas (6%)

Circa l'80% dei carcinomi del colon-retto insorge da lesioni pre-cancerose (adenomi con componente displastica via-via crescente).

Gli stili di vita e la familiarità sono da tempo chiamati in causa quali fattori di aumento del rischio di incidenza di queste lesioni. Tra i primi spiccano fattori dietetici quali il consumo di carni rosse, insaccati, farine e zuccheri raffinati, il sovrappeso e la ridotta attività fisica, il fumo e l'eccesso di alcol [6]. Ulteriori condizioni di rischio sono costituite dalla malattia di Crohn e dalla rettocolite ulcerosa [7,8]. Circa un terzo dei tumori del colon-retto presenta caratteristiche di familiarità ascrivibile a suscettibilità ereditarie [9]: il rischio di ammalarsi di carcinoma del colon-retto è circa doppio rispetto alla popolazione generale, nel caso in cui sia presente un consanguineo di 1° grado, ed è più che triplo qualora quest'ultimo abbia contratto il tumore prima dei 50 anni di età [10]. Solo una parte di questo rischio familiare (2-5%) è riconducibile a sindromi conosciute in cui sono state identificate mutazioni genetiche associate ad aumento del rischio di carcinoma coloretale [11]. Fra queste, le sindromi caratterizzate dall'insorgenza di polipi, quali la poliposi adenomatosa familiare (FAP) e quelle non poliposiche come la sindrome di Lynch. L'ereditarietà è per lo più su base autosomico-dominante.

### **1.1.3. Cancerogenesi**

#### **1.1.3.1. Omeostasi cellulare**

Le cripte di Lieberkuhn sono strutture funzionali costituite da una porzione apicale costituita da diversi tipi cellulari differenziati quali enterociti, cellule di Paneth e cellule mucipare. Nella porzione basale è presente il compartimento proliferativo costituito da cellule staminali totipotenti e cellule progenitrici che hanno la capacità di differenziarsi in tutti i tipi epiteliali [12,13]. Le cellule staminali si dividono prevalentemente in modo asimmetrico, generando cioè una cellula con uguali caratteristiche staminali (*self-renewal*) che rimane alla base della cripta, mentre l'altra gradualmente si differenzia e si divide ulteriormente migrando verso il lume intestinale, formando un continuo strato in perpetuo movimento. Questa tipologia di divisione permette di mantenere costante il

numero di cellule staminali nelle cripte [14]. L'aumento di cellule alla base della cripta è compensato in apice dall'apoptosi delle cellule differenziate, in un processo che richiede circa 2-3 giorni. Il rinnovo cellulare è molto rapido ed è costituito da una serie coordinata di eventi (proliferazione, migrazione, differenziazione e apoptosi) verso il lume intestinale.

L'omeostasi cellulare è mantenuta dalla WNT (*wingless-type MMTV integration site family*) pathway che tramite fattori di crescita permette il controllo della localizzazione nucleare della  $\beta$ -catenina (il fattore trascrizionale per TCF, *T-cell factor/lymphocyte enhancer factor*), mediante il legame con la proteina *adenomatosis polyposis coli* protein (APC). Questo previene l'accumulo nucleare di  $\beta$ -catenina e quindi la trascrizione di proteine favorevoli al ciclo cellulare [12,13].

### **1.1.3.2. Lesioni pre-neoplastiche**

#### 1.1.3.2.1. Aberrant Cripte Foci (ACF)

Le cellule staminali sembrano essere i bersagli ideali per l'iniziazione del processo di cancerogenesi, date le caratteristiche di longevità e di autorinnovamento, dal momento che esse possono accumulare mutazioni *de novo* rispetto alle cellule differenziate [15]. La prima alterazione neoplastica visibile sono le *foci delle cripte aberranti* (ACF) in cui le alterazioni genetiche portano all'iperproliferazione, caratterizzate da cellule iperplastiche e/o displastiche. La continua espansione clonale porta alla formazione di neoformazioni nella mucosa intestinale chiamati *polipi* [16].

#### 1.1.3.2.2. Polipi ed adenomi

Il polipo coloretale è una massa ben circoscritta di cellule epiteliali più o meno displastiche con iperproliferazione cellulare nelle cripte, che si proietta sopra la superficie di una mucosa normale. Singole cellule trasformate possono generare subcloni attraverso nuove mutazioni con la possibilità di aumentare il tasso di crescita.

Si distinguono 4 diversi tipi di polipo:

- *infiammatorio* dove è presente flogosi e si manifesta in conseguenza di malattie infiammatorie quali Morbo di Chron, Colite Ulcerosa, Diverticolosi;
- *iperplastico* è una piccola lesione piatta caratterizzata dalla presenza di polipi sessili, raramente peduncolati e generalmente inferiori a 0.5 cm di diametro. Questo è considerato come un fenomeno rigenerativo e non neoplastico;
- *amartomatoso* in genere si presentano nell'infanzia e nell'adolescenza a causa di malattie ereditarie quali la Sindrome di Peutz-Jeghers e la Poliposi Giovanile;
- *adenomatoso* (generalmente chiamato *adenoma*) rappresenta la lesione benigna asintomatica in cui le cellule presentano iperproliferazione e displasia; una larga maggioranza può evolvere in tumore maligno.

L'adenoma istologicamente si manifesta macroscopicamente come piatto, sessile o peduncolato, secondo la base di impianto; microscopicamente come serrato, tubulare, villosa o tubulo-villosa, secondo la struttura architettonica del tessuto. A livello citologico invece si distingue la tipologia di alterazione cellulare con il grado di displasia lieve, moderato o severo. Molto importante è anche la numerosità degli adenomi presenti nel colon: unico, multipli ( $\leq 100$ ) o poliposi ( $> 100$ ).

La progressione da adenoma a carcinoma è segnata dal graduale aumento delle dimensioni tumorali, dallo sviluppo dell'architettura villosa, da una displasia di sempre più alto grado e dalla capacità di invadere la membrana basale (Figura 2).

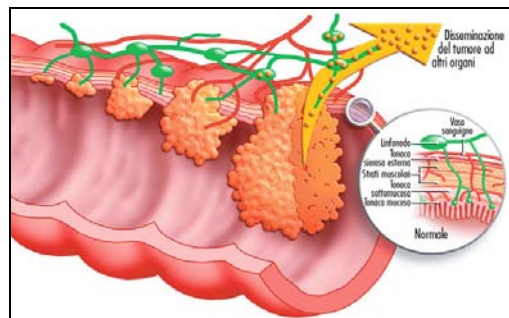


Figura 2. Schema della progressione del CRC.

#### 1.1.3.4. Sequenza adenoma-carcinoma

La tumorigenesi del cancro del colon si sviluppa attraverso alterazioni multi-step e multi-genetiche (Figura 3) che correlano con la comparsa di alterazioni morfologiche, in una *stepwise progression* della durata di circa 10-15 anni [17].

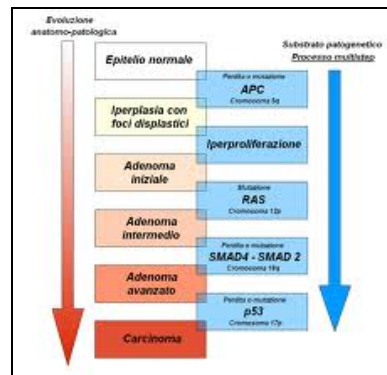


Figura 3. Principali alterazioni molecolari nello sviluppo del CRC.

Vogelstein e collaboratori [18], già nel 1988, ha introdotto il modello genetico multi-step di cancerogenesi coloretale. Tale modello assume che il primo evento nel processo di cancerogenesi sia una mutazione -anche intesa come perdita di eterozigotità (LOH)- del gene oncosoppressore *Adenomatous Polyposis Coli* (APC). Tale mutazione è presente nell'80% negli adenomi e adenocarcinomi e nel 43% nelle ACF. La proteina APC è il principale fattore intracellulare della WNT pathway e regola diversi processi cellulari attraverso il legame e la successiva degradazione del fattore di trascrizione  $\beta$ -catenina. La proteina APC mutata non è in grado di legare la proteina  $\beta$ -catenina, che trasloca nel nucleo dove si lega al fattore di trascrizione TCF, che a sua volta attiva la trascrizione di geni come *c-myc*, *cyclin D1* e *c-jun*, che promuovono la proliferazione cellulare.

L'evento cancerogenetico che si assume essere successivo nel CRC, è una mutazione dell'oncogene *K-ras*. La proteina tradotta KRAS è un membro della famiglia delle proteine RAS leganti GTP/GDP (guanosina 5'-trifosfato/guanosina 5'-difosfato) e possiede un'attività GTPasica intrinseca. Il segnale originato da *Epidermal Growth Factor Receptor* (EGFR) è trasmesso mediante il complesso KRAS-GTP che lega la proteina citoplasmatica RAF e permette l'attivazione a cascata delle *Mitogen-Activated Protein Kinase* (MAPK), trasducendo il segnale

all'interno del nucleo e attivando fattori di trascrizione che promuovono la proliferazione e differenziazione cellulare [19]. Mutazioni attivanti di *K-ras*, che portano alla proliferazione, trasformazione e differenziazione cellulare, sono state identificate nel 58% dei casi di adenoma maggiori di 1 cm e nel 47% dei casi di carcinoma [18]. Dato che la mutazione è stata ritrovata sia nelle lesioni di adenoma che di adenocarcinoma degli stessi pazienti, si deduce che essa sia un evento precedente la trasformazione maligna.

Più del 90% dei carcinomi primitivi presentano una perdita di eterozigotità (LOH) del cromosoma 18q con delezione del gene oncosoppressore *Deleted in Colorectal Carcinoma* (DCC); infatti alcuni studi (Itoh, 1993; Goi, 1998) correlano la LOH di cr. 18q con la riduzione dei livelli di espressione dell'mRNA e della proteina codificata da DCC, un recettore transmembrana implicato nell'apoptosi cellulare [20].

Il punto che determina la linea di confine tra adenoma e adenocarcinoma è la mutazione del gene oncosoppressore *p53*; la sua mutazione, o la LOH, è presente nel 75% dei casi di carcinoma, ma raramente è riscontrata negli adenomi. La proteina p53 agisce come un sensore allo stress cellulare e l'aumento del suo livello causa l'arresto del ciclo cellulare in fase G1, la senescenza cellulare o l'apoptosi per induzione di vari geni target. L'arresto in G1 è parte di una risposta al *checkpoint* cellulare alla presenza di danni al DNA, che ne permette il riparo, prima della progressione del ciclo cellulare. Tale meccanismo limita la propagazione di potenziali mutazioni genetiche. In alternativa, la *pathway* apoptotica p53-dipendente può essere attivata tramite Caspasi, in quei tipi cellulari che proliferano in modo inappropriato [21,22].

La perdita di tale controllo facilita l'accumulo ulteriore di alterazioni genetiche nella tumorigenesi quali amplificazioni, delezioni, mutazioni puntiformi di specifici geni e riarrangiamenti genomici.

## 1.1.4. Screening, Diagnosi e Terapia

### 1.1.4.1. Screening del CRC

Lo screening del CRC, condotto su una popolazione asintomatica e a rischio standard, ha due obiettivi principali:

1. l'identificazione precoce delle forme neoplastiche invasive;
2. la rimozione terapeutica dei polipi adenomatosi per interrompere lo sviluppo naturale della malattia cancerosa.

A tale scopo sono proposte tre metodiche di screening:

- l'*identificazione di soggetti ad alto rischio* mediante un'accurata anamnesi della storia clinica e familiare.
- la *ricerca di sangue occulto nelle feci (FOBT)* il cui limite principale del test è la falsa positività, dovuta a cause non neoplastiche quali emorroidi, gastriti, coliti ecc., infatti la sensibilità del test si aggira attorno al 33-50%.
- la *rettosigmoidoscopia flessibile* e la *colonscopia* che rappresentano un'indagine invasiva, ma con migliore affidabilità diagnostica (sensibilità >95%), e che permette di effettuare delle biopsie in corso d'opera.

Le linee guida internazionali suggeriscono di effettuare, a partire dai 50 anni, la ricerca di sangue occulto nelle feci ogni anno e una colonscopia periodica ogni 5 anni [23,24].

### 1.1.4.2. Diagnosi di CRC

Il CRC è diagnosticato sulla base dei risultati di colonscopia e prelievo bioptico della massa sospetta. Successivamente, è necessario valutare nel modo più accurato possibile l'estensione della malattia rilevando anche eventuali tumori metacroni/sincroni o metastasi a distanza, per indirizzare nella scelta migliore del trattamento. La stadiazione tumorale (Tabella 3) è descritta secondo il sistema di classificazione TNM (Tumore, linfonodi, Metastasi), che include quella clinica pre-trattamento (cTNM) e quella istopatologica post-operatoria (pTNM). Tale distinzione è fondamentale, perché le due valutazioni si effettuano in momenti diversi e si basano su tipi di esami ed obiettivi diversi [25].



**Tabella 3. Schema della classificazione TNM del cancro coloretale.**

STADIO	T	N*	M	Descrizione del tumore	
0	T0 o Tis	N0	M0	assenza di tumore primitivo o presente in situ	
I	T1	N0	M0	tumore primitivo limitato alla parete colica	
	T2				
II	T3	N0	M0	tumore primitivo esteso oltre la parete colica	
	T4				
III	A	N1	M0	coinvolgimento dei linfonodi pericolici* ( $\geq$ a 3)	
					T2
	B				T3
					T4
C	ogni T	N2 o N3	M0	coinvolgimento dei linfonodi pericolici* ( $\leq$ a 4)	
IV	ogni T	ogni N	M1	metastasi a distanza	

\* valutazione effettuata su almeno 12 linfonodi.

La sopravvivenza globale a 5 anni è del 60-65% ed è significativamente condizionata dallo stadio.

#### 1.1.4.3. Trattamento del CRC

La chirurgia è la base della terapia del CRC, che richiede l'asportazione ampia del segmento intestinale sede del tumore e la rimozione delle rispettive aree di drenaggio linfatico. Il tipo di intervento è condizionato dalla sede, dimensioni, estensione del tumore e dalle condizioni del paziente.

L'utilizzo di trattamenti integrati chemioradioterapici pre-operatori (*terapia neoadiuvante*) permette la riduzione del volume e l'estensione della massa tumorale, consentendo l'esecuzione di interventi chirurgici più conservativi.

La chemioterapia può essere utilizzata inoltre dopo l'intervento chirurgico (*terapia adiuvante*) a scopo precauzionale (per evitare la ricomparsa di recidive) o per il trattamento della malattia in fase metastatica, per migliorare la sopravvivenza e la qualità di vita, oltre che per la reseccabilità della metastasi.

Il farmaco utilizzato di base è il *5-fluoruracile* (5-FU) o il suo profarmaco capecitabina, a volte usato in associazione con oxaliplatino (schema FOLFOX) o irinotecano (schema FOLFIRI) e generalmente somministrati per via endovenosa.

La ricerca farmacologica ha sviluppato nuovi farmaci antitumorali biotecnologici progettati secondo le caratteristiche biologiche delle cellule tumorali e somministrati in associazione alla chemioterapia tradizionale di cui sembrano potenziare l'efficacia, senza aggravare gli effetti collaterali. Il *Bevacizumab* è un

anticorpo monoclonale che lega il recettore del Vascular Endothelial Growth Factor (VEGF) impedendo la neo-angiogenesi nella massa tumorale, con conseguente blocco della crescita. Il *Centuximab* è un anticorpo monoclonale chimerico diretto contro il recettore di EGF, mentre il *Panitumumab* è il suo equivalente umano [25,26].

#### **1.1.4.4 Follow-up**

Un aspetto molto importante della terapia del CRC è la sorveglianza clinica che si deve mantenere nel tempo dopo il trattamento primario, per garantire l'individuazione di recidiva locale o di metastasi in modo poter intervenire tempestivamente. E' fortemente consigliata per gli stadi I e II, in quanto permette una sopravvivenza globale a 5 anni, dopo intervento chirurgico curativo, del 55-75%.

#### **1.1.5. Marcatori Tumorali**

Idealmente un marcatore tumorale dovrebbe dare un segnale esclusivo di presenza della neoplasia ed essere facilmente e costantemente individuabile. Tuttavia, dato che la cellula tumorale è in grado di mimetizzarsi molto bene e presenta caratteristiche molto simili alla cellula normale, nessuno dei marcatori tumorali che oggi si conoscono sono prerogativa peculiare di uno specifico tumore.

Sia la Società Americana di Oncologia Clinica (ASCO) che il Gruppo Europeo per i Marker Tumorali (EGTM) forniscono delle linee guida per l'uso di marcatori tumorali come test per la prevenzione, lo screening, il trattamento e la sorveglianza del CRC, ma concludono che non ci sono effettive evidenze di rilievo per individuare marcatori tumorali di certa accuratezza e precisione [27,28].

Attualmente il marcatore di elezione del CRC è la valutazione della concentrazione sierica del *carcinoembryonic antigen* (CEA), ma non è raccomandato per lo screening; esso può essere indicativo per la prognosi in associazione con gli esiti chirurgici e la stadiazione del tumore.

Per quanto riguarda i livelli sierici di CA19.9 (*Carbohydrate Antigen 19-9*) e l'espressione tumorale di p53, TS (*Timidilato sintasi*), DCC non esistono dati sufficienti per dar loro un ruolo nell'iter gestionale del paziente con CRC.

#### **1.1.5.1. Fattori prognostici e di predizione di risposta al trattamento**

I fattori prognostici sono associati alla sopravvivenza del paziente e sono indipendenti dagli effetti del trattamento. I principali fattori associati a una cattiva prognosi sono [29]:

- elevati livelli pre-operatori di Carcinoembryonic Antigen (CEA)
- metastasi epatiche alla diagnosi clinica
- invasione intraparietale del tumore
- positività tumorale dei margini chirurgici
- grado di differenziazione istologica (Grading)
- invasione vascolare
- coinvolgimento di linfonodi regionali
- micrometastasi linfonodali
- aneuploidia
- perdita di eterozigosi del cr.18q
- residui tumorali dopo trattamento.

I marcatori tumorali di predizione indicano invece i benefici del trattamento terapeutico ed alcuni possono avere anche un senso prognostico. Per la chemioterapia con 5-FU si indagano enzimi del metabolismo del farmaco quali *Timidilato Sintasi* (TS), *Diidropirimidina Deidrogenasi* (DPD) e *Timidina Fosforilasi* (TP), presenti all'interno della massa tumore. Per la chemioterapia con Oxaliplatino e Irinotecano si valuta invece la concentrazione intratumorale di ERCC1 (*Excision Repair Cross-Complementing*) e dell'enzima *Topoisomerasi 1* rispettivamente [26]. Infine, per la chemioterapia effettuata con farmaci anti-EGFR è stata validata la ricerca dello status mutazionale di *K-ras* per il CRC metastatico, in quanto circa il 40% dei pazienti presenta mutazioni somatiche attivanti. Ma non tutti i pazienti *KRAS wild-type* rispondono alla terapia, infatti, anche mutazioni di BRAF (fattore a valle di KRAS), PI3KCA (*Phosphatidylinositol-4, 5-bisphosphate 3-kinase catalytic subunit alpha*, fattore della *pathway* di Akt,

attivata in alternativa alla pathway di MAPK) e perdita di eterozigoti per PTEN (*Phosphatase and tensin homolog*, inibitore di PI3KCA) sono fattori che contribuiscono alla resistenza alla terapia monoclonale [25].

## **1.2. Proteomica**

Il proteoma di una cellula o di un organismo è un sistema dinamico nel tempo e al momento dello studio esso riflette l'immediata situazione ambientale. Le proteine non solo sono rapidamente sintetizzate o degradate in risposta a stimoli esterni o interni, ma possono subire modificazioni post-traduzionali od essere traslocate in altri compartimenti cellulari, dando l'immagine di un ambiente proteico in quello specifico momento. Per comprendere la complessità del proteoma, si deve considerare inoltre che un unico genoma può potenzialmente dare origine a infiniti proteomi, grazie a fenomeni di splicing alternativo e alle modificazioni post-traduzionali [30]. Infatti, la scoperta che il genoma umano è costituito da molti meno geni rispetto al previsto (circa 35.000), ha reso evidente che le sole informazioni genomiche sono insufficienti a spiegare tutti gli eventi biologici ed ad individuare nuovi bersagli farmacologici, inoltre non sempre i profili di espressione di mRNA correlano con i livelli di espressione proteica [31]. Infatti, la proteomica mostra almeno due livelli di complessità che mancano al genomica: mentre il genoma è definito da una sequenza di nucleotidi, il proteoma non si limita alla somma delle sequenze amminoacidiche (struttura primaria), ma prevede la conoscenza delle conformazioni delle proteine (struttura secondaria, terziaria e quaternaria) e le interazioni funzionali tra le proteine stesse.

### **1.2.1. Proteomica clinica**

La proteomica riveste un ruolo importante nella pratica clinica, soprattutto attraverso due vie diverse.

#### **1.2.1.1. Ricerca di biomarcatori tumorali**

Le recenti tecnologie genomiche e proteomiche quali l'analisi di array (DNA, RNA o proteici), il completamento dello Human Genome Project, la spettrometria di

massa e la bioinformatica hanno fortemente aumentato il numero di potenziali biomarcatori da investigare [32].

La ricerca di biomarcatori proteici si focalizza sull'identificazione di:

- nuovi *target terapeutici* analizzando direttamente le cellule tumorali;
- marcatori per la *diagnosi precoce*, la stadiazione della malattia e di valore prognostico, che possono essere prodotti direttamente dal cancro o che riflettano una reazione dell'ospite al tumore. Ciò avviene esaminando principalmente i fluidi corporei;
- marcatori per il *monitoraggio della risposta alla terapia*, soprattutto per l'utilizzo di farmaci non citotossici che non possono essere monitorati con solo parametri clinici, vedi anticorpi monoclonali, inibitori dell'angiogenesi o del ciclo cellulare.

Tale ricerca affronta diverse grandi sfide ed una di esse è la variabilità biologica dei campioni che riflette l'immagine complessa delle variabili derivanti dalle differenze di età, sesso, diete, attività fisica, ritmo circadiano ecc. Un altro problema è l'enorme intervallo dinamico della concentrazione delle proteine nei tessuti e nei fluidi biologici, che incide sulla preparazione del campione prima dell'analisi.

Infine, allo scopo di poter essere usato quotidianamente nella routine clinica, il test di determinazione/quantificazione dei marcatori proteici deve essere non invasivo, semplice, robusto, affidabile, sensibile, rapido e con possibilità di automazione [33,34].

Per la ricerca di un potenziale biomarcatore si ricorre usualmente a un gruppo limitato di campioni spesso ben selezionato, perciò si rende necessaria, successivamente, una validazione estensiva su alcune centinaia di campioni di pazienti [35]. Emerge dalle esigenze cliniche la necessità di un pannello di marcatori sufficientemente sensibile e specifico per la diagnosi e la prognosi del cancro, in quanto l'utilizzo di un singolo marcatore si è dimostrato essere inadatto alla pratica medica a causa della complessità molecolare ed eterogeneità cellulare della cancerogenesi [36].

### 1.2.1.2. Identificazione del profilo proteico

Le tecniche che si basano sulla ricerca di un singolo marcatore si scontrano tutt'ora con la difficoltà di identificare inequivocabilmente il cancro, poichè i potenziali biomarcatori tumorali sono spesso espressi anche nelle cellule normali. Per effettuare la predizione, lo screening o la prognosi di malattia è fondamentale quindi, come primo *step*, effettuare l'analisi di profili proteici per poter ottenere in un'unica analisi dati quali/quantitativi di centinaia di proteine [37].

Il profilo proteico si basa sulla rilevazione di *pattern* identificativo di proteine in un alto numero di campioni. Tale *pattern* o profilo proteico discriminante è un sottogruppo di proteine che varia per espressione di intensità nelle diverse popolazioni studiate, e che permette di classificare e distinguere differenti popolazioni (per esempio sano o malato, responsivo o non responsivo ad un trattamento, etc.) [38]. L'utilizzo della proteomica nella pratica clinica avrà in futuro un notevole impatto su tutti gli elementi della cura e gestione del paziente: la diagnosi precoce della malattia con profili proteici su campioni di fluidi biologici, la diagnosi basata sulla "firma" proteica come complemento della diagnosi istopatologica, la selezione personale di combinazioni di farmaci scelte sulla base di specifiche vie proteiche attivate nella malattia, l'accertamento in tempo reale dell'efficacia e della tossicità della terapia e l'eventuale aggiustamento della stessa sulla base delle vie proteiche di resistenza ai farmaci [39]. Inoltre, dal momento che la crescita e la de-differenziazione tumorale sono fattori che garantiscono la presenza del cancro, è ragionevole esplorare la differente espressione del repertorio proteico del tumore rispetto alla controparte normale. Ciò vale anche per altre differenze biologiche come la sensibilità e la resistenza alla terapia che possono manifestarsi non solo qualitativamente, ma anche quantitativamente. I cambiamenti dei profili proteici sono strettamente associati all'espansione clonale delle cellule cancerose e potrebbero permettere di predire meglio la classificazione tumorale e di fornire migliori indicazioni per un trattamento appropriato [40]. Questo comunque potrà andare ad implementare i dati forniti da singoli e specifici biomarcatori che

permettono la diagnosi tumorale o che sono coinvolti nella risposta alla chemioterapia e, di conseguenza, lo sviluppo di nuovi agenti terapeutici.

### 1.2.1.3. La frazione peptidica plasmatica e peptidi a basso peso molecolare

All'interno del proteoma, un interessante contenuto informativo sembra risiedere nelle proteine a basso peso molecolare (LMW) e nei peptidi che rappresentano la frazione meno abbondante nei fluidi biologici (Figura 4).

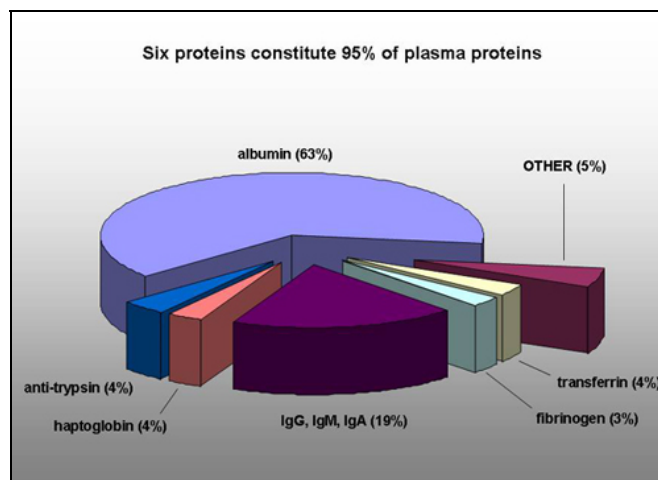
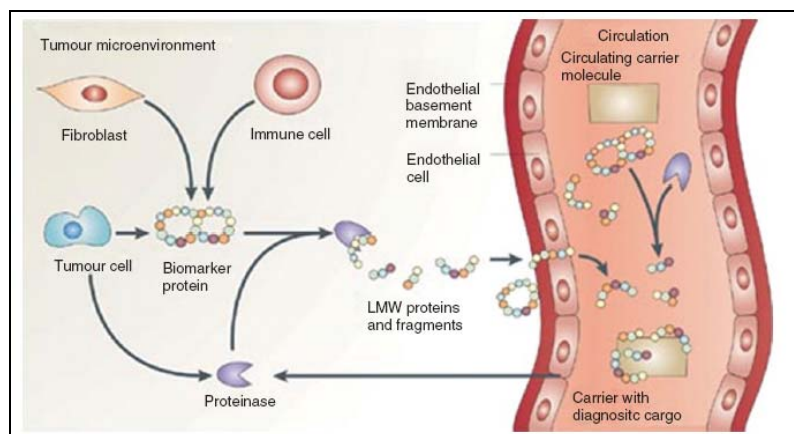


Figura 4. Composizione proteica del plasma.

Tali molecole sono rilasciate nel microambiente tissutale (Figura 5), a causa del *turnover* proteico mediato dalla proteolisi enzimatica endogena che crea una ricca miscela interstiziale di frammenti proteici che riflettono i processi fisiologici/patologici tissutali e una parte di questi possono essere riversati direttamente nel compartimento vascolare o indirettamente attraverso il sistema linfatico [41].





**Figura 5. Peptidi circolanti e frammenti proteici sono rilasciati da tutti i tipi cellulari nel microambiente che poi diffondono nel sistema circolatorio.**

Mentre le proteine dei tessuti sono generalmente troppo grandi per diffondere passivamente nel sistema circolatorio e solo le proteine secrete sono rilevabili, i peptidi, anche di diversa localizzazione cellulare, possono essere rilasciati dalle cellule. Essi sono riconducibili alle proteine progenitrici e in questo senso possono avere una capacità diagnostica potente, perché la loro origine può essere ricondotta a un evento specifico del tumore.

Anche se i peptidi possono derivare non solo da una popolazione di cellule tumorali (che è un piccolo volume rispetto alla totalità dei tessuti), possono originare anche dal microambiente circostante. È stato dimostrato che attraverso l'aberrante processo di crescita, invasione, alterazioni della funzione del sistema immunitario e fenomeni di neo-angiogenesi, si presentano alterate vie di degradazione e quindi di livelli dei peptidi [42].

L'identificazione di potenziali marcatori peptidici dipende da vari fattori tra cui: la concentrazione della molecola nel sangue, la sensibilità e il *range* dinamico di rilevazione della tecnica analitica usata.

Per questo motivo, esistono diverse tecniche di preparazione del campione per migliorare la qualità dell'analisi proteomica ed alcune di queste sono:

- precipitazione con solventi organici come acetone [43] o etanolo [44] o acetonitrile [45];
- gel filtrazione [46];
- elettroforesi in fase solida [47] o in fase liquida [48];

- ultrafiltrazione con cartucce a cut-off di peso molecolare [49-52];
- estrazione su fase solida con differenti proprietà cromatografiche: idrofobicità, idrofilicità, scambio ionico, affinità [38,53].

## 1.2.2. Nanoproteomica

Le innovazioni tecnologiche che sono avvenute nel campo della proteomica, affiancate a metodologie tradizionali, sono utilizzate ampiamente nella ricerca di nuovi biomarcatori tumorali. La *nanoproteomica* nasce come applicazione delle nano(bio)tecnologie alla proteomica, per migliorare gli attuali protocolli di purificazione, di visualizzazione e di identificazione di proteine presenti in bassa quantità nei campioni di studio. L'uso di dispositivi nell'ordine dei nanometri, ha migliorato l'analisi di proteine e peptidi che possono essere isolate e rilevate anche in presenza di scarse quantità di materiale, come possono essere le biopsie.

I nuovi approcci nanotecnologici (Tabella 4) hanno permesso di ottenere nanodispositivi di elevato impatto, sia sulla fase pre-analitica (estrazione, purificazione, separazione) che su quella analitica (utilizzo di rivelatori molecolari che permettano di determinare singole molecole o loro singoli complessi) [54].

**Tabella 4. Esempi di nanotecnologia applicata alla proteomica.**

<b>Nanotecnologia</b>	<b>Descrizione</b>
Sistemi per nanoelettroforesi	Per facilitare la micropurificazione e il recupero delle proteine.
Sistemi per nanocromatografia	
Sistemi a biglie magnetiche	
Biochips	Cattura selettiva di proteine da una miscela e di concentrarle mediante un legame su una superficie.
Protein-array	Permettono di unire numerosi biochips in un unico dispositivo, per aumentare l'efficienza del biochip stesso.
Sistemi per la microscopia a scansione e ottica	Per la rilevazione di specifiche proteine, la loro conformazione, il loro avvolgimento e il loro stato funzionale.
Cryomassdetectors	Per la registrazione e l'identificazione di molecole proteiche singole e i loro complessi (>1MDa) senza trispinolisì.

La nanoproteomica ha sviluppato principalmente le superfici nanostrutturate e i materiali nanoporosi per un legame selettivo e un frazionamento di proteine e peptidi da campioni complessi per migliorare l'analisi proteomica in spettrometria di massa (MS) [55,56] e nella *Reverse Phase Protein-array* (RPPA) [57].

Anche nel campo dell'analisi di peptidi e proteine a basso peso molecolare, sono stati sviluppati diversi dispositivi nanotecnologici per l'arricchimento degli analiti. Tra essi, di particolare interesse sono quelli che utilizzano la presenza di nanopori su un materiale di supporto per effettuare l'estrazione in fase solida dei peptidi (Tabella 5).

**Tabella 5. Recente nanotecnologia applicata al recupero di LMW.**

<b>Nanotecnologia</b>	<b>Riferimento bibliografico</b>
Superficie nanoporosa in silice	Gaspari <i>et al</i> , 2006 [58]
Supporto di silice nanoporoso e biglie in vetro con nanopori	Geho <i>et al</i> , 2006 [55]
Particelle nanoporose in silice sintetizzate con 2 diverse procedure	Terracciano <i>et al</i> , 2006 [56]
Microsfere magnetiche funzionalizzate di polimero C-8	Chen <i>et al</i> , 2008 [59]
Biglie in silice mesoporosa	Terracciano <i>et al</i> , 2010[60]
Chips in silice mesoporoso	Bouamrani <i>et al</i> , 2010[61]
Alluminosilicali mesoporosi	Preianò <i>et al</i> , 2012[62]



## 2. SCOPO DEL PROGETTO

La comunità scientifica ha esplorato strategie innovative per rilevare sottili e fini cambiamenti biologici che possono essere associati con l'avvio o essere precursori di sviluppo neoplastico. Nell'era post-genomica, il campo della proteomica è attivamente indagato per il potenziale clinico che assumono proteine e peptidi in quanto possono essere ricca fonte di marcatori biologici. Durante la transizione delle cellule tumorali da strutture monostrato a masse multistrato, è prodotta una serie unica di proteine associate all'angiogenesi, alla matrice extracellulare, all'inibizione dell'apoptosi, alla crescita cellulare e all'invasione tissutale che successivamente è riversata nella circolazione sanguigna.

La sfida tecnica più impegnativa che ostacola la scoperta di nuovi candidati biomarcatori proteici è la capacità di accedere al proteoma circolante nel sangue, che sembra avere più rilevanza clinica. Oltre il 90% del contenuto proteico nel sangue è determinato dalla presenza di poche specie molto abbondanti e con pesi molecolari relativamente elevati, come albumina (67 kDa) e immunoglobuline G (50 kDa) che mascherano la presenza di proteine con abbondanza più bassa.

Con l'obiettivo di discriminare i peptidi a basso peso molecolare da campioni complessi, quali siero e plasma, il gruppo del Prof. Mauro Ferrari (The Methodist Hospital Research Institute, Houston, TX, USA) ha sviluppato un dispositivo di silicio (*Nanoporous Silica Chip*, NSC) che utilizza una superficie nanoporosa (3-10 nm) per selezionare tale frazione.

I progressi tecnologici nella spettrometria di massa (MS) hanno permesso la scoperta e l'identificazione di importanti proteine, anche da campioni complessi come il siero, il plasma e gli altri fluidi corporei e di fornire approcci per nuove strategie affidabili di diagnosi precoce.

Si è scelto di iniziare questo studio preliminare sulla patologia del cancro colorettale, in quanto la mancanza di marcatori precoci di diagnosi, la diffusione dei fattori di rischio, l'anticipazione diagnostica e l'aumento dell'età media della

popolazione, sono alla base della progressiva crescita dell'incidenza di questo tumore negli ultimi decenni.

L'arruolamento dei pazienti e la raccolta dei campioni di sangue è stata possibile grazie al personale e alla presenza della Biobanca Tissutale della Clinica Chirurgica I (Direttore Prof. Donato Nitti).

Obiettivi principali del progetto sono:

- Messa a punto di un protocollo per selezione di peptidi a basso peso molecolare dal plasma mediante *Nanoporous Silica Chip*;
- Valutazione del profilo peptidico plasmatico mediante spettrometria di massa (MALDI-TOF) nello sviluppo e progressione del cancro coloretale;
- Identificazione di peptidi e/o vie metaboliche alterate nel cancro coloretale.

### 3. MATERIALE E METODI

#### 3.1. Casistica dei pazienti

Lo studio è stato condotto su una casistica (Tabella 6) di 37 soggetti sani ("*clean colon*" con endoscopia e storia familiare negativa per cancro), 27 pazienti con adenoma del colon (lesioni pre-cancerose) e 67 pazienti con adenocarcinoma del colon, di cui 33 di stadio precoce (I e II) e 34 di stadio tardivo (III e IV). Per tutti i casi arruolati sono stati raccolti e registrati tutti i dati anagrafici, anamnestici, clinici, anatomopatologici e consenso informato.

Tabella 6. Caratteristiche dei casi arruolati nello studio.

Gruppo: Descrizione:	Controlli	Adenomi	CRC Early	CRC Late	
	<i>Clean Colon</i>	<i>Lesione pre-cancerosa</i>	<i>Stadio I-II</i>	<i>Stadio III-IV</i>	
Età (anni)	Media	61	66	71	66
	Min-Max	43-77	41-80	47-85	43-85
Genere (n°)	Maschio	19	15	21	21
	Femmina	15	12	12	13
Campioni totali		34	27	33	34

Il sangue periferico è stato raccolto in provette da 6 ml con EDTA come anticoagulante (DB Vacutaineer™, BD, USA). La frazione plasmatica è stata ottenuta mediante centrifugazione a 3000 rpm per 10 minuti, aliquotata in provette criogeniche (T309 Cryovial, Simport) e conservata in azoto liquido, secondo le procedure della Biobanca Tissutale della Clinica Chirurgica I (Direttore Prof D. Nitti).

#### 3.2. Nanoporous Silica Chip (NSC): fabbricazione e caratteristiche

Il chip di silicio nanoporoso (NSC) è un prototipo prodotto e brevettato dal Laboratorio del Prof. M. Ferrari al Dipartimento di Nanomedicina del The Methodist Hospital Research Institute, Houston, Texas, USA.

Il chip è costituito da un supporto circolare di silicio su cui è depositato un film copolimerico anfipatico di silice che presenta dei pori di dimensioni controllate tra 2-50 nm.

La produzione del chip prevede una procedura di auto-assemblaggio evaporazione-indotta in condizioni acide:

- preparazione della soluzione di *coating* costituita da 14 ml di tetraetilortosilicato (TEOS), 17 ml di etanolo, 2 ml di acido cloridrico 2 M e 6.5 ml di H<sub>2</sub>O deionizzata, in forte agitazione e riscaldata per 2 ore a 75°C;
- aggiunta della *soluzione di copolimero* costituita da Pluronic TriBlock (un copolimero anfifilico costituito da derivati di ossido di polietilene-ossido di polipropilene-ossido di polietilene (PEO-PPO-PEO) che portano ad una struttura con le estremità idrofile e la porzione centrale idrofoba) in 10 ml di etanolo a temperatura ambiente, che permette la formazione di micelle che fungono da stampo per la struttura dei pori;
- rivestimento di un disco di silice, Si(100)wafer, di 10 cm con 1 ml della soluzione di *coating* mediante un movimento rotatorio (2000 rpm per 20 sec );
- essiccamento del rivestimento in forno a 80° C per 12 ore, seguito da un aumento della temperatura con un tasso di 1° C/min fino a 425° C, per il mantenimento della stessa per 5 ore. Il raffreddamento è lento ed avviene a temperatura ambiente.

Lo spessore del film può essere controllata regolando la concentrazione del polimero nella soluzione, mentre la porosità è in relazione al rapporto tra copolimero e TEOS nella soluzione di *coating*. Viene anche eseguito un trattamento con plasma ad ossigeno per eliminare eventuali contaminazioni superficiali e garantire un migliore adsorbimento. Inoltre, aggiungendo un agente rigonfiante nella soluzione di copolimero, è possibile variare accuratamente la dimensione dei pori [63].

L'aspetto finale del NSC è visibile in Figura 6.





Figura 6. NSC con Pluronic Triblock L121.

Diverse tecniche sono state utilizzate per la caratterizzazione dell'architettura dei pori e delle proprietà chimico-fisiche della sottile superficie nanoporosa di NSC:

- *Microscopia elettronica a trasmissione* (TEM, FEI Technai; FEI Co.) per acquisire immagini della superficie;
- *Ellipsometria spettroscopica ad angolo variabile* (M-2000DI; J.A. Woollam Co., Inc.) per valutare la porosità e lo spessore della superficie;
- *Spettroscopia fotoelettrica a raggi X* (XPS, PHI Quantera; Physical Electronics, Inc.) per valutare la purezza e concentrazione atomica di Silicio, Ossigeno e Carbonio del film nanoporoso;
- *Diffrazione a raggi X* (XRD, Philips X'PERT MPD) per caratterizzare la superficie.
- Misurazione dell'*angolo di contatto* per valutare l'idrofilia del film nanoporoso

### **3.3. Frazionamento del plasma con NSC**

Il protocollo di selezione e purificazione ottimizzato su NSC prevede che a 15  $\mu$ l di plasma si aggiungono 5  $\mu$ l di una soluzione di Acetonitrile (ACN, Sigma Aldrich) 25% e Acido Trifluoroacetico (TFA, Sigma Aldrich) 5% (v/v) in rapporto 1:4 per acidificare il pH del campione. La miscela è incubata 30 minuti in agitazione a bassa velocità ed a temperatura ambiente. Sulla superficie di NSC costituita da polimero Pluronic L121 con pori di dimensione di 6 nm, si depositano 5  $\mu$ l della soluzione, in triplicato per ogni campione di plasma; poi NSC è incubato per 1 ora a temperatura ambiente in camera umida. Su ogni campionamento, si seguono 5 lavaggi con 10  $\mu$ l di H<sub>2</sub>O deionizzata ciascuno. Al termine i campioni vengono eluiti dal NSC con 7  $\mu$ l di soluzione composta di ACN 100% e TFA 0.1% (v/v) in rapporto 1:1. I campioni sono conservati a -80° C fino alle successive analisi.

### **3.4. Analisi del profilo peptidico mediante MALDI-TOF**

Per l'analisi del profilo peptidico si è utilizzato uno strumento MALDI-TOF (matrix-assisted laser desorption/ionization-time of flight). Questo strumento è stato scelto per la semplicità e la facilità con cui si prepara il campione, per l'ampio range dinamico che permette l'analisi di proteine con un peso molecolare relativamente alto e per la possibilità di eseguire effettuare il sequenziamento peptidico.

Sulla piastra del MALDI si depongono 0,5  $\mu$ l di campione eluito dal NSC, sul quale si aggiunge la matrice di acido  $\alpha$ -Ciano-4-idrossicinnamico (CHCA, Sigma Aldrich) in soluzione satura 10 mg/ml con ACN 100% e TFA 0.1% (v/v) in rapporto 1:1.

In posizioni prestabilite, è depositato il calibrante esterno "*Mass Standard Kit for Calibration of AB Sciex TOF/TOF Instruments*" (AB Sciex) composto da una miscela di peptidi a peso molecolare noto: Bradichinina m/z 905,05, Angiotensina I m/z 129,51, Glu-Fibropeptide B m/z 1571,61, ACTH [1-17] m/z 2094,46, ACTH [18-39] m/z 2466,72 e ACTH [7-38] m/z 3660,19.

La piastra è stata analizzata con lo strumento MALDI-TOF 4700 *Mass Spectrometer* (Applied Biosystems) che si trova presso la Protemic Facility

dell'istituto MD Anderson Cancer Center, Houston, TX, USA. I dati sono acquisiti automaticamente sia in modalità lineare che in *reflectron* a ioni positivi.

Nella Tabella 7 sottostante sono riportati i principali parametri di acquisizione:

**Tabella 7. Parametri di acquisizione al MALDI-TOF.**

<b>MS Setting</b>	<b>Reflectron Mode</b>	<b>Linear Mode</b>
Mass Range (Da)	600-4000	800-1000
Focus Mass (Da)	2000	5000
Total Shot/Spectrum	2500	3000
Laser Intensity	4000	4200

### **3.4.1. Processamento dati ottenuti dall'analisi in MALDI-TOF**

I file dei dati ottenuti dall'analisi in modalità reflectron dei campioni di plasma frazionati sono stati convertiti dal formato originale (Raw .t2d) al formato .txt per permetterne il processamento mediante il software open source *mMass* ([www.mmass.org](http://www.mmass.org)).

Gli spettri di ogni campione, che sono acquisiti in triplicato, sono stati ricalibrati internamente, utilizzando una lista di peptidi di riferimento, e successivamente mediati per ottenere un singolo spettro per ogni campione. Agli spettri così ottenuti è stato eliminato il rumore di fondo dell'analisi (processo di *denoising*) ed è stato applicato un algoritmo di *smoothing* (Savitzky–Golay,  $m/z$  0.15).

E' stata ottenuta la lista delle specie  $m/z$  presenti nei campioni considerando per ciascun peptide solo l'intensità assoluta dello ione monoisotopico ( $M+H^+$ ). I dati così ottenuti sono stati normalizzati sulla corrente ionica totale di ciascuno spettro (*TIC normalization*) prima di essere sottoposti all'analisi statistica.

## **3.5 Identificazione dei peptidi mediante sequenziamento MALDI-TOF/TOF**

I campioni sono utilizzati ulteriormente per identificare i peptidi che corrispondono alle specie ioniche di interesse.

5  $\mu$ l di campione eluito dal NSC, non trattato con digestione enzimatica, è mescolato con la matrice CHCA in soluzione satura di ACN 100% e TFA 0.1% (v/v) in rapporto 1:1.

L'identificazione dei peptidi è stata effettuata mediante sequenziamento in *tandem mass* utilizzando lo strumento *Ultraflex II TOF/TOF* (Bruker Daltonics, Bremen, Germany). La calibrazione dello strumento è stata effettuata nell'intervallo di massa 700-3200Da utilizzando una miscela di peptidi noti. Brevemente, per ogni ione ( $m/z$ ) da identificare è stato acquisito lo spettro del precursore (specie ionica di interesse) e quello dei frammenti prodotti mediante il processo di *laser-induced dissociations* (LID). I dati acquisiti sono uniti a dare un unico spettro finale (spettro di frammentazione) che contiene sia le informazioni relative al precursore, che quelle relative ai frammenti.

Gli spettri di frammentazione sono stati processati mediante il software *flexAnalysis v3.3*, che è in grado di effettuare automaticamente sia il *denoising* che l'identificazione dei segnali monoisotopici utilizzando uno specifico algoritmo di rilevazione (Snap<sup>TM</sup>). L'identificazione dei peptidi è stata infine ottenuta utilizzando i motori di ricerca open source *Mascot v2.3* ([www.matrixscience.com](http://www.matrixscience.com); Matrix Science, London, UK) e *MS-Tag* ([prospector.ucsf.edu](http://prospector.ucsf.edu); UCSF, CA, USA). I parametri utilizzati per la ricerca sono elencati di seguito:

- Database: UniprotKB
- Enzima: nessuno
- Tassonomia: Homo sapiens
- Errore ione precursore:  $\pm 0.4$  Da
- Errore ioni prodotto:  $\pm 0.8$  Da

Ai peptidi identificati dai motori di ricerca è associato un punteggio di probabilità, nel caso di Mascot, oppure un punteggio di valore assoluto, nel caso di MS-Tag. In entrambi i casi è stata effettuata la validazione manuale dei risultati ottenuti per verificare la plausibilità dell'identificazione. Per i peptidi non identificati, è stata effettuata una seconda ricerca prendendo in considerazione l'eventuale presenza delle modificazioni post-trasduzionali più comuni (come ossidazione di metionina, istidina e triptofano ed acetilazione della porzione N terminale).

Come esempio dei risultati ottenuti, si riportano le identificazioni dello ione a  $m/z$  1060.6 usando Mascot (Figura 7) e dello ione a  $m/z$  1498.7 usando MS-Tag (Figura 8).

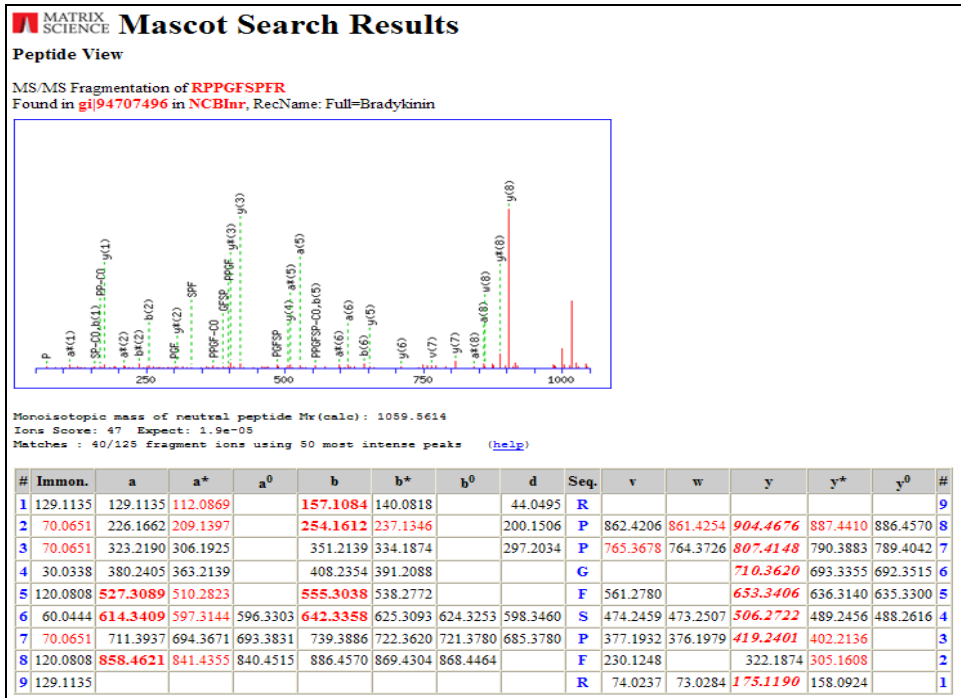


Figura 7. Attribuzione dei frammenti MS/MS (in rosso) alla sequenza RPPGFSPFR (Bradichinina) usando Mascot.

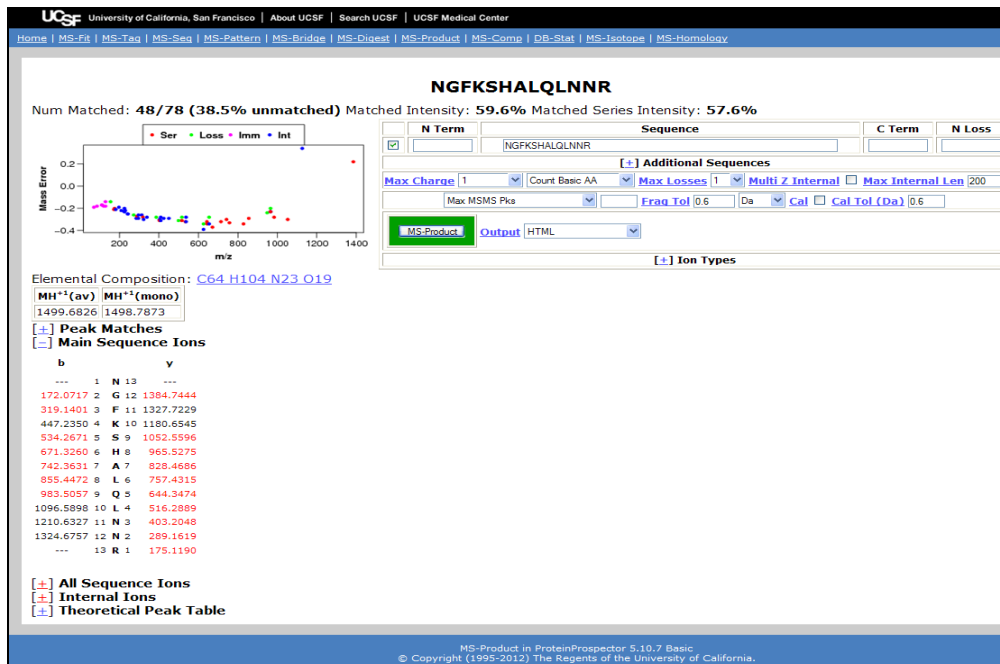


Figura 8. Attribuzione dei frammenti MS/MS (in rosso) alla sequenza NGFKSHALQLNRR (Complemento C4 A/B) usando MS-Tag.

### **3.6. Analisi statistica dei dati e metodi statistici**

I dati processati ottenuti dal profilo peptidico sono stati sottoposti a un'analisi statistica approfondita per ricercare possibili specie ioniche che siano coinvolte nella progressione del cancro e possano permettere una predizione della classificazione dei campioni analizzati in gruppi omogenei.

Una volta selezionate le specie ioniche che maggiormente identificano, caratterizzano e distinguono un gruppo di studio dagli altri presi in considerazione, queste saranno successivamente identificate e ulteriormente studiate.

Sono stati utilizzati sia metodi statistici univariati che multivariati per ridurre la dimensionalità dei dati e permettere la predizione della classificazione. A tale scopo è stato utilizzato il software open source *MetaboAnalyst v2.0* che offre una grande varietà di procedure comunemente usate per l'elaborazione dei dati in particolare di metabolomica [64,65].

La matrice di dati è stata sottoposta a normalizzazione per ridurre eventuali *bias* mediante la mediana dei campioni e trasformazione logaritmica. Sono state utilizzate analisi univariate come calcolo del Fold Change, t-tests, Volcano plot, one-way ANOVA con post hoc test e analisi di Correlazione; analisi multivariate unsupervised come Principal Component Analysis (PCA), analisi di Clustering gerarchico con dendrogramma e heatmap, K-means e Self Organizing Map (SOM) e analisi supervised come Partial Least Squares-Discriminant Analysis (PLS-DA), Random Forest e Support Vector Machine (SVM).

Per la rappresentazione grafica di Box plot e curve Receiver operating characteristic (ROC) si è utilizzato il software *GraphPad Prism* (GraphPad Software, Inc.).

## 4. RISULTATI

### 4.1. Ottimizzazione del processo di frazionamento con NSC

#### 4.1.1. Valutazione di NSC con diversa dimensione dei pori

I dispositivi NSC sono prodotti con un procedimento di autoassemblaggio evaporazione-indotto in condizioni acide, usando il copolimero Pluronic Triblock (Figura 9) come stampo della struttura nanoporosa.

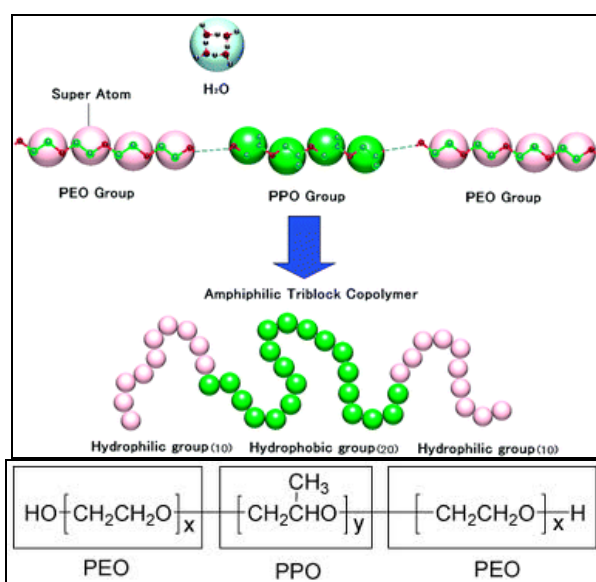
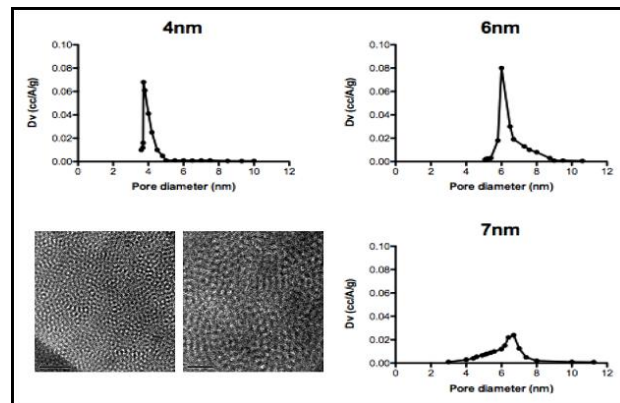


Figura 9. Rappresentazione schematica di copolimero Pluronic Triblock (in alto) e sua formula bruta (in basso).

Diverse nanostrutture, possono essere prodotti variando alcuni parametri quali: il peso molecolare del copolimero, il rapporto della porzione PEO (idrofilica) e PPO (idrofobica) nella composizione del copolimero anfifilico e la quantità di copolimero nel materiale di partenza.

Sono stati sintetizzati 3 diversi NSC utilizzando due tipici copolimeri Triblock con diverso peso molecolare e composizione: il Pluronic F127 (PEO<sub>106</sub>-PPO<sub>70</sub>-PEO<sub>106</sub>) e Pluronic L121 (PEO<sub>5</sub>-PPO<sub>7</sub>-PEO<sub>5</sub>). Due tipi di NSC sono stati prodotti con pori di 4nm e 6nm rispettivamente, mentre un terzo NSC è stato ottenuto usando il Pluronic L121 in presenza di PPG (polipropilenglicole) in rapporto 1:1, che agisce come agente di rigonfiamento, per ottenere dei pori di 7 nm.

La distribuzione del diametro dei pori è stata valutata mediante la curva di adsorbimento/desorbimento di un gas inerte, quale l'azoto ( $N_2$ ), sulla superficie di NSC. Per il calcolo della curva è stata utilizzata la teoria di Brunauer–Emmett–Teller (BET) che descrive l'adsorbimento multistrato di molecole di gas su superfici solide.



**Figura 10. Grafico di distribuzione della dimensione dei pori di 3 diverse tipologie di superfici porose. Immagini TEM di NSC con Pluronic L121 a 6nm.; la barra rappresenta 20 nm.**

La curva ottenuta mostra la distribuzione della dimensione dei pori calcolata dalla isoterma di desorbimento e permette di verificare l'omogeneità dei pori prodotti (Figura 10). La distribuzione dei pori è buona per NSC con Pluronic F127 (4 nm) e L121 (6 nm) poiché la maggior parte dei pori ha la dimensione attesa (picco della curva), mentre risulta essere più scarsa per NSC con Pluronic L121 e PGG (7 nm), in quanto la distribuzione dei pori è dispersa in un range da 3 a 8 nm e non si presenta una significativa presenza della dimensione del poro attesa. Infine è stato osservato come la distribuzione dei pori del NSC con Pluronic F127 (4 nm) sia molto asimmetrica a differenza di quella di NSC con Pluronic L121 (6 nm) che, invece, risulta con distribuzione normale.

#### **4.1.2. Valutazione della rimozione di proteine dalla superficie di NSC**

Per caratterizzare e quantificare il fenomeno di adsorbimento delle proteine sulla superficie del NCS è stato utilizzato il dosaggio colorimetrico Micro BCA Protein Assay Kit (Pierce, Thermo Scientific). 5 ul di plasma sono stati depositati sulla



superficie del NSC e dopo 1 ora di incubazione sono stati effettuati dei lavaggi con 10 ul di H<sub>2</sub>O deionizzata per rimuovere le proteine plasmatiche. L'efficacia del lavaggio è stata misurata quantificando le proteine rilasciate dal NSC nell'eluato di 5 lavaggi successivi. L'assorbanza delle soluzioni di lavaggio è stata interpolata con i dati di assorbimento di una curva di standard di Albumina (range: 0.5-20µg/ml) misurata a λ=562nm con spettrofotometro UV/Vis (Spectra Max M2, Molecular Devices), secondo le istruzioni del kit.

Il test è stato ripetuto 3 volte in sedute analitiche distinte.

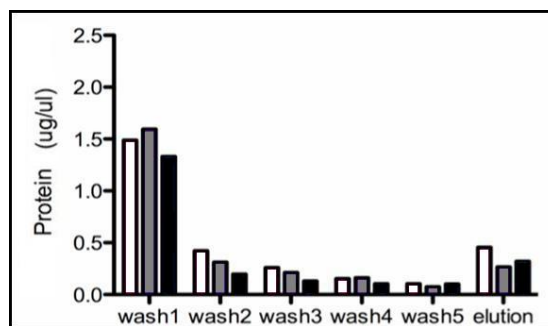


Figura 11. Quantificazione delle proteine presenti nella soluzione acquosa dopo ogni step di lavaggio.

La quantità di proteine rimosse diminuiscono incrementando il numero di lavaggi e dopo 5 lavaggi le proteine sono quasi assenti nella soluzione (Figura 11).

#### 4.1.3. Valutazione del tempo di incubazione del campione

È stata analizzata la cinetica di assorbimento dei peptidi nei pori del NSC in funzione del tempo di contatto con la superficie stessa. Il campione di plasma (5 ul) è stato incubato per i tempi: 5, 15, 30, e 60 minuti sulla superficie del NSC e dopo 5 lavaggi della superficie con H<sub>2</sub>O il campione è stato eluito con soluzione di eluizione ed è stata misurata la quantità di peptidi presenti in soluzione con Micro BCA Protein Assay Kit (Pierce, Thermo Scientific).

Il test è stato ripetuto 3 volte in sedute analitiche distinte.

L'assorbimento risulta veloce, con la metà della quantità di peptidi adsorbiti ( $Q_{max}=84.46\mu\text{g}$ ) in circa 6 minuti e dopo circa 30 minuti si verifica il raggiungimento dell'equilibrio di adsorbimento.

I dati sperimentali ottenuti sono riportati in Figura 12. Essi sono in linea con il modello isotemico di Langmiur, che è il modello usato nella cinetica di assorbimento delle proteine.

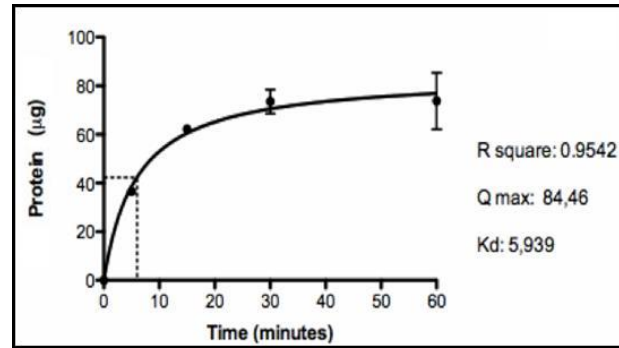


Figura 12. Curva di assorbimento delle proteine plasmatiche.

#### 4.1.4. Valutazione della temperatura di incubazione del campione

Poiché la temperatura influisce sulla cinetica di adsorbimento dei peptidi su NSC sono state investigate tre condizioni diverse di temperatura di incubazione del campione: 4°C, 25°C e 37°C.

Il test è stato ripetuto 3 volte in sedute analitiche distinte.

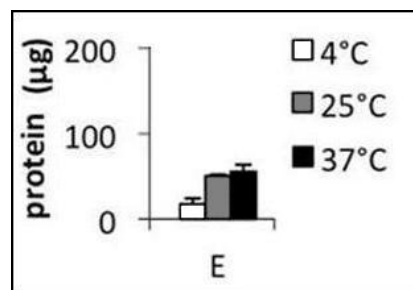


Figura 13. Quantificazione delle proteine nell'eluato dopo incubazione a diversa temperatura.

I peptidi eluiti sono stati quantificati (Figura 13) e la loro la concentrazione è risultata essere più bassa nel campione incubato a 4°C rispetto a quelli incubati a 25°C e 37°C che, invece, hanno dato risultati comparabili.

## 4.2. Valutazione del profilo peptidico plasmatico nel CRC

Dati i risultati ottenuti nella caratterizzazione dei NSC, per le successive applicazioni con campioni biologici si è scelto di operare con il seguente protocollo:

- NSC usati: Pluronic Triblock L121 con pori di 6 nm
- Tempo incubazione: 60 minuti
- Numero di lavaggi: 5
- Temperatura di incubazione: 25°C

Con il protocollo di frazionamento ottimizzato precedentemente riportato, si è proceduti all'isolamento di peptidi a basso peso molecolare (LMW) da campioni di plasma selezionati nel gruppo di studio.

Le caratteristiche di porosità e spessore del film nanoporoso dei cinque diversi lotti di NSC utilizzati per questo studio sono visibili in Tabella 8.

**Tabella 8. Caratteristiche di NSC utilizzati per lo studio.**

NSC	pore size (nm)	porosity (%)	thickness (nm)
		<i>mean ± std dev</i>	<i>mean ± std dev</i>
L121-1	6	49,9 ± 0,7	810 ± 15
L121-2	6	48,3 ± 0,7	806 ± 20
L121-3	6	48,9 ± 0,7	817 ± 8,3
L121-4	6	49,1 ± 0,7	810 ± 16
L121-5	6	48,0 ± 0,7	779 ± 8,6

15 ul di plasma sono mescolati a 5 ul di una soluzione di ACN 25% e TFA 5% (v/v) in rapporto 1:4. Della soluzione ottenuta, 5 ul sono stati depositati sulla superficie del NSC in triplicato e, dopo i lavaggi, l'eluato ottenuto è stato analizzato mediante MALDI-TOF in modalità reflectron nel range di 600-4000 m/z. Gli spettri ottenuti sono stati pre-processati con software *mMass* per calibrarli, allinearli e sottrarre il background; inoltre l'intensità dei picchi è stata normalizzata sulla TIC (Total Ion Current) di ciascuno spettro, per permettere il confronto tra i diversi gruppi.

## 4.2.1. Spettri di massa

Lo spettro di massa è rappresentato da un istogramma definito da un'intensità (asse y) e dal rapporto massa/carica (asse x). Gli spettri di massa in Figura 14 rappresentano lo spettro medio dei diversi gruppi in studio. Lo spettro ottenuto è un ingrandimento nel range di m/z da 900 a 2700, in quanto fino a m/z 900 lo spettro presenta solo segnali riconducibili ad addotti della matrice CHCA (verificati con l'analisi dello spettro di massa della matrice priva di campione), mentre per valori superiori a m/z 2700 non sono rilevati segnali significativi.

Le specie ioniche rilevate dagli spettri di massa, che potenzialmente sono tutte riconducibili ai peptidi plasmatici isolati dal NSC, sono in totale 77.

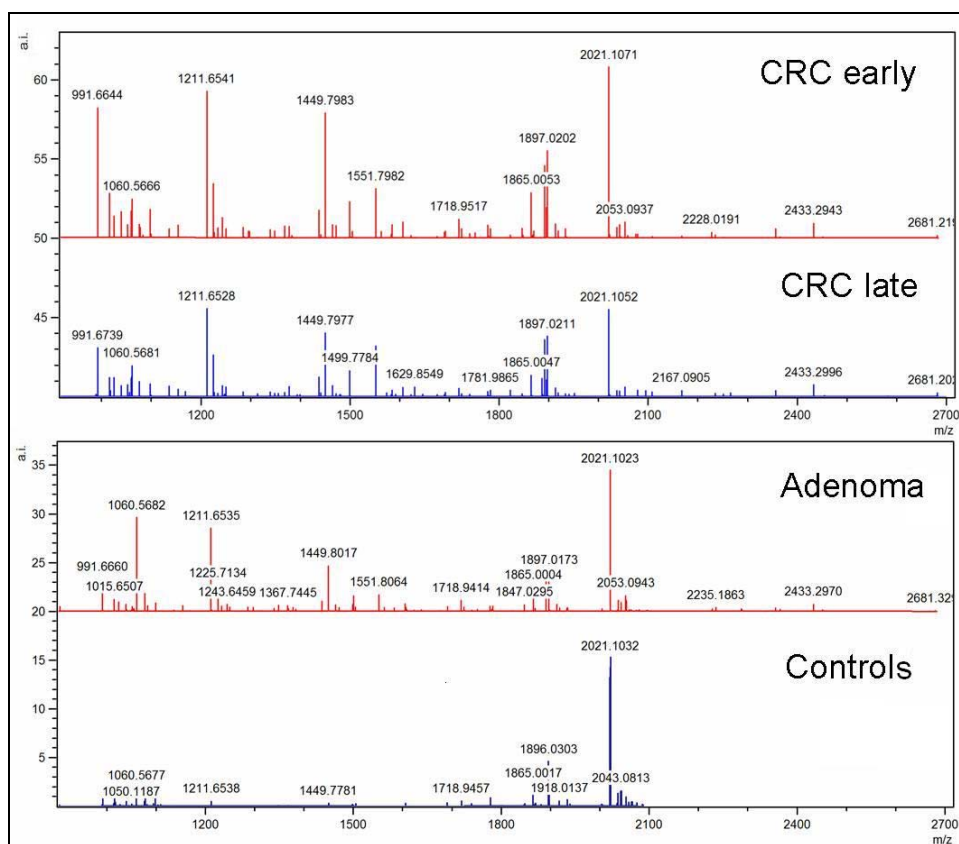


Figura 14. Spettro mediato di ogni singolo gruppo in studio.

Nello spettro medio del gruppo di Controlli, a differenza degli spettri patologici, si nota la quasi totale mancanza di segnali nell'intervallo tra 1000-1600 m/z e nella regione superiore a 2100 m/z.

Sebbene tra gli spettri siano presenti delle differenze d'intensità tra le diverse specie ioniche, per tutti gli spettri la specie ionica più caratteristica e più abbondante è la m/z 2021.1.

### 4.3. Fase esplorativa

Inizialmente si è andati a valutare il profilo dei peptidi negli spettri MALDI confrontando tra loro i gruppi in studio per ottenere una visione globale e generale dei dati ottenuti.

#### 4.3.1. Analisi dei 4 gruppi di soggetti in studio

La valutazione è stata inizialmente condotta sui dati ottenuti considerando tutti i 4 gruppi di studio: Controlli, Adenomi, CRC con stadio precoce (I-II) e CRC con stadio tardivo (III-IV) per un totale di 128 campioni.

Dopo una preliminare analisi univariata mediante one-way ANOVA e post hoc test (test di Tukey) si sono osservate 54 specie ioniche statisticamente significative con  $p\text{-value} < 0.05$ . Tutte queste specie presentano una significatività nel confronto con il gruppo di Controlli, ad eccezione di quelle a m/z 2021.10 e 2037.09 che sono espresse in modo significativamente diverso anche nel confronto tra Adenomi e CRC con stadio precoce.

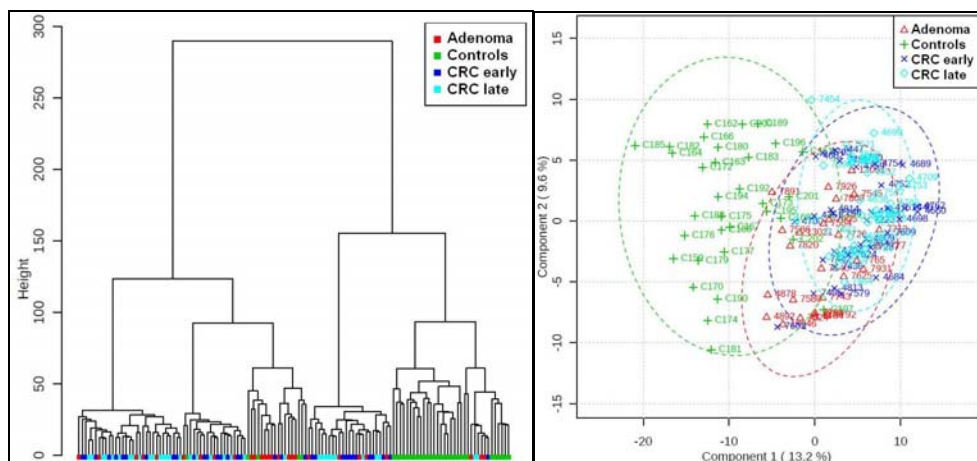


Figura 15. Dendrogramma dell'analisi di Cluster (sinistra) e Score plot di PLS-DA (destra).

L'analisi di Cluster gerarchico mostra la classificazione dei 4 gruppi di studio: il gruppo dei Controlli è ben separato dagli altri 3 gruppi di pazienti e si osservano solo 5 campioni che si distaccano dal gruppo. Gli altri campioni dei gruppi di Adenomi e di CRC con stadio precoce e tardivo invece sono tutti mescolati tra loro (Figura 15 a sinistra).

La PLS-DA conferma il raggruppamento dei Controlli rispetto agli altri gruppi che invece mostrano una sovrapposizione pressoché totale (Figura 15 a destra).

Anche utilizzando un altro algoritmo di classificazione, Random Forest, si confermano i raggruppamenti precedenti ottenendo un errore di classificazione pari allo 0% solo per il gruppo di Controlli, mentre si evidenziano alti errori classificativi nei confronti dei gruppi rimanenti (Figura 16)

	adenoma	control	stage I-II	stage III-IV	class.error
adenoma	2.00	1.00	10.00	14.00	0.93
control	0.00	34.00	0.00	0.00	0.00
stage I-II	5.00	0.00	15.00	13.00	0.55
stage III-IV	3.00	1.00	12.00	18.00	0.47

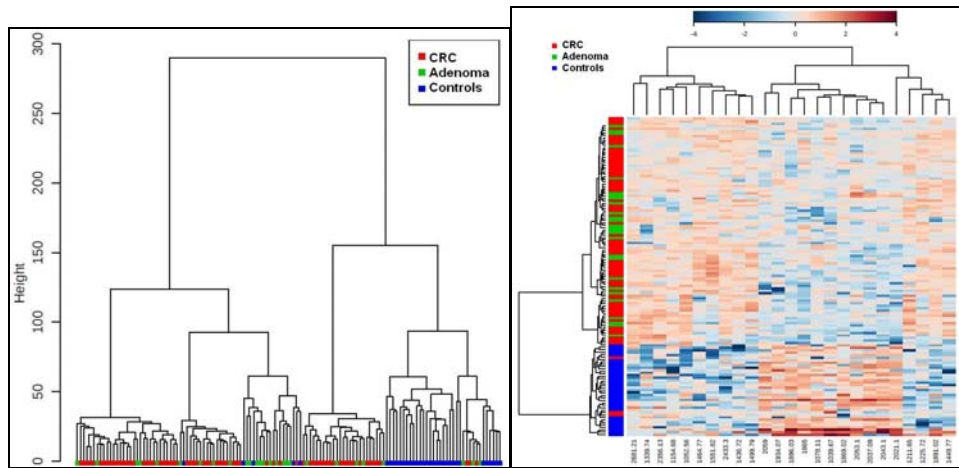
**Figura 16. Performance di classificazione dell'analisi con Random Forest.**

Da questa prima analisi si deduce che con la valutazione dell'intero profilo peptidico ottenuto è possibile classificare correttamente i campioni del gruppo di Controllo, ma non i gruppi di pazienti con Adenoma e con CRC a stadio precoce e tardivo.

#### **4.3.2. Analisi a 3 gruppi di soggetti**

Per cercare di enfatizzare differenze di specie ioniche che si presentano durante la cancerogenesi del cancro si è voluto unire i 2 gruppi dei campioni di pazienti con CRC a stadio precoce e tardivo in un unico gruppo denominato CRC.

L'analisi one-way ANOVA con post hoc test (test di Tukey) ha rilevato 56 specie ioniche statisticamente significative ( $p\text{-value} < 0.05$ ) nel confronto con i Controlli, confermando che le specie ioniche a  $m/z$  2021.1, 2037.09 e 2451.12 sono significative anche nel confronto tra Adenomi e CRC.



**Figura 17. Dendrogramma dell'analisi di Cluster (sinistra) e conseguente analisi di Cluster con solo le prime 25 specie ioniche con p-value significativo ottenuto alla one-way ANOVA (destra).**

L'analisi di Cluster conferma una buona similitudine tra i campioni del gruppo di Controlli con solo 5 campioni che si accoppiano con campioni appartenenti al gruppo di Adenomi o di CRC (Figura 17 a sinistra). Tuttavia, anche in questo caso le specie ioniche selezionate non permettono la separazione del gruppo degli Adenomi dai CRC. Per migliorare la separazione, sono state utilizzate solo le specie ioniche selezionate per valore di p-value più basso ( $p\text{-value} \leq 1.55 \times 10^{-5}$ ) ottenuto all'ANOVA. Con queste prime 25 specie non si ha più l'errore di classificazione dei casi di Adenoma con i Controlli, ma rimangono tuttavia 3 campioni del gruppo di CRC che si classificano con gruppo di Controlli (Figura 17 a destra).

Anche con l'algoritmo di classificazione Random Forest, si confermano i dati precedenti osservando errore di classificazione pari allo 0% per il solo gruppo di Controlli ed evidenziando un bassissimo errore di classificazione nei confronti del gruppo di CRC (6%). Tuttavia permane un alto errore nei confronti del gruppo di Adenomi (96%) (Figura 18) i cui campioni vengono prevalentemente classificati come CRC (Ca).

	Ca	adenoma	control	class.error
Ca	63.00	4.00	0.00	0.06
adenoma	25.00	1.00	1.00	0.96
control	0.00	0.00	34.00	0.00

**Figura 18. Performance di classificazione dell'analisi con Random Forest.**

Quindi si può dedurre dall'analisi esplorativa che è possibile classificare correttamente il gruppo di Controlli, ma distinguere il gruppo degli Adenomi rispetto al gruppo di CRC è difficile.

Allo scopo di trovare il miglior approccio statistico per ottimizzare la classificazione, si è voluto ridurre la dimensionalità delle variabili in considerazione, selezionando solo le specie ioniche statisticamente significative ad almeno 2 test eseguiti. Questo ha permesso di ottenere una lista di 11 variabili a m/z 1078.11, 1436.72, 1869.02, 1891.02, 1896.03, 1918.01, 2021.1, 2037.09, 2043.1, 2059.0 e 2433.3, con le quali si è proceduto a ri-analizzare i dati.

Tra queste 11 specie ioniche solo 9 risultano statisticamente significative alla one-way ANOVA e post hoc test (test LSD di Fisher) con p-value<0.05 nei confronti del gruppo di Controlli e solo le specie ioniche a m/z 2021.1, 2037.09 e 1891.02 sono anche significative nel confronto tra il gruppo di Adenomi e CRC.

Sia osservando l'analisi di Cluster che la classificazione con Random Forest, la distinzione con il gruppo di Controlli è buona, ma rimane scarsa tra Adenomi e CRC.

Si è effettuata anche l'analisi di correlazione con coefficiente di Pearson (Figura 19) tra le 11 m/z in considerazione per valutare l'esistenza di una relazione tra le variabili, e si nota la presenza di 2 raggruppamenti: le specie ioniche a m/z 1436.72, 1891.02, 2433.3 (con forte correlazione positiva tra loro) e le rimanenti 8, che presentano correlazione negativa con le precedenti. Si presenta una forte relazione inversa tra le specie ioniche a m/z 1896.03 con 1436.72 e 2433.3, tra m/z 2043.1 con 1436.72 e 1891.02, e una più debole tra m/z 2021.1 con 1436.72 e 2433.3 e tra 2037.09 e 1891.02.



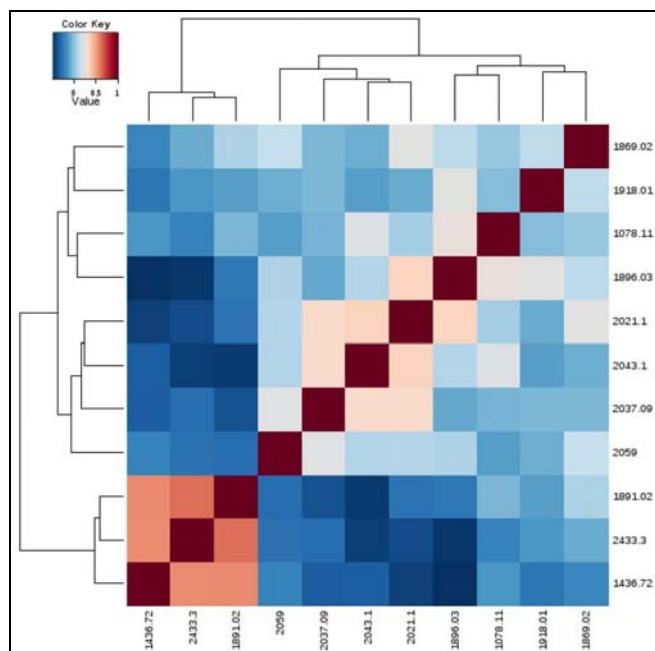


Figura 19. Heatmap dell'analisi di correlazione.

Nonostante i tentativi fatti, si osserva solo l'identificazione del gruppo di Controlli. Per tale motivo si è eseguita un'analisi statistica solo confrontando i gruppi di campioni di Adenoma, CRC con stadio precoce e CRC con stadio tardivo, con l'obiettivo di verificare la presenza di specie ioniche discriminanti la progressione della malattia.

Si sono rilevate statisticamente significative all'one-way ANOVA e post hoc test (test di Tukey) solo 7 specie ioniche con  $p\text{-value} > 0.05$ : 1436.72, 1605.87, 2021.1, 2037.09, 1050.12, 2451.12 e 2681.21 m/z.

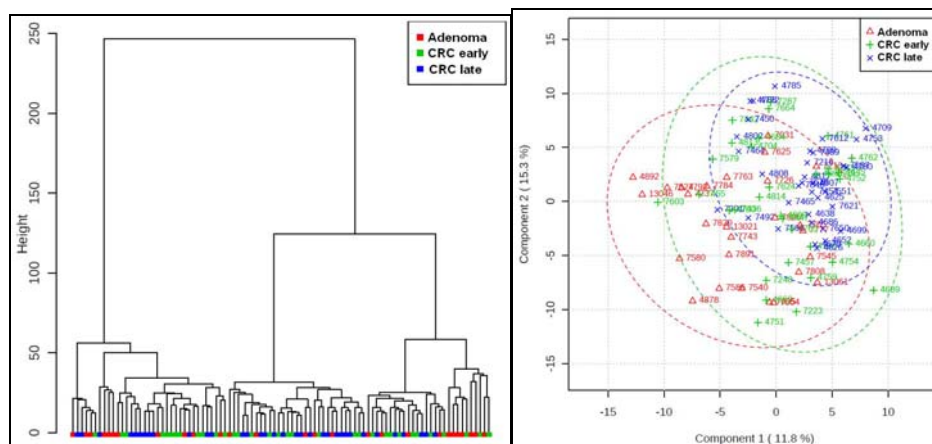


figura 20. Dendrogramma dell'analisi di Cluster (sinistra) e Score plot di PLS-DA (destra).

Tuttavia, l'analisi di Cluster, la PLS-DA (Figura 20) e la Random Forest (errore di classificazione tra 50 e 85%) mostrano anche in questo caso una forte dispersione dei campioni.

#### 4.4. Confronto diretto

Poichè la visione generale dei dati ottenuti ha dimostrato una netta distinzione dei gruppi in studio solamente tra i campioni di Controlli e soggetti malati (sia con soggetti con rischio molto alto di sviluppare il cancro colorettaale che con quelli con malattia già diagnosticata), si è reso necessario approfondire nei dettagli la classificazione, mediante il confronto diretto dei campioni di Controlli con Adenomi e successivamente con CRC, con l'obiettivo di ricercare le specifiche specie ioniche che discriminano i gruppi in studio.

##### 4.4.1. Controlli vs Adenomi

L'analisi univariata effettuata sul confronto di tutte le specie ioniche acquisite tra Controlli e Adenomi ha evidenziato la presenza di 15 specie statisticamente significative al Volcano plot (Figura 21) All'analisi multivariata effettuata con metodi unsupervised i campioni in esame si disperdono in modo casuale, mentre con quelli supervised è possibile discriminare il gruppo di Controlli da quello degli Adenomi (Figura 22).

	Compounds	FC	log2(FC)	p.value	-log10(p)
1	2021.1	2.5356	1.3423	7.5883e-09	8.1199
2	2433.3	0.40769	-1.2945	4.4225e-08	7.3543
3	1865	2.4117	1.2701	4.0092e-07	6.3969
4	2043.1	2.0079	1.0057	1.2836e-06	5.8916
5	2037.09	2.3067	1.2058	2.136e-05	4.6704
6	1211.65	0.47235	-1.0821	3.3723e-05	4.4721
7	1464.77	0.34378	-1.5405	0.00010281	3.988
8	1934.07	6.1594	2.6228	0.00011204	3.9506
9	1897.02S	0.46398	-1.1079	0.00065467	3.184
10	1918.01	2.1687	1.1169	0.01394	1.8557
11	1015.66	7.8285	2.9687	0.014577	1.8363
12	1777.98	4.6842	2.2278	0.01496	1.8251
13	1076.56	2.2221	1.1519	0.019282	1.7149
14	2235.21	2.4512	1.2935	0.03867	1.4126
15	1740.95	3.3157	1.7293	0.045687	1.3402

Figura 21. Specie ioniche identificate al Volcano plot.

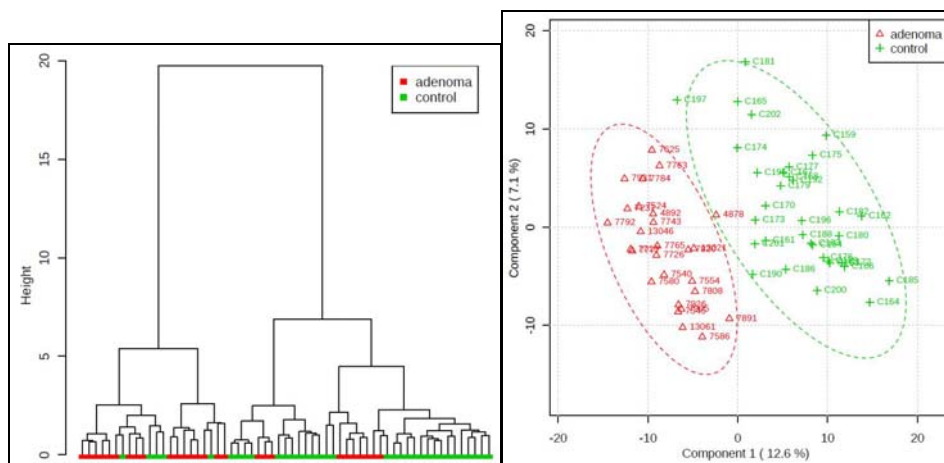


Figura 22. Dendrogramma dell'analisi di Cluster (sinistra) e Score plot di PLS-DA (destra).

Inoltre nella classificazione ottenuta con Random Forest, la performance di errore di classificazione è solo dell'1% per gli Adenomi e 0% per il gruppo di Controlli.

Considerando solo le 5 specie ioniche risultate significative ad almeno 3 dei test statistici sopra riportati e rianalizzando la casistica, otteniamo un Volcano plot con 4 variabili statisticamente significative al t-test ( $p\text{-value} < 0.05$ ) e Fold Change (FC, threshold=2) (Figura 23).

	Compounds	FC	$\log_2(\text{FC})$	p.value	$-\log_{10}(p)$
1	2021.1	2.8022	1.4866	2.9725e-10	9.5269
2	1934.07	4.6251	2.2095	6.8033e-07	6.1673
3	1211.65	0.30412	-1.7173	2.5314e-06	5.5966
4	1464.77	0.3626	-1.4635	0.00068489	3.1644

Figura 23. Specie ioniche identificate al Volcano plot.

All'analisi multivariata le 5 variabili hanno permesso di ottenere una buona classificazione dei campioni di Adenomi con l'analisi di Cluster: solo 2 campioni di Adenomi si raggruppano in con quelli dei Controlli, mentre la PLS-DA mostra una leggera sovrapposizione dei gruppi (Figura 24).

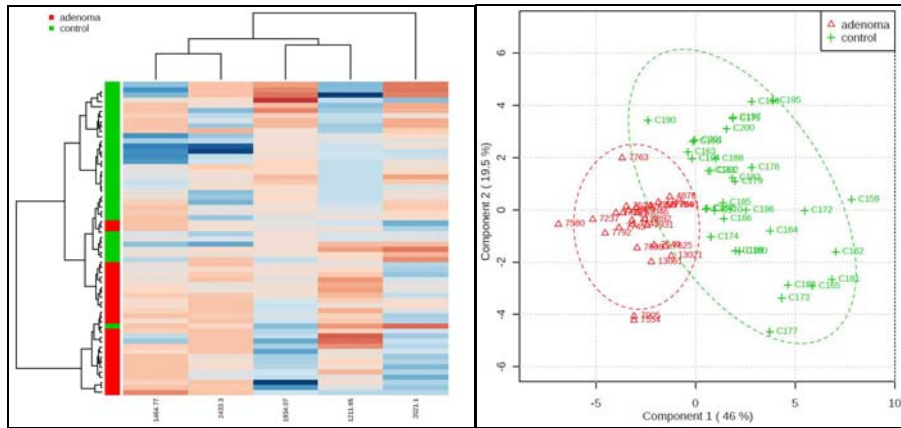


Figura 24. Analisi di Cluster (sinistra) e Score plot di PLS-DA (destra).

Alla classificazione con Random Forest l'errore di classificazione è del 6% per il gruppo di Controllo e dell'11% per quello degli Adenomi.

Valutando la correlazione delle 5 variabili (Figura 25) si nota un andamento diretto tra le specie ioniche a m/z 1934.07 e 2021.1, mentre è presente una forte relazione inversa tra m/z 1211.65 e 1934.07.

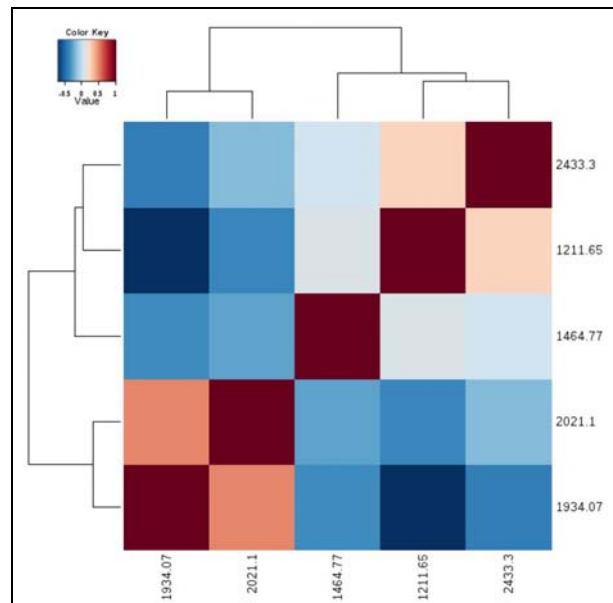


Figura 25. Heatmap dell'analisi di correlazione.

#### 4.4.2. Controlli vs CRC

All'analisi univariata, il Volcano plot (Figura X destra) permette di individuare le specie ioniche che sono statisticamente significative sia al t-test ( $p\text{-value} < 0.05$ ) che al Fold Change (FC,  $\text{threshold} = 2$ ). Inoltre l'analisi multivariata con PLS-DA e Random Forest confermano la buona discriminazione dei 2 diversi gruppi presi in considerazione (Figura 26).

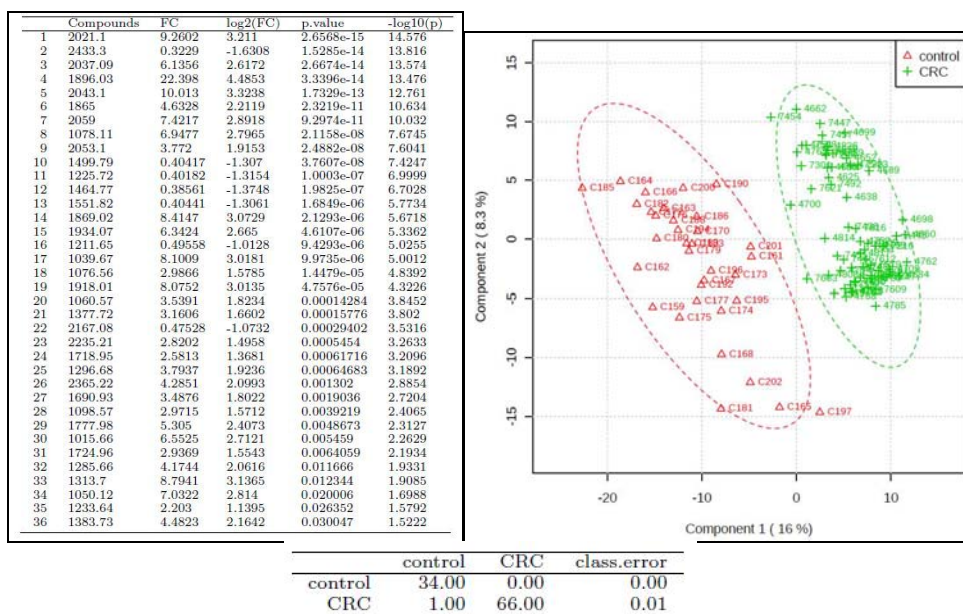


Figura 26. Specie ioniche identificate al Volcano plot (sinistra), score plot di PLS-DA (destra) e performance di classificazione dell'analisi con Random Forest (in basso).

Quindi il profilo peptidico tra il gruppo di Controlli e quello di pazienti affetti da CRC è pressoché diverso.

Considerando invece le analisi effettuate con solo le 21 specie ioniche risultate significative ad almeno 3 test statistici, il Volcano plot seleziona 16 m/z (Figura 27) e la classificazione unsupervised è buona anche con PCA e analisi di Cluster (Figura 28). L'errore di classificazione con Random Forest è dello 0% per il gruppo di Controlli e 1% per CRC.

	Compounds	FC	log2(FC)	p-value	-log10(p)
1	2433.3	0.13986	-2.838	9.8721e-25	24.006
2	1436.72	0.226	-2.1456	2.3116e-15	14.636
3	1225.72	0.25385	-1.9779	3.5848e-13	12.446
4	1551.82	0.12604	-2.9881	3.2944e-12	11.482
5	2356.13	0.30622	-1.7074	5.7918e-12	11.237
6	1464.77	0.19573	-2.3531	1.5749e-11	10.803
7	1052.56	0.21836	-2.1952	3.447e-10	9.4626
8	2021.1	3.7092	1.8911	7.0712e-10	9.1505
9	1896.03	6.8618	2.7786	1.991e-09	8.7009
10	2167.08	0.2149	-2.2182	5.7328e-07	6.2416
11	2043.1	3.3432	1.7412	2.9432e-06	5.5312
12	2037.09	3.0762	1.6212	1.2987e-05	4.8865
13	2059	2.8227	1.4971	0.00012169	3.9147
14	1078.11	2.5126	1.3292	0.00077837	3.1088
15	1934.07	2.5027	1.3235	0.023344	1.6318
16	1869.02	2.5095	1.3274	0.032408	1.4894

Figura 27. Specie ioniche identificate al Volcano plot.

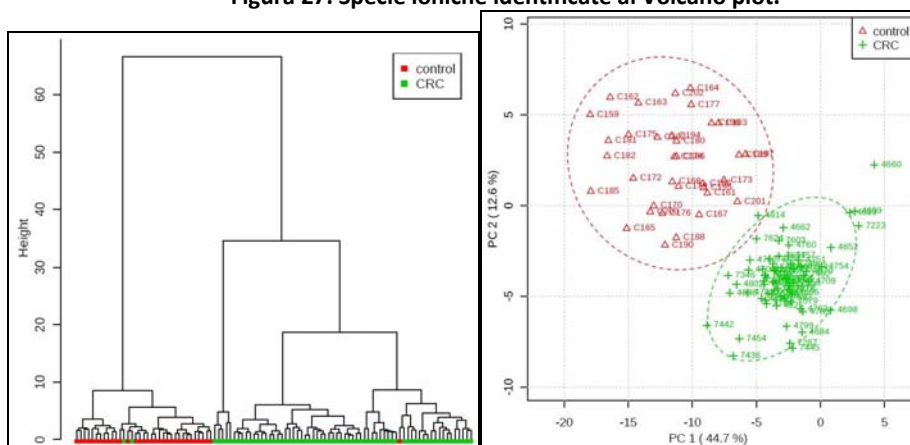


Figura 28. Dendrogramma dell'analisi di Cluster (sinistra) e Score plot di PCA (destra).

L'analisi di correlazione delle 21 variabili prese in considerazione (Figura 29) mostra 2 raggruppamenti principali in cui 8 specie ioniche presentano una relazione positiva tra loro e sono m/z 1052.56, 1225.72, 1436.72, 1464.77, 1551.82, 2167.08, 2356.13 e 2433.3. È da notare la correlazione diretta presente tra le specie ioniche a m/z 1060.57 e 1076.56 e tra 1865, 1934.07 e 2037.09 m/z. Inoltre si presenta una forte correlazione inversa tra le specie ioniche a m/z 2021.1 e 2433.3.

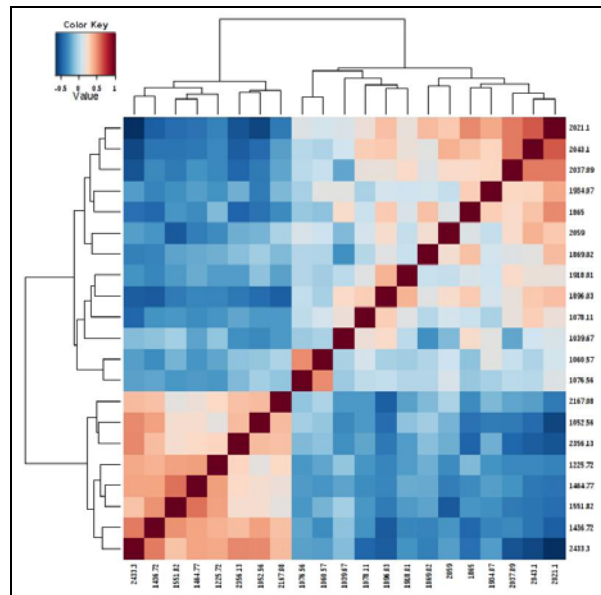


Figura 29. Heatmap dell'analisi di correlazione.

#### 4.5. Identificazione delle specie ioniche

Si è scelto di effettuare il sequenziamento in *tandem mass* e la successiva identificazione peptidica di quelle specie ioniche che risultavano significative ad almeno 10 test statistici (30 su 77 specie ioniche totali acquisite dallo spettro). Di queste 30 specie ioniche solo 18 sono state identificate finora.

Con l'obiettivo di ottenere maggiori informazioni possibili da questo progetto si sta cercando di sequenziare il maggior numero possibile di specie ioniche. Attualmente 45 specie ioniche sono state correttamente identificate (Tabella 9)

**Tabella 9. Lista delle specie ioniche identificate. Quelle con \* sono le 18 specie ioniche significative ad almeno 10 test statistici.**

m/z	PEPTIDE IDENTIFICATION	SEQUENCE
904.46	Des-Arg9 Bradykinin	RPPGFSPF
991.67	2x Lysophosphatidylcholine	2x C16:0
1015.66	2x Lysophosphatidylcholine	C16:1 + C18:2
*1052.56	C4-A/B fragment	SHALQLNNR
*1060.57	Bradykinin	RPPGFSPFR
1076.56	oxidated Bradykinin	RPPGFSPFR
1098.57	C3f fragment	HWESASLLR
*1136.57	C3f fragment	THRIHWESA
1154.68	C4-A/B fragment	LQLNNRQIR
*1211.65	C3f fragment	IHWESASLLR
*1225.72	C4-A/B fragment	ALQLNNRQIR
1249.65	C3f fragment	ITHRIHWESA
1285.66	Alpha-2-Antiplasmin fragment	MEPLGRQLTSGP
1348.63	Kininogen fragment	RHDWGHEKQR
1367.77	C3f fragment	RIHWESASLLR
1377.72	C3f fragment	KITHRIHWESA
*1436.72	C4-A/B fragment	GLEEELQFSLGSK
*1449.77	C4-A/B fragment	SHALQLNNRQIR
*1464.77	C3f fragment	SKITHRIHWESA
*1499.79	deamidated C4-A/B fragment	NGFKSHALQLNNR
*1551.82	C3f fragment	SKITHRIHWESA
1605.87	C3f fragment	THRIHWESASLLR
1606.83	Fibrinogen Alpha chain fragment	SSHHPGIAEFPSRGK
1675.80	ITIH4 fragment	GPPDVPDHAAYHPFR
1718.95	C3f fragment	ITHRIHWESASLLR
1777.98	C3f fragment	SKITHRIHWESASLL
1782.00	C4-A/B fragment	GFKSHALQLNNRQIR
1847.04	C3f fragment	KITHRIHWESASLLR
*1865.00	C3f fragment	SSKITHRIHWESASLL
*1891.02	C4-A/B fragment	GLEEELQFSLGSKINVK
*1896.03	C4-A/B fragment	NGFKSHALQLNNRQIR
1897.02	deamidated C4-A/B fragment	NGFKSHALQLNNRQIR
1913.01	C4-A/B fragment with NH4+	NGFKSHALQLNNRQIR(NH4+)
*1934.07	C3f fragment	SKITHRIHWESASLLR
*2021.10	C3f	SSKITHRIHWESASLLR
*2037.09	oxidated C3f	SSKITHRIHWESASLLR
2043.10	C3f with Na+	SSKITHRIHWE(Na+)SASLLR
2059.00	C3f with K+	SSKITHRIHWE(K+)SASLLR
*2167.08	ITIH4 fragment (Gln->Pyroglu)	QLGLPGPPDVPDHAAYHPFR
2228.03	Fibrinogen Beta chain fragment	QGVNDNEEGFFSARGHRPLD
2235.21	Fibrinogen Beta chain fragment	KREEAPSLRPAPPPISGGGYR
2356.13	Fibrinogen Beta chain fragment	QGVNDNEEGFFSARGHRPLDK
2365.22	Kininogen HMW fragment	KHNLGHGHKHERDQGHGHQR
*2433.30	C4-A/B fragment	GLEEELQFSLGSKINVKVGGNSK
2681.21	Fibrinogen Alpha chain fragment	SSSYSKQFTSSTSYNRGDSTFESK

Molti dei peptidi riconosciuti sono frammenti di proteine fisiologicamente presenti nel circolo sanguigno e appartenenti alle proteine del sistema immunitario, mediatori della flogosi come Chininogeno, Bradichinina, Fibrinogeno e fattori del Complemento C3 e C4.



Il *Chininogeno* è una proteina codificata dal gene *KNG1* (cr. 3.q27.3) in 3 diverse isoforme prodotte mediante processo di *splicing* alternativo. Le sequenze identificate sono relative all'isoforma 1 (chiamata High Molecular Weight, HMW) di 644 amminoacidi (Figura 30). Il *Chininogeno* è secreto dal fegato ed è costituito da una catena pesante di 361 aa e da una leggera di 254 aa. A livello plasmatico subisce l'azione enzimatica di Callicreina (serin-proteasi) che libera Bradichinina, un peptide attivo di 9 aa.

**P01042 (KNG1\_HUMAN) Kininogen High Molecular Weight**

*MKLITLFLCSRLLLSLTQESQSEEDCNDKDLFKAVDAALKKYNSQNQSNQFVLYRIT  
 EATKTVGSDTFYSFKYEIKEGDCPVQSGKTWQDCEYKDAAKAATGECTATVGKRSSTKFS  
 VATQTCQITPAEGPVVTAQYDCLGCVHPISTQSPDLEPILRHGIQYFNNTQHSSLFMLN  
 EVKRAQRQVVAGLNFIRITYSIVQTNCSKENFLFLTPDCKSLWNGDTGECTDNAYIDIQLR  
 IASFSQNCDIYPGKDFVQPPTKICVGCPRDIPTNSPELEETLHTITKLN AENNATFYFK  
 IDNVKKARVQVVAGKKYFIDFVARETTCSKESNEELTESCETKKGQSLDCNAEVYVVPW  
 EKKIYPTVNCQPLGMI~~SLMKRPPGFSPFR~~SSRIGEIKEETTVSPPHSMAPAQDEERDSG  
 KEQGHTRRHDWGHEKQ~~RKHNLGHGHKHERDQGHGHQR~~GHGLGHGHEQQHGLGHGHKFKLD  
 DDLEHQGGHVL~~DHGHKHKHGHGHGKHNKGKKNKGKHNKGWTEHLASSE~~STTPSAQTQE  
 KTEGPTPIPSLAKPGVTVTFSDFQSDLIATMMPPI~~SPAPIQSDDDWIPDIQIDPNGLSF~~  
 NPISDFPDTTSPKCPGRPWKSVSEINPTTQMKEYSYFDLTDGLS*

**Figura 30. Sequenza amminoacidica in formato FASTA del Chininogeno HMW (KNG1). In italico: peptide segnale, in grassetto: catena pesante, nel riquadro: sequenza peptidica di frammenti precursori identificati, nel riquadro in grigio: Bradichinina.**

Il Chininogeno è un inibitore delle proteasi con diversi ruoli nella cascata coagulativa del sangue e nell'aggregazione piastrinica, mentre il suo peptide attivo, Bradichinina, ha numerose funzioni come mediatore dell'infiammazione e del dolore sia a livello endoteliale (causando vasodilatazione e extravasamento di proteine plasmatiche) che a livello della contrazione della muscolatura liscia.

Il *Fibrinogeno* è un eteroesamero, costituito da 2 subunità formate da 3 diverse catene: alfa, beta e gamma tutte legate tra loro da ponti disolfuro. È sintetizzato dal fegato e secreto nel plasma. Le sequenze identificate sono frammenti peptidici della catena alfa (FIBA), costituita da 866 aa, e della catena beta (FIBB) di 491 aa (Figura 31 e 32).

**P02671 (FIBA HUMAN) Fibrinogen alpha chain**

MF~~SMRIVCLVLSVVGTA~~**WTADSGEGD**~~FLAEGGGV~~RGRVVERHQSACKDSDWPFCSDE  
DW  
NYKCP~~SGCRMKGLIDEVNQDF~~TNRINKLKN~~SLFEYQKNNKDSHSLTTNIMEILRGDFSSA~~  
NNRDNTYNRVSEDLRSRIEVLKRKVI~~EKVQHIQLLQKNVRAQLVDMKRLEVDIDIKIRSC~~  
RGSCSRALAREVDLKDYE~~DQQKQLEQVI~~AKDLLPSRDRQHLPLIKMKPVPDLVPGNFKSQ  
LQKVPPEWKALDMPQMRMELERPGGNEITRGGSTSYGTGSETESPRNPSSAGSWNSGSS  
GPGSTGNRNPGSSGTGGTATWKP~~GSSGPGSTGSWNSGSSGTG~~STGNQNP~~GS~~PRPGSTGTW  
NPGSSERGSAGHWTSESSVSGSTGQWHSESGSFRPDS~~PGSGNARPNNDWGT~~FEEVSGNV  
SPGTRREYHTEKLVTSKGD~~KELRTGKEKVTSGSTTTTRR~~SCKTVTKTVIGPDGHKEVTK  
EVVTSE~~DGSDCPEAMD~~LGTLSGIGTL~~DGFRHRHPDEAAFFD~~TASTGKTFPGFFSPMLGEF  
VSE~~TESRGSESGIF~~TNTKESSSHHPGIAEFP~~SRGKSSSYSKQFTS~~SYNRGDSTFESK~~S~~  
YKMADEAGSEADHEGTHSTKRGHAKSRPVRDCDDVLQTHPSGTQSGIFNIKLP~~GSSKIFS~~  
VYCDQETS~~LGWLLIQQRMDGSLN~~FNRTWQDYKR~~GFGSLNDEGE~~GEF~~WLGN~~DYLHLLTQR  
GSVLRVELEDWAGNEAYAEYHFRV~~GSEAE~~GYALQVSSYEGTAGDALIEGSVEEGA~~EY~~TSH  
NNMQFSTFDRDADQWEENCAEVYGGGWYNNCQAANLNGIYYPGGSYDPRNNSPYEIE~~ENG~~  
VVVVSFRGADYSLRAVRMKIRPLVTQ

**Figura 31. Sequenza amminoacidica in formato FASTA del Fibrinogeno catena alfa (FIBA). In** *italico: peptide segnale, in grassetto: Fibrinopeptide A, nel riquadro: sequenza peptidica di frammento identificato.*

**P02675 (FIBB HUMAN) Fibrinogen beta chain**

MKRMV~~SWSFHKLKTMKHL~~LLLLLLCVFLVKS**QGVNDNEEGFFSARG**H~~RPLDK~~KREEAPSLR  
PAPPPISGGGYRARPAAKAAATQKKVERKAPDAGGCLHADPDLGVLCPTGCQLQEALLQQE  
RP~~IRNSVDELNNNVEAVSQTSSSS~~FQYMYLLKDLWQKRQKQVKDNENVVNEYSSELEKHQ  
LYIDETVNSNIPTNLRVLR~~SILENLR~~SKIQKLESDVSAQMEYCRTPCTVSCNIPVVS~~GKE~~  
CEEIIRKGGGETSEMYLIQPDSSVKPYRVYCDMNTENGGWTVIQNRQDGSVDFGRKWD~~PK~~  
QFGGNVATNTDGNK~~NYCGLPGEY~~WLGNDKISQLTRMGPT~~ELLIEMED~~WKGDKVKAHYGGFT  
VQNEANKYQISV~~NKYRGTAGNALMDGASQLMGENR~~TMTIHNGMFFSTYDRDNDGWLTS~~DP~~  
RKQCSKEDGGGWYNRCHAANPNGRYYWGGQYTWDMAKHGTTDDGVVWMNWKGSWYS~~MRKM~~  
SMKIRPFFFPQQ

**Figura 32. Sequenza amminoacidica in formato FASTA del Fibrinogeno catena beta (FBB). In** *italico: peptide segnale, in grassetto: Fibrinopeptide B, nel riquadro: sequenza peptidica di frammento precursore identificato.*

Il Fibrinogeno è una proteina caratterizzata da una doppia funzione: quella di produrre i monomeri di Fibrina per azione della Trombina (che taglia nel sito Arg-Lys liberando i Fibrinopeptidi A e B), che polimerizzando permettono la formazione del coagulo e quella di agire come cofattore di aggregazione piastrinica. Inoltre i prodotti di scissione del Fibrinogeno e la Fibrina stessa regolano la diffusione e l'adesione cellulare, presentano attività chemotattica, di vasocostrizione e sono mitogeni per diversi tipi cellulari coinvolti nella risposta infiammatoria.

Il C3f è un frammento di 17 aa derivante dal Complemento C3 (di 1641 aa) (Figura 33), il quale gioca un ruolo chiave nella risposta infiammatoria e immunitaria (Sistema del Complemento) in quanto la cascata proteolitica porta alla formazione di C3b con conseguente formazione dell'enzima C5 Convertasi. In particolare, il frammento C3f è generato durante l'inattivazione del C3b per azione proteolitica del Fattore del Complemento I.

**P01024 (CO3 HUMAN) Complement C3**

MGPTSGPSLLLLLLTHLPLALGSPMYSIITPNILRLESEETMVLEAHDAQGDVPVTVTVH  
 DFPGKKLVLSSEKTVLTPATNHMGNVFTTIPANREFKSEKGRNKFVTVQATFGTQVVEKV  
 VLVSLQSGYLFIQTDKTIYTPGSTVLYRIFTVNHKLLPVGRVVMVNIENPEGIPVKQDSL  
 SSQNQLGVLPLSWDIPELVNMGQWKIRAYYENSPQQVFSTEFVKEYVLPSEFVIVEPTE  
 KFYIYNEKGLEVTITARFLYGKKVEGTAFVIFIGIQDGEQRISLPESLKRIPIEDGSGEV  
 VLSRKVLLDGVQNPRAEDLVGKSLYVSATVILHSGSDMVQAERSGIPIVTSPYQIHFTKT  
 PKYFKPGMPFDLMVFVTNPDGSPAYRVPVAVQGEDTVQSLTQGDGVAKLSINTHPSQKPL  
 SITVTRTKQELSEAEQATRMTQALPYSTVGNNSNNYLHLSVLRTELRPGETLNVNFLLRMD  
 RAHEAKIRYYTYLIMNKGRLKAGRQVREPGQDLVVLPLSITTDIFIPSFRLVAYYTLIGA  
 SGQREVADSVWVDVKDSCVGSLLVVKSGQSEDRQPVPGQQMTLKIIEGDHGARVVLVAVDK  
 GVFVNLKKNKLTQSKIWDVVEKADIGCTPGSGKDYAGVFS DAGLFTTSSSGQQAQRAEL  
 QCPQPAARRRSVQLTEKRMVKYKPKELRKCCEDEGMRENPMRFSCQRTRFISLGEAC  
 KKVFLDCCNYITELRRQHARASHLGLARSNLDEDI~~IAEENIVSRSEFPESWLWNVEDLKE~~  
 PPKNGISTKLMNIFLKDSITTTWEILAVSMSDKKGCVADPFVTVMQDFFIDRLRPLYSVV  
 RNEQVEIRAVLYNYRQNELKVRVELLHNPAFCSLATTKRRHQQTVTIPPKSSLSVPYVI  
 VPLKTGLQEVEVKAAVYHHFISDGVRSKLVVPEGIRMNKTVAVRTLDPERLREGVQKE  
 DIPPADLSDQVPDTESETRILLQGTTPVAQMTEDA~~VDAERLKH~~LIVTPSGCGEQNMIGMTP  
 TVIAVHYLDETEQWEKFGLEKRQGALELIKKGYTQQLAFRQPSSAFAAFVKRAPSTWLTA  
 YVVKVFSLAVNLI~~AIDSQVLCGAVKWLILEKQKPDGVFQEDAPVIHQEMIGGLRNNNEKD~~  
 MALTAFLVLI~~SLQEAKDICEEQVNSLPGSITKAGDFLEANYMNLQRSYTVAIAGYALAQMG~~  
 RLKGPLLNKFLTTAKDKNRWEDPGKQLYNVEATSYALLALLQLKDFDFVPPVVRWLNEQR  
 YYGGGYGSTQATFMVFQALAQYQKDAPDHQELNLDVSLQLPSRS~~SKITHRIHWESASLLR~~  
 SEETKENEGFTVTAEGKGQGTLSVVTMYHAKAKDQLTCNKFDLKVTIKPAPETEKRPQDA  
 KNTMILEICTRYRGDQDATMSILDISMMTGFAPDTDDLKQLANGVDRIYSKYELDKAFSD  
 RNTLIIYLDKVSHSEDDCLAFKVH~~QYFNVELIQPGAVKVYAYYNLEESCTRFYHPEKEDG~~  
 KLNKLCRDELRC~~CAEENC~~FIQKSDDKVTL~~EEERLDKACEPGVDYVYKTRLVKVQLSNDFDE~~  
 YIMAIEQTIKSGSDEVQVGGQRTFISPIKCREALKLEEKHYLMWGLSSDFWGEKPNLSY  
 IIGKDTWVEHWPEEDECQDEENQKQCQDLGAFTESMVVFGCPN

Figura 33. Sequenza amminoacidica in formato FASTA del Complemento C3 (CO3). In italico: peptide segnale, sottolineato: C3a (Anafilotossina), in grassetto: C3b, nel riquadro: sequenza peptidica di frammento precursore identificato come C3f.

Il Complemento C4 (di 1744 aa), invece, è secreto e circola nel sangue come trimero di 3 diverse catene (alfa, beta e gamma) legate da ponti disolfuro. Sono presenti due isoforme proteiche, C4-A e C4-B, che differiscono solo per 10

amminoacidi dispersi nella catena alfa tra amminoacido 907 e il 1286 (Figura 34).  
 In questa regione sono presenti in particolare le sequenze dei frammento peptidico C4b e C4d. I 2 frammenti identificati corrispondono a 2 diverse sequenze, tra loro successive nella regione omologa della catena alfa di C4.

**P0C0L4 (CO4A HUMAN) Complement C4-A e P0C0L5 (CO4B HUMAN) Complement C4-B**

C4-A	1	MRLWGLIWASSFFTL <del>SLQKPRLLLF</del> SPSVVHLGVPLSVGVQLQDVPRGQVVKGSVFLRN	60
C4-B	1	MRLWGLIWASSFFTL <del>SLQKPRLLLF</del> SPSVVHLGVPLSVGVQLQDVPRGQVVKGSVFLRN	60
C4-A	61	PSRNNVPCSPKVDFTLSSERDFALLSLQVPLKDAKSCGLHQLLRGPEVQLVAHSPWLKDS	120
C4-B	61	PSRNNVPCSPKVDFTLSSERDFALLSLQVPLKDAKSCGLHQLLRGPEVQLVAHSPWLKDS	120
C4-A	121	LSRTTNIQGINLLFSSRRGHFLQTDQPIYNPQQRVRYRVFALDQKMRPSTDTITVMVEN	180
C4-B	121	LSRTTNIQGINLLFSSRRGHFLQTDQPIYNPQQRVRYRVFALDQKMRPSTDTITVMVEN	180
C4-A	181	SHGLRVRKKEVYMPSSIFQDDFVIPDISEPGTWKISARFSDGLESNSSTQFEVKKYVLPN	240
C4-B	181	SHGLRVRKKEVYMPSSIFQDDFVIPDISEPGTWKISARFSDGLESNSSTQFEVKKYVLPN	240
C4-A	241	FEVKITPGKPYILTVPGHLDQMQLDIQARYIYGKPVQGVAYVRFGLLDEDGKKTFFRGGLE	300
C4-B	241	FEVKITPGKPYILTVPGHLDQMQLDIQARYIYGKPVQGVAYVRFGLLDEDGKKTFFRGGLE	300
C4-A	301	SQTKLVNGQSHISLSKAEFQDALEKLNMGITDLQGLRLYVAAAIIESPGGEMEEAELTSW	360
C4-B	301	SQTKLVNGQSHISLSKAEFQDALEKLNMGITDLQGLRLYVAAAIIESPGGEMEEAELTSW	360
C4-A	361	YFVSSPFLDLSKTKRHLVPGAPFLLQALVREMSGSPASGIPVKVSATVSSPGSVPEVQD	420
C4-B	361	YFVSSPFLDLSKTKRHLVPGAPFLLQALVREMSGSPASGIPVKVSATVSSPGSVPEVQD	420
C4-A	421	IQQNTDGSQVSIPIIIPQTISELQLSVAGSPHPAIARLTVAAPPSSGGPGFLSIERPDS	480
C4-B	421	IQQNTDGSQVSIPIIIPQTISELQLSVAGSPHPAIARLTVAAPPSSGGPGFLSIERPDS	480
C4-A	481	RPPRVGDTLNLNLRVAVGSGATF <del>SHYYMI</del> LSRGQIVFMNREP <del>KRTL</del> TSVSVFVDHHLAPS	540
C4-B	481	RPPRVGDTLNLNLRVAVGSGATF <del>SHYYMI</del> LSRGQIVFMNREP <del>KRTL</del> TSVSVFVDHHLAPS	540
C4-A	541	FYFVAFYYHGDHPVANSLRVDVQAGACEGKLELSVDGAKQYRNGESVKLHLETDSLALVA	600
C4-B	541	FYFVAFYYHGDHPVANSLRVDVQAGACEGKLELSVDGAKQYRNGESVKLHLETDSLALVA	600
C4-A	601	LGALDTALYAAGSKSHKPLNMGKVF <del>EAMNSYDL</del> GCGPGGGDSALQVFQAAGLAFSDGDQW	660
C4-B	601	LGALDTALYAAGSKSHKPLNMGKVF <del>EAMNSYDL</del> GCGPGGGDSALQVFQAAGLAFSDGDQW	660
C4-A	661	TL <del>SRKRL</del> SCPK <del>ETTR</del> K <del>RNVNF</del> QKAIN <del>EKL</del> GQYASPTAKRCCQDGV <del>TRLP</del> MMRSCEQRA	720
C4-B	661	TL <del>SRKRL</del> SCPK <del>ETTR</del> K <del>RNVNF</del> QKAIN <del>EKL</del> GQYASPTAKRCCQDGV <del>TRLP</del> MMRSCEQRA	720
C4-A	721	ARVQQPDCREPF <del>LSCCQFAESLR</del> KKSRDKGQAGLQRALEILQEEDLIDEDDIPVRSFFPE	780
C4-B	721	ARVQQPDCREPF <del>LSCCQFAESLR</del> KKSRDKGQAGLQRALEILQEEDLIDEDDIPVRSFFPE	780
C4-A	781	NWLWRVETVDRFQILTLWLPDLSLTWEIHGLSLSKTKGLCVATPVQLRVFREFHLHLRLP	840
C4-B	781	NWLWRVETVDRFQILTLWLPDLSLTWEIHGLSLSKTKGLCVATPVQLRVFREFHLHLRLP	840
C4-A	841	MSVRRFEQLELRPVLYNYLDKNLTVSVHVS <del>PVEGL</del> CLAGGGGLAQQVLPAGSARPVAFS	900
C4-B	841	MSVRRFEQLELRPVLYNYLDKNLTVSVHVS <del>PVEGL</del> CLAGGGGLAQQVLPAGSARPVAFS	900
C4-A	901	VVPTAAAVSLKVVARGSF <del>FEFPV</del> GD <del>AVSKVLQIE</del> KEGAIHREELVYELNPLDHRGRTLEI	960
C4-B	901	VVPTAAAVSLKVVARGSF <del>FEFPV</del> GD <del>AVSKVLQIE</del> KEGAIHREELVYELNPLDHRGRTLEI	960
C4-A	961	PGNSDPNMIPDGFNSYVRVTASDPLDTL <del>GSE</del> GALS <del>PGG</del> VASLLR <del>LR</del> PRGCGEQ <del>TM</del> IY <del>LP</del>	1020
C4-B	961	PGNSDPNMIPDGFNSYVRVTASDPLDTL <del>GSE</del> GALS <del>PGG</del> VASLLR <del>LR</del> PRGCGEQ <del>TM</del> IY <del>LP</del>	1020
C4-A	1021	TLAASRYLDKTEQWSTLPPETKDHAVDLIQKGYMRIQQFRKADGSYA <del>AWLSR</del> DSSTWLTA	1080
C4-B	1021	TLAASRYLDKTEQWSTLPPETKDHAVDLIQKGYMRIQQFRKADGSYA <del>AWLSR</del> DSSTWLTA	1080
C4-A	1081	FVLKVL <del>SLAQEQV</del> GGSP <del>EKLQ</del> ETS <del>SNWLLS</del> QQQADG <del>SFQD</del> PC <del>PVLD</del> DRSMQGGLVGND <del>ETVA</del>	1140
C4-B	1081	FVLKVL <del>SLAQEQV</del> GGSP <del>EKLQ</del> ETS <del>SNWLLS</del> QQQADG <del>SFQD</del> PC <del>PVLD</del> DRSMQGGLVGND <del>ETVA</del>	1140
C4-A	1141	LTAFVTIALHHGLAVFQDEGA <del>EPLKQ</del> RVEASISKANS <del>F</del> LGEKASAGLLGAHAAAITAYAL	1200
C4-B	1141	LTAFVTIALHHGLAVFQDEGA <del>EPLKQ</del> RVEASISKANS <del>F</del> LGEKASAGLLGAHAAAITAYAL	1200
C4-A	1201	TLTKAPVDLLGVAHNNLMAMAQETGDNL <del>Y</del> WGSV <del>TGSQ</del> SN <del>AVSPT</del> PAPRNP <del>SD</del> PMQ <del>APAL</del>	1260
C4-B	1201	TLTKAPVDLLGVAHNNLMAMAQETGDNL <del>Y</del> WGSV <del>TGSQ</del> SN <del>AVSPT</del> PAPRNP <del>SD</del> PMQ <del>APAL</del>	1260
C4-A	1261	WIETTAYALLHLLHHEGKAEMADQAA <del>AWL</del> TRQGSFQGGFRSTQDTVIALDAL <del>SAY</del> WIASH	1320
C4-B	1261	WIETTAYALLHLLHHEGKAEMADQAA <del>AWL</del> TRQGSFQGGFRSTQDTVIALDAL <del>SAY</del> WIASH	1320
C4-A	1321	TTEERGLNVTLSSTGRNGFKSHALQ <del>LNNRQIRGLEEELQ</del> FSLGSKINVKVGGNSK <del>G</del> TLKV	1380
C4-B	1321	TTEERGLNVTLSSTGRNGFKSHALQ <del>LNNRQIRGLEEELQ</del> FSLGSKINVKVGGNSK <del>G</del> TLKV	1380
C4-A	1381	LRTYNVLD <del>MKN</del> TT <del>CQDLQIE</del> VTVKGHVEY <del>TMEANEDY</del> EDY <del>EYDEL</del> PAKDD <del>PD</del> APLQ <del>PVTP</del>	1440
C4-B	1381	LRTYNVLD <del>MKN</del> TT <del>CQDLQIE</del> VTVKGHVEY <del>TMEANEDY</del> EDY <del>EYDEL</del> PAKDD <del>PD</del> APLQ <del>PVTP</del>	1440
C4-A	1441	LQ <del>L</del> FEGRNRRRRREAPKVVEEQESRVHYTVC <del>IWR</del> NGKVGLSGMAIADVTLLSGFHALRAD	1500
C4-B	1441	LQ <del>L</del> FEGRNRRRRREAPKVVEEQESRVHYTVC <del>IWR</del> NGKVGLSGMAIADVTLLSGFHALRAD	1500
C4-A	1501	LEKLTSLSDRYVSHFETEGPHVLLYFDSVPTSREC <del>VGF</del> EAVQEV <del>PVGLVQ</del> PASATLYDY	1560
C4-B	1501	LEKLTSLSDRYVSHFETEGPHVLLYFDSVPTSREC <del>VGF</del> EAVQEV <del>PVGLVQ</del> PASATLYDY	1560

C4-A	1561	NPERRCSVFGAPSKSRLLATLCSAEVCQCAEGKCPQRRALERGLQDEDEDGYRMKFACY	1620
C4-B	1561	NPERRCSVFGAPSKSRLLATLCSAEVCQCAEGKCPQRRALERGLQDEDEDGYRMKFACY	1620
C4-A	1621	PRVEYGFQVKVLRSDSRAAFRLFETKITQVLHFTKDVKAAANQMRNFLVRASCRLRLEPG	1680
C4-B	1621	PRVEYGFQVKVLRSDSRAAFRLFETKITQVLHFTKDVKAAANQMRNFLVRASCRLRLEPG	1680
C4-A	1681	KEYLIMGLDGATYDLEGHPOYLSDSNWIEEMPSERLCRSTRQRAACAQLNDFLQYEGTQ	1740
C4-B	1681	KEYLIMGLDGATYDLEGHPOYLSDSNWIEEMPSERLCRSTRQRAACAQLNDFLQYEGTQ	1740
C4-A	1741	GCQV	1744
C4-B	1741	GCQV	1744

**Figura 34. Allineamento di sequenze amminoacidiche del Complemento C4-A e C4-B ottenuto con BlastP. In grigio: amminoacidi differenti, in grassetto: Catena Alfa, nel riquadro: sequenza peptidica di frammenti precursori identificati.**

L'*Alfa-2-Antiplasmina* è una proteina di 491 aa, espressa dal fegato e secreta nel plasma. È un inibitore delle serin-proteasi, principalmente di Plasmina e Tripsina. La sequenza identificata corrisponde ai 12 aa del propeptide.

L'*ITI4* (Inter-alpha-trypsin inhibitor heavy chain H4) è una proteina di 930 aa espressa esclusivamente dal fegato e secreta in circolo sanguigno dove è scissa dalla Callicreina in 2 catene, la 70 kDa da 633 aa e la 35 kDa da 242 aa, più un propeptide di 27 aa (Figura 35). Le sequenze identificate corrispondono agli ultimi 15 e 20 amminoacidi del propeptide.

**Q14624 (ITI4 HUMAN) Inter-alpha-trypsin inhibitor heavy chain H4**

*MKPPRPVRTCSKVLVLLSLLAIHQTTTAEKNGIDIYSLTVDSRVSSRFAHTVVTSRVNR  
ANTVQEATFQMELPKKAFITNFSMIDGMTYPGIIKEKAEAQAQYSAAVAKGKSAGLVKA  
TGRNMEQFQVSVVAPNAKITFELVYEELLKRRLLGVYELLLKVRPQQLVKHLQMDIHIFE  
PQGISFLETSTFMTNQLVDALTTWQNKTKAHIRFKPTLSQQQKSPEQQETVLDGNLIIR  
YDVDRAISGGSIQIENGYFVHYFAPEGLTTPKNNVVFVIDKSGSMSGRKIQQTREALIKI  
LDDLSPRDQFNLI VFSTEATQWRPSLVPASAENVNKARSFAAGIQALGGTNINDAMLMV  
QLLDSSNQEERLPEGSVSLIILLTDGDPTVGETNPRS IQNNVREAVSGRYSFLCLGFGFD  
VSYAFLEKLALDNGGLARRIHEDSDSALQLQDFYQEVANPLLTAVTFEYPSNAVEEVTQN  
NFRLLFKGSEM VVAGKLQDRGPDVLTATVSGKLPQNTITFQTESSVAEQEAQFQSPKYIF  
HNFMERLWAYLTIQQLLEQTVSASDADQALRNQALNLSLAYSFVTPLTSMVVTKPDDQE  
QSQVAEKPMEGESRNRNVHSGSTFFKYLLQGAIKPKPEASFSPRRGWNROAGAAGSRMNF  
RPGVLSRRLGLPGPPDVPDHAAYHPFRRLAILPASAPPATSNPDPVSRVMMMKIEETT  
MTTQTPAPIQAPSAILPLPGQSVRLCVDPRHRQGPVNLLSDPEQGVVETGQYEREKAGF  
SWIEVTFKNPLVWVHASPEHVVTNRNRSSAYKWKETLFSVMPGLKMTMDKTGLLLSDP  
DKVTIGLLFWDGRGEGRLRLRLDRFSSHVGGTLGQFYQEVWLGSPAASDDGRRTLRVQ  
GNDHSATRERRLDYQEGPPGVEISCWSVEL*

**Figura 35. Sequenza amminoacidica in formato FASTA di ITI4. In italico: peptide segnale, in grassetto: propeptide, nel riquadro: sequenza peptidica di frammento precursore identificato.**

L'ITIH4 è un inibitore delle serin-proteasi, con funzione anti-infiammatoria e coinvolta nelle reazioni di fase acuta di tipo 2. Inoltre sembra avere un ruolo nella stabilità della matrice extracellulare mediante il legame covalente con acido ialuronico e nello sviluppo e rigenerazione del fegato.

La *lipofosfatidilcolina* è prodotta dall'idrolisi della fosfatidilcolina per azione della fosfolipasi A<sub>2</sub>, la quale favorisce l'aumento delle proprietà emulsionanti delle micelle biliari, favorendo la cessione degli acidi grassi in fase acquosa.

## 5. Discussione

La valutazione del profilo peptidico per la ricerca di biomarcatori su campioni di plasma di soggetti sani e di pazienti affetti da CRC, è stato possibile dopo una attenta messa a punto della metodica pre-analitica utilizzando il dispositivo prototipo progettato dal gruppo di lavoro del Prof Mauro Ferrari. Tale dispositivo e' costituito da una superficie nanoporosa che mediante un processo di adsorbimento ed esclusione per dimensione, è in grado di estrarre e purificare le molecole a basso peso molecolare (ovvero meno di 3000 Da) presenti nel plasma. L'analisi in spettrometria di massa ha permesso l'acquisizione del profilo ionico generato da peptidi e frammenti di peptidi, che successivamente sono stati identificati.

Fase cruciale è l'analisi informatizzata, necessaria per processare la matrice di dati ottenuta dagli spettri MALDI e che permette di correlare l'intensità delle specie ioniche osservate alle informazioni cliniche dei campioni in studio. A tale scopo si sono utilizzati diversi strumenti statistici multivariati distinguibili in metodi supervisionati, che nel calcolo utilizzano l'appartenenza del campione ad un certo gruppo, ed in metodi non supervisionati, che non lo adottano. L'uso di entrambi è importante sia al fine di ottimizzare la capacità informativa del dato finale (metodi supervisionati), sia al fine di evitare l'introduzione di *bias* che possono generare un modello di previsione adatto allo studio in corso, ma non generalizzabile a dati futuri (metodi non supervisionati).

L'analisi di predizione di classificazione, effettuata mediante analisi di Cluster, PCA, PLS-DA e Random Forest, ha permesso una buona discriminazione tra soggetti sani da soggetti con lesione pre-cancerosa ad alto rischio e da pazienti malati di CRC sia a stadio precoce (stadio I e II) che tardivo (III e IV).

La capacità discriminante risulta minore tra i gruppi di soggetti malati (fase pre-neoplastica, stadio precoce e stadio tardivo), tanto che si è ritenuto opportuno di riunire le casistiche dei pazienti con cancro in un unico gruppo.

L'identificazione delle specie ioniche ha permesso di valutare più criticamente i dati forniti dall'analisi di correlazione delle variabili. Questo ha permesso di verificare la valenza biologica delle correlazioni dirette (stesso trend di aumento

o diminuzione) e inverse (diverso trend di aumento o diminuzione) osservate. Ad esempio si è potuto giudicare coerenti le relazioni dirette presenti tra le specie a  $m/z$  1060.57 e 1076.56 poiché esse sono state identificate essere rispettivamente la Bradichinina e la Bradichinina ossidata. Allo stesso modo, è stata identificata l'origine biologica della correlazione osservata tra le specie a  $m/z$  1934.07, 2021.1 e 2037.09 (identificate come un frammento di C3f, il C3f intatto ed il C3f ossidato), e tra le specie  $m/z$  2433.3 e 1891.02: (identificate come il frammento precursore di C4-A/B ed il relativo frammento più corto di C4-A/B). Interessanti sono invece le relazioni inverse, in cui all'aumento di intensità di una specie, si è osservata la diminuzione dell'intensità di un'altra. Ad esempio tra la specie a  $m/z$  2021.1 e le specie 2433.3 e 1436.72 (ovvero tra il C3f e due frammenti di un'altra proteina del complemento, il C4-A/B). Infine, in maniera inaspettata è stata osservata una correlazione inversa anche tra due specie ioniche riconducibili allo stesso precursore iniziale di C3f, ovvero quelle a  $m/z$  1211.65 e 1934.07. Questo fatto può essere giustificato dalla complessità della cascata proteolitica, che chiama in causa diverse endo ed esoproteasi, le quali possono agire indipendentemente sullo stesso precursore.

Da queste analisi preliminari sono state individuate specie ioniche che presentano negli spettri MALDI differenze di intensità (Figura 36) tali da permettere di ottenere una buona capacità discriminante tra Controlli e Adenomi e CRC (Tabella 10) con un'accuratezza rappresentata da ROC-AUC rispettivamente compresa tra 0.7-0.8 e  $>0.8$ .



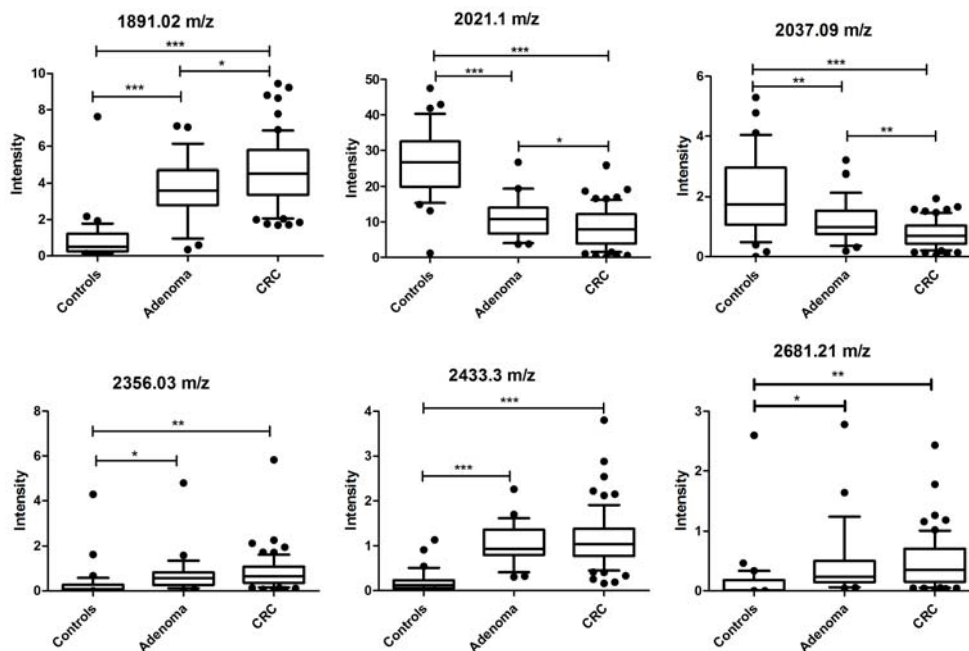


Figura 36. Box plot delle specie ioniche più significative.

Tabella 10. Dati delle curve ROC delle specie ioniche più significative. Confronto tra Controlli e Adenomi (in alto) e tra Controlli e CRC (in basso).

Controls vs Adenoma						
m/z	AUC	95% CI	p-value	Cut-off	Sensitivity%	Specificity%
2433.3	0.9586	0.9087 - 1.008	< 0.0001	> 0.3145	96.3	88.24
2021.1	0.9237	0.851-0.9957	<0.0001	< 18.45	88.89	82.35
1891.02	0.9118	0.8299 - 0.9936	< 0.0001	> 2.219	85.19	97.06
1896.03	0.8856	0.7902 - 0.9810	< 0.0001	< 2.435	85.19	88.24
1436.72	0.8595	0.7689 - 0.9501	< 0.0001	> 0.5060	81.48	76.47
2356.03	0.8355	0.7321 - 0.9389	< 0.0001	> 0.2035	85.19	73.53
2681.21	0.8219	0.7174 - 0.9264	< 0.0001	> 0.1205	81.48	70.59
2037.09	0.7397	0.6143 - 0.8650	0.00141	< 1.305	66.67	70.59

Controls vs CRC						
m/z	AUC	95% CI	p-value	Cut-off	Sensitivity%	Specificity%
1891.02	0.9680	0.9150 - 1.021	< 0.0001	> 1.662	100	91.18
2433.3	0.9592	0.9180 - 1.000	< 0.0001	> 0.5500	89.55	94.12
2021.1	0.9539	0.8987 - 1.009	< 0.0001	< 16.12	91.04	88.24
1436.72	0.9085	0.8424 - 0.9745	< 0.0001	> 0.8965	80.6	82.35
1896.03	0.8854	0.7945 - 0.9764	< 0.0001	< 2.759	83.58	82.35
2356.03	0.8628	0.7762 - 0.9494	< 0.0001	> 0.3170	80.6	85.29
2681.21	0.8374	0.7491 - 0.9256	< 0.0001	> 0.1490	77.61	73.53
2037.09	0.8338	0.7373 - 0.9304	< 0.0001	< 0.9565	71.64	82.35
2167.08	0.7796	0.6745 - 0.8847	< 0.0001	> 0.0825	79.1	70.59

La capacità discriminata tra i gruppi patologici (Adenomi e CRC) è tuttavia scarsa e presenta ROC-AUC comprese tra 0.5 e 0.7.

Hanno suscitato particolare attenzione i *cluster* di peptidi di C3f e C4-A/B che si presentano come un set frammenti sovrapposti, ma con troncamento a scaletta alle estremità N- o C-terminale (Figura 37)

<i>N-term</i>	<i>Sequence</i>	<i>C-term</i>	<i>m/z</i>
	HWESA	SLLR	1098.6
I	HWESA	SLLR	1211.65
RI	HWESA	SLLR	1367.77
THRI	HWESA	SLLR	1605.87
ITHRI	HWESA	SLLR	1718.95
KITHRI	HWESA	SLLR	1847.04
SKITHRI	HWESA	SLLR	1934.07
SSKITHRI	HWESA	SLLR	2021.10
SSKITHRI	HWESA	SLLR	2037.09
SSKITHRI	HW $E(Na^+)$ SA	SLLR	2043.1
SSKITHRI	HW $E(K^+)$ SA	SLLR	2059.0
SSKITHRI	HWESA	SLL	1865.00
SKITHRI	HWESA	SLL	1777.98
SSKITHRI	HWESA		1551.82
SKITHRI	HWESA		1464.77
KITHRI	HWESA		1377.72
ITHRI	HWESA		1249.65
THRI	HWESA		1136.57

<i>N-term</i>	<i>Sequence</i>	<i>C-term</i>	<i>m/z</i>
	LQLNNR	QIR	1154.68
A	LQLNNR	QIR	1225.72
SHA	LQLNNR	QIR	1449.77
GFKSHA	LQLNNR	QIR	1782.00
NGFKSHA	LQLNNR	QIR	1896.03
NGFKSHA	LQLNNR	QIR	1897.02
NGFKSHA	LQLNNR	QIR(NH $4^+$ )	1913.0
NGFKSHA	LQLNNR		1499.79
SHA	LQLNNR		1052.56

<i>Sequence</i>	<i>C-term</i>	<i>m/z</i>
GLEEELQFSLGSK	INVKVGNSK	2433.30
GLEEELQFSLGSK	INVK	1891.02
GLEEELQFSLGSK		1436.72

Figura 37. Rappresentazione dei tagli esoproteasici dei frammenti identificati. Cluster di C3f (in alto) e di C4-A/B del frammento [1337-1352] (al centro) e del frammento [1353-1375] (in basso). In grigio: frammenti precursore.

A differenza della Bradichinina, del frammento di Alfa-2-Antiplasmina, del frammento di ITIH4 e del complemento C3f, che sono generati da enzimi plasmatici attraverso un taglio endoproteolitico operato in punti ben precisi della sequenza amminoacidica, i *cluster* peptidici osservati hanno una diversa origine. Essi sono derivati dall'attività di una famiglia specifica di proteasi, ovvero le *esoproteasi*. Questa famiglia di enzimi della classe delle idrolasi contiene tutte le proteasi che idrolizzano il legame peptidico terminale della catena amminoacidica in maniera totalmente aspecifica. A seconda che il residuo idrolizzato si trovi sulla parte N-terminale o C-terminale del peptide, si distinguono rispettivamente in aminopeptidasi e carbossipeptidasi.

Valutando il rapporto dell'intensità dei frammenti derivanti dal precursore C3f e C4-A/B nel gruppo di Adenomi e CRC rispetto ai Controlli (Figura 38), si può

notare la differenza di abbondanza dei singoli frammenti, che porta ad identificare un diverso *pattern* di frammenti in grado di caratterizzare ciascun gruppo.

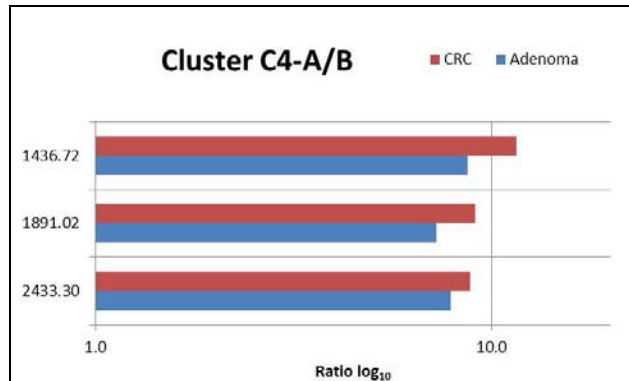
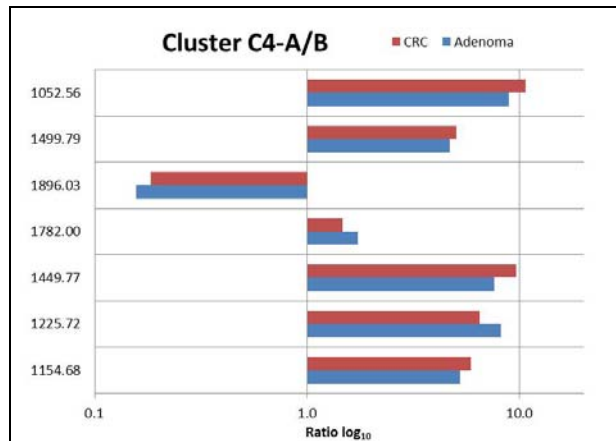
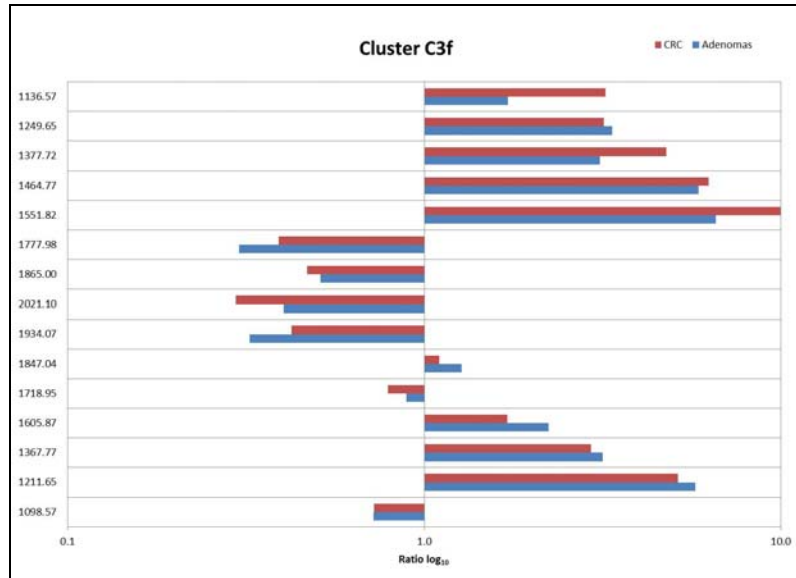


Figura 38. Grafico dei rapporti delle mediane delle intensità (caso/controlli), per il cluster di C3f (in alto) e di C4-A/B del frammento [1337-1352] (al centro) e del frammento [1353-1375] (in basso).

Alcuni frammenti peptidici presentano un rapporto tra l'intensità delle mediane inferiore negli Adenomi e nei CRC rispetto ai Controlli, come le specie a m/z 1777.98, 1865.0, 2021.1, 1934.07, 1718.95 e 1098.6 per il cluster di C3f o la specie a m/z 1896.03 per il cluster C4-A/B. A parte per la specie a m/z 1136,57, che sembra essere presente con maggior intensità nel gruppo del cancro, non si nota sostanziale differenza tra i rapporti delle specie ioniche tra gli Adenomi e i CRC.

Per visualizzare la variazione di intensità del precursore, è utile valutare la mediana dell'intensità espressa come percentuale sul totale dei frammenti prodotti (Figura 39). Questo permette di visualizzare le variazioni riscontrate nella loro globalità e di verificare se, e con che impatto, sia variata la quantità di precursore circolante.

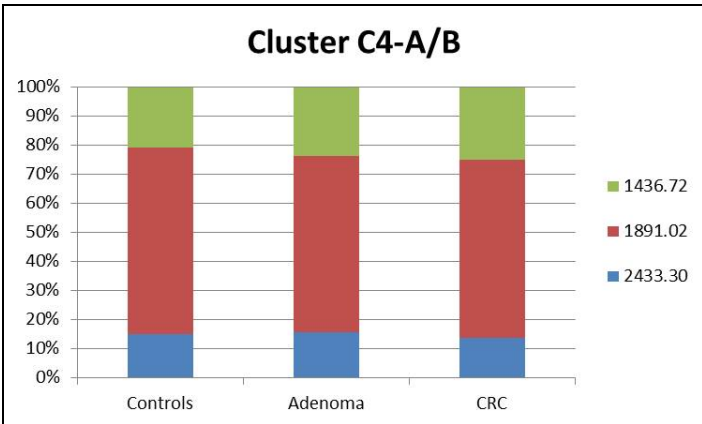
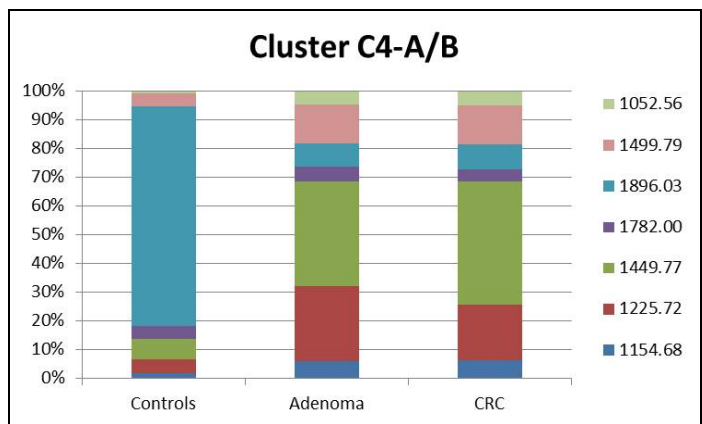
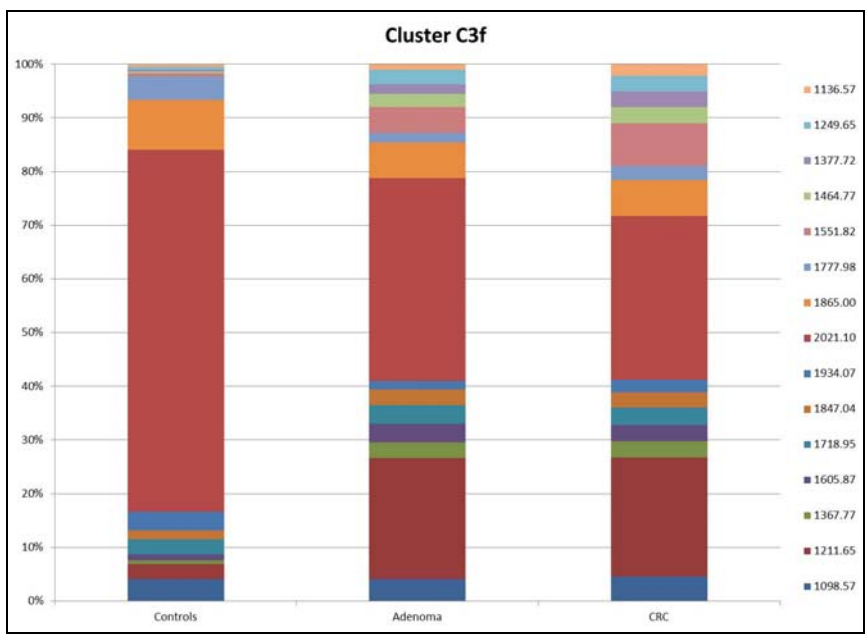


Figura 39. Grafico della percentuale relativa delle intensità dei frammenti del cluster di C3f (in alto) e di C4-A/B del frammento [1337-1352] (al centro) e del frammento [1353-1375] (in basso).

Il *cluster* di C3f nei Controlli presenta un'alta percentuale di frammento precursore a m/z 2021,1 (rappresenta circa il 65% del totale), che cala drasticamente negli Adenomi (< del 40%) e nei CRC (circa il 30%). A questa diminuzione corrisponde un aumento delle specie ioniche a m/z 1136.57, 1249.65, 1337.72, 1464.77 e 1551.82. Le prime quattro specie rappresentano frammenti troncamenti a scaletta sull'amminoacido N-terminale a partire dalla specie m/z 1551.82, derivante dal precursore C3f. Sul lato C-terminale è interessante invece osservare l'aumento notevole della specie ionica a m/z 1211.65, che passa dal 3% circa dei controlli al 20% degli Adenomi e dei CRC. Questa specie deriva direttamente dal frammento precursore C3f, ed al suo drastico aumento si accompagna solo una lieve variazione delle percentuali dei frammenti intermedi. Questo dato potrebbe essere dovuto sia ad un aumento dell'attività proteasica che porta al frammento 1211.65, sia ad una diminuzione del suo catabolismo. Tuttavia, in mancanza di altri dati non è possibile verificare queste ipotesi.

Per quanto riguarda i due *cluster* generati da due frammenti di C4-A/B, sono stati osservati due andamenti opposti. In un caso è stata osservata un'estesa variazione percentuale della specie ionica m/z 1896.03 (frammento avente sequenza NGFKSHALQLNNRQIR[1337-1352]) che è passa da circa l'80% d'intensità nei controlli al 10% circa negli Adenomi e nei CRC. A questa sua diminuzione è associato un aumento generale dell'intensità di tutte le specie, ed in particolare dei frammenti a m/z 1225.72 e 1449.77.

Nel secondo caso, ovvero quello relativo al peptide a m/z 2433.3 avente sequenza GLEELQFSLGSKINVKGGNSK[1353-1375], nessuna sostanziale variazione di percentuale è stata osservata. Questo dato conferma nuovamente la complessità del sistema endo/esoproteasico, per cui peptidi diversi che però derivano da uno stesso precursore e si trovano nello stesso medium biologico, sono sottoposti a vie di metaboliche completamente differenti.

Questi pattern di diversi peptidi, che si sono rilevati negli spettri MALDI analizzati sembrano essere causati da un aumento quantitativo del frammento precursore, oppure da un aumento della quantità di specifiche esoproteasi o riflettere una

diversa attività proteasica. Tale attività è presente normalmente nel sangue durante la coagulazione e attivazione del Complemento, generando instabilità delle proteine e dei peptidi anche durante il prelievo di sangue e successivamente durante la preparazione del campione, soprattutto per la preparazione del siero. Questa attività proteasica *ex vivo* intrinseca del campione è presente, sebbene in minor quantità, anche nel plasma, in quanto i comuni anticoagulanti (eparina, citrato ed EDTA) hanno una parziale azione inibente, perchè specifici per una molecola della cascata coagulativa [66].

Ciò non toglie che le cellule tumorali possano contribuire con il rilascio di proteasi uniche che determinano una alterazione delle centinaia di peptidi che possono essere rilevati nel campione. Le proteasi hanno un ruolo importante in diversi processi biologici come la differenziazione cellulare, il rimodellamento tessutale, l'omeostasi, l'immunità, l'angiogenesi e l'apoptosi [67]

Soprattutto per i tumori solidi è nota la relazione tra le proteasi e il cancro: la disregolazione dell'attività delle proteasi è un fattore che promuove la progressione locale e le metastasi a distanza [68]

Il tumore promuove la funzione proteasica come parte di una complessa rete di interazioni con altre importanti vie di segnale coinvolgendo chemochine, citochine e chinasi [69,70]. Le proteasi in molti tumori sono caratterizzate da frequente sovra-espressione, come le *metalloproteasi* (MMP) nel rimodellamento della matrice extracellulare per favorire angiogenesi, invasività e metastatizzazione [71], in contrasto l'inattivazione di distinte proteasi, come le *caspiasi*, possono contribuire alla progressione tumorale inibendo l'apoptosi [72].

Le proteasi associate al tumore sono secrete direttamente dal tumore o dal tessuto connettivo circostante e dai leucociti infiltranti a seguito dell'interazione presente tra stroma e tumore. Diversi meccanismi portano a concentrare l'attività delle proteasi nel microambiente circostante la massa tumorale come l'espressione di proteasi ancorate alla membrana cellulare, l'attivazione di proteasi solubili mediante legame con recettori di membrana, la presenza di recettori sulla superficie cellulare che attivavano le proteasi circostanti, la regolazione localizzata di recettori e proteasi in determinati domini sub-cellulari

e l'incremento locale della concentrazione di proteasi sulla specifica componente della matrice extracellulare [73].

Lo specifico pattern di proteasi può riflettere il potenziale fenotipo biologico e metastatico del tessuto tumorale e lo studio per l'identificazione di proteasi specifiche del tumore è molto promettente, sebbene investigare lo stesso pattern nel campione di siero o plasma è più facile accesso alla pratica di laboratorio. Infatti, lo studio di Villanueva e collaboratori [74] ha evidenziato che l'attività esoproteasica sovrapposta a quella *ex vivo* della cascata coagulativa e del Complemento, contribuiscono a generare sia peptidi cancro-specifici che cancro-tipo-specifici, identificando pattern peptidici da analisi MALDI-TOF del profilo sierico. Durante le cascate proteolitiche si generano peptidi da tagli endoproteasici intrinseci e alcuni di questi sono conosciuti come molecole bioattive (Bradichinina), altri rappresentano propeptidi tagliati (ITIH4C3f, Alfa-2-Antiplasmina) e altri ancora sembrano frammenti casuali di proteine precursore (C4-A/B, Chininogeno, Fibrinogeno). Osservando i siti di taglio sono generalmente consistenti in attività simili alla tripsina e chimotripsina per azione di serin-protesi conosciute come Callicreina, Plasmina, Trombina. Una volta generato il peptide, per azione di esoproteasi, subisce un taglio a scaletta.

Studi preliminari sono stati fatti per identificare il profilo funzionale di proteasi utilizzando peptidi reporter esogeni su campioni di siero [75] oltre che studi per identificare i siti di taglio e i substrati di proteasi associate al tumore [76]

Inoltre interessante è l'identificazione dei frammenti di ITIH4. Alcuni dei suoi frammenti sono già stati descritti da altri autori e il diverso pattern di frammentazione è stato associata a diverse patologie come infarto [77] tumore, alla mammella [78], alla prostata, alla vescica [74], all'ovaio, al pancreas e nel diabete [79]. La frammentazione osservata da Song e collaboratori [79] sembra avvenire in vivo e non essere relativa alla formazione del coagulo. Sono presenti diversi frammenti di ITIH4 anche nei soggetti sani, ma solo alcuni frammenti aumentano o diminuiscono in pazienti con diversa patologia, ma non in quelli con cancro del colon. Differenti tumori presentano diverso pattern proteolitico e quindi tale frammentazione mantiene importanti informazioni diagnostiche.



## 6. CONCLUSIONI

Il progetto sviluppato in questa tesi di dottorato si è focalizzato sull'utilizzo del *Nanoporous Silica Chip (NSC)*, un dispositivo prototipo, per l'estrazione selettiva e la purificazione del peptidoma da campioni biologici complessi, quali il plasma. Per gli obiettivi preposti si è reso necessario lo sviluppo e la standardizzazione di vari prototipi di NSC, nonché la valutazione della nanostruttura più adatta al tipo di campione usato in questo studio. Il protocollo sviluppato ha seguito criteri di semplicità, velocità e specificità nella selezione della frazione a basso peso molecolare del plasma.

Lo studio del profilo peptidico applicato a campioni di soggetti con cancro coloretale è stato mirato ad approfondire ed identificare eventuali candidati peptidici come marcatori di diagnosi e di progressione tumorale.

I profili peptidici ottenuti hanno permesso di eseguire un'analisi di predizione di classificazione, che ha condotto a distinguere e raggruppare correttamente un gruppo di soggetti sani rispetto al gruppo dei soggetti malati, sia con adenoma (lesione pre-cancerosa, ad alto rischio di sviluppare neoplasia), sia con tumore al colon conclamato a diverso stadio. Alcune delle specie ioniche che sono risultate essere differenzialmente espresse nei gruppi di studio sono state identificate e sono riconducibili alla Bradichinina ed ai frammenti del Complemento C3f e C4-A/B.

I risultati ottenuti pongono nuovi obiettivi da approfondire riguardo alla presenza dei *pattern* proteolitici ottenuti, suggerendo la presenza ed il coinvolgimento di proteasi specifiche del tumore.

Di sicuro interesse è l'identificazione di un frammento di ITIH4 nel plasma dei campioni patologici. Se da un lato è nota l'associazione di alcuni frammenti di ITIH4 con diversi stati patologici, dall'altro, lo specifico peptide da noi osservato (una porzione di 20 amminoacidi) non è mai stato descritto in associazione al cancro coloretale. Doveroso è quindi approfondire lo studio del ruolo ricoperto da questa proteina, espressa esclusivamente dal fegato, nella cancerogenesi e nello sviluppo del cancro coloretale, nonché della eventuale implicazione nello sviluppo di metastasi epatiche.

Considerando tutti i risultati ottenuti in questo progetto, la metodica sviluppata sembra essere molto promettente per le analisi del profilo peptidico nei campioni di plasma.

Il più ampio potenziale applicativo dei prototipi di NSC testati è sicuramente quello di poter essere utilizzato per l'estrazione e la purificazione di peptidi in diverse tipologie di campioni sia di fluidi biologici (come siero, plasma e liquor) sia di estratti cellulari e tissutali.

## 7. BIBLIOGRAFIA

1. Fiocca Silvio. Fondamenti di Anatomia e Fisiologia Umana. Ed. Sorbona, 1991.
2. Mc Clintic, J. Robert. Fisiologia del corpo umano. Ed. Zanichelli, 1994.
3. Parkin D. M., Bray F., Ferlay J., and Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin*, 55: 74-108, 2005.
4. Ferlay J., Autier P., Boniol M., Heanue M., Colombet M., and Boyle P. Estimates of the cancer incidence and mortality in Europe in 2006. *Ann Oncol*, 18: 581-592, 2007.
5. AIOM, CCM, AIRTUM. I numeri del cancro in Italia, 2013.
6. Chan A.T., Giovannucci E.L., Primary Prevention of Colorectal Cancer. *Gastroenterology*, 138: 2029–2043, 2010.
7. Freeman HJ. Colorectal cancer risk in Crohn disease. *World J Gastroenterol*, 14: 1810-11, 2008.
8. Mellekjær L, Olsen JH, Frisch M, et al. Cancer in ulcerative colitis. *Int J Cancer*, 60: 330-33, 1995.
9. Czene K, Lichtenstein P, Hemminki K. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *Int J Cancer*; 99: 260-66, 2002.
10. Butterworth AS, Higgs JP, Pharoah P. Relative and absolute risk of colorectal cancer for individuals with a family history: a meta-analysis. *Eur J Cancer*; 42: 216-27, 2006.
11. Jaspersion KW, Tuohy Tm, Neklason DW, et al. Hereditary and familial colon cancer. *Gastroenterology*; 138: 2044-58, 2010.
12. Reya, T. and Clevers, H. Wnt signalling in stem cells and cancer. *Nature*, 434: 843-850, 2005.
13. Clevers, H. Wnt/beta-catenin signaling in development and disease. *Cell*, 127: 469-480, 2006.
14. Humphries, A. and Wright, N. A. Colonic crypt organization and tumorigenesis. *Nat Rev Cancer*, 8: 415-424, 2008.
15. Ricci-Vitiani L, Fabrizio E, Palio E, De Maria R. Colon cancer stem cells. *J Mol Med*, 87: 1097-1104, 2009.
16. Kauh J, Umbreit J. Colorectal cancer prevention. *Curr Probl Cancer*, 28: 240-264, 2004.
17. Fodde R, Smits R, Clevers H. APC, signal transduction and genetic instability in colorectal cancer. *Nat Rev Cancer*, 1: 55-67, 2001.

18. Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, Nakamura Y, White R, Smits AM, Bos JL. Genetic alterations during colorectal-tumor development. *N Engl J Med*, 319: 525-532, 1988.
19. Roberts PJ, Der CJ. Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene*, 26: 3291-3310, 2007.
20. Forcet C, Ye X, Granger L, Corset V, Shin H, Bredesen DE, Mehlen P. The dependence receptor DCC (deleted in colorectal cancer) defines an alternative mechanism for caspase activation. *Proc Natl Acad Sci USA*, 98: 3416-3421, 2001.
21. Vogelstein B, Lane D, Levine AJ. Surfing the p53 network. *Nature*, 408: 307-310, 2000.
22. Vousden KH. Activation of the p53 tumor suppressor protein. *Biochim Biophys Acta*, 1602: 47-59, 2002.
23. Levin B, Lieberman DA, McFarland B, Smith RA, Brooks D, Andrews KS, Dash C, Giardiello FM, Glick S, Levin TR, Pickhardt P, Rex DK, Thorson A, Winawer SJ. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *CA Cancer J Clin*, 58: 130-160, 2008.
24. Lieberman DA. Clinical practice. Screening for colorectal cancer. *N Engl J Med*, 361: 1179-1187, 2009.
25. Cunningham D, Atkin W, Lenz HJ, Lynch HT, Minsky B, Nordlinger B, Starling N. Colorectal cancer. *Lancet*, 375: 1030-1047, 2010.
26. Weitz J, Koch M, Debus J, Hohler T, Galle PR, Buchler MW. Colorectal cancer. *Lancet*, 365: 153-165, 2005.
27. Locker GY, Hamilton S, Harris J, Jessup JM, Kemeny N, Macdonald JS, Somerfield MR, Hayes DF, Bast RC Jr. ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. *J Clin Oncol*, 24: 5313-5327, 2006.
28. Duffy MJ, van Dalen A, Haglund C, Hansson L, Holinski-Feder E, Klapdor R, Lamerz R, Peltomaki P, Sturgeon C, Topolcan O. Tumour markers in colorectal cancer: European Group on Tumour Markers (EGTM) guidelines for clinical use. *Eur J Cancer*, 43: 1348-1360, 2007.
29. Cappell MS. From colonic polyps to colon cancer: pathophysiology, clinical presentation, screening and colonoscopic therapy. *Minerva Gastroenterol Dietol*, 53: 351-373, 2007.
30. Graves PR, Haystead TA. Molecular biologist's guide to proteomics. *Microbiol Mol Biol Rev*, 66: 39-63; table of contents, 2002.
31. Anderson L, Seilhamer J. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis*, 18: 533-537, 1997.

32. Kulasingam V, Diamandis EP. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nat Clin Pract Oncol*, 5: 588-599, 2008.
33. Kolch W, Mischak H, Pitt AR. The molecular make-up of a tumour: proteomics in cancer research. *Clin Sci (Lond)*, 108: 369-383, 2005.
34. Hanash SM, Pitteri SJ, Faca VM. Mining the plasma proteome for cancer biomarkers. *Nature*, 452: 571-579, 2008.
35. Latterich M, Abramovitz M, Leyland-Jones B. Proteomics: new technologies and clinical applications. *Eur J Cancer*, 44: 2737-2741, 2008.
36. Tainsky MA. Genomic and proteomic biomarkers for cancer: a multitude of opportunities. *Biochim Biophys Acta*, 1796: 176-193, 2009.
37. Wouters BG. Proteomics: methodologies and applications in oncology. *Semin Radiat Oncol*, 18: 115-125, 2008.
38. Callesen AK, Madsen JS, Vach W, Kruse TA, Mogensen O, Jensen ON. Serum protein profiling by solid phase extraction and mass spectrometry: a future diagnostics tool? *Proteomics*, 9: 1428-1441, 2009.
39. Petricoin EF, Liotta LA. Clinical applications of proteomics. *J Nutr*, 133: 2476S-2484S, 2003.
40. Shau H, Chandler GS, Whitelegge JP, Gornbein JA, Faull KF, Chang HR. Proteomic profiling of cancer biomarkers. *Brief Funct Genomic Proteomic*, 2: 147-158, 2003.
41. Geho DH, Luchini A, Garaci E, Belluco C, Petricoin EF, Liotta LA. Nanotechnology in clinical proteomics. *Nanomedicine (Lond)*, 2: 1-5, 2007.
42. Petricoin EF, Belluco C, Araujo RP, Liotta LA. The blood peptidome: a higher dimension of information content for cancer biomarker discovery. *Nature Reviews*, 6: 961-967, 2006.
43. Chertov O, Biragyn A, Kwak LW, Simpson JT, Boronina T, Hoang VM, Prieto DA, Conrads TP, Veenstra TD, Fisher RJ. Organic solvent extraction of proteins and peptides from serum as an effective sample preparation for detection and identification of biomarkers by mass spectrometry. *Proteomics*, 4: 1195-1203, 2004.
44. Villanueva J, Philip J, Entenberg D, Chaparro CA, Tanwar MK, Holland EC, Tempst P. Serum peptide profiling by magnetic particle-assisted, automated sample processing and MALDI-TOF mass spectrometry. *Anal Chem*, 76: 1560-1570, 2004.
45. Kay R, Barton C, Ratcliffe L, Matharoo-Ball B, Brown P, Roberts J, Teale P, Creaser C. Enrichment of low molecular weight serum proteins using acetonitrile precipitation for mass spectrometry based proteomic analysis. *Rapid Commun Mass Spectrom*, 22: 3255-3260, 2008.

46. Rose K, Bougueleret L, Baussant T, Bohm G, Botti P, Colinge J, Cusin I, Gaertner H, Gleizes A, Heller M, Jimenez S, Johnson A, Kussmann M, Menin L, Menzel C, Ranno F, Rodriguez-Tome P, Rogers J, Saudrais C, Villain M, Wetmore D, Bairoch A, Hochstrasser D. Industrial-scale proteomics: from liters of plasma to chemically synthesized proteins. *Proteomics*, 4: 2125-2150, 2004.
47. Marshall J, Jankowski A, Furesz S, Kireeva I, Barker L, Dombrovsky M, Zhu W, Jacks K, Ingratta L, Bruin J, Kristensen E, Zhang R, Stanton E, Takahashi M, Jackowski G. Human serum proteins pre-separated by electrophoresis or chromatography followed by tandem mass spectrometry. *J Proteome Res*, 3: 364-382, 2004.
48. Moritz RL, Skandarajah AR, Ji H, Simpson RJ. Proteomic analysis of colorectal cancer: prefractionation strategies using two-dimensional free-flow electrophoresis. *Comp Funct Genomics*, 6: 236-243, 2005.
49. Chernokalskaya E, Gutierrez S, Pitt AM, Leonard JT. Ultrafiltration for proteomic sample preparation. *Electrophoresis*, 25: 2461-2468, 2004.
50. Tirumalai RS, Chan KC, Prieto DA, Issaq HJ, Conrads TP, Veenstra TD. Characterization of the low molecular weight human serum proteome. *Mol Cell Proteomics*, 2: 1096-1103, 2003.
51. Orvisky E, Drake SK, Martin BM, Abdel-Hamid M, Resson HW, Varghese RS, An Y, Saha D, Hortin GL, Loffredo CA, Goldman R. Enrichment of low molecular weight fraction of serum for MS analysis of peptides associated with hepatocellular carcinoma. *Proteomics*, 6: 2895-2902, 2006.
52. Greening DW, Simpson RJ. Low molecular weight plasma proteome analysis using centrifugal ultrafiltration. *Methods Mol Biol*, 728: 109-124, 2011.
53. Ahmed FE. Sample preparation and fractionation for proteome analysis and cancer biomarker discovery by mass spectrometry. *J Sep Sci*, 32: 771-798, 2009.
54. Ivanov YD, Govorun VM, Bykov VA, Archakov AI. Nanotechnologies in proteomics. *Proteomics*, 6: 1399-1414, 2006.
55. Geho D, Cheng MM, Killian K, Lowenthal M, Ross S, Frogale K, Nijdam J, Lahar N, Johann D, Herrmann P, Whiteley G, Ferrari M, Petricoin E, Liotta L. Fractionation of serum components using nanoporous substrates. *Bioconjug Chem*, 17: 654-661, 2006.
56. Terracciano R, Gaspari M, Testa F, Pasqua L, Tagliaferri P, Cheng MM, Nijdam AJ, Petricoin EF, Liotta LA, Cuda G, Ferrari M, Venuta S. Selective binding and enrichment for low-molecular weight biomarker molecules in human plasma after exposure to nanoporous silica particles. *Proteomics*, 6: 3243-3250, 2006.
57. Johnson CJ, Zhukovsky N, Cass AE, Nagy JM. Proteomics, nanotechnology and molecular diagnostics. *Proteomics*, 8: 715-730, 2008.

58. Gaspari M, Ming-Cheng M, Terracciano R, Liu X, Nijdam AJ, Vaccari L, di Fabrizio E, Petricoin EF, Liotta LA, Cuda G, Venuta S, Ferrari M. Nanoporous surfaces as harvesting agents for mass spectrometric analysis of peptides in human plasma. *J Proteome Res*, 5: 1261-1266, 2006.
59. Chen H, Xu X, Yao N, Den C, Yang P, Zhang X. Facile synthesis of C8-functionalized magnetic silica microspheres for enrichment of low-concentration peptides for direct MALDI-TOF MS analysis. *Proteomics*, 8: 2778-2784, 2008.
60. Terracciano R, Casadonte F, Pasqua L, Candeloro P, Di Fabrizio E, Urbani A, and Savino R Enhancing plasma peptide MALDI-TOF-MS profiling by mesoporous silica assisted crystallization. *Talanta*, 80: 1532-1538, 2010.
61. Bouamrani A, Hu Y, Tasciotti E, Li L, Chiappini C, Liu X, Ferrari M. Mesoporous silica chips for selective enrichment and stabilization of low molecular weight proteome. *Proteomics*, 10: 496-505, 2010.
62. Preianò M, Pasqua L, Gallelli L, Galasso O, Gasparini G, Savino R, Terracciano R. Simultaneous extraction and rapid visualization of peptidomic and lipidomic body fluids fingerprints using mesoporous aluminosilicate and MALDI-TOF MS. *Proteomics*, 12: 3286–3294, 2012.
63. Hu Y, Bouamrani A, Tasciotti E, Li L, Liu X, Ferrari M. Tailoring of the nanotexture of mesoporous silica films and their functionalized derivatives for selectively harvesting low molecular weight protein. *ACSNANO*, 4 (1): 439–451, 2010.
64. Xia J, Psychogios N, Young N, Wishart D.S. MetaboAnalyst: a web server for metabolomics data analysis and interpretation. *Nucleic Acids Research*, 37: 652-660, 2009.
65. Xia J, Mandal R, Sinelnikov I.V., Broadhurst D, Wishart D.S. MetaboAnalyst 2.0 - a comprehensive server for metabolomic data analysis. *Nucleic Acids Research*, 40: 127-133, 2012.
66. Yi J, Kim C, Gelfand CA. Inhibition of intrinsic proteolytic activities moderates preanalytical variability and instability of human plasma. *J Proteome Res*, 6: 1768-1781, 2007.
67. Lopez-Otin C, Bond JS. Proteases: multifunctional enzymes in life and disease. *J Biol Chem*, 283: 30433-304037, 2008.
68. Joyce JA, Pollard JW. Microenvironmental regulation of metastasis. *Nat Rev Cancer*, 9: 239-252, 2009.
69. Lopez-Otin C, Hunter T. The regulatory crosstalk between kinases and proteases in cancer. *Nat Rev Cancer*, 10, 278-292, 2010.
70. Mason SD, Joyce JA. Proteolytic networks in cancer. *Trends Cell Biol*, 21,228-237, 2011.
71. Hua H, Li M, Luo T, Yin Y, Jiang Y. Matrix metalloproteinases in tumorigenesis: an evolving paradigm. *Cell Mol Life Sci*, 68, 23, 3853-3868, 2011.

72. Ghavami S, Hashemi M, Ande SR, Yeganeh B, Xiao W, Eshraghi M, Bus CJ, Kadkhoda K, Wiechec E, Halayko AJ, Los M. Apoptosis and cancer: mutations within caspase genes. *J Med Genet*, 46, 8, 497-510, 2009.
73. Findeisen P, Neumaier M, Fuctional protease profiling for diagnosis of malignant disease. *Proteomics Clin Appl*, 6, 60-78, 2012.
74. Villanueva J, Shaffer RD, Philip J, Chaparro CA, Erdjument-Bromage H, Olshen AB, Fleisher M, Lilja H, Brogi E, Boyd J, Sanchez-Carbayo M, Holland EC, Cordon-Cardo C, Scher HI, Tempst P. Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J Clin Invest*, 116, 1, 271-84, 2006.
75. Findeisen P, Costina V, Yepes D, Hofheinz R, Neumaier M. Fuctional protease profiling with reporter peptides in serum specimens of colorectal cancer patients: demonstration of its routine diagnostic applicability. *J Exp Clin Cancer Res*, 8: 31-56, 2012.
76. Overall CM, Blobel CP. In search of partners: linking extracellular proteases to substrate. *Nat Rev Mol Cell Biol* 8,245-257, 2007; Diamond SL. Methods for mapping protease specificity. *Curr Opin Chem Biol*, 11: 46-51, 2007.
77. Kashyap RS, Nayak AR, Deshpande PS, Kabra D, Purohit HJ, Taori GM, Dagainawala HF. Inter-alpha-trypsin inhibitor heavy chain 4 is a novel marker of acute ischemic stroke. *Clin Chim Acta*, 402: 1-2, 160-163, 2009.
78. van den Broek I, Sparidans RW, Schellens JH, Beijnen JH. Specific Investigation of Sample Handling Effects on Protease Activities and Absolute Serum Concentrations of Various Putative Peptidome Cancer Biomarkers. *Clin Proteomics*, 6 (4): 115-127, 2010.
79. Song J, Patel M, Rosenzweig CN, Chan-Li Y, Sokoll LJ, Fung ET, Choi-Miura NH, Goggins M, Chan DW, Zhang Z. Quantification of fragments of human serum inter-alpha-trypsin inhibitor heavy chain 4 by a surface-enhanced laser desorption/ionization-based immunoassay. *Clin Chem*, 52 (6): 1045-1053, 2006.



## 8. PUBBLICAZIONI

1. Front Biosci (Schol Ed). 2014 Jan 1;6:110-9.

### **Predictive response biomarkers in rectal cancer neoadjuvant treatment.**

Agostini M, Crotti S, Bedin C, Cecchin E, Maretto I, D'Angelo E, Pucciarelli S, Nitti D.

#### **Abstract**

Locally advanced rectal cancer (RC) treatment is a challenge, because RC has a high rate of local recurrence. To date preoperative chemoradiotherapy (pCRT) is widely accepted as standard protocol of care for middle-low RC, but complete tumour response rate ranges from 4 to 44% and 5-year local recurrence rate is 6%. Better understanding of molecular biology and carcinogenesis pathways could be used both for pre-neoplastic lesions and locally recurrence diagnosis, and for tumour response prediction to therapy. Circulating molecules, gene expression and protein signature are promising sources to biomarker discovery. Several studies have evaluated potential predictors of response and recently, cell-free Nucleic Acid levels have been associated to tumour response to neoadjuvant therapies. Alternative method is the serum or plasma proteome and peptidome analysis. It may be ideally suited for its minimal invasiveness and it can be repeated at multiple time points throughout the treatment in contrast to tissue-based methods which still remain the most reliable and specific approach. Many studies have analyzed preoperative rectal tissue prognostic factor, but data are controversial or not confirmed.

PMID: 24389265 [PubMed - in process]

2. Dis Markers. 2013;34(4):269-78. doi: 10.3233/DMA-130971.

**Multivariate analysis approach to the serum peptide profile of morbidly obese patients.**

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.

**Abstract**

**BACKGROUND:** Obesity is currently epidemic in many countries worldwide and is strongly related to diabetes and cardiovascular disease. Mass spectrometry, in particular matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) is currently used for detecting different pattern of expressed protein. This study investigated the differences in low molecular weight (LMW) peptide profiles between obese and normal-weight subjects in combination with multivariate statistical analysis.

**MATERIALS:** Serum samples of 60 obese patients and 10 healthy subjects were treated by cut-off membrane (30000 Da) to remove the most abundant proteins. The filtrates containing the LMW protein/peptides were analyzed by MALDI-TOF mass spectrometry. Dataset was elaborated to align and normalize the spectra. We performed cluster analysis and principal component analysis to detect some ionic species that could characterize and classify the subject groups.

**RESULTS:** We observed a down-expression of ionic species at m/z 655.94 and an over-expression of species at m/z 1518.78, 1536.77, 1537.78 and 1537.81 in obese patients. Furthermore we found some ionic species that can distinguish obese patients with diabetes from those with normal glucose level.

**CONCLUSION:** Serum peptide profile of LMW associate with multivariate statistical approach was revealed as a promising tool to discriminate and characterize obese patients and it was able to stratify them in relation to comorbidity that usually are associated with this disease. Further research involving a larger sample will be required to validate these findings.

PMID: 23396294 [PubMed - indexed for MEDLINE]

PMCID: PMC3810107 Free PMC Article

3. Mol Biol Rep. 2012 Oct;39(10):9307-10. doi: 10.1007/s11033-012-1729-2. Epub 2012 Jul 11.

**Soft tissue sarcoma and the hereditary non-polyposis colorectal cancer (HNPCC) syndrome: formulation of an hypothesis.**

Urso E, Agostini M, Pucciarelli S, Bedin C, D'angelo E, Mescoli C, Viel A, Maretto I, Mammi I, Nitti D.

**Abstract**

Hereditary non-polyposis colorectal cancer (HNPCC) is a genetic disorder caused by mutation in one of the mismatch repair (MMR) genes (MLH1, MSH2, MSH6, PMS2) which predisposes to colorectal cancer and other malignances, that not yet include sarcomas. For sustaining that soft tissue sarcomas could be HNPCC related malignances, we report on a HNPCC patient with leiomyosarcoma and review the English literature. Overall, we report on eleven cases of soft tissue malignant tumors involving HNPCC patients, with a mean age of 34 years at diagnosis of sarcomas. In the majority of these tumors loss of MSH2 expression can be found at immunohistochemistry (IHC) and in 10 patients a germline mutation in one of the MMR genes was found (7 cases were MSH2 defective and 3 cases MLH1 defective). Data for supporting our hypothesis are also experimental, epidemiologic, histopathological: excess of sarcomas in PMS2 defective mice; sporadic soft tissue sarcomas are rare, with mean age at onset of 56 years and normal IHC for MMR proteins. In conclusion, the data collected support the hypothesis that soft tissue sarcomas could be included in the spectrum of tumors that, even if rarely, depend on MMR genes deficiency.

PMID: 22782591 [PubMed - indexed for MEDLINE]

4. Histopathology. 2012 Aug;61(2):247-56. doi: 10.1111/j.1365-2559.2012.04217.x. Epub 2012 Mar 14.

**Survivin and laryngeal carcinoma prognosis: nuclear localization and expression of splice variants.**

Marioni G, Agostini M, **Bedin C**, Blandamura S, Stellini E, Favero G, Lionello M, Giacomelli L, Burti S, D'Angelo E, Nitti D, Staffieri A, De Filippis C.

**Abstract**

**AIMS:** Aberrant survivin expression in cancer cells has been associated with tumour progression, radiation/drug resistance and shorter patient survival. The aim of the present study was to investigate survivin expression in laryngeal carcinoma (LSCC) tissue and - for the first time at this site - the expression of survivin splice variants. P53 was also studied.

**METHODS AND RESULTS:** Survivin and p53 expression was determined immunohistochemically in 86 consecutive patients operated for LSCC. Survivin mRNA expression was assessed by quantitative real-time polymerase chain reaction (PCR). Hot-spot mutations in exons 5, 6, 7 and 8 of the TP53 gene were studied by sequencing analysis. A nuclear localization for survivin predominated. There was a significant association between a higher nuclear survivin expression and LSCC recurrence (P = 0.046). Disease-free survival (DFS) for LSCC patients with a nuclear survivin expression >7.0% was shorter than in cases whose expression was ≤7.0% (P = 0.05). Wild-type survivin correlated significantly with nuclear survivin expression (P = 0.02). p53 expression was associated with the co-expression of wild-type survivin and survivin-2B (P = 0.01).

**CONCLUSIONS:** Nuclear expression of survivin appears to influence LSCC aggressiveness, a higher nuclear survivin expression correlating with a higher recurrence rate and a shorter DFS. Wild-type survivin was the most frequently detected splice variant in LSCC tissues.

© 2012 Blackwell Publishing Ltd.

PMID: 22416874 [PubMed - indexed for MEDLINE]

5. Tumour Biol. 2012 Jun;33(3):857-64. doi: 10.1007/s13277-011-0312-0. Epub 2012 Jan 26.

**Clinical and molecular detection of inherited colorectal cancers in northeast Italy: a first prospective study of incidence of Lynch syndrome and MUTYH-related colorectal cancer in Italy.**

Urso E, Agostini M, Pucciarelli S, Rugge M, Bertorelle R, Maretto I, **Bedin C**, D'Angelo E, Mescoli C, Zorzi M, Viel A, Bruttocao G, Ferraro B, Erroi F, Contin P, De Salvo GL, Nitti D.

**Abstract**

The reported incidence of hereditary colorectal cancers (CRCs) is widely variable. The principal aim of the study was to prospectively evaluate the incidence of familial CRCs in a region of northern Italy using a standardized method. Consecutive CRC patients were prospectively enrolled from October 2002 to December 2003. Patients underwent a structured family history, the microsatellite instability (MSI) test and a screen for MUTYH mutations. Following family history patients were classified as belonging to high, moderate and mild risk families. Immunohistochemistry for MLH1, MSH2, MSH6 and PMS2 proteins and investigation for MLH1/MSH2 mutations, for MLH1 promoter methylation and for the V600E hotspot BRAF mutation were performed in high MSI (MSI-H) cases. Of the 430 patients enrolled, 17 (4%) were high risk [4 hereditary non-polyposis colorectal cancer (HNPCC), 12 suspected HNPCC and 1 MUTYH-associated adenomatous polyposis coli (MAP)], 53 moderate risk and 360 mild risk cases. The MSI test was performed on 393 tumours, and 46 (12%) of them showed MSI-H. In these patients, one MLH1 pathogenetic mutations and two MSH2 pathogenetic mutations were found. Thirty-two (70%) MSI-H cases demonstrated MLH1 methylation and/or BRAF mutation: None of them showed MLH1/MSH2 mutation. Two biallelic germline MUTYH mutations were found, one with clinical features of MAP. A strong family history of CRC was present in 4% of the enrolled cases; incidence of MLH1/MSH2 or MUTHY mutations was 1.3% and of MSI-H phenotype was 12%. MLH1 methylation and BRAF mutation can exclude 70% of MSI-H cases from gene sequencing.

PMID: 22278153 [PubMed - indexed for MEDLINE]

6. Int J Biol Markers. 2012 Jan-Mar;27(1):13-9. doi: 10.5301/JBM.2011.8908.

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

Agostini M, Bedin C, Pucciarelli S, Enzo M, 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.

**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.

PMID: 22180177 [PubMed - indexed for MEDLINE]

7. Cancer Biomark. 2012;11(2-3):89-98.

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

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.

**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 woman 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  $\beta$ -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 247/115 and 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.

PMID: 23011155 [PubMed - indexed for MEDLINE]

8. Ann Surg Oncol. 2011 Sep;18(9):2461-8. doi: 10.1245/s10434-011-1638-y. Epub 2011 Mar 17.

**Circulating cell-free DNA: a promising marker of pathologic tumor response in rectal cancer patients receiving preoperative chemoradiotherapy.**

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.

**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  $\beta$ -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.

PMID: 21416156 [PubMed - indexed for MEDLINE]



9. Genet Med. 2011 Feb;13(2):115-24. doi: 10.1097/GIM.0b013e3182011489.

**Integrated analysis of unclassified variants in mismatch repair genes.**

Pastrello C, Pin E, Marroni F, **Bedin C**, Fornasarig M, Tibiletti MG, Oliani C, Ponz de Leon M, Urso ED, Della Puppa L, Agostini M, Viel A.

**Abstract**

**PURPOSE:** Lynch syndrome is a genetic disease that predisposes to colorectal tumors, caused by mutation in mismatch repair genes. The use of genetic tests to identify mutation carriers does not always give perfectly clear results, as happens when an unclassified variant is found. This study aimed to define the pathogenic role of 35 variants present in MSH2, MLH1, MSH6, and PMS2 genes identified in our 15-year case study.

**METHODS:** We collected clinical and molecular data of all carriers, and then we analyzed the variants pathogenic role with web tools and molecular analyses. Using a Bayesian approach, we derived a posterior probability of pathogenicity and classified each variant according to a standardized five-class system.

**RESULTS:** The MSH2 p.Pro349Arg, p.Met688Arg, the MLH1 p.Gly67Arg, p.Thr82Ala, p.Lys618Ala, the MSH6 p.Ala1236Pro, and the PMS2 p.Arg20Gln were classified as pathogenic, and the MSH2 p.Cys697Arg and the PMS2 p.Ser46Ile were classified as likely pathogenic. Seven variants were likely nonpathogenic, 3 were nonpathogenic, and 16 remained uncertain.

**CONCLUSION:** Quantitative assessment of several parameters and their integration in a multifactorial likelihood model is the method of choice for classifying the variants. As such classifications can be associated with surveillance and testing recommendations, the results and the method developed in our study can be useful for helping laboratory geneticists in evaluation of genetic tests and clinicians in the management of carriers.

PMID: 21239990 [PubMed - indexed for MEDLINE]

## **8.1. Sottomissioni a revisione editoriale**

1. JOURNAL: Clinical Chemistry  
MS ID#: CLINCHEM/2014/221655  
MS TITLE: "Circulating peptides in rectal cancer patients underwent neoadjuvant chemoradiotherapy"  
AUTHORS: Sara Crotti, Maria Vittoria Enzo, **Chiara Bedin**, Salvatore Pucciarelli, Isacco Maretto, Paola Del Bianco, Cosimo De Filippis, Pietro Traldi, Ennio Tasciotti, Mauro Ferrari, Donato Nitti
2. JOURNAL: IEEE/ACM Transactions on Computational Biology and Bioinformatics  
MS ID#: TCBB-2013-10-0308  
MS TITLE: "Proteomic Signature Discovery From Mass Spectrometry Data"  
AUTHORS: Kong Ao; Gupta Chinmaya; Ferrari Mauro; Agostini Marco; **Bedin Chiara**; Bouamrani Ali; Tasciotti Ennio; Azencott Robert
3. JOURNAL: Journal of Molecular Biology  
MS ID#: MOLE-8176  
MS TITLE: "Integrative computational biology and molecular determinants of rectal cancer resistance to chemoradiotherapies: XRCC3 a crossroads for the chemosensitivity?"  
AUTHORS: Marco Agostini, Andrea Zangrando, Chiara Pastrello, Giuseppe Romano, Roberto Giovannoni, Alessandro Ambrosi, Claudia Gaspar, Isacco Maretto, **Chiara Bedin**, Maura Digito, Gianni Esposito, Edoardo D'Angelo, Maria Luisa Friso, Sara Lonardi, Claudia Mescoli, Riccardo Fodde, Giuseppe Basso, Marialuisa Lavitrano, Flavio Rizzolio, Igor Jurisica, Salvatore Pucciarelli, Donato Nitti