

UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Scuola di Medicina e Chirurgia

Dottorato di Ricerca in Tecnologie Biomediche XXVIII Ciclo

Coordinatore: Chiar.ma Prof.ssa Marina DEL PUPPO



**RICERCA DI UNA FIRMA MOLECOLARE TIPIZZANTE
TUMORI RENALI MEDIANTE APPROCCIO PROTEOMICO DI
PROFILING IN SPETTROMETRIA DI MASSA.**

Tutor: Chiar.mo Prof. Fulvio MAGNI

Tesi di:

Dr.ssa Marta Cazzaniga

matricola n° 708611

Anno Accademico 2014/2015

INDICE

1. INTRODUZIONE	1
1.1. I tumori renali	1
1.1.1. I tumori benigni	3
1.1.2. I tumori maligni	4
1.1.3. Il carcinoma a cellule renali (RCC)	5
1.1.3.1. Epidemiologia	5
1.1.3.2. Eziologia	6
1.1.3.3. Classificazione	7
1.1.3.4. Citogenetica	10
1.1.3.5. Diagnosi	12
1.1.3.6. Stadiazione	13
1.1.3.7. Prognosi	15
1.1.3.8. Strategie di intervento	16
1.1.3.9. Patogenesi del carcinoma renale a cellule chiare (ccRCC)	17
1.2. La Proteomica	20
1.2.1. Proteomica clinica	22
1.2.2. <i>Protein Profiling</i>	23
1.2.3. Il peptidoma	26
1.2.4. Urine	28
1.2.5. Peptidoma/proteoma urinario del RCC	30
1.2.6. Proteomica quantitativa	33
1.2.6.1. Tecniche di quantificazione senza marcatura o <i>label-free</i>	35
1.2.7. Proteomica integrativa	37
1.2.7.1. Integrazione di campioni biologici	37
1.2.7.2. Integrazione di tecniche	39
1.3. <i>Data Mining</i>	39
1.3.1. <i>Biomarker discovery</i>	40
1.3.2. Sistemi di classificazione	43
1.3.2.1. Risultati dei modelli di classificazione	45
1.4. La spettrometria di massa	49
1.4.1. Sorgente ionica	51

1.4.1.1.	Ionizzazione per desorbimento laser assistita da matrice (MALDI)	51
1.4.1.2.	Ionizzazione <i>elettrospray</i> (ESI)	52
1.4.2.	Analizzatore	53
1.4.2.1.	Analizzatori quadrupolari	54
1.4.2.2.	Analizzatori a trappola ionica (<i>ion trap</i>)	54
1.4.2.3.	Analizzatori a tempo di volo	55
1.4.2.4.	Analizzatori Qq-TOF	55
1.4.2.5.	Analizzatori <i>ultra high resolution time of flight</i> (UHR-TOF)	55
1.4.3.	Rivelatore	56
1.4.4.	La spettrometria di massa MALDI-TOF	57
1.4.5.	La spettrometria di massa tandem (MS/MS)	60
1.4.6.	Lo spettro di massa	62
1.5.	<i>High performance liquid chromatography</i> (HPLC)	63
1.5.1.	Nano-cromatografia liquida (nano-LC)	63
1.5.2.	Cromatografia liquida a prestazioni ultra elevate (UHPLC)	65
1.5.3.	La cromatografia liquida associata alla spettrometria di massa	66
2.	SCOPO	68
3.	MATERIALI E METODI	70
3.1.	Disegno sperimentale	70
3.2.	Raccolta e processamento dei campioni di urine	71
3.3.	Analisi del peptidoma urinario	71
3.3.1.	Estrazione del peptidoma urinario mediante biglie magnetiche	71
3.3.2.	Deposito su target ed analisi MALDI-TOF	72
3.3.3.	Processamento dei profili spettrali	73
3.3.4.	Analisi statistica dei dati sperimentali	73
3.3.5.	Identificazione dei peptidi endogeni mediante tecnica nLC-ESI-MS/MS	78
3.4.	Analisi del proteoma urinario	80
3.4.1.	Digestione enzimatica con metodica FASP	80
3.4.2.	Analisi nLC-ESI-MS/MS	81
3.4.3.	Quantificazione relativa <i>label-free</i>	82
3.5.	Validazione di una proteina di interesse mediante <i>western blot</i>	85

4. RISULTATI	89
4.1. Dati clinici	89
4.2. Analisi del peptidoma urinario	90
4.2.1. Analisi dei profili peptidici e costruzione di <i>cluster</i> diagnostici	90
4.2.1.1. <i>Cluster</i> discriminante tra ctrl <i>vs</i> ccRCC	98
4.2.1.2. <i>Cluster</i> discriminante tra tumori renali maligni <i>vs</i> masse renali benigne e controlli	100
4.2.2. Analisi dei profili peptidici e individuazione di segnali correlati alla progressione tumorale	104
4.2.2.1. Variazioni peptidiche secondo la dimensione tumorale	104
4.2.2.2. Variazioni peptidiche secondo lo stadio tumorale	105
4.2.2.3. Alterazioni peptidiche rispetto al grado tumorale	106
4.2.2.4. Peptidi urinari differentemente espressi rispetto ai dati istologici tumorali	106
4.2.3. Identificazione dei peptidi endogeni urinari	108
4.2.3.1. Identificazione dei segnali presenti nei <i>cluster</i> discriminanti	109
4.2.3.2. Identificazione dei peptidi correlati alla progressione tumorale	116
4.3. Analisi del proteoma urinario	119
4.3.1. Individuazione di alterazioni proteomiche	119
4.4. Integrazione delle informazioni peptidomiche e proteomiche	122
4.5. Validazione dei risultati ottenuti mediante <i>western blot</i>	124
5. DISCUSSIONE	125
6. CONCLUSIONI	135
7. BIBLIOGRAFIA	137

1. INTRODUZIONE

1.1. I TUMORI RENALI

Nel rene possono svilupparsi tumori sia benigni che maligni. Queste due classi presentano varie differenze sotto diversi punti di vista: clinico, prognostico, diagnostico e terapeutico. L'incidenza e la mortalità di questa patologia variano in base al sesso (Figura 1).

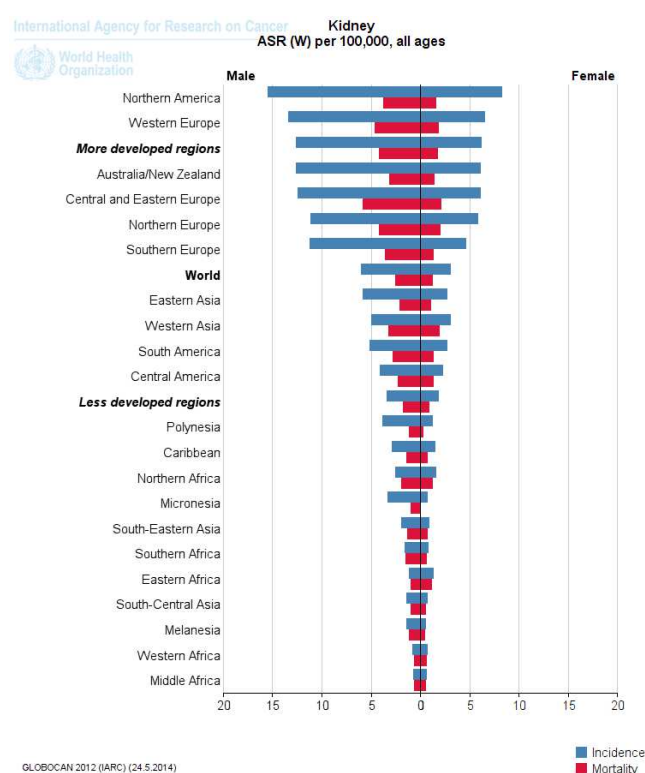


Figura 1. Incidenza e mortalità del tumore renale; ultimo aggiornamento nel 2012 [1].

È stato stimato che nel 2016 negli Stati Uniti sono stati diagnosticati più di 61000 nuovi casi di tumore renale e i decessi dovuti a questa patologia sono stati circa 14000 [2]. Negli ultimi decenni la progressiva diffusione della diagnostica per immagini ha determinato l'aumento delle diagnosi precoci di lesioni renali asintomatiche, con una ricaduta sui più importanti indicatori epidemiologici ed i loro andamenti nel tempo. La ricerca di una classificazione semplice e appropriata dei tumori renali ha impegnato anatomopatologi e urologi per tutto il secolo. Il problema di una corretta classificazione rimane attualmente aperto a causa della notevole varietà degli istotipi conosciuti e del continuo evolversi delle metodiche di studio adottate dagli anatomopatologi. Questo spinge ad una continua revisione della classificazione dei tumori renali. Attualmente le neoplasie del rene vengono classificate secondo il sistema redatto dalla World Health

Organisation (WHO) nel 2004 (Tabella 1) [3]. La classificazione WHO del 2004 definisce numerosi sottotipi di tumori, che presentano aspetti istopatologici, alterazioni genetiche e comportamenti clinici distinti. In questa classificazione sono inclusi sia tumori benigni sia tumori maligni ed un gruppo indicato come "carcinoma a cellule renali inclassificabile", nel quale sono incluse tutte le neoplasie renali che non rientrano in nessuna delle categorie ben definite (Tabella 1).

WHO: classificazione dei tumori renali	
Neoplasie a cellule renali	Neoplasie miste mesenchimali ed epiteliali
<u>Maligne</u>	Nefroma cistico
Carcinoma a cellule renali a cellule chiare	Tumore misto epiteliale e stromale
Carcinoma a cellule renali a cellule chiare multiloculare	Sarcoma sinoviale
Carcinoma a cellule renali papillare	
Carcinoma a cellule renali cromofobo	Neoplasie nefroblastiche
Carcinoma dei dotti collettori del Bellini	Residui nefrogenici
Carcinoma a cellule renali midollare	Nefroblastoma
Carcinoma con traslocazione Xp11	Nefroblastoma cistico
Carcinoma associato a neuroblastoma	
Carcinoma mucinoso a cellule tubulari e fusate	Neoplasie neuroendocrine
Carcinoma a cellule renali inclassificabile	Carcinoide
<u>Benigne</u>	Carcinoma neuroendocrino
Adenoma papillare	Tumore neuroectodermico primitivo
Oncocitoma	Neuroblastoma
Neoplasie metanefriche	Altri tumori
Adenoma metanefrico	Neoplasie mesenchimali
Adenofibroma metanefrico	Neoplasie ematopoietiche e linfoidi
Tumore stromale metanefrico	Neoplasie a cellule germinali
	Neoplasie metastatiche

Tabella 1. Classificazione anatomopatologica dei tumori del rene [3]

1.1.1. I TUMORI BENIGNI

I tumori benigni del rene possono essere divisi in due tipologie: neoformazioni di natura epiteliale e neoformazioni di natura connettivale. Della prima tipologia fanno parte le cisti, gli adenomi e gli oncocitomi renali; mentre tra le neoformazioni di natura connettivale sono da annoverare gli angiomiolipomi, i fibromi e i fibroangiomi. Complessivamente i tumori benigni renali non sono frequenti e la loro incidenza è compresa tra 3-7% [4] dei tumori renali. Possono insorgere da qualsiasi tipo di cellula presente nel rene, possono raggiungere grandi dimensioni e non sono sempre facilmente differenziabili dalle forme maligne anche con le moderne tecniche radiologiche. Nella maggior parte dei casi, la diagnosi delle forme benigne è data dall'esame istologico di un pezzo operatorio e quindi solo dopo intervento chirurgico. Solo nel caso dell'angiomiolipoma, le immagini all'ecografia e alla TAC sono caratteristiche.

Per anni l'**oncocitoma renale** è stato oggetto di pareri contrastanti riguardo alla sua benignità o malignità; la diatriba venne risolta nel 1985 quando fu chiaramente identificata la variante cromofoba del carcinoma a cellule renali e l'oncocitoma venne classificato come tumore benigno [5]. Ha una maggiore incidenza nel sesso maschile con un rapporto maschi/femmine pari a 2,5/1. Questo tumore benigno di tipo epiteliale si presume derivi dalle cellule intercalari dei dotti collettori e rappresenta il 3-5% delle neoplasie renali primarie [6]. Macroscopicamente, appare di colore brunastro, relativamente omogeneo, ben capsulato e può raggiungere dimensioni cospicue (più di 12 cm di diametro) ma non sono frequenti fenomeni di necrosi (Figura 2). Dal punto di vista istologico è costituito in prevalenza da grosse cellule epiteliali eosinofile disposte a "nido". La perdita dei cromosoma 1 e/o 4 è frequente, così come la presenza di mutazioni a livello del DNA mitocondriale [7]. La variante eosinofila del carcinoma renale cromofobo è ancora molto difficile da distinguere dall'oncocitoma attraverso una semplice osservazione della sezione istologica colorata con ematossilina e eosina. Questi due tumori possono essere differenziati con maggiore confidenza attraverso la colorazione con ferro colloidale di Hale a cui sono positivi il 100% dei carcinomi renali cromofobi, mentre gli oncocitomi sono negativi o presentano una colorazione atipica [8].

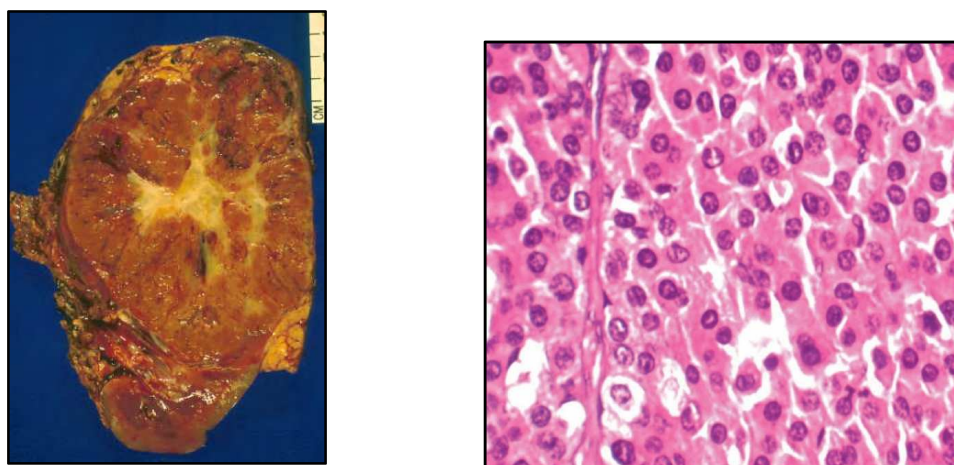


Figura 2. A sinistra: aspetto macroscopico di oncocitoma renale [9]. A destra: tipico aspetto di cellule oncocitiche con abbondante citoplasma eosinofilo, nuclei tondeggianti e uniformi e nucleoli poco visibili. Colorazione ematossilina e eosina.

L'**angiomiolipoma** è la neoplasia mesenchimale più frequente del rene (Figura 3). Gli angiomiolipomi classici sono tumori benigni composti, principalmente, da cellule adipose, cellule muscolari lisce e vasi sanguigni [10]; proprio l'abbondanza lipidica tipica di questo tumore forma aree ipodense che possono essere facilmente identificate con tecniche di *imaging*.

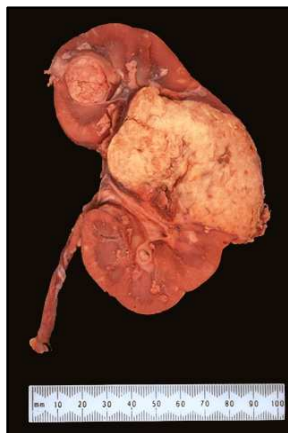


Figura 3. Immagine macroscopica di angiomiolipoma in cui si può notare la massa tumorale solida di colorito giallastro indice dell'alto contenuto in lipidi [11].

L'angiomiolipoma può riscontrarsi sia in forma sporadica che nella sindrome della Sclerosi Tuberosa, malattia genetica rara. La cellula di origine di questo tumore è ritenuta essere la cellula epitelioida perivascolare (PEC). Approssimativamente il 55-75% di pazienti affetti da sclerosi tuberosa sviluppa a livello renale l'angiomiolipoma; questa percentuale elevata è associata a mutazioni genetiche ereditarie a carico dei geni TSC1 (9q34) e TSC2 (16p13.3) tipiche in questi pazienti [12]. Tuttavia, nonostante la loro benignità, in letteratura sono riportati solo 3 casi di sarcoma sviluppatosi a partire da angiomiolipoma sporadico [13]. Nella maggior parte dei casi si tratta di tumori asintomatici che variano da pochi millimetri fino 20 centimetri di diametro.

1.1.2. I TUMORI MALIGNI

Il carcinoma renale rappresenta 4% di tutte le neoplasie dell'adulto ed è considerato la forma più maligna delle neoplasie urologiche comuni, in quanto caratterizzato da prognosi nefasta [14]. Il carcinoma renale (o adenocarcinoma) non rappresenta una unica entità patologica, ma uno spettro di malattia al quale concorrono differenti neoplasie primitive del rene, ognuna con peculiari caratteristiche istologiche, differente decorso clinico, differente risposta alla terapia e differente modificazione genetica alla base della sua insorgenza [15]. Per ciascun carcinoma renale è possibile identificare il tipo di cellule specifiche, all'interno del nefrone o nei dotti collettori, da cui origina la massa tumorale [16].

La maggior parte dei carcinomi renali è sporadica, tuttavia esistono alcune forme di cancro familiare autosomico dominante, che colpiscono generalmente gli individui con un'età sostanzialmente più giovane e che rappresentano solo il 4% dei tumori renali [4]:

- **Sindrome di Von Hippel-Lindau (VHL):** circa due terzi dei pazienti con questa sindrome sviluppano carcinomi a cellule renali bilaterali e multipli.
- **Carcinoma a cellule chiare ereditario (familiare):** è una forma tumorale strettamente confinata al rene, ma comporta comunque anomalie riguardanti il gene VHL o geni associati.
- **Carcinoma capillifero ereditario:** questa forma è caratterizzata da tumori multipli bilaterali con istologia papillare e coinvolge mutazioni nel proto-oncogene MET.

La bilateralità e/o la multicentricità sono caratteristiche dei carcinomi ereditari.

1.1.3. IL CARCINOMA A CELLULE RENALI (RCC)

Allo stato attuale delle conoscenze il carcinoma a cellule renali può essere considerato non un singolo tumore ma come una eterogenea famiglia di neoplasie con differenti caratteristiche istopatologiche e genetiche, che hanno come elemento comune una stessa derivazione: il tubulo renale [17]. I carcinomi a cellule renali (RCC), derivando dall'epitelio tubulare renale, sono quindi localizzati prevalentemente a livello della corticale [18].

1.1.3.1. EPIDEMIOLOGIA

Il carcinoma a cellule renali costituisce circa l'85% dei tumori renali primari e il 3-6% dei cancri più frequenti che insorgono in età adulta [19], ha un tasso di incidenza europea di 16.7 tra gli uomini e 7.6 nelle donne per 100000 persone/anno [20]. È una malattia tipicamente dell'adulto e in tutto il mondo la neoplasia è più frequente nel sesso maschile rispetto a quello femminile, con un rapporto maschio-femmina di 3-2. Il carcinoma a cellule renali ha un andamento crescente in base all'età, è infatti raro prima dei vent'anni, rappresentando solo il 2% dei tumori pediatrici; l'incidenza aumenta poi progressivamente raggiungendo un picco tra i sessanta ed i settanta anni (Figura 4). Essa occupa il terzo posto tra i tumori urologici più frequenti, dopo il carcinoma prostatico e quello a cellule di transizione della vescica. Ogni anno, vengono diagnosticati in Europa circa 100.000 nuovi casi e in Italia circa 11.300 (4.160 uomini e 7.140 donne) [21](Figura 4). Nel corso degli ultimi decenni, si è osservato un aumento dei casi nei paesi a più alta incidenza, che risultano essere quelli economicamente più sviluppati. In Italia, per esempio l'incidenza è in crescita, mentre la mortalità è in riduzione [22, 23]. Questo trend sfavorevole di incidenza è probabilmente dovuto alla maggior diffusione della diagnostica per immagini, che tuttavia ha migliorato la percentuale di diagnosi precoce, con un effetto positivo sulle possibilità terapeutiche.

Europe renal cell cancer				
Year	Estimated number of new cancers (all ages)	Male	Female	Both sexes
2012		71790	43462	115252
	ages < 65	35965	18401	54366
	ages >= 65	35825	25061	60886
2015		74631	44966	119597
	ages < 65	36681	18860	55541
	ages >= 65	37950	26106	64056
Demographic change		2841	1504	4345
ages < 65		716	459	1175
ages >= 65		2125	1045	3170

GLOBOCAN 2012 (IARC) - 7.6.2014

Figura 4. Confronto tra il numero di nuovi casi stimati in Europa nel 2012 rispetto a quelli previsti per il 2015.

1.1.3.2. EZIOLOGIA

Sebbene siano stati individuati nei modelli animali numerosi fattori eziologici [17], nell'uomo l'eziologia è tuttora poco conosciuta, ma alcuni fattori ambientali e comportamentali sembrano influire sull'insorgenza del carcinoma a cellule renali (Tabella 2).

Fattore di rischio	Effetti	Referenze
Fumo di sigaretta	1) raddoppia il rischio di ammalarsi di tumore del rene 2) eleverebbe anche il rischio di progressione del tumore di quasi 3 volte dopo la chirurgia	[14, 24-28]
Sovrappeso ed obesità	Aumento del rischio relativo di morte per carcinoma renale	[29]
Sedentarietà	Aumento del rischio del 25% di ammalarsi di RCC	[30]
Cancerogeni chimici (cadmio, asbesto, pesticidi ed erbicidi)	Aumento dell'incidenza di RCC	[31-33]
Malattie pregresse (ipertensione)	Aumentato rischio di sviluppare l'RCC; in particolare si ipotizza che l'ipertensione possa causare un danno renale sia diretto che mediante cambiamenti funzionali e metabolici	[14]
Dialisi e trapianto di rene	Un aumento dell'incidenza di malattia cistica acquisita del rene, che a sua volta predispone al carcinoma renale	[34-38]
Alimentazione	Il consumo di cibi fritti, soprattutto carne di vitello e pollo, costituisce un fattore di rischio in virtù della produzione durante la cottura di amine eterocicliche aromatiche, dal provato potere cancerogeno.	[30]
Forme di cancro familiare autosomico dominante	1) Anomalie riguardanti il gene VHL o geni associati 2) Mutazioni nel proto-oncogene MET	[4]

Tabella 2. Fattori di rischio per il carcinoma a cellule renali.

1.1.3.3. CLASSIFICAZIONE

I carcinomi a cellule renali sono una classe eterogenea di tumori che includono diverse sottotipologie istologiche di RCC (Tabella 3).

Istologia	Frequenza	Cellule di origine
Cellule chiare	60%-80%	Tubulo prossimale
Papillare	5%-15%	Tubulo prossimale
Cromofobo	5%-10%	Dotto collettore corticale

Tabella 3. Classificazione istologica del RCC.

➤ Il **carcinoma a cellule renali a cellule chiare** rappresenta circa il 60% dei tumori renali ed è quindi la neoplasia renale più frequente [39]. La lesione può essere sporadica o a carattere familiare, associata a sindromi genetiche quali la sindrome di *Von Hippel-Lindau*. Studi al microscopio elettronico stabilirono che le cellule dell'epitelio tubulare renale sono la normale sede di origine di questo carcinoma. Le cellule dei tubuli prossimali posseggono numerosi microvilli di superficie, che conferiscono loro l'aspetto a orletto a spazzola, e contengono un citoplasma più complesso di quello delle cellule del tubulo distale. Le caratteristiche ultrastrutturali delle cellule prossimali si riscontrano in vario grado nella maggior parte dei carcinomi renali; tuttavia, l'orletto a spazzola non è pienamente sviluppato ed è presente come tale solo in alcune cellule neoplastiche.

La variante a cellule chiare presenta uno stroma particolarmente vascolarizzato, spesso con aree emorragiche. La tipica superficie giallastra del tumore è causata dal contenuto lipidico della cellula: infatti colesterolo, lipidi neutri e fosfolipidi risultano essere abbondanti (Figura 5). Solitamente le neoplasie a cellule chiare (Figura 5) sono lesioni solitarie e unilaterali, masse di pochi millimetri o molto grandi (la dimensione media è di 7 cm), che arrivano a pesare diversi chilogrammi; sono composte da tessuto grigiobianco-giallastro che alterano la configurazione del rene. Non sono dotati di una vera e propria capsula, ma posseggono sempre una pseudocapsula composta da parenchima compresso e tessuto fibroso. La variante a cellule chiare del carcinoma a cellule renali è caratterizzata microscopicamente da distese, da acini e da alveoli di cellule neoplastiche unite tra loro da una rete di delicati sinusoidi vascolari e di fibre di reticolina. Le cellule chiare sono rotondeggianti o poligonali con citoplasma abbondante, contenente colesterolo, esteri del colesterolo, fosfolipidi e glicogeno (Figura 6). La cellularità chiara deriva da un artefatto causato dalla procedura istologica di routine che, rimuovendo il glicogeno ed i lipidi all'interno dei citoplasmici li fa apparire otticamente vuoti al microscopio dopo colorazione con ematossilina ed eosina.



Figura 5. Aspetto macroscopico di un carcinoma renale a cellule chiare. Il colore giallo di alcune aree è causato dalla presenza di lipidi e colesterolo nelle cellule tumorali [16].

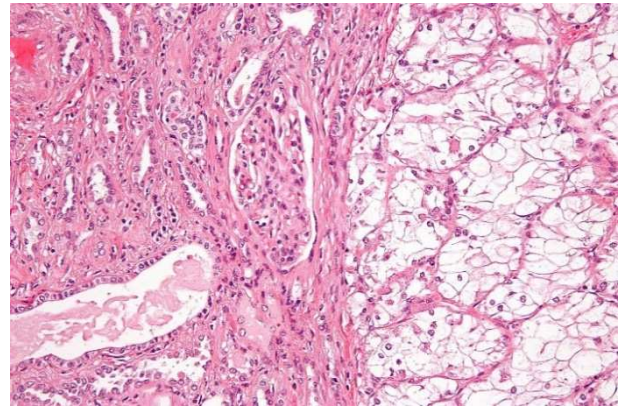


Figura 6. Aspetto microscopico di un carcinoma renale a cellule chiare [16].

Le cellule possono talora, presentare citoplasma granulare e questo può causare difficoltà diagnostiche, ma forma e dimensioni cellulari sono d'aiuto, come pure possono esserlo indagini immunohistochimiche o di biologia molecolare. Solo un limitato numero di tumori contiene unicamente cellule chiare; di solito, infatti, è presente anche una quota di cellule granulari (cellule scure) che talvolta può rappresentare l'unica componente cellulare della neoplasia. Le cellule granulari hanno un citoplasma eosinofilo e abbondanti mitocondri. All'analisi immunohistochimica, circa il 50% delle cellule carcinomatose, soprattutto se di grado elevato, esprime la vimentina. Cellule vimentino-positive sono più frequenti nei carcinomi renali sarcomatoidi e più rare nelle forme prive di zone fusiformi. La quasi totalità degli adenocarcinomi renali esprime la cheratina 8, la cheratina 18 e la maggioranza anche l'antigene epiteliale di membrana.

Alcuni RCC a cellule chiare hanno un aspetto cistico dovuto alla presenza di necrosi (pseudo-cisti), che è associata con l'aumento dell'aggressività tumorale [16]. Alcuni carcinomi di questo tipo formano cisti neoplastiche; nel caso in cui vi siano cisti ma non una componente tumorale solida si parla di carcinoma a cellule renali cistico multiloculare [40]. Questo è considerato un sottotipo del RCC a cellule chiare ed ha un'eccellente prognosi.

È possibile che all'interno del tumore avvengano cambiamenti sarcomatoidi che sono associati ad una prognosi sfavorevole [41, 42]. La trasformazione sarcomatoide avviene nel 5% circa di tale neoplasia [43] ed è evento più frequente nei tumori a cellule chiare, ma non esclusivo di questo istotipo.

La maggior parte dei carcinomi a cellule chiare generano una debole risposta infiammatoria, tuttavia vi può essere la presenza di infiltrati linfocitari o neutrofilici con cellule *Natural Killer* [44] e vi è un'associazione tra la forte infiltrazione linfocitaria e l'*outcome* sfavorevole [45]. Un'altra caratteristica peculiare di questa neoplasia è la sua tendenza a invadere la vena renale, formando un blocco di cellule neoplastiche all'interno del vaso; l'ulteriore accrescimento può portare ad un trombo neoplastico nella cava inferiore. I tumori di dimensioni maggiori ed a malignità più elevata possono invadere localmente le strutture e gli organi contigui. Gli RCC a cellule chiare metastatizzano comunemente attraverso la vena cava fino al polmone. Metastasi

retrograde possono svilupparsi attraverso le vene paravertebrali, la vena testicularis o ovarii, le vene intrarenali o lungo l'uretere. Il carcinoma a cellule renali a cellule chiare è conosciuto per la metastatizzazione in siti inusuali e in tempi lunghi, anche dopo 10 anni dalla diagnosi/trattamento. Tutti i tumori renali a cellule chiare sono considerati maligni, indipendentemente dalle loro dimensioni [43].

➤ Il **carcinoma a cellule renali di tipo papillare** comprende circa il 10% dei carcinomi a cellule renali, sono caratterizzati da una crescita papillare e si manifestano sia in forma familiare che sporadica [16, 46, 47]. La distribuzione in termini di età di insorgenza di questo tipo di tumore è simile a quella del tumore renale a cellule chiare, con età media alla diagnosi compresa tra i 50 e i 65 anni [48].

Macroscopicamente i carcinomi renali di tipo papillare spesso presentano necrosi e in alcuni casi è possibile identificare una pseudo-capsula. Tumori bilaterali e multifocali sono più frequenti in questo sottotipo rispetto ad altri tumori renali maligni. All'indagine istologica, è caratterizzato da cellule epiteliali che formano papille e tubuli. Le papille tumorali contengono un delicato core fibro-vascolare e spesso sono presenti aggregati di macrofagi e corpi di Psammoma, ossia micro calcificazioni concentriche e laminate. Sono stati descritti due tipi morfologici di carcinoma a cellule renali papillare [49]. Il tipo 1 è spesso multifocale ed è caratterizzato da papille ricoperte da piccole cellule con scarso citoplasma. Il tipo 2 presenta cellule di grado nucleare più elevato con citoplasma eosinofilo e nuclei pseudo stratificati. Il RCC di tipo papillare può andare incontro a differenziamento sarcomatoide ed in questo caso è spesso associato a prognosi infausta (Figura 7). Le cellule fusiformi ricordano le cellule mesenchimali pleomorfe e la differenziazione dal fibrosarcoma può essere difficile.

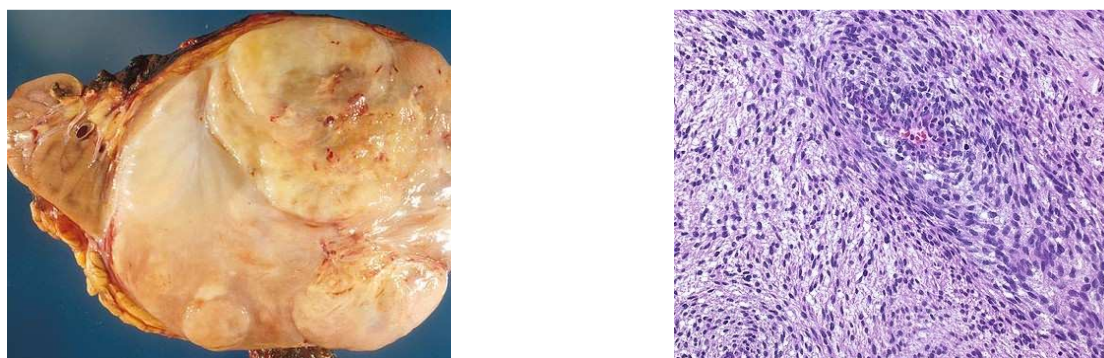


Figura 7. Aspetto macroscopico e microscopico di tumore sarcomatoide del rene.

Sono stati inoltre descritti carcinomi papillari interamente composti da oncociti [50] e questo tipo di tumori mostra caratteristiche clinico-patologiche diverse da quelle dei carcinomi di tipo 1 e 2 e sono quindi considerati come una terza tipologia di tumori renali papillari.

➤ Il **carcinoma a cellule renali di tipo cromofobo** costituisce circa il 5% di tutti i tumori renali. Hanno una prognosi migliore rispetto ai tumori renali a cellule chiare [51, 52]. Il tasso di mortalità è infatti inferiore al 10%. Raramente si può verificare una trasformazione di tipo sarcomatoide ed è un segno diagnostico di prognosi infausta [53]. Questo tipo di tumore può presentare metastasi a distanza a livello di polmoni, fegato e pancreas. È stato suggerito che le metastasi epatiche siano più frequenti nei carcinomi

renali di tipo cromofobo che negli altri sottotipi istologici. La superficie di taglio del carcinoma a cellule renali di tipo cromofobo appare omogenea e di colore grigio o grigio-marrone (Figura 8).

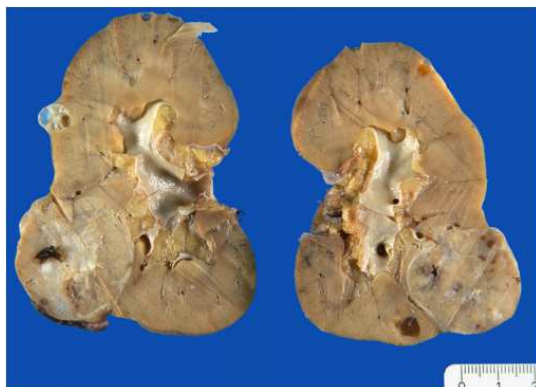


Figura 8. Aspetto macroscopico di un tumore renale di tipo cromofobo.

Il tumore è caratterizzato da grandi cellule poligonali con citoplasma reticolato e membrane cellulari prominenti. Alcune cellule sono di forma irregolare e multinucleate ed è comune la presenza di aloni perinucleari. Un tratto diagnostico tipico è dato dalla reazione diffusa del citoplasma alla colorazione con ferro colloidale [54]. La variante eosinofila del carcinoma renale di tipo cromofobo è costituita solo da cellule eosinofile. Alla microscopia elettronica, il citoplasma delle cellule tumorali è caratterizzato da depositi di glicogeno e da numerose vescicole. La principale diagnosi differenziale del tumore renale cromofobo è l'oncocitoma renale, che, come già detto, è considerato una neoplasia benigna [6]. Per ora vengono definiti tumori ibridi quelli che condividono caratteristiche istopatologiche del carcinoma cromofobo e dell'oncocitoma.

1.1.3.4. CITOGENETICA

Il carcinoma a cellule renali comprende un gruppo di tumori eterogeneo anche a livello clinico-patologico con prognosi diversificata; le differenze tra i sottotipi di RCC sono verosimilmente dovute ad una diversa citogenesi e progressione molecolare [55]. Sono state identificate differenti alterazioni genetiche nei diversi sottotipi di neoplasie renali. Tali alterazioni sono utili nel formulare una corretta diagnosi e definire una più precisa prognosi, oltre che a selezionare appropriatamente i pazienti per differenti opzioni terapeutiche. I tumori renali si presentano solitamente come malattie sporadiche, ma possono insorgere anche in forma familiare. Il pattern delle mutazioni somatiche è stato a lungo studiato nei tumori renali ed è diventato il principale criterio di classificazione [56].

La delezione del cromosoma 3 (LOH 3p) è la più tipica anomalia genetica nei carcinomi a cellule renali a cellule chiare sporadici ed è considerata uno dei principali passaggi dell'insorgenza del tumore. Sono stati individuati diversi geni sul braccio corto del cromosoma 3 che probabilmente sono coinvolti nel carcinoma renale. Nel 98% di casi di questo tumore si ha una delezione o una traslocazione cromosomica non bilanciata (3;6, 3;8, 3;11) con perdita della più piccola regione sovrapposta del cromosoma 3 (3p14 fino a

3p26). In questa regione è collocato il gene VHL (3p25.3) [57, 58], gene soppressore tumorale della malattia di *von Hippel-Lindau* [57, 59].

La mutazione del gene VHL è stata identificata esclusivamente in questo tipo di tumore e non appare associata né al grado né allo stadio tumorale. La malattia di *von Hippel-Lindau* (VHL) è una malattia ereditaria molto rara (colpisce un individuo ogni 36.000 circa), presenta una trasmissione autosomica dominante e predispone all'insorgenza di una varietà di tumori benigni e maligni quali neoplasie e cisti renali bilaterali e multifocali, tumori cerebellari e del midollo spinale (emangioblastoma), emangioma retinico, feocromocitoma, tumori endocrini del pancreas e cistoadenoma dell'epididimo. I tumori renali sono carcinomi a cellule chiare con tendenza all'invasione e alla metastatizzazione [60] e se non diagnosticati e trattati precocemente, rappresentano la causa di morte nel 35-45% dei pazienti affetti da questa sindrome. Per la maggior parte dei casi la patologia esordisce in età adulta e richiede una diagnosi tempestiva affinché si instauri una adeguata terapia. In questa malattia un allele VHL malato viene ereditato, mentre l'altro va incontro a inattivazione indotta da ipermetilazione in circa l'80% dei ccRCC, evidenziando che il gene VHL si comporta come un gene soppressore tumorale sia nelle forme sporadiche che familiari. La proteina funzionante codificata dal gene VHL forma, in associazione con elonghina B, elonghina C e cullina 2 (CUL2), un complesso multiproteico proteico (VCB-CUL2). In condizioni normali, questo complesso si lega alle proteine che devono essere degradate dalla cellula, quando queste sono danneggiate o non più necessarie. In particolare il complesso VCB-CUL2 lega una proteina chiamata HIF (*hypoxia-inducible factor*), che è in grado di indurre la trascrizione di molti importanti geni che, quando espressi, portano alla produzione di proteine quali VEGF, PDGF-B, GLUT-1 e eritropoietina; nel complesso risultano essere proteine coinvolte nella divisione cellulare e nella formazione di nuovi vasi sanguigni. HIF è anche coinvolta nella formazione della matrice extracellulare. Conseguenze delle mutazioni del gene VHL sono l'assenza (sindrome di *von Hippel-Lindau* di tipo 1) o la produzione di forme anormali della proteina VHL (sindrome di *von Hippel-Lindau* di tipo 2). Quando la proteina VHL è alterata o assente, il complesso VCB-CUL2 non è più in grado di riconoscere HIF o altre proteine che devono essere degradate. Questo porta all'accumulo nelle cellule di HIF, che stimola le cellule a dividersi in maniera anormale e innesca la produzione di vasi sanguigni [61]. La rapida e incontrollata divisione cellulare, insieme con l'anormale formazione di vasi, può portare allo sviluppo di cisti e tumori in soggetti affetti dalla sindrome di *von Hippel-Lindau*.

È stato inoltre suggerito che, nel carcinoma renale, intervenga l'attivazione molecolare di alcuni proto-oncogeni alla base delle alterazioni citogenetiche. La sovra-espressione di c-myc e dell'mRNA del recettore per l'*epidermal growth factor* (EGFR) (Erb B-1) e la sotto-espressione dell'mRNA dell'HER-2 (Erb B-2) si osservano nella maggioranza dei pazienti affetti da carcinoma a cellule renali [62, 63].

I carcinomi a cellule renali di tipo papillare non sono invece caratterizzati da delezioni a livello del 3p ma sono associati a trisomie o polisomie dei cromosomi 3q, 7, 8, 12, 16, 17 e 20 e perdita del cromosoma Y [64-66]. La mutazione del proto-oncogene c-MET sul cromosoma 7 caratterizza i carcinomi papillari ereditari ma è rara nelle forme sporadiche [65, 67]. MET è un proto-oncogene che serve come recettore delle tirosino-chinasi per il fattore di crescita epatocitario; quest'ultimo media la crescita cellulare, la

mobilità, l'invasione e la differenziazione morfogenetica. Sono state identificate entrambe le mutazioni della linea germinale e somatica nel dominio delle tirosino-chinasi del gene MET, rendendo lo stesso mutato un oncogene candidato nel cancro. Un secondo gene, PRCC (per carcinoma renale papillare) sul cromosoma 1 è risultato coinvolto nei tumori sporadici, soprattutto nei bambini che manifestano traslocazioni X,1 [68]. Sono stati descritte inoltre alterazioni molecolari nella proteine delle cellule del carcinoma renale come p53, metalloproteinasi e telomerasi [69-71]. Le analisi del cariotipo, in fluorescenza, l'ibridazione in situ e gli *array* CGH (*comparative genomic hybridization*) hanno rivelato che perdite cromosomiche multiple e non in maniera *random* dei cromosomi 1, 2, 6, 10, 13, 17, 21 e dei cromosomi X e Y sono frequenti nei carcinomi a cellule renali di tipo cromofobo, mentre gli oncocitomi presentano solo riarrangiamenti o traslocazioni che coinvolgono l'11q13 o perdite parziali dei cromosomi 1, 14 e dei cromosomi sessuali [72]. La perdita del braccio corto del cromosoma 3 è invece un evento raro in questi sottotipi tumorali [73].

1.1.3.5. DIAGNOSI

Il carcinoma renale rimane asintomatico finché le sue dimensioni sono ridotte ritardando la sua diagnosi che risulta spesso successiva all'invasione locale o a metastasi a distanza. L'utilizzo sempre più frequente di tecniche di diagnostica strumentale mediante analisi dell'immagine ha migliorato e aumentato la rilevazione del carcinoma renale in fase asintomatica. Tuttavia, la natura di alcune particolari lesioni renali cistiche e masse solide, anche benigne, non può essere distinta in maniera confidentiale con le sole tecniche di *imaging*; dunque risulta urgente la scoperta di nuove procedure diagnostiche e predittive per la ricerca di efficaci marcatori dell'RCC. I sintomi della presenza di carcinoma renale fanno parte della così detta triade classica e sono: ematuria, dolori lombari e presenza di una massa addominale alla palpazione; questi tre sintomi però si manifestano come triade solamente nel 10% dei casi. Spesso il sospetto della presenza di carcinoma renale è dato dalla presenza di segnali quali:

- colore anormale delle urine (dovuto alla presenza di piccole quantità di sangue)
- perdita di peso (circa 5% del peso corporeo) ed aspetto emaciato
- ingrossamento di un testicolo dovuto al blocco della vena gonadale e causato dall'invasione del tumore nella vena renale sinistra
- fenomeni paraneoplastici dovuti alla secrezione anomala di sostanze con attività ormonale.

Inoltre spesso si presentano sintomi dovuti a patologie metastatiche (ad esempio fratture patologiche dovute a metastasi alle ossa). Data la mancanza di sintomi soprattutto nella fase iniziale della neoplasia, molto spesso questa viene diagnosticata in maniera accidentale tramite indagini all'addome (ad esempio la tomografia assiale computerizzata o TAC) eseguite per propositi non correlati al carcinoma renale. Oltre ad un corretto ed attento esame clinico, rivestono dunque un ruolo fondamentale le tecniche di diagnostica strumentale come l'ecografia e la tomografia computerizzata; tramite queste due tecniche il tumore viene studiato e monitorato. L'ecografia permette di discriminare tra una massa solida (sintomo di una probabile neoplasia) ed una massa cistica; essa ha una sensibilità maggiore dell'85% nell'evidenziare i tumori di maggiori dimensioni, mentre solo del 60% circa per i piccoli tumori. La TAC permette, invece, di

visualizzare accuratamente le masse di dimensioni ridotte e viene utilizzata anche per valutare l'eventuale estensione del tumore ad altri organi; essa consente di esaminare il volume della neoplasia, la sua localizzazione e la presenza di linfadenopatie. Inoltre, associando la TAC alla Risonanza Magnetica Nucleare (RMN) è possibile ottenere un'immagine tridimensionale della massa tumorale e quindi programmare in maniera ottimale l'eventuale intervento chirurgico. In seguito alla somministrazione di liquido di contrasto intravenoso (sia nella tomografia computerizzata oppure nelle immagini di risonanza magnetica), aumenterà il contrasto delle zone densamente irrorate (abbondante vascolarizzazione), mettendo in evidenza il tumore rispetto al parenchima normale. L'arteriografia e la TAC sono equivalenti nel descrivere i rapporti con la vena renale, ma la tomografia computerizzata risulta la metodica migliore per dimostrare un interessamento linfonodale. Tutte e tre le strumentazioni diagnostiche (ecografia, TAC e RMN) forniscono informazioni adeguate sull'invasione della vena cava. In ogni caso, nessuna di queste metodiche può essere considerata adeguata per tutti i pazienti. Per questo motivo, le tecniche vengono spesso combinate per ottenere informazioni più complete, soprattutto quando si intende procedere con l'asportazione di tumori di grandi dimensioni. La biopsia percutanea non è una tecnica particolarmente usata a scopo diagnostico, perché la possibilità di una diagnosi errata ed il rischio per il paziente di complicazioni mediche la rendono sfavorevole come metodica.

1.1.3.6. STADIAZIONE

Nel valutare la sopravvivenza a distanza, il rischio di progressione e di recidiva nei pazienti con carcinoma a cellule renali è necessario considerare diversi ordini di fattori prognostici, clinici, di laboratorio, anatomici, istologici e molecolari [74]. L'insieme di queste informazioni è contemplato nel sistema di classificazione tumorale TNM (*Tumor, Node, Metastasis*) [75]:

TUMORE PRIMITIVO (T)

T_x Tumore primitivo non definibile

T₀ Tumore primitivo non evidenziabile

T₁ Tumore della dimensione massima di 7 cm, confinato al rene

T_{1a} Tumore ≤ 4 cm

T_{1b} Tumore > 4 cm e < 7 cm

T₂ Tumore della dimensione massima oltre i 7 cm, confinato al rene

T_{2a} Tumore > 7 cm e ≤ 10 cm, confinato al rene

T_{2b} Tumore > 10 cm, confinato al rene

T₃ Tumore che invade le maggiori vene o i tessuti perinefrici, ma non la ghiandola surrenale ipsilaterale, e che non va oltre la fascia di Gerota

T_{3a} Tumore che invade la vena renale e le sue ramificazioni o il tessuto perineale, ma che non va oltre la fascia di Gerota

T_{3b} Tumore che invade macroscopicamente la vena cava al di sotto del diaframma

T_{3c} Tumore che invade macroscopicamente la vena cava al di sopra del diaframma o la parete della vena cava

T₄ Tumore che invade oltre la fascia di Gerota (includendo adiacente invasione della ghiandola surrenale ipsilaterale)

LINFONODI REGIONALI (N)

Nx Linfonodi regionali non valutabili

N0 Linfonodi regionali liberi da metastasi

N1 Metastasi in un singolo linfonodo regionale

N2 Metastasi in più linfonodi regionali

N3 Metastasi in un singolo linfonodo >5 cm (dimensione massima)

METASTASI A DISTANZA (M)

M0 Metastasi a distanza assenti

M1 Metastasi a distanza presenti

RAGGRUPPAMENTO IN STADI

STADIO I: T1, N0, M0

STADIO II: T2, N0, M0

STADIO III: T1, N1, M0-T2, N1, M0-T3, N0, N1, M0

STADIO IV: T4, N0, N1, M0-Ogni T, N2, M0-Ogni T, Ogni N, M1.

Nella classificazione TNM, T si riferisce alle dimensioni del tumore primitivo, N indica lo stato dei linfonodi vicini al tumore e M la presenza di metastasi a distanza. La categoria T3 comprende le neoplasie che infiltrano la capsula e quelle che si estendono alla vena renale o cava; essa è suddivisa in tre sottocategorie (T3a, T3b e T3c). La stadiazione regionale può essere eseguita mediante TAC, ma talvolta solamente l'anatomopatologo è in grado di definire con certezza l'esatta estensione locale e regionale di una neoplasia renale. La valutazione della presenza di metastasi a distanza è molto importante. Le metastasi ematogene del carcinoma renale possono interessare qualsiasi organo, ma le sedi più comuni sono il polmone, il fegato, le ossa, i tessuti sottocutanei ed il sistema nervoso centrale. Sembra pertanto ragionevole che una appropriata valutazione pre-operatoria comprenda un radiogramma standard del torace, gli esami di funzionalità epatica, la misurazione del calcio sierico, l'anamnesi e l'esame obiettivo.

1.1.3.7. PROGNOSE

La prognosi di questi pazienti è chiaramente correlata allo stadio della malattia al momento della diagnosi. Oltre a questo gli altri fattori prognostici importanti nel RCC sono il diametro della neoplasia e il grado di differenziazione nucleare.

La classificazione più largamente impiegata rispetto al grado di differenziazione nucleare è quella di Fuhrman *et al.* (1982) [76] che distingue 4 differenti categorie (Figura 9).

Grado 1	Nuclei piccoli (10 mm circa), rotondeggianti, uniformi con nucleoli poco evidenti o assenti
Grado 2	Nuclei di medie dimensioni (15 mm circa) con contorno irregolare e nucleoli piccoli (visibili con l'ingrandimento $\times 400$)
Grado 3	Nuclei voluminosi (20 mm circa) con marcate irregolarità del contorno e evidenti nucleoli
Grado 4	Come nel Grado 3 con l'aggiunta di nuclei bizzarri, spesso plurilobulati e con aggregati di cromatina

Modificata da Fuhrman SA e coll.: Am J Surg Pathol 1982; 6:655-663.

Figura 9. Metodo di Fuhrman per la definizione del grado nucleare nell'adenocarcinoma renale

Negli ultimi anni sono stati suggeriti numerosi sistemi prognostici integrati. Fra questi il più importante ed accurato sembra essere lo UISS (*University of California Los Angeles, Ucla Integrated Staging System*), lo *score* SSIGN (*Stage, Size, Grade and Necrosis*) proposto dalla Mayo Clinic ed il nomogramma di Karakiewicz [77-79].

L'accuratezza prognostica di questi sistemi integrati si dimostra più elevata rispetto a quella del solo TNM e raggiunge circa l'80%. Tuttavia l'applicazione di tali sistemi nella pratica clinica di routine risulta difficile, poiché il calcolo di uno *score* o l'assegnazione di un paziente ad una categoria di rischio attraverso una combinazione matematica di numerose variabili richiederebbe strumentazioni apposite, che non sono sempre facilmente disponibili, e una maggiore tempistica. Per questo motivo i sistemi integrati non sono ancora stati approvati nelle linee-guida e non sono raccomandati nella pratica clinica.

Dunque, lo stadio patologico è l'indice prognostico più importante e nel complesso la sopravvivenza causa-specifica a 5 anni varia dall' 88 al 99% negli RCC pT1, dal 70,5% all'82% nei pT2, dal 10% al 60% nei pT3 e fino al 20% nei pT4. Inoltre, nei pazienti con malattia metastatica la sopravvivenza causa-specifica a 5 anni varia dal 10% al 30% dei casi [75]. Tuttavia gli esiti del carcinoma risultano significativamente peggiori con l'aumentare dell'età nei pazienti con tumore di media grandezza tra i 4 e i 7 cm; la sopravvivenza relativa a 5 anni è inferiore nei soggetti malati di età più avanzata (tra 50 e 79 anni) rispetto ai pazienti più giovani (tra 30 e 39 anni) [80]. Altri fattori prognostici sfavorevoli sono l'interessamento della vena renale, l'estensione ai linfonodi regionali, il coinvolgimento della *Fascia Di Gerota*, l'interessamento degli organi contigui e le metastasi a distanza [81]. L'invasione della vena renale (IIIA) non ha effetti marcatamente negativi sulla prognosi, ma per i pazienti allo stadio IIIB (con invasione dei linfonodi regionali), la

sopravvivenza a 5 anni scende radicalmente. L'interessamento dei linfonodi regionali che drenano la linfa dal parenchima renale è un segno prognostico negativo che si associa ad una sopravvivenza a 5 anni dello 0-30%. L'estensione della disseminazione linfatica è senza dubbio molto importante e, solamente, i pazienti con un interessamento linfatico limitato sembrano sopravvivere. La maggioranza degli studi mostra una miglior sopravvivenza in quei pazienti a cui sia stata diagnosticata una malattia metastatica e che presentano le seguenti condizioni:

- Un lungo intervallo di tempo libero da malattia tra la nefrectomia e la comparsa di metastasi
- La presenza di metastasi solo polmonare
- Buon *performance status*
- L'asportazione radicale del tumore primitivo.

1.1.3.8. STRATEGIE DI INTERVENTO

Attualmente la nefrectomia radicale rappresenta l'unica possibilità di guarigione nelle forme localizzate di carcinoma renale; essa consiste nell'asportazione del rene, del tessuto perirenale contenuto nella fascia di Gerota, del surrene e dei linfonodi loco-regionali. Per neoplasie di piccole dimensioni (diametro inferiore a 4 cm) in sede periferica e che non coinvolgano la via escrettrice, oppure in condizioni di necessità quale la presenza di un unico rene, la cui asportazione obbligherebbe ad una terapia dialitica, rimane comunque l'opzione di una chirurgia conservativa. Nei casi di tumori localizzati, ma in cui il paziente non sia un buon candidato per una procedura chirurgica, è possibile ricorrere a terapie non invasive; queste consistono nella distruzione della massa tumorale tramite calore (ablazione mediante radiofrequenze) oppure mediante crioterapia. Queste tecniche richiedono l'inserimento di una sonda il cui posizionamento viene monitorato utilizzando tecnologie di *imaging* in tempo reale. Per quanto riguarda chemioterapia e radioterapia, queste danno scarsi risultati in quanto questo tipo di tumore è sia radio che chemio resistente [82]; la spiegazione della scarsità dei risultati della chemioterapia potrebbe risiedere nella presenza del gene MDR-1 (*human multidrug resistant-1*) nelle cellule di carcinoma renale. Questo gene, espresso anche in altre tipologie di cellule tumorali, codifica per una proteina di membrana, la glicoproteina P, che funziona da pompa per l'espulsione dalla cellula dei farmaci citotossici. Scarsi risultati, ottenuti nel 20-25% dei casi, derivano dall'immunoterapia utilizzando interferone- γ e/o interleuchina-2 [83] in associazione con Vinblastina e/o Floxuridina; questi farmaci presentano inoltre molti effetti collaterali che possono tuttavia essere ridotti diminuendo il dosaggio oppure somministrandoli per via sottocutanea. Nonostante la scarsità di risultati e gli effetti collaterali, attualmente l'interleuchina-2 risulta essere l'unica terapia approvata dalla FDA (*Food and Drug Administration*) per il trattamento dei casi di RCC con metastasi. In aggiunta, non sono ancora conosciuti efficaci biomarcatori che possano consentire una diagnosi precoce del carcinoma renale e/o che permettano il controllo continuo dell'evoluzione del tumore ed il *follow-up* dei pazienti; per la caratterizzazione di questa tipologia di tumore e la scoperta di eventuali biomarcatori, si stanno compiendo numerosi ed onerosi sforzi in diversi campi della ricerca (genetica molecolare, citogenetica, immunoistochimica e proteomica).

1.1.3.9. PATOGENESI DEL CARCINOMA RENALE A CELLULE CHIARE (ccRCC)

Ruolo di HIF nella Patogenesi del ccRCC

Il fattore inducibile dall'ipossia (HIF) riveste un ruolo centrale nella patogenesi del carcinoma renale a cellule chiare. L'unica associazione stretta fra il ccRCC e la presenza di mutazione nel gene VHL consiste nella stabilizzazione costitutiva dei fattori inducibili da ipossia HIF-1 α e HIF-2 α . L'interazione fra HIF e pVHL è mediata dalla modificazione enzimatica post-traduzionale d'idrossilazione su residui conservati di prolina della proteina HIF. La maggior parte delle mutazioni nel gene VHL sono predittive dell'abolizione di questa interazione. Oltre a regolare programmi trascrizionali in favore dell'adattamento fisiologico e cellulare all'ipossia. Inoltre, HIF-1 α è stato associato anche a regolazione genica di *pathway* associati al metabolismo glicolitico, mentre HIF-2 α riveste un ruolo maggiore nell'attivazione di geni associati a proliferazione e differenziamento tra i quali il gene TGF α (Figura 10).

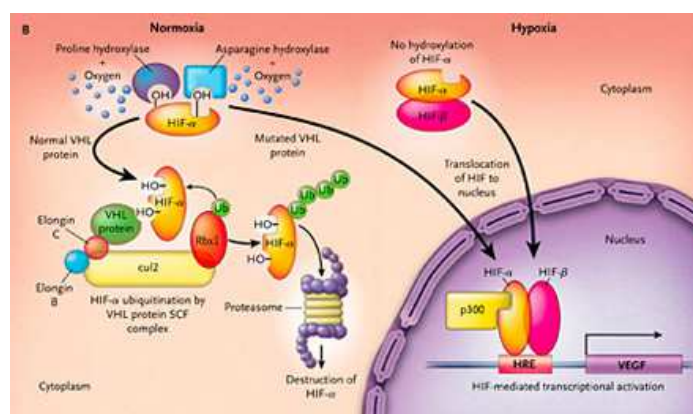


Figura 10. Pathway di HIF.

I fattori di crescita trasformanti (*Transforming Growth Factors*, TGF) alfa e beta sono fattori di crescita regolatori prodotti dal tumore, che possono essere correlati allo sviluppo del carcinoma a cellule renali [84]. Il TGF α si lega all'EGFR e, poiché sia il TGF α che l'EGFR (*Epidermal Growth Factor Receptor*) sono sovra-espressi nelle neoplasie renali, è possibile che la loro interazione svolga un ruolo significativo nella promozione della trasformazione e/o proliferazione delle cellule tumorali, probabilmente attraverso un meccanismo autocrino. Il TGF β viene prodotto in una forma biologicamente inattiva da linee cellulari del carcinoma renale umano in coltura e, dopo incubazione con TGF β esogeno, la proliferazione di tali cellule risulta inibita [85]. Sembra, quindi, che alterazioni nella produzione di fattori di crescita con azione stimolante o inibente e la risposta a tali sostanze possano determinare l'insorgenza di uno squilibrio nei meccanismi di crescita, la perdita dell'inibizione autocrina della proliferazione e lo sviluppo del cancro.

Alterazioni genetiche in regolatori epigenetici

Un altro gene PBRM1 (*Protein polybromo-1*), localizzato sul cromosoma 3p, è stato identificato mutato circa nel 41% dei ccRCC; codifica per la proteina Baf180 (*BRG1-associated factor 180*). Questa proteina è una delle subunità del complesso di rimodellamento cromatinico SWI/SNF, implicato nell'attivazione trascrizionale

attraverso l'interazione con l'istone H3 [86-89]. In effetti, alcune evidenze mostrano che nella biologia del ccRCC possa avere un ruolo importante una disregolazione epigenetica mediata da metilasi/demetilasi che agiscono sull'istone H3. Oltre al già citato PBRM1, i geni KDM6A/UTX (*lysine (K)-specific demethylase 6A*), SETD2 (*SET domain containing 2*) e KDM5C/JARID1C (*lysine (K)-specific demethylase 5C*) che codificano rispettivamente per la demetilasi della lisina 27 dell'istone H3 (H3K27), la metiltransferasi di H3K36 e la demetilasi di H3K4, sono stati identificati mutati nel ccRCC. In particolare, il gene SETD2 si trova anch'esso sul cromosoma 3p suggerendo che la delezione riscontrata in questa regione cromosomica in associazione con la patologia ccRCC possa aver avuto conseguenze anche su questo gene [90, 91]. L'attività modificata di specifici regolatori epigenetici, singole proteine o complessi proteici, può condurre ad effetti pleiotropici che possono alterare l'omeostasi cellulare. È stato dimostrato infatti che la ri-espressione della proteina VHL (pVHL) in linee cellulari renali VHL-deficienti ha come conseguenza l'aumento dei livelli di metilazione sull'istone H3 associata a diminuzione di JARID1C, un target di HIF2 α . Inoltre, è stato osservato che il silenziamento di JARID1C in linee cellulari tumorali fomenta la crescita tumorale in modelli murini suggerendo il ruolo di soppressore tumorale per JARID1C [92].

Regolazione della matrice extracellulare nella patogenesi ccRCC

La regolazione della matrice extracellulare è un altro importante processo la cui de-regolazione può contribuire allo sviluppo del ccRCC. La distruzione dell'architettura regolare della matrice extracellulare (ECM) è stata associata a crescita cellulare anormale, angiogenesi e metastatizzazione. pVHL gioca un ruolo importante nella regolazione di ECM. È stato mostrato infatti che l'interazione diretta di pVHL con la fibronectina e il collagene IV porta al loro assemblaggio in ECM con conseguente soppressione della tumorigenesi, dell'angiogenesi e dell'invasione tumorale [93-96]. Molte varianti alterate della proteina VHL falliscono nel legare e degradare HIF- α . Tuttavia, è stato osservato che il fallimento del legame di questa proteina con la fibronectina e con il collagene IV, con conseguente perdita della capacità di assemblare ECM, è una prerogativa di tutte le varianti mutate di VHL ad oggi testate [97, 98]. In effetti, è stato evidenziato che l'inattivazione del *pathway* VHL-ECM porta a tumori molto vascolarizzati, con una matrice lassa di fibronectina e collagene IV e con una aumentata capacità invasiva [95]. La perdita di pVHL può portare, inoltre, a sotto-regolazione di proteine delle giunzioni strette come occludina (OCLN) e claudina (CLDN) [99].

Un ulteriore meccanismo con il quale il *pathway* VHL regola il *turnover* di ECM implica l'attivazione di metalloproteinasi (MMPs), una famiglia di enzimi coinvolti nella degradazione della matrice. Linee cellulari di RCC mancanti del gene VHL infatti mostrano una iper-produzione di alcuni membri di questa famiglia e sviluppano un fenotipo morfologico del tessuto renale caratteristico dei tumori invasivi [100]. È stato ipotizzato che il meccanismo molecolare mediante il quale il rimodellamento proteolitico di ECM da parte di MMPs avviene potrebbe coinvolgere l'esposizione di siti critici del collagene IV, normalmente interni alla sua struttura a tripla elica, portando all'interazione con specifiche integrine che stimolano l'angiogenesi [101].

Alterazioni nel Metabolismo associate a ccRCC

Uno stretto legame fra alterazioni nel processo metabolico e cancro renale è stato ipotizzato sulla base dell'osservazione che la trasfezione di geni codificanti per enzimi del ciclo di Krebs in linee cellulari sopprimono la formazione tumorale in cellule renali [102, 103]. Mutazioni legate a perdita di funzione in SDH (*Succinate dehydrogenase*) e in FH (*fumarate hydratase*) sono state infatti associate a RCC [104] e sembra causino l'interruzione del ciclo di Krebs con conseguente accumulo di intermedi di reazione come fumarato e succinato [105, 106]. Quindi, l'accumulo di queste molecole potrebbe condurre all'inibizione di enzimi implicati nell'idrossilazione di HIF impedendo, di fatto, il riconoscimento di quest'ultimo da parte di pVHL. Inoltre, l'interruzione di questo importante ciclo metabolico nei mitocondri di cellule tumorali ccRCC potrebbe portare anche ad aumento compensatorio dell'*uptake* di glucosio oltre che ad incremento nell'attività di enzimi glicolitici [107, 108]. Queste scoperte hanno condotto allo sviluppo di agenti farmacologici che hanno come bersaglio molecolare il complesso VHL, come potenziale approccio terapeutico nei pazienti affetti da RCC avanzato [109].

Un'altra via metabolica chiave che sembra essere implicata nella patogenesi RCC è rappresentata dal *pathway* di mTORC1 (*mammalian target of rapamycin complex 1*, *target* terapeutico del RCC approvato dalla FDA (*Food and Drug Administration*) (Figura 11). Il meccanismo molecolare con il quale mTORC1 trasduce i segnali provenienti da riserve di energia [110] ed ossigeno [111], e da fattori di crescita [112] passa attraverso il complesso formato dalle proteine TSC1-2 (*Tuberous sclerosis proteins*). Per queste ultime sono state identificate mutazioni somatiche nel 5% dei casi di ccRCC ed è stato mostrato che le alterazioni a loro carico garantiscono maggiore sensibilità della massa tumorale ai farmaci che hanno come *target* molecolare mTORC1 [113]. Infine, mTORC1 svolge funzioni chiave nell'attivazione di HIF-1 fungendo da collegamento fra la necessità di nutrienti e l'angiogenesi [114-116]. Rimane da chiarire come tali alterazioni di espressione contribuiscano allo sviluppo del RCC.

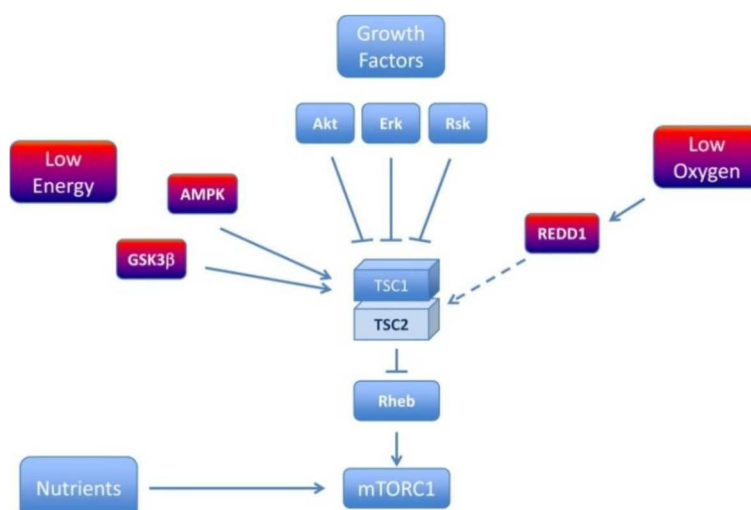


Figura 11. Connessioni del *pathway* di mTOR: in condizioni di ipossia viene bloccata l'espressione di HIF attraverso l'attivazione del *pathway* TSC1.

1.2. LA PROTEOMICA

I *pathway* e i sistemi biologici che sottendono all'insorgenza di malattie o al successo di trattamenti farmacologici possono essere caratterizzati da una complessità tale da rendere inefficaci, in molti casi, degli approcci analitici basati su metodi riduzionistici. Pertanto è preferibile, e necessario, un approccio basato sull'impiego di scienze così dette *omiche*, fra le quali la proteomica [117]. Il termine proteomica è stato coniato, nel 1996, in analogia al termine genomica, disciplina rispetto alla quale la proteomica costituisce il passo successivo, essendo una disciplina più complessa.

La proteomica è l'analisi su larga scala del complesso di proteine codificate ed espresse dall'intero genoma di un organismo (proteoma) in determinate condizioni ambientali. Il proteoma, infatti, rappresenta l'insieme di tutti i possibili prodotti proteici espressi in una cellula, incluse tutte le isoforme e le modificazioni post-traduzionali a cui queste proteine vanno incontro in determinate condizioni fisiopatologiche, non tanto singolarmente, quanto piuttosto come complesso proteico. A differenza del genoma che può essere virtualmente statico, il proteoma è dinamico nel tempo, varia in risposta a fattori esterni e differisce sostanzialmente tra i diversi tipi cellulari di uno stesso organismo. Dunque, la proteomica mira a colmare il *gap* tra la sequenza genomica e la fisiologia della cellula, e a studiare la dinamica dei prodotti genici e delle loro interazioni [118]. Questa scienza può essere a sua volta suddivisa in differenti discipline:

- **Proteomica sistematica:** si occupa di identificare e caratterizzare il proteoma e di costruire *database* al fine di creare mappature di riferimento delle proteine espresse in vari tessuti e fluidi biologici.
- **Proteomica classica** (o di espressione): studia il *pattern* di espressione delle proteine espresse dalle cellule, dai tessuti o presenti all'interno di fluidi biologici.
- **Proteomica funzionale:** studia le interazioni fisiche e funzionali fra le proteine ed i *pathway* nei quali sono coinvolte.
- **Proteomica strutturale:** si occupa di determinare la struttura tridimensionale delle proteine.
- **Proteomica differenziale:** si occupa dell'identificazione di proteine associate ad una malattia attraverso l'alterazione dei loro livelli di espressione e delle loro modificazioni post-traduzionali.

Oltre a ciò, il termine “proteomica” identifica anche l'insieme delle tecnologie e degli approcci sviluppati per lo studio del proteoma [119]. La proteomica ha aperto prospettive di notevole interesse in diversi campi della medicina, della biologia, della biochimica, della farmacologia e della tossicologia in quanto consente di ottenere un quadro completo relativo sia alle proteine presenti all'interno del sistema biologico in studio, che alle loro possibili modificazioni avvenute in seguito a stimoli di diversa natura. L'obiettivo a lungo termine della proteomica è non solo quello di identificare una proteina, ma anche di caratterizzarla, associando ad essa una serie di peculiarità strutturali, biochimiche e funzionali, in modo da poter valutare il quadro complessivo di situazioni fisiologiche o patologiche complesse mediante l'analisi di fluidi biologici

come liquor, siero ed urine, e materiali biologici più complessi come cellule, tessuti e organi [120] (Figura 12).

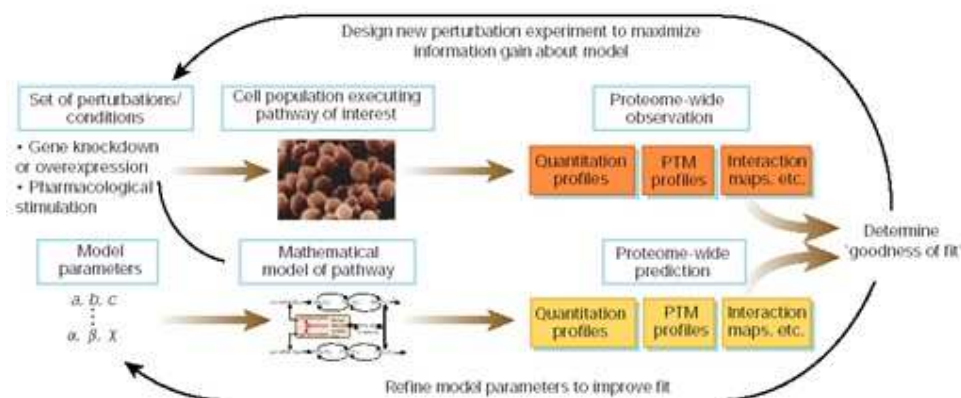


Figura 12. Schema del paradigma dei sistemi biologici

Le proteine di interesse identificate possono essere utilizzate ad esempio come nuovi *biomarkers* o per lo studio della funzione dei corrispettivi geni; inoltre possono essere utilizzate per osservare gli effetti di specifici trattamenti o stimoli ambientali [121].

Un uso più sofisticato della proteomica permette di studiare le proteine che in un determinato sistema possono essere modificate, sempre in termini quantitativi e qualitativi, da fattori endogeni quali modulazione della trascrizione, modificazioni post-traduzionali, *splicing* alternativi e interazioni con altre proteine o con acidi nucleici. Tutti questi fattori modificano tutte quelle caratteristiche di una proteina (per esempio l'emivita) che non sono direttamente prevedibili dallo studio del gene o dell'mRNA da cui la proteina è codificata, rendendo così la proteomica uno strumento unico nelle sue potenzialità [122].

È importante comunque sottolineare che, se per gli acidi nucleici esiste una tecnica (*Polymerase Chain Reaction, PCR*) che consente di amplificare una determinata sequenza genica, per le proteine non si dispone ancora di alcun metodo di espansione selettiva. È necessario, quindi, che le tecniche di analisi della proteomica presentino una sensibilità elevata, in modo da permettere la rilevazione di proteine o peptidi poco rappresentati e sconosciuti, ma utili dal punto di vista della fisiopatologia del distretto in esame, senza che siano "nascosti" da altre proteine abbondantemente espresse [123]. Il campo della proteomica si è notevolmente sviluppato in questi ultimi anni, anche grazie alla tecnologia ed alla ricerca applicata nello studio di tecniche o strumentazioni che hanno permesso di rilevare quantità di proteine nell'ordine dei picogrammi e di separare proteine nell'ordine di $\mu\text{g}/\text{ng}$ oppure ancora di poterle identificare mantenendo l'informazione morfologica. Dunque, è riduttivo associare il termine "proteomica" ad una sola e specifica metodica di rilevazione, poiché essa si avvale di più tecnologie, integrabili tra loro, per comprendere l'espressione e la funzione delle proteine nella regolazione dei sistemi biologici [124].

1.2.1 PROTEOMICA CLINICA

La definizione di proteomica clinica è apparsa per la prima volta sulla rivista “Proteomics – Clinical Applications”, come *applicazione delle tecniche proteomiche al campo della medicina, allo scopo di risolvere un determinato quesito medico all'interno di uno studio clinico* [125]. Sebbene attualmente una delle principali applicazioni della proteomica clinica consista nella ricerca diagnostica e nella scoperta di nuovi biomarcatori, tale disciplina è anche applicabile all'identificazione di nuovi *target* terapeutici, nuovi farmaci e nuovi vaccini, dunque nel trattamento e nella prevenzione di numerose patologie. Negli ultimi anni la proteomica ha assunto sempre più importanza nel campo della ricerca clinica, mirando all'identificazione di marcatori biologici utili al fine di caratterizzare e predire una patologia multifattoriale, come il cancro. Il termine “*marcatore*” è stato introdotto nel 1989 ed è inteso come parametro biologico misurabile, utile ad individuare variazioni dallo stato fisiologico oppure il rischio o la presenza di una patologia. Nel 2001, il *National Institute of Health* statunitense standardizzò la definizione di biomarcatore come “*caratteristica misurabile capace di definire un processo come fisiologico oppure morboso e/o in grado di stabilire la risposta biologica ad un trattamento farmacologico*”.

Un marcatore ideale possiede determinati requisiti di specificità e sensibilità: avere un'utilità clinica superiore rispetto ai marcatori già in uso ed essere presente in un campione biologico ottenibile attraverso procedure non invasive [126]. Dal punto di vista biochimico, i biomarcatori sono spesso proteine, misurabili con anticorpi, la cui presenza o quantità alterata è indice di patologia [127]. Nello scenario delle patologie tumorali umane, sono necessari biomarcatori efficaci sia in campo diagnostico che in campo terapeutico. Nel primo caso, la diagnosi precoce di malattia grazie all'uso di questi segnali permetterebbe di aumentare notevolmente l'aspettativa di vita dei pazienti fino a raggiungere altissime probabilità di sopravvivenza dopo 5 anni; nel secondo caso, i nuovi biomarcatori usati come *target* terapeutici permetterebbero una migliore discriminazione fra cellule cancerose e cellule sane con minore effetto collaterale o indesiderato per i pazienti e maggiore effetto terapeutico [128]. Un ostacolo alla ricerca di biomarcatori è costituito dalle proteine specifiche per altre patologie concomitanti al tumore, come l'infiammazione sistemica. Per questo motivo è spesso necessaria una validazione con modelli *in vivo* per assicurarsi della distribuzione/localizzazione del biomarcatore candidato nel tessuto tumorale [129, 130].

Esistono due strategie per la ricerca di biomarcatori proteici nell'organismo. La prima prevede di effettuare la ricerca direttamente nei tessuti dove la concentrazione di tali *markers* è più elevata, rendendo più facile l'analisi in spettrometria di massa. L'altro valido approccio è costituito, invece, dalla ricerca dei marcatori direttamente in fluidi periferici; lo studio su questi ultimi si adatta perfettamente all'analisi del proteoma in ambito clinico diagnostico in quanto questa tipologia di campioni può essere facilmente ottenuta mediante procedure non invasive ed inoltre consente di studiare le numerose proteine secrete dai tessuti patologici. La scoperta di biomarcatori in tessuti e fluidi biologici umani ha un enorme potenziale. A tal fine, la proteomica differenziale (o comparativa) permette di comparare soggetti sani e pazienti alla ricerca di biomarcatori senza però fornire informazioni circa la quantificazione delle proteine in studio [131]. Purtroppo la diagnosi basata su marcatori singoli è ormai ampiamente riconosciuta come non risolutiva poiché la maggior parte delle malattie deriva da alterazioni poligeniche [132, 133]. Ciò suggerisce, pertanto,

che solo l'utilizzo di una combinazione di biomarcatori possa, con buona sensibilità e specificità, essere realmente utile a questo scopo [134].

Negli ultimi anni abbiamo assistito ad una crescita esponenziale nelle applicazioni di proteomica. Inizialmente, la proteomica si era focalizzata sulla creazione di mappe proteiche di estratti cellulari mediante l'utilizzo della tecnica di *elettroforesi bidimensionale (2D-PAGE)*, che separa le proteine prima in base al punto isoelettrico e poi in base al peso molecolare, consentendo un'elevata risoluzione e riproducibilità. Grazie a questa metodica è possibile evidenziare, con potenti tecniche di analisi d'immagine, le differenze nell'espressione proteica per esempio tra campioni patologici e campioni sani fornendo informazioni sia qualitative che quantitative [135]. Con questa tecnica è altresì possibile studiare una particolare proteina che si considera essere un potenziale marcatore patologico. L'elettroforesi bidimensionale continua quindi ad essere una tecnica di eccellenza per lo studio del profilo di espressione proteica e per l'analisi quantitativa, pur risultando una metodica limitata in campo clinico a causa della laboriosità delle procedure e del dispendio di tempo che il protocollo richiede. Le recenti innovazioni tecnologiche nel campo della proteomica hanno permesso di aumentare la sensibilità e specificità del metodo grazie allo sviluppo della spettrometria di massa (MS) [121, 136]. Infatti, le proteine individuate dalla comparazione delle mappe proteiche, risultate essere sovra- o sotto-espresse in differenti stati fisiologici o in seguito a trattamento, vengono estratte dal gel, al fine di eseguire un'identificazione e una caratterizzazione utilizzando la tecnica chiamata *peptide-mass fingerprinting (PMF)* e la spettrometria di massa [137, 138]. Non tutte le proteine possono però essere identificate mediante PMF. Con il passare del tempo le tecniche utilizzate per lo studio del proteoma/peptidoma hanno avuto un grande sviluppo e hanno cercato di migliorare i tempi di analisi passando da sistemi semi manuali, come le tecniche *in gel*, a sistemi semi-automatici *off-gel* che consentono analisi più veloci, che mirano ad individuare segnali biologici a valle di processi fisio-patologici e che meglio si adattano a studi in ambito diagnostico. A questo proposito, sono stati di grande rilevanza non solo gli avanzamenti strumentali ma anche bioinformatici che hanno contribuito alla messa a punto di programmi in grado di elaborare le centinaia di migliaia di informazioni ottenute con l'analisi strumentale e che hanno reso gli approcci proteomici compatibili con la ricerca clinica.

1.2.2 PROTEIN PROFILING

Recentemente una nuova area di applicazione, che va oltre la proteomica di espressione classica, è sorta con l'uso della spettrometria di massa per il *profiling* di fluidi biologici a scopo diagnostico. L'idea alla base del *profiling* è una diagnosi precoce e più specifica rispetto ai classici approcci come l'ELISA, in quanto determinate condizioni fisio-patologiche dovrebbero essere rilevabili come tracce molecolari lasciate nei fluidi biologici. In contrasto con le tecnologie basate su anticorpi, il *protein profiling* non si basa su un singolo biomarcatore, ma esegue un'analisi multivariata di un intero pannello di diverse specie di segnali proteici. Inoltre tale approccio è possibile anche senza che si abbia una identificazione preventiva di quelli che potrebbero essere peptidi biomarcatori. In questo campo, l'accuratezza e la sensibilità della spettrometria di massa devono essere combinati con le esigenze di analisi ad alto rendimento, caratterizzate da velocità e

automazione [139]. Dall'analisi mediante spettrometria di massa dei campioni in esame si ottiene uno spettro proteico/peptidico, dove ciascun picco corrisponde ad una determinata proteina/peptide. In questo modo, è disponibile per ogni analita un profilo spettrale specifico e, mediante confronto dei diversi spettri, è possibile notare le eventuali differenze di espressione tra il proteoma dei soggetti sani e quello dei pazienti affetti da tumore [121]. Dunque il *protein profiling* permette di ottenere profili caratteristici, formati da un gran numero di biomarcatori putativi, che si comportano da impronte digitali di determinate condizioni fisio-patologiche e/o farmacologiche.

È comunque importante sottolineare che i campioni biologici utilizzati sono costituiti da una matrice con un'elevata complessità, quindi è essenziale eseguire un pre-frazionamento (frazioni subcellulari, cromatografia [140], estrazione sequenziale, ecc.) al fine di effettuare un'analisi proteomica robusta e di alta qualità mediante spettrometria di massa. La purificazione dei fluidi biologici ha lo scopo di eliminare ad esempio i sali e le piccole molecole che potrebbero interferire con l'analisi spettrometrica [141]; inoltre, data la notevole quantità di specie molecolari che costituiscono il proteoma dei fluidi biologici, si possono avere effetti di soppressione del segnale sulle proteine di interesse [142]. Due sono le tecniche automatizzabili che sono state sviluppate per individuare potenziali marcatori biologici all'interno dei fluidi corporei in grado di caratterizzare e predire una patologia e per la creazione di profili proteici che permettano una diagnosi precoce e un chiarimento dei meccanismi di insorgenza della malattia [143]: SELDI e frazionamento con biglie magnetiche funzionalizzate.

La **tecnologia SELDI** (*surface-enhanced laser desorption/ionization*), si basa sulla combinazione di un sistema di pre-frazionamento del campione tramite superfici funzionalizzate (*ProteinChip*) con l'analisi MALDI-TOF (*Matrix Assisted Laser Desorption Ionization-Time of Flight*). Questa tecnologia utilizza un *target* modificato, dotato di una superficie arricchita con molecole ad affinità biochimica elevata per le proteine/peptidi del campione. Esistono diverse superfici di pre-frazionamento che permettono di catturare specifici gruppi di proteine e peptidi: a fase inversa, a fase normale, a scambio ionico, ad affinità metallica oppure sfruttando substrati biologici come anticorpi, proteina A, recettori, ligandi, enzimi e acidi nucleici (Figura 13).

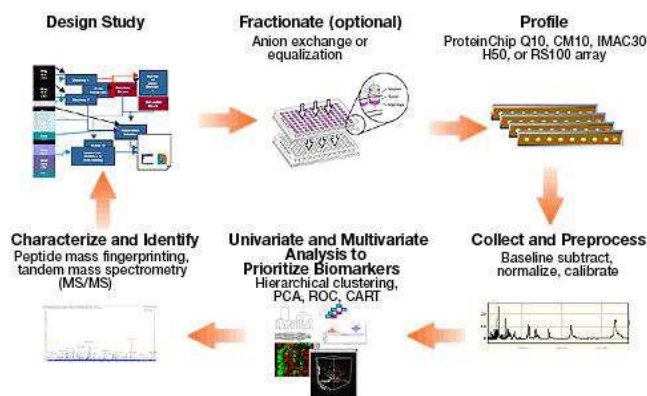


Figura 13. Principio della tecnologia *ProteinChip* SELDI-TOF.

È stato osservato che le piattaforme SELDI sono dotate di alta sensibilità e sono relativamente facili da usare [144]. Le analisi SELDI presentano alcuni svantaggi fra cui la minore risoluzione e riproducibilità rispetto ad altre tecnologie come le piattaforme MALDI [145].

Una valida alternativa al SELDI consiste nell'accoppiamento della spettrometria di massa MALDI-TOF con un sistema cromatografico su fase solida del campione basato sull'utilizzo di superfici o **biglie magnetiche funzionalizzate** che permettono l'arricchimento specifico di proteine e peptidi, applicabile a tutti i fluidi biologici [140, 146, 147]. A differenza dell'approccio SELDI, questo metodo permette di individuare differenze nei profili proteici utilizzando solo una frazione del campione per l'analisi al MALDI-TOF, cosicché una parte dell'analita resti disponibile per ulteriori approcci analitici con altre strumentazioni e altre tecnologie (ad esempio nLC-ESI-MS/MS) [148].

A tale scopo, esistono diverse tipologie di biglie magnetiche [119]:

- Biglie magnetiche C₃, C₈ e C₁₈, che permettono una separazione a fase inversa.
- Biglie magnetiche IMAC, grazie alle quali si esegue una cromatografia di "affinità ionica" con metalli immobilizzati, quali Fe e Cu.
- Biglie magnetiche a scambio cationico (WCX) e anionico (WAX), che permettono invece interazioni ioniche.
- Biglie funzionalizzate con lectine (es. concavalina A).
- Biglie magnetiche attivate con anticorpi, che vengono solitamente utilizzate nelle analisi focalizzate su una singola specifica proteina.

Tutte le procedure possono essere automatizzate mediante l'utilizzo di una piattaforma robotica. Le proteine e i peptidi estratti con questa tecnologia vengono successivamente analizzati mediante spettrometria di massa MALDI-TOF e gli spettri ottenuti processati con opportuni strumenti statistici che consentono di individuare diversi *pattern* proteici specifici della patologia (Figura 14) [149].

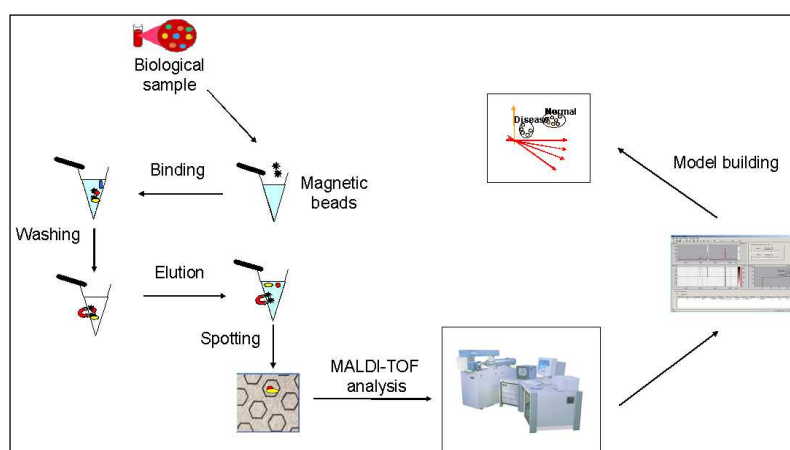


Figura 14. Schema generale della tecnologia del frazionamento con biglie magnetiche funzionalizzate associato alla spettrometria di massa MALDI-TOF.

Questi due approcci, appena descritti, rappresentano soluzioni interessanti per la ricerca di marcatori proteici nei fluidi biologici in quanto consentono di ovviare ai problemi legati alla 2D-E. Il successo in ambito clinico è stato permesso da una caratteristica unica condivisa da queste due tecniche: l'elevato *throughput*. Infatti, la rapidità d'analisi e l'*high-throughput* le rendono compatibili con l'utilizzo clinico.

L'aspettativa, dunque, è quella di poter utilizzare la proteomica nell'individuazione di nuovi marcatori biologici per la diagnosi precoce dei tumori [150]. Alcuni potenziali biomarcatori sono già stati identificati e tra questi vi sono, ad esempio, il *PSA*, associato al tumore alla prostata e dal ruolo ancora molto discusso, oppure il *CA-125* per il tumore all'ovaio [151]. Tuttavia, da molti anni ormai, risulta difficile trovare singole proteine coinvolte in una patologia specifica; l'obiettivo attuale è quindi cercare di confrontare l'intero proteoma/peptidoma e identificare più marcatori biologici, analizzando le differenze con sofisticati sistemi di analisi statistica [142]. In supporto a questo complesso lavoro di analisi, sono stati creati *software* ed algoritmi matematici che facilitano la visualizzazione e l'elaborazione statistica dei dati, creando potenziali modelli diagnostici da utilizzare in campo clinico [152].

Sono già descritte, in letteratura, alcune interessanti applicazioni di queste tecnologie, come ad esempio nel carcinoma della mammella [153].

1.2.3 IL PEPTIDOMA

I peptidi endogeni sono polimeri amminoacidici che possono variare in dimensione da dipeptidi a molecole di circa 20 kDa naturalmente presenti *in vivo* [154]. Essi possono essere il risultato dell'azione proteolitica di enzimi coinvolti nella regolazione del *turn-over* di specifiche proteine oppure essere prodotti a partire da un mRNA codificante per peptidi e svolgono differenti funzioni biologiche come nel caso degli ormoni, dei neurotrasmettitori e degli agenti antimicrobici [155]. I peptidi endogeni sono implicati in un vasto numero di processi molecolari e di conseguenza sono di grande interesse in campo biomedico. Infatti, è possibile valutare tramite differenti tecnologie analitiche come i livelli di tali peptidi siano regolati in condizioni fisiologiche e patologiche e identificare così potenziali biomarcatori per le principali malattie umane (Figura 15) [156].

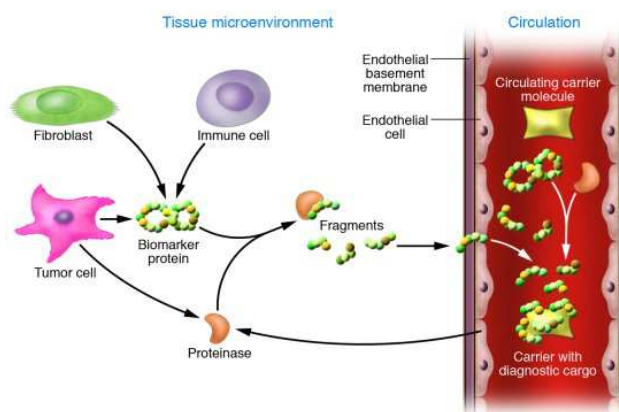


Figura 15. Il peptidoma. Frammenti proteici circolanti generati da un dato microambiente tissutale possono servire come biomarcatori proteici diagnostici.

Recentemente, lo sviluppo delle metodiche associate alla spettrometria di massa ha garantito nuove opportunità nello studio del peptidoma. I cambiamenti nell'attività enzimatica endogena, ad esempio, possono essere studiati sperimentalmente attraverso l'identificazione di *patterns* di segnali generati da frammenti proteolitici associati ad una patologia o valutando in modo diretto la regolazione fisio-patologica dell'attività di specifiche proteasi [157]. Rispetto al proteoma, il peptidoma, fornisce una visione più ampia dell'intero *pool* peptidico di un campione biologico [158, 159].

Nonostante l'efficacia del *profiling* peptidico ottenuto mediante spettrometria di massa MALDI-TOF nella scoperta di cluster di segnali peptidici discriminanti, a pochi di questi peptidi endogeni è possibile assegnare un'identità [160]. Sebbene il significato diagnostico/prognostico di questi segnali sia indipendente dalla determinazione della loro sequenza amminoacidica, l'identificazione di tali potenziali biomarcatori potrebbe chiarire il loro ruolo biologico incrementando le conoscenze a nostra disposizione sulle funzioni molecolari, sul grado di degradazione endogena delle proteine da cui derivano e sul ruolo fisiologico nel quale sono implicati [147]. Inoltre, in assenza di informazioni sulla sequenza amminoacidica di questi peptidi, il loro utilizzo sarebbe limitato alla tecnologia con la quale sono stati individuati e ogni ulteriore passaggio di validazione, ad esempio mediante saggio immunologico, e di applicazione clinica potrebbe essere precluso. Dunque, potrebbe essere vantaggioso combinare approcci di *profiling* proteico con tecniche di identificazione di potenziali biomarcatori. Dal momento che la frammentazione dei precursori isolati è spesso difficoltosa per le analisi MALDI principalmente a causa della loro tendenza a formare ioni monocarica [161], è possibile utilizzare approcci di peptidomica ad alta sensibilità e accurati per il sequenziamento peptidico come LC-MS/MS o CE (*Capillary Electrophoresis*)-MS/MS o una combinazione di queste due tecniche [147, 162, 163]. L'accurato allineamento delle masse permette una corretta assegnazione delle sequenze amminoacidiche identificate con i picchi rilevati tramite MALDI. Per migliorare il potere identificativo è poi possibile modificare parametri di acquisizione in MS/MS aumentando ad esempio l'energia di frammentazione e il tempo di scansione [164] o utilizzare tecniche di frammentazione alternative al CID, ad esempio nel caso dell'ETD (*Electron transfer dissociation*) [165]. Nonostante i numerosi vantaggi ottenibili dalle piattaforme CE-MS quali la sensibilità e la riproducibilità rispetto alle tecniche LC accoppiate a spettrometria di massa, l'uso dei capillari elettroforetici presenta una forte limitazione nella capacità di caricamento del campione se confrontato con altre strategie analitiche come la cromatografia liquida associata a spettrometria di massa LC-MS. Infatti, mentre nelle colonne cromatografiche LC possono essere iniettati milli-litri (mL) di soluzione, nei capillari elettroforetici possono esserne analizzati generalmente 10-100 μ L. Ciò può, purtroppo, diminuire fortemente la sensibilità dello spettrometro di massa al quale l'apparato CE è accoppiato [145]. Le identificazioni di peptidi ottenuti dall'impiego delle due metodiche risultano essere comunque complementari [147, 162, 163].

Ogni risultato ottenuto tramite un metodo analitico necessita di una seconda tecnica differente per essere validato. Allo stesso modo, i dati sui peptidi/proteine differentemente espressi fra soggetti controllo e pazienti affetti da una specifica patologia, ottenuti attraverso queste tecnologie di purificazione e analisi, necessitano di essere validati al fine di poterne definire l'efficacia come biomarcatori di malattia.

Generalmente, la validazione dei dati di spettrometria di massa avviene tramite rilevamento di proteine/molecole *target* con anticorpi. In alcuni casi, però, questo tipo di validazione non è possibile in quanto i saggi immunologici possono avere problemi di specificità se usati per il rilevamento di peptidi endogeni o presentare fenomeni di cross-attivazione che potrebbero falsare i risultati ottenuti. Recentemente, è stato proposto un nuovo approccio di validazione dei dati, ottenuti con tecnologie MALDI, tramite una differente sorgente ionica quale l'ESI [166]. In particolare, in un lavoro del 2012 sono stati confrontati i dati provenienti da MALDI *profiling* con i risultati ottenuti da analisi nLC-ESI-MS/MS quantitativa, con approccio *label-free* (LF). Questo metodo è stato utilizzato su campioni di siero pre-frazionati con biglie magnetiche C18, provenienti da soggetti sani e da pazienti con carcinoma a cellule renali (RCC); per ciascun frammento identificato in ESI e presente nel profilo MALDI, tramite allineamento accurato delle masse, sono stati messi a confronto i *ratio* peptidici (rapporto dei valori di abbondanza relativa dei peptidi provenienti dal siero di soggetti malati e sani). In questo confronto, la maggior parte dei rapporti peptidici presentavano valori di abbondanza relativa coerenti e l'analisi statistica dei dati allineati ha mostrato l'assenza di differenze statisticamente significative fra i risultati delle due diverse piattaforme ($p > 0.05$) [166].

1.2.4 URINE

La proteomica dei biofluidi sta diventando negli ultimi anni un importantissimo campo di ricerca di biomarcatori grazie anche allo sviluppo delle piattaforme tecnologiche. In particolare, l'“Urinomica”, ovvero lo studio del proteoma urinario, sta subendo una progressiva crescita in questo settore. Le proteine contenute nelle urine sono per lo più derivate dall'ultrafiltrazione glomerulare del sangue da parte del rene oppure sono secrete dai principali organi dell'apparato urinario. Un uomo adulto sano espelle, attraverso l'apparato urinario, fino a 150 mg di proteine e da 1 a 4 g di peptidi al giorno [167].

In condizioni fisiologiche è stato stimato che il contenuto di proteine urinarie proviene circa il 70% dal rene e dal tratto urinario, mentre il restante 30% proviene dal plasma [168] (Figura 16).

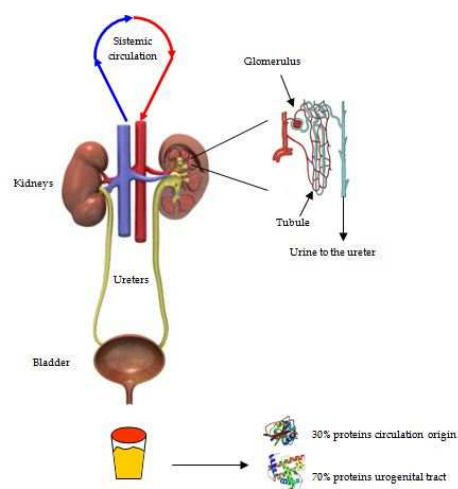


Figura 16. Origine delle proteine presenti nelle urine.

È stato dimostrato che le concentrazioni e la composizione di questi peptidi e proteine possono essere influenzate dalle condizioni di salute dei soggetti; di conseguenza, il proteoma urinario può essere considerato in qualche modo la rappresentazione fisiologica e/o patologica dei tessuti urologici. Al contrario di quanto ritenuto in passato, le urine dunque rappresentano una fonte ricchissima di informazioni da usare nella *biomarker discovery*, in particolare per i tumori urologici. Esse possono essere facilmente disponibili senza l'impiego di personale medico specializzato, in modo assolutamente non invasivo ed in grandi quantità. Inoltre, è stato dimostrato che il proteoma urinario è molto più stabile di quanto si pensasse in passato, soprattutto grazie alla standardizzazione dei protocolli di raccolta.

L'urina del mattino, infatti, permane per un tempo sufficientemente lungo all'interno della vescica prima di essere raccolta permettendo un processamento completo da parte delle proteasi endogene e questo garantisce una minore complessità e una composizione più omogenea. I fattori che potrebbero influire sulle variazioni di concentrazione proteica in questo fluido sono soprattutto legate alla dieta e all'esercizio fisico. Uno studio del 2011 condotto da Sauer *et al.* mediante tecnica *gel-based* su campioni urinari di volontari sani ha dimostrato che il 10% circa del proteoma urinario può essere considerato "*proteoma personale*" a causa della sua fortissima variabilità in termini di composizione e abbondanza relativa [169]. A questi inconvenienti si può ovviare attraverso la standardizzazione della raccolta del campione [170], con l'utilizzo di uno standard interno a concentrazione nota e/o alla normalizzazione dei rapporti di abbondanza relativa degli analiti con la creatinina urinaria, una proteina la cui concentrazione urinaria è nota per essere fortemente dipendente da sesso, età e razza dei soggetti da cui derivano i campioni d'interesse [171, 172]. Confrontando l'uso delle urine nella *biomarker discovery* con altri biofluidi come il siero possiamo osservare similarità e differenze:

- per quanto riguarda le similarità, entrambi i fluidi presentano un ampio *range* dinamico. Ciò è principalmente dovuto alla forte presenza di uromodulina, una proteina tipica di questo tipo di fluido che può sopprimere i segnali di proteine meno abbondanti; questo problema è, però, molto meno accentuato nelle urine rispetto al siero;
- per quanto riguarda le differenze, le urine presentano una più elevata presenza di sali rispetto al siero e ciò potrebbe interferire con l'analisi in spettrometria di massa; per questo motivo è sempre necessario un passaggio preliminare di desalificazione del campione.

Le potenzialità del proteoma/peptidoma urinario umano nel mettere in luce diversi processi patologici a carico dei reni e dell'apparato urogenitale sono già state evidenziate da diversi studi [147, 173]. L'urina umana contiene un proteoma particolarmente ricco ed eterogeneo che può rappresentare un'importante fonte di potenziali biomarcatori da utilizzare per caratterizzare malattie e tumori che alterano la fisiologia renale come il carcinoma a cellule renali [157]. Per queste ragioni le urine sono definite come "fluido biopsia" del rene e del tratto urogenitale in quanto molti dei cambiamenti nel rene e del tratto urogenitale possono essere rilevati attraverso lo studio del proteoma urinario. La NHGRI (*National Human Genome Research Institute*) ha affermato che lo studio della proteomica dei fluidi è uno degli strumenti più promettenti per lo sviluppo di strumenti non invasivi per la diagnosi precoce di malattie umane [174]. Ad oggi, migliaia

di proteine e peptidi diversi sono stati sequenziati nell'urina umana, fornendo in tal modo informazioni sempre più complete sul contenuto endogeno urinario e suggerendo la necessità di ulteriori studi, su specifiche patologie, al fine di un loro potenziale utilizzo nella pratica clinica. In particolare, da un punto di vista proteomico, le urine possono essere divise in tre compartimenti principali: proteine solubili, proteine associate agli exosomi e peptidi endogeni e ciascuno di questi compartimenti racchiude potenziali informazioni utili in campo diagnostico.

1.2.5 PEPTIDOMA/PROTEOMA URINARIO DEL RCC

Numerosi studi sono stati condotti su campioni urinari per la scoperta di biomarcatori del carcinoma a cellule renali [147, 175-177], allo scopo di ottenere un valido strumento diagnostico non invasivo. In questo contesto, una serie di proteine sono state nel tempo indicate come potenziali biomarcatori dell'RCC, tra cui le citocheratine, la vimentina, PAX2, PAX8, HIF(s), lipocaina associata alla gelatinasi dei neutrofilo, VEGF(R), CD10, caderina E, caderina specifica del rene, claudina 7, claudina 8, alfa metilacil coenzima A racemasi, TFE3, MMP-2 e MMP-9, glicoproteina B7H1, famiglia di inibitori della apoptosi e l'anidrasi carbonica IX; ma, nonostante le promettenti aspettative, ad oggi nessun biomarcatore viene usato nella routine clinica a scopo diagnostico o prognostico per il carcinoma a cellule renali.

Per studiare come il proteoma urinario fosse influenzato dalla presenza del carcinoma a cellule renali sono stati adottati approcci proteomici tradizionali, come l'elettroforesi bidimensionale su gel di poliacrilammide (2DE) accoppiata alla spettrometria di massa MALDI-TOF, e/o più innovativi, come la cromatografia liquida accoppiata alla spettrometria di massa tandem (LC-MS/MS). Dalla letteratura risulta chiaramente evidente la complementarità dei diversi approcci analitici; infatti è stato dimostrato che quando si combinano approcci *in gel* con approcci *off gel* i risultati sono molto più robusti e aprendo nuove opportunità di ricaduta clinica (Figura 17).

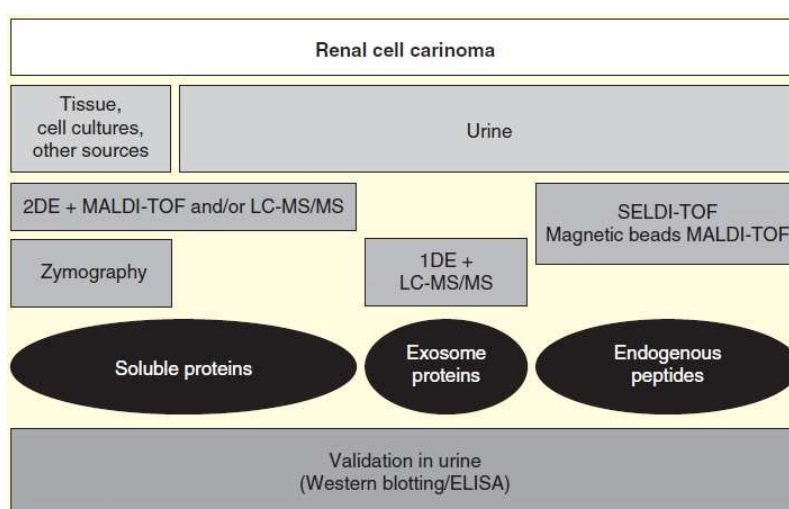


Figura 17. Panoramica sui diversi approcci utilizzati per lo studio del carcinoma renale [178].

Tali approcci sono stati usati sia per lo studio a livello del proteoma (Tabella 4) che di peptidoma (Tabella 5) correlato al carcinoma a cellule renali.

Studio (anno)	Tipo di pazienti	Campione di urine	Approccio analitico	Proteine di interesse	Referenza
Pieper <i>et al.</i> (2004)	ccRCC	<ul style="list-style-type: none"> • Pre-intervento • Post-intervento 	2DE in combinazione con: <ul style="list-style-type: none"> • MALDI-TOF • LC-MS/MS 	Retinol binding protein, carbonic anhydrase I, β 2 microglobulin MASP-2, Kininogen	[179]
Vasudev <i>et al.</i> (2009)	<ul style="list-style-type: none"> • ccRCC • Patologie urologiche benigne • Soggetti sani 	<ul style="list-style-type: none"> • Pre-intervento 	2DE in combinazione con MALDI-TOF on conditioned media Western blotting/validazione con ELISA in urine di pazienti	<ul style="list-style-type: none"> • \uparrow Cathepsin D nelle urine di pazienti ccRCC: • Nessun valore diagnostico • Correlazione con la prognosi 	[180]
Minamida <i>et al.</i> (2011)	<ul style="list-style-type: none"> • RCC • Altre patologie renali • Soggetti sani 	<ul style="list-style-type: none"> • Pre-intervento 	RP-HPLC + agarose-2DE in combinazione con LC-MS/MS su fluido cistico di pazienti RCC. Validazione con Western Blot su urine di pazienti.	\uparrow 14-3-3 protein b/a in urine di pazienti RCC: potenziale biomarker per diagnosi precoce	[176]
Morrissey <i>et al.</i> (2010, 2013, 2014)	<ul style="list-style-type: none"> • ccRCC • RCC papillare • Patologie non renali • Altre patologie renali • Soggetti sani 	<ul style="list-style-type: none"> • Pre-intervento • Post-intervento 	Western Blotting	<ul style="list-style-type: none"> • \uparrow Aquaporin-1 and Perilipin-2 In RCC a cellule chiare e papillare: • Discriminazione di RCC a cellule chiare e papillare da altre lesioni renali • Valore prognostico: correlazione con stage e dimensione tumorale. 	[181-183]
Di Carlo (2012, 2013, 2014)	<ul style="list-style-type: none"> • ccRCC • Oncocitoma • Soggetti sani 	<ul style="list-style-type: none"> • Pre-intervento 	Zimografia su gelatina ELISA	MMP-2, MMP-9, NGAL: No correlazione con tipo, stage o grade tumorali	[184-186]
Zhang <i>et al.</i> (2013)	<ul style="list-style-type: none"> • ccRC • RCC papillare • RCC cromofobo 	<ul style="list-style-type: none"> • Pre-intervento • Post-intervento 	ELISA	\uparrow KIM-1 in urine pre-intervento di pazienti RCC a cellule chiare e papillare: potenziali marcatori prognostici	[187]
Sandim <i>et al.</i> (2015)	ccRCC	Pre-intervento	1DE LC-MS/MS; 1D LC-MSE; 2DE MALDI-TOF/TOF; LC-MS/MS	<ul style="list-style-type: none"> • \downarrow kininogen-1, uromodulin, apolipoprotein D, polyubiquitin, and CD5. • \uparrow apolipoprotein A, fibrinogen, and haptoglobin 	[188]

Tabella 4. Panoramica delle procedure analitiche utilizzate per indagare il proteoma urinario di pazienti affetti da carcinoma a cellule renali.

Studio (anno)	Tipo di pazienti	Campione di urine	Approccio analitico	Proteine di interesse	Referenza
Rogers <i>et al.</i> (2003)	<ul style="list-style-type: none"> • ccRCC • RCC benigni • Soggetti sani 	Pre-intervento	SELDI-TOF	No identificazioni	[189]
Wu <i>et al.</i> (2008)	<ul style="list-style-type: none"> • ccRCC • Altre patologie renali • Soggetti sani 	Pre-intervento	SELDI-TOF con IMAC-Cu ProteinChip	No identificazioni	[190]
Alves <i>et al.</i> (2013)	<ul style="list-style-type: none"> • ccRCC • Soggetti sani 	Pre-intervento	SELDI-TOF (CM10 chip) SDS-PAGE combinato con MALDI-TOF/TOF	↓ 22 picchi in pazienti: no dati su performance diagnostiche ↓ MASP-2 e proteina K12	[191]
Sim <i>et al.</i> (2009)	<ul style="list-style-type: none"> • ccRCC • donatori viventi di rene 	<ul style="list-style-type: none"> • Pre-intervento • Post-intervento 	SELDI-TOF LC-MS/MS	65 picchi differentemente espressi ↓ β 2 microglobulin e bdefensin-1 in urine post-intervento di pazienti RCC e donator: correlazione con rimozione rene	[192]
Bosso <i>et al.</i> (2008)	<ul style="list-style-type: none"> • ccRCC • Soggetti sani 	• Pre-intervento	Biglie magnetiche C8/MALDI-TOF	Cluster di 3 ioni in grado di discriminare pazienti da controlli, sensibilità del 100% per pazienti allo stadio 1: potenziali biomarker per la diagnosi precoce di RCC. ↓ frammento di Tamm-Horsfall Protein in pazienti RCC	[147]
Frantzi <i>et al.</i> (2014)	<ul style="list-style-type: none"> • ccRCC • Soggetti sani • Altre patologie renali 	• Pre-intervento	CE-MS	Cluster di 86 peptidi in grado di discriminare I pazienti dai controlli, 80% sensibilità e 87% specificità. Validazione su 1153 campioni di urine	[193]

Tabella 5. Panoramica delle procedure analitiche utilizzate per indagare il peptidoma urinario di pazienti affetti da carcinoma a cellule renali.

1.2.6 PROTEOMICA QUANTITATIVA

La quantificazione è uno strumento fondamentale per la comprensione degli effetti trascrizionali, traduzionali e post-traduzionali che condizionano la produzione e la funzione delle proteine [194]. Nell'ambito della proteomica quantitativa la Spettrometria di Massa (MS) riveste un ruolo importante per l'identificazione e la quantificazione delle proteine differenzialmente espresse in stadi fisiologici o patologici [195]. Con il passaggio dalle tecniche *gel-based* a quelle *gel-free* si è reso necessario sviluppare nuovi metodi di quantificazione di miscele complesse di peptidi e non più di singole proteine [196].

Tale quantificazione può avvenire principalmente in due modalità:

- Quantificazione relativa: i campioni sono analizzati separatamente in spettrometria di massa e gli spettri ottenuti sono confrontati per determinare l'abbondanza di ogni peptide nelle condizioni oggetto di studio (ad esempio confronto controllo vs patologico o trattato)
- Quantificazione assoluta (*Absolute QUAntification*, AQUA): viene determinata l'abbondanza assoluta di una proteina o peptide. A differenza della quantificazione relativa, in genere richiede una precedente valutazione qualitativa in modo da definire le proteine/peptidi di interesse. La concentrazione assoluta di un peptide/proteina può essere determinata tramite dosaggio con aggiunta al campione di uno standard interno, marcato o non, che mimi il comportamento della loro controparte endogena durante il processo analitico [197] o tramite tecniche che richiedono la marcatura dei campioni, come nel caso della metodica VICAT (*Visible isotope-code affinity tags*) [198]. Recentemente sono state implementate nel campo della proteomica metodiche basate sul principio della diluizione isotopica: in questi approcci al campione vengono aggiunte quantità definite di analoghi marcati con isotopi di peptidi proteolitici specifici (strategie AQUA e QconCAT) o di proteine (strategia PSAQ) [199, 200]. Il principale problema di questi ultimi approcci riguarda la difficoltà nell'ottenere tali standard interni, soprattutto nel caso in cui sia necessario quantificare simultaneamente un elevato numero di proteine.

Nonostante i numerosi vantaggi di entrambe le forme di quantificazione, negli ultimi anni la proteomica quantitativa mediante spettrometria di massa si è principalmente focalizzata sulla determinazione quantitativa differenziale dell'espressione proteica e non sulla misurazione assoluta poiché molte applicazioni relative alla scoperta di biomarcatori, *target* farmacologici e alla valutazione di eventi segnale sono legate principalmente alla variazione della concentrazione delle proteine piuttosto che alla determinazione della loro abbondanza. Le proteine possono essere presenti nel campione in concentrazioni molto differenti (ampio *range* dinamico), quindi è fondamentale l'implementazione costante di tecnologie che permettano una quantificazione relativa unita ad una calibrazione molto fine dei protocolli sperimentali per consentire che tutte le proteine possano essere risolte ed evidenziate contemporaneamente.

Esistono due principali approcci di quantificazione, sia relativa che assoluta (Figura 18):

- Approcci *Label-based*, si basano sull'analisi di una miscela di differenti campioni marcati con tag isotopici o fluorescenti; la quantificazione dei peptidi è eseguita direttamente sullo spettro MS

o sullo spettro MS/MS in quanto la presenza del marcatore causa uno spostamento di m/z per i segnali dello spettro di massa relativi ai peptidi a cui è legato.

- Approcci *Label-free*, confidano su una riproducibilità a lungo termine del segnale peptidico e una linearità dell'intensità/area del segnale rispetto alla concentrazione molecolare delle specie di interesse.

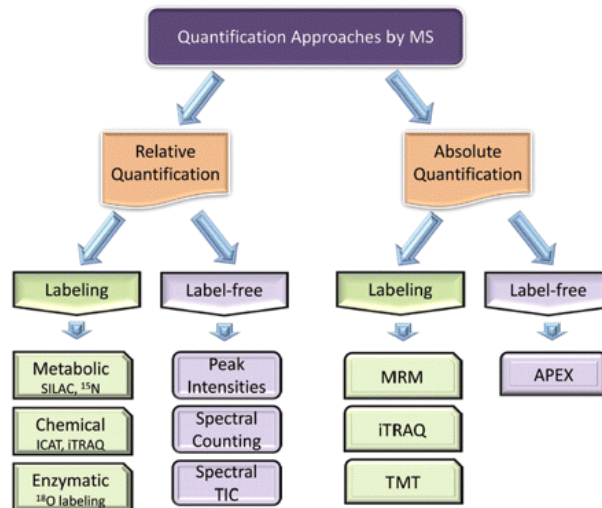


Figura 18. Schema riassuntivo degli approcci di quantificazione Relativa (a sinistra) e Assoluta (a destra)

Le principali metodiche di quantificazione *label-based* sono:

- marcatura isotopica metabolica: prevede l'incorporazione di marcatori isotopici durante i processi di metabolismo cellulare e sintesi proteica; uno delle più usate è la SILAC (*Stable Isotope Labelling by amino acids in cell culture*) (Figura 19) [201].
- marcatura chimica: i più usati sono la tecnologia ICAT (*Isotope-Coded Affinity Tag*) (Figura 19), ICPL (*Isotope Coded Protein Labelling*), ITRAQ (*Isobaric Tags For Relative And Absolute Quantification*) e l'incorporazione di O^{18} durante taglio enzimatico della proteina.

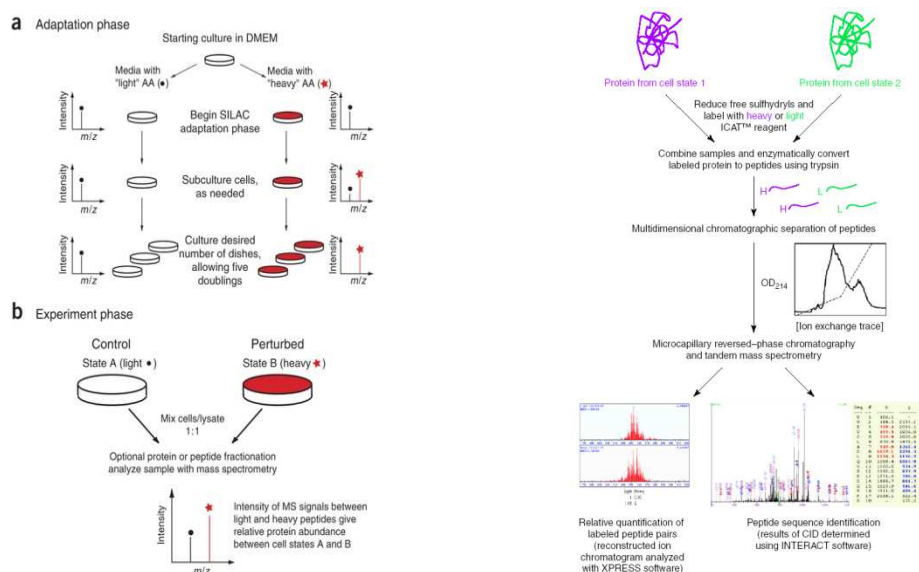


Figura 19. Approcci SILAC (a sinistra) e ICAT (a destra) a confronto.

Le metodiche *label-based* conferiscono notevoli vantaggi in quanto la pre-miscelazione dei due campioni studiati garantisce una minima variabilità sperimentale. Molti fattori, però, come il numero limitato di campioni analizzabili contemporaneamente, la complessità delle procedure sperimentali, l'efficienza di marcatura e i costi elevati dei reagenti limitano l'applicabilità di queste tecniche nelle analisi *high throughput*. Una valida alternativa è rappresentata dalle tecniche *label-free*.

1.2.6.1 TECNICHE DI QUANTIFICAZIONE SENZA MARCATURA O LABEL-FREE

Diversamente dalle altre metodiche di quantificazione, le tecniche *label-free* non richiedono la marcatura isotopica dei campioni in esame [202-205], sono impiegate sia negli studi di proteine che di metaboliti, oltre che applicabili a tutti i fluidi biologici e tessuti.

Questo tipo di quantificazione relativa si basa principalmente su tre categorie di misurazione:

1. variazioni nello *spectral total-ion-current* (TIC);
2. variazioni nell'intensità dello ione, misurabili attraverso l'area o l'altezza del picco cromatografico, che sono stati osservati correlare linearmente con la concentrazione proteica anche in matrici altamente complesse;
3. variazioni nello *spectra counting* di protein identificate dopo analisi MS/MS.

Il metodo di quantificazione relativa *label-free* più utilizzato è lo *spectral counting* (SC), in cui l'abbondanza relativa di una proteina è calcolata in base al numero degli spettri MS/MS corrispondenti ai peptidi assegnati alla proteina. Un'altra strategia è basata sulla somma della corrente ionica totale (TIC) degli spettri MS/MS assegnati ai peptidi di una determinata proteina [206]. In studi recenti, l'intensità dello

ione precursore (MS1) è stata utilizzata come misura quantitativa [207]. La metodica SC è semplice da utilizzare, riproducibile ed altamente sensibile per le piccole variazioni nell'abbondanza proteica. Ciononostante, è considerata inaccurata per quanto riguarda il calcolo del *fold change* delle proteine poco abbondanti, in quanto è influenzata dalla dimensione delle proteine [206]. L'utilizzo della TIC permette di analizzare un range dinamico maggiore rispetto alla metodica precedente, di avere una maggiore sensibilità per piccole variazioni, ma risulta essere poco riproducibile e risente della saturazione del segnale delle proteine altamente abbondanti. La quantificazione basata sull'utilizzo dell'intensità dello ione precursore è altamente accurata, ma necessita di misurazioni ad alta risoluzione al fine di poter discriminare picchi vicini [208]. Dato che tutti questi metodi presentano vantaggi e limitazioni, il loro utilizzo combinato dovrebbe aumentare l'affidabilità delle variazioni proteiche identificate [206, 208].

Recentemente, la quantificazione label-free è stata applicata per determinare la concentrazione proteica utilizzando il *protein abundance index* (PAI), che è definito come il rapporto fra il numero di peptidi identificati e il numero di peptidi triptici teorici totali osservati per ciascuna proteina [127].

Inoltre, una variante della strategia *spectral counting* chiamata *absolute protein expression profiling* (APEX) permette la misurazione della concentrazione proteica sfruttando la proporzionalità tra l'abbondanza proteica e il numero di peptidi osservati [209]. Questo approccio utilizza appropriati fattori di correzione che rendono la frazione del numero atteso di peptidi proporzionale a quella del numero di peptidi osservati e viceversa. Il metodo di quantificazione APEX inoltre restituisce uno *score* di abbondanza calcolato a partire dall'insieme degli spettri di massa di un peptide osservato e corretto per un fattore che stima *a priori* il numero di peptidi unici attesi per la proteina a cui appartiene il peptide [127]. La quantificazione *label-free* è di facile applicazione, ha basso costo ed è automatizzabile. Inoltre, essa presenta il grande vantaggio di poter studiare più condizioni in un unico esperimento ed è altamente efficiente se si utilizza uno spettrometro di massa ad alta risoluzione munito di analizzatori quali TOF, FT-ICR o *Orbitrap* in modo da garantire una migliore rilevazione dell'intensità ionica [210].

Molti studi hanno utilizzato la quantificazione di tipo *label-free* nella ricerca di biomarcatori: ad esempio, uno studio di Fatima N. *et al.*, mediante *profiling* proteomico di campioni di siero derivato da pazienti con poliposi adenomatosa ha permesso di rilevare proteine modulate dal farmaco *celecoxib* [211]. Inoltre, l'analisi LC-MS/MS condotta su un estratto proteico derivato da melanoma ha permesso l'identificazione di centinaia di proteine che correlano in modo significativo con la metastatizzazione [212]. In uno studio di Levin Y. *et al.*, la quantificazione *label-free* tramite LC-MS/MS condotta su 22 campioni sierici di pazienti con schizofrenia e 33 soggetti controllo ha permesso l'identificazione di proteine seriche differenzialmente espresse ed evidenziato la presenza di nuovi potenziali biomarcatori [213].

1.2.7 PROTEOMICA INTEGRATIVA

Il progresso scientifico sia delle tecnologie che degli strumenti ha permesso lo sviluppo di piattaforme sperimentali ad *high throughput* e la nascita delle scienze “-omiche” in cui una grande mole di indagini molecolari sono condotte in un unico esperimento [214], come nel caso della genomica (studio di tutti i geni di un organismo), della trascrittomica (studio della composizione di tutti gli RNAs di un tessuto), della proteomica (studio della composizione in proteine di un tessuto o un compartimento dell'organismo), e della metabolomica (studio dei piccoli metaboliti che caratterizzano lo stato funzionale di un tessuto). Nonostante gli straordinari sviluppi legati a queste scienze omiche, spesso la comprensione della complessità biologica, come quella alla base delle malattie appare limitata, quando applicate singolarmente.

Le cellule infatti sono composte da una grande varietà di molecole di diversa natura che agiscono in modo coordinato attraverso numerosi e intricati meccanismi molecolari al fine di mettere in atto processi anche molto complessi come la replicazione o l'adattamento cellulare a stimoli esterni.

Dunque, la forte complessità dei processi molecolari e cellulari, insieme alla elevata numerosità dei dati ottenuti, può essere di ostacolo nell'identificazione e interpretazione delle cause molecolari di malattia [215]. Una soluzione a questi inconvenienti potrebbe essere rappresentata da una visione integrativa dei processi molecolari, che crei un'unica mappa d'interazioni cellulari o di funzionalità tissutale. Tuttavia, la gestione di grandi banche-dati, contenenti informazioni provenienti da differenti tipologie di analisi – omiche, e la loro visualizzazione e comunicazione necessitano di personale specializzato in bioinformatica, di *softwares* di elaborazione statistica e sistemi di *back-up* particolarmente costosi [214].

L'integrazione può avvenire a diversi livelli:

- Target: differenti *target* (o classi di molecole) provenienti da una stessa tipologia di campione possono essere analizzati utilizzando una medesima strategia d'indagine e l'integrazione multi-omica dei dati ottenuti può essere utile a fini di ricerca, come suggeriscono numerose pubblicazioni a riguardo [216, 217].
- Campione di origine: molecole di egual natura (ad esempio, le proteine) provenienti da differenti compartimenti corporei (es. tessuti o fluidi diversi) possono essere analizzate con la medesima tecnica.
- Tecniche di preparazione e/o analisi: uno stesso tipo di campione può subire differenti condizioni di preparazione ed essere analizzato parallelamente con la medesima tecnica oppure, dopo una medesima preparazione, può essere sottoposto a differenti tecniche di analisi.

1.2.7.1 INTEGRAZIONE DI CAMPIONI BIOLOGICI

Lo scopo dell'integrazione di informazioni derivanti dall'analisi di differenti campioni mediante la stessa metodica è spesso quello di ricreare un *network* seguendo il destino di una proteina e rincorrendo un flusso di informazioni fra i differenti compartimenti corporei studiati. Più completa e dettagliata è la

caratterizzazione delle proteine (come peso molecolare, punto isoelettrico, eventuali modificazioni post-traduzionali, quantificazione), più informativa sarà l'integrazione dei dati in possesso. Alcuni studi di proteomica integrativa di questo tipo sono stati applicati alla comprensione delle funzioni fisiologiche del rene. Infatti, i proteomi di plasma, urine e tessuto renale possono essere considerati come strettamente interdipendenti: un grande volume di sangue è filtrato dal rene per generare circa 150-180 L/al giorno di ultrafiltrato, la maggior parte del quale subisce un riassorbimento, mentre circa l'1% viene escreto definitivamente sotto forma di urina [168]. Dunque, una strategia che sembra promettente per studi di proteomica integrativa sulla funzionalità renale potrebbe essere rappresentata dalla teoria del "black box": il rene può essere infatti considerato come una scatola nera, in cui il sangue rappresenta l'*input* d'informazioni che raggiungono l'organo e le urine l'*output* che lo abbandona [218] (Figura 20). In un lavoro del 2009, i dati ottenuti dall'analisi e dall'identificazione delle proteine plasmatiche e urinarie provenienti da differenti esperimenti e da differenti campioni umani sono stati raccolti in un'unica libreria d'identificazioni. L'intera lista di informazioni raccolte è stata elaborata al fine di confrontare il subproteoma specifico del plasma, quello delle urine e, infine, quello comune fra i due compartimenti [218].

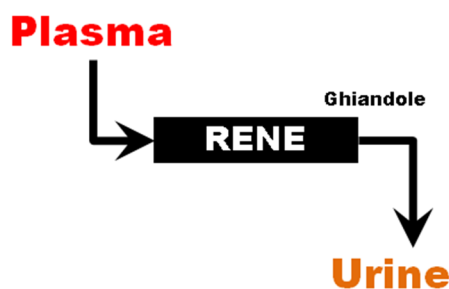


Figura 20. Rene descritto come una "scatola nera" [218].

Nuove opportunità poi potrebbero aprirsi a seguito del calcolo del *rate* di filtrazione di ciascuna proteina attraverso il rene e dalla comprensione della relazione tra variazioni di tale tasso e la presenza di danno renale o condizione patologica.

Un altro interessante approccio per lo studio della funzionalità renale è stato proposto da Ferrah *et al.* [219]. Gli autori hanno effettuato analisi LC-MS/MS su campioni derivati da glomeruli renali, da plasma e da urine di volontari sani al fine di confrontare dal punto di vista qualitativo le informazioni ottenute. Inoltre, è stata eseguita una ricerca *gene ontology* (GO) per approfondire le conoscenze sulle funzioni molecolari delle proteine identificate e una quantificazione di tipo *label-free* per evidenziare i livelli di espressione relativi di ciascuna proteina fra i compartimenti oggetto di studio.

1.2.7.2 INTEGRAZIONE DI TECNICHE

La spettrometria di massa è oggi molto utilizzata e trova spazio per molte applicazioni quali scoperta di target farmacologici, diagnostici e nello studio dei meccanismi molecolari fisio- patologici. Date le enormi potenzialità delle tecniche proteomica come la spettrometria di massa, accoppiare differenti tecniche MS o differenti protocolli di preparazione complementari del campione ha il vantaggio di aumentare la sensibilità e l'*high throughput*.

Recentemente è stata effettuata un'analisi combinata utilizzando due differenti piattaforme tecnologiche (MALDI-TOF e LC-ESI MS/MS) per l'analisi del peptidoma sierico e urinario di pazienti affetti da RCC e di soggetti controllo. Mediante tale approccio di peptidomica integrativa, è stato possibile ottenere informazioni accurate sulla massa molecolare dei peptidi sierici o urinari attraverso l'analisi MALDI-TOF e una maggiore sensibilità e capacità identificativa utilizzando una separazione nLC accoppiata *on-line* all'analisi ESI [166, 220].

Un ulteriore e interessante esempio di integrazione di tecniche proteomiche è rappresentato dal lavoro di Cole et al. [221] eseguito su tessuto tumorale (fibrosarcoma) e tessuto sano proveniente da modelli murini. In questo lavoro, è stata effettuata un'analisi combinata di MALDI-Imaging e HPLC accoppiata a spettrometria di massa che ha permesso di aggiungere alle informazioni quali-quantitative dei dati LC-MS/MS anche indicazioni sulla localizzazione spaziale delle proteine nel tessuto in esame.

1.3. DATA MINING

Contrariamente a quanto il suo nome sembrerebbe suggerire il Data Mining non consiste nella semplice "query" (o "mining") di dati in un database, ma nella ricerca di relazioni complesse, nascoste all'interno dell'insieme dei dati disponibili. Il Data Mining è il processo di individuazione di informazioni utilizzabili da grandi set di dati che prevede l'utilizzo di analisi matematiche per dedurre modelli e tendenze esistenti nei dati. In genere, questi modelli non possono essere individuati tramite l'esplorazione tradizionale dei dati perché le relazioni sono troppo complesse o perché sono presenti troppi dati. Questo spiega come mai, nel corso degli ultimi anni, si siano sviluppate una serie di tecniche mirate a ricercare in un campione multivariato le caratteristiche rilevanti per spiegare la classificazione clinica. Il processo di Data Mining su dati proteomici è ormai largamente applicato al fine di identificare biomarcatori rilevanti per la diagnosi molecolare di malattie complesse, come il tumore [222-224]. L'identificazione di un modello capace di discriminare campioni biologici (che condividono alcune caratteristiche come ad esempio l'*m/z* degli ioni originati da peptidi o proteine) è un grande obiettivo della proteomica clinica. L'analisi di dati in spettrometria di massa al fine di risolvere problemi di classificazione, si articola in tre fasi principali:

- 1) *preprocessing* dei dati;
- 2) classificazione vera e propria;
- 3) identificazione dei biomarcatori da associare ai composti proteici coinvolti nel processo biologico in esame.

Uno schema esemplificativo del processo di lavoro è riportato in Figura 21.

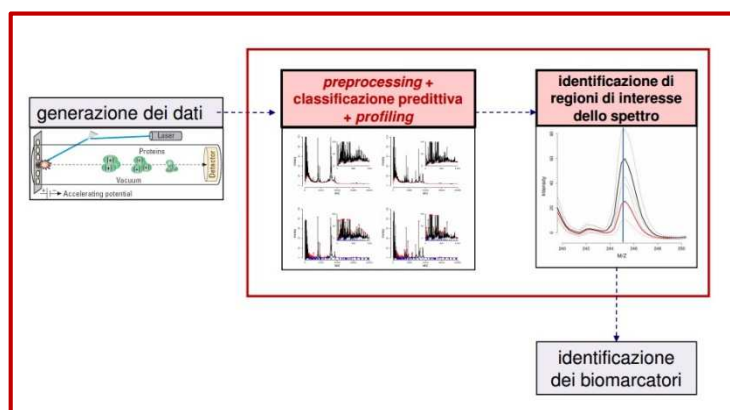


Figura 21. *Work Flow* del processo di Data Mining su dati proteomici.

Una volta ottenuti i dati, prima di procedere con il vero e proprio processo di Data Mining, è necessario iniziare a fare un *preprocessing* (pre-elaborazione) sugli spettri grezzi al fine di cercare di ridurre eventuali *bias* e variabili che potrebbero andare ad inficiare sulla robustezza del risultato finale. In un problema di *profiling* proteico, risulta infatti cruciale questa fase iniziale che consiste in una sorta di “pulizia” e organizzazione dei dati ottenuti dalle analisi. Durante il *preprocessing*, gli spettri grezzi subiscono, dunque, alcune operazioni finalizzate a facilitare il processo di riconoscimento e a renderli compatibili con le successive procedure di apprendimento automatico; quanto detto può essere ottenuto a partire da un *dataset* di spettri grezzi attraverso processi quali la sottrazione del rumore di fondo, la normalizzazione, l'estrazione dei picchi, etc. Il segnale viene esaltato ed adattato in funzione della finalità del processo di classificazione/riconoscimento. Dopo questa fase preliminare, può avere inizio il processo di Data Mining: l'obiettivo finale di questo processo, nel suo complesso, è la definizione di potenziali marcatori, ovvero l'identificazione dei picchi che verranno considerati come variabili del processo di classificazione.

1.3.1. BIOMARKER DISCOVERY

Nell'ambito della proteomica clinica, con il termine *biomarker discovery* si intende l'identificazione di variabili ottimali che risultano significativamente differenti tra due classi e che possono essere usate per una previsione accurata della classe di appartenenza. Un biomarcatore può essere formato da una sola variabile, o alternativamente da un insieme di variabili [225]. Il processo algoritmico che permette di identificare un sottoinsieme ottimale di variabili, tra tutte quelle possibili, in grado di discriminare e classificare i campioni, è definito “selezione delle variabili” (*feature selection*); questo processo si configura come primo step del vero e proprio processo di *Data Mining* ed è seguito dal processo di classificazione (Figura 22).

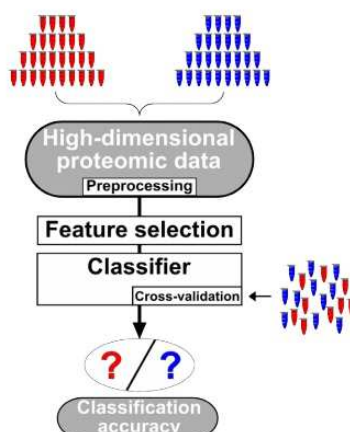


Figura 22. Data mining: feature selection e classificazione.

Un errore comune tra gli studi pionieristici nella ricerca di biomarcatori è stato quello di utilizzare test statistici univariati per cercare proteine che avessero caratteristiche quali/quantitative significativamente differenti in due popolazioni di soggetti messe a confronto dal punto di vista statistico. Il limite principale di questo approccio è quello di considerare le variabili singolarmente correlate con una malattia, ignorando completamente l'esistenza di possibili relazioni tra di loro. Alcuni ricercatori hanno tentato di modificare questo approccio, non cercando più un solo biomarcatore ma un insieme di biomarcatori/variabili che meglio correlassero con una data condizione (nel campo proteomico questo significava di cercare non più una sola proteina ma un insieme di esse che meglio correlasse con una data condizione); approccio definito multivariato. Solo recentemente questo approccio multivariato ha iniziato ad essere considerato migliore rispetto al precedente approccio univariato e con il passare del tempo sta acquisendo sempre più favore nella ricerca proteomica. Dunque, all'interno del processo di Data Mining possono essere identificate, più specificamente, diverse fasi sequenziali:

- 1) *Feature selection*
- 2) *Learning Process*
- 3) *Training Process*
- 4) *Testing*
- 5) Valutazione del modello
- 6) Interpretazione e rappresentazione dei risultati.

Le molteplici applicazioni e l'interesse biologico, medico e clinico legati alla possibilità di conoscere i biomarcatori di patologie complesse come i tumori hanno fatto sì che negli ultimi anni siano state proposte numerosissime tecniche per la ricerca e la selezione di *features* caratteristiche a partire da dati di MALDI *profiling*.

Tali tecniche di *feature selection*, consentono di individuare, tra le decine di migliaia di variabili monitorate, un sottoinsieme di *features* che siano più predittive rispetto all'insieme di partenza. Inoltre alcune variabili inutili possono comportarsi in maniera simile al rumore e impattare in modo negativo nella classificazione [226]. Nella maggior parte dei casi si tratta di tecniche di *feature selection* proposte e sviluppate in altri ambiti, come ad esempio nel campo di *machine learning*, e poi modificate e riadattate per poter essere applicate a dati di origine biologica. Tipicamente il numero di variabili (proteine) monitorate è di gran lunga superiore al numero di campioni disponibili per ciascuna variabile, che corrisponde al numero di soggetti di ciascuna classe (Figura 23).

ID	var 1	var 2	var 3	•	var j	Group
Sample 1	value 1.1	value 1.2	value 1.3	•	value 1.j	1
Sample 2	value 2.1	value 2.2	value 2.3	•	value 2.j	1
Sample 3	value 3.1	value 3.2	value 3.3	•	value 3.j	0
Sample 4	value 4.1	value 4.2	value 4.3	•	value 4.j	0
•	•	•	•	•	•	•
Sample i	value i.1	value i.2	value i.3	•	value i.j	0

Figura 23. Esempio di dati di *profiling* ottenuti da gruppo di controlli e da un gruppo di pazienti. Le righe rappresentano i campioni, mentre le colonne indicano le loro *features*. In ciascuna cella è riportato il valore corrispondente al parametro associato alla *feature*.

Dunque gli algoritmi di *feature selection* prevedono un approccio che punta ad individuare un sottoinsieme minimo e ottimale di variabili. Gli algoritmi di *feature selection* possono essere divisi in due categorie principali: la categoria dei *filter model* e quella dei *wrapper model*. I *filter model* tentano di rimuovere gli attributi irrilevanti dall'insieme delle *features* prima dell'esecuzione dell'algoritmo di apprendimento (fase di *learning*) ed è quindi indipendente da esso; questo non accade nell'approccio *wrapper*. Alla fine della fase di *feature selection*, si arriva ad avere uno o più potenziali biomarcatori. Generalmente avere un numero di biomarcatori/variabili elevato tende all'*overfitting* che si verifica quando un modello si adatta perfettamente ai dati osservati, utilizzati per la sua costruzione, a discapito della generalizzazione, ovvero tende a non essere efficiente quando utilizzato su nuovi dati. Dall'altra parte un sottoinsieme di biomarcatori troppo piccolo può non possedere un potere discriminatorio sufficiente. La selezione del modello e la scelta del biomarcatore dovrebbero arrivare ad un compromesso tra l'*overfitting* e la generalizzazione. In questo senso risulta importante considerare il numero di variabili usate come biomarcatori. Biomarcatori che siano veramente multivariati dovrebbero preferibilmente essere formati da non più di dieci variabili. Se le classi che devono essere differenziate sono particolarmente eterogenee, allora più variabili potrebbero essere necessarie per costruire un biomarcatore. In ogni caso, se il biomarcatore contiene più di venti variabili sarebbe opportuno cercare un'altra strada che permetta di identificare un numero minore di variabili [227]. Le variabili così selezionate

possono essere utilizzate, in un secondo momento, come variabili di ingresso ad un classificatore per suddividere i soggetti nelle classi di appartenenza.

1.3.2. SISTEMI DI CLASSIFICAZIONE

La creazione di modelli di classificazione è uno degli aspetti più studiati e sviluppati del processo di Data Mining. La classificazione è infatti un problema molto complesso sia per la difficoltà intrinseca della stessa, sia per la quantità di dati che bisogna analizzare per la creazione dei modelli. Al fine di identificare un modello capace di discriminare campioni biologici, negli ultimi anni, sono stati messi a punto svariati algoritmi.

Un sistema di classificazione, o classificatore, è il risultato dell'applicazione di un algoritmo di apprendimento automatico per identificare il rapporto fra un set di variabili e le classi. Le classi corrispondono alle popolazioni che noi vogliamo differenziare; mentre le variabili, o meglio, le *features* sono caratteristiche che descrivono i campioni.

I seguenti algoritmi sono quelli più comunemente utilizzati dai sistemi di classificazione:

- *Discriminant analysis*
- *Support Vector Machines*
- *Random Forests and other tree-based classifiers*
- *k-Nearest Neighbors*
- *Artificial Neural Networks.*

Nella bioinformatica clinica, gli algoritmi *Support Vector Machine* (SVMs) hanno permesso la costruzione di potenti modelli sperimentali diagnostici nello studio di neoplasie basati su valori proteici con migliaia di variabili e con poche decine di campioni [228]. In numerosi studi, differenti autori hanno applicato il SVM a dati sperimentali ottenuti dall'analisi di fluidi biologici attraverso le tecnologie MALDI e SELDI, mentre, ad oggi, un numero ristretto di autori ha cercato di analizzare i dati ottenuti da approcci LC-MS.

Il *Support Vector Machine* (SVM) (Figura 24) seleziona una combinazione di picchi che permettano di separare le popolazioni in esame attraverso l'utilizzo di un vettore. Si basa sul principio di "minimizzazione del rischio strutturale (SRM)", ovvero ha lo scopo di identificare la combinazione migliore e con il minor rischio di fallimento.

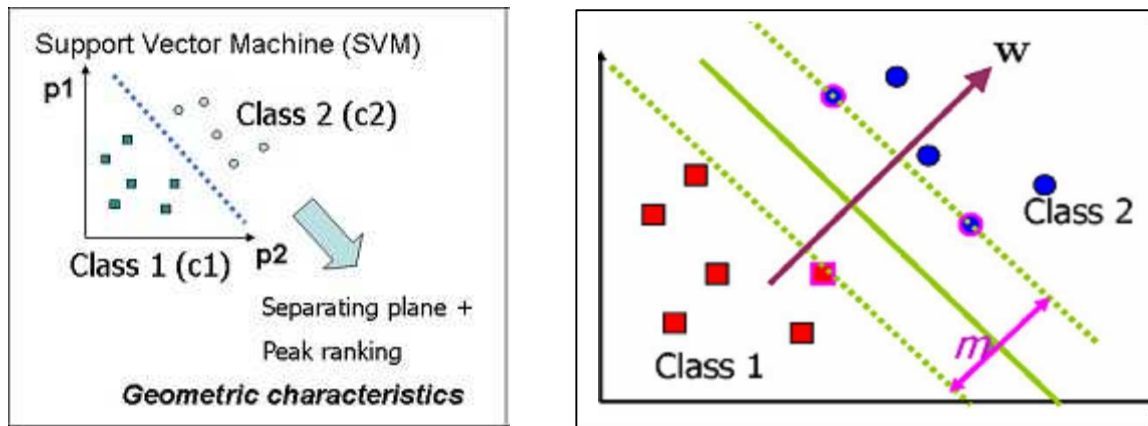


Figura 24. Meccanismo di funzionamento dell'algoritmo *Support Vector Machine* (SVM).

Esistono due tipologie di algoritmi, quelli parametrici e quelli non parametrici. Nel primo caso si tratta di algoritmi che fanno delle precise assunzioni circa la distribuzione dalla quale sono stati campionati i dati proteomici osservati; in questo caso la struttura del modello viene stabilita *a priori*, contrariamente a quanto succede in quelli non parametrici in cui la struttura del modello viene determinata dai dati stessi. Infatti, quando non sono disponibili informazioni circa la distribuzione dei dati o quando è difficile verificare le assunzioni a causa della scarsità di campioni disponibili, è possibile ricorrere all'applicazione di metodi non parametrici, o *model-free*. Questi metodi non fanno alcuna ipotesi, o fanno ipotesi meno stringenti, circa la distribuzione statistica dei dati. La classificazione, inoltre può essere di tipo *supervised*, nel qual caso l'insieme delle classi è definito *a priori*, oppure *unsupervised*, in cui è il classificatore che deve individuare le classi naturali presenti nei dati. La ricerca di un biomarcatore ideale e la progettazione di un classificatore non può essere considerata completa fino a quando non è stata effettuata una stima dell'errore del modello di classificazione. Il principale obiettivo è di ottenere un modello di classificazione generalizzabile, così che possa essere esteso a una popolazione di gran lunga più numerosa rispetto a quella presa in esame (ossia un modello che possa avere un basso errore nel processo di classificazione quando viene applicato a nuovi campioni provenienti da popolazioni differenti). Infatti, la classificazione dei dati avviene in un processo a due fasi. Nella prima si costruisce un modello per descrivere un set di dati; a questo segue un secondo passaggio in cui il modello viene usato per la classificazione e ne permette la sua validazione.

Di seguito sono elencati i metodi di validazione e cross validazione usati durante il processo di Data Mining:

- *Reclassification*
- *Leave-One-Out cross-validation*
- *K-Fold cross-validation*
- *Holdout method of validation*
- *Validation based on classification of out-of-bag samples*
- *Validation on an independent test data set.*

Sono tutti metodi che permettono di ottenere una stima dell'errore di generalizzazione del classificatore. La *cross validation* è un metodo statistico per validare un modello predittivo. Preso un campione di dati esso viene suddiviso in sottoinsiemi, alcuni dei quali usati per la costruzione del modello (*training sets*) e altri da confrontare con le predizioni del modello (*validation sets*). Una completa validazione del modello permette di disaccoppiare la fase di selezione delle variabili da quella di validazione, evitando la sottostima della *performance* dovuta all'errore di selezione. Tra questi metodi la validazione incrociata *K-Fold* (*K-Fold cross-validation*) permette di dividere in modo casuale i dati di *training* in K sottoinsiemi non sovrapposti e di grandezza approssimativamente uguale. Il modello di classificazione viene addestrato su $K-1$ sottoinsiemi e poi utilizzato per classificare i campioni dei rimanenti sottoinsiemi. Questo processo è ripetuto K volte affinché ciascun campione venga classificato dal modello che è stato costruito senza quel campione. Vengono comunemente utilizzati valori di K compresi tra 5 e 10. La *cross validation K-fold* può essere interna o esterna. Nella validazione interna, il processo di selezione delle variabili fornisce un biomarcatore ideale (fatto da n variabili) e in seguito i dati usati per l'addestramento (*training set*) vengono *splittati* applicando lo schema del *k-fold*; ciascuno di essi è addestrato su un sottoinsieme di campioni di addestramento e poi usato per classificare i rimanenti campioni. Tuttavia ciascun classificatore utilizza lo stesso insieme di variabili. Nella validazione esterna i dati vengono a priori divisi in dati per il *training* (addestramento) e in dati per la validazione. In questa validazione vengono utilizzati degli schemi indipendenti dal processo di selezione delle variabili. Se si suppone di utilizzare gli stessi dati sia per la fase di addestramento che per la fase di verifica, questa scelta espone il classificatore al problema dell'*overfitting*. Infatti si ottiene così un classificatore perfettamente addestrato a riconoscere i dati a disposizione a discapito della generalità, ovvero a sfavore della capacità di classificare dati mai visti. Infine l'ultimo processo a cui è bene sottoporre il modello costruito è la sua validazione su un insieme di dati possibilmente ampio e indipendente; bisognerebbe sempre poter verificare il modello su un *set* di dati indipendente così da ottenere una validazione finale del biomarcatore.

1.3.2.1. RISULTATI DEI MODELLI DI CLASSIFICAZIONE

Per ottenere una stima dell'efficienza del sistema di classificazione è possibile calcolare la sua accuratezza, definita come proporzione di campioni testati che sono correttamente classificati all'interno della loro vera classe di appartenenza; sottraendo questo valore ad uno si otterrà l'indice di *misclassification*. Al termine del processo di classificazione, il risultato ottenuto può essere valutato attraverso una tabella, detta matrice di confusione o tabella di classificazione, qui riportata a come esempio (Tabella 6).

		Classe predetta	
		Malattia (Positivi)	No Malattia (Negativi)
Classe vera	Malattia (Positivi)	Veri Positivi (VP)	Falsi Negativi (FN)
	No Malattia (Negativi)	Falsi Positivi (FP)	Veri Negativi (VN)

Tabella 6. Tabella di classificazione o matrice di confusione: le righe rappresentano le classi reali dei dati utilizzati nel test. Le colonne rappresentano le classi predette dal classificatore.

In questa tabella sono visibili i campioni realmente appartenenti a ciascuna classe (classe vera) e i campioni assegnati a ciascuna classe dal modello (classe predetta). Le celle della matrice di confusione rappresentano le combinazioni possibili tra classi predette e reali. In un problema di classificazione binaria l'insieme dei dati da classificare è suddiviso in due classi che possono essere convenzionalmente indicate come positivi (P) e negativi (N). I risultati possono, in una classificazione binaria, essere presentati mediante la matrice di confusione 2 x 2, (vedi Tabella 6), dove:

- **Veri Positivi (VP):** il numero di casi positivi al test che sono correttamente classificati come positivi.
- **Veri Negativi (VN):** numero di casi negativi al test che sono correttamente classificati come negativi.
- **Falsi Positivi (FP):** numero di casi negativi che sono erroneamente classificati come positivi.
- **Falsi Negativi (FN):** numero di casi positivi che sono erroneamente classificati come negativi.

Il confronto tra i risultati del test in esame e la vera appartenenza ad una data classe consente di stimare due importanti parametri: la sensibilità e la specificità.

La **sensibilità** si definisce come la proporzione di casi positivi al test che sono classificati correttamente come positivi (la probabilità che un test ha di predire la malattia quando la classe vera è la malattia):

$$\text{Sensibilità} = \frac{VP}{VP+FN} = \frac{VP}{P} \quad (\text{dove } P = VP + FN).$$

La **specificità** si definisce come la proporzione di casi negativi al test che sono classificati correttamente come negativi (la probabilità che un test ha di predire la non malattia quando la classe vera è la non malattia):

$$\text{Specificità} = \frac{VN}{VN+FP} = \frac{VN}{N} \quad (\text{dove } N = VN + FP).$$

Con il termine **False Discovery rate (FDR)** si indica la proporzione di casi classificati come positivi che sono falsi positivi:

$$\text{FDR} = \frac{FP}{FP+VP}.$$

Inoltre si può valutare l'**accuratezza** o **efficienza** del test. L'accuratezza corrisponde al numero di pazienti correttamente classificati, ovvero indica la proporzione di tutti i casi che sono correttamente classificati e permette di avere un'idea della capacità del test di discriminare tra la malattia e la non malattia:

$$\text{Accuratezza} = \frac{VP+VN}{VP+VN+FP+FN} = \frac{VP+VN}{P+N}.$$

In aggiunta a tutti questi indici si possono definire altre due misure che danno un'idea del valore predittivo del risultato di un classificatore: il **valore predittivo positivo (VPP)** e il **valore predittivo negativo (VPN)**. Il VPP indica la probabilità che un paziente con test positivo abbia realmente la malattia, ovvero la proporzione di casi classificati come positivi che sono realmente positivi:

$$\text{VPP} = \frac{VP}{VP+FP} = 1 - \text{FDR}.$$

Il VPN indica la probabilità che un soggetto sano con un test negativo effettivamente non abbia la malattia, ovvero la proporzione di casi classificati come negativi che sono realmente negativi:

$$\text{VPN} = \frac{VN}{VN+FN}.$$

Tuttavia per poter valutare in modo completo la capacità discriminante/diagnostica di un test non è possibile fare affidamento ai soli indici appena elencati; infatti bisogna tenere sempre presente il problema del *cut-off*. Il *cut-off* è un valore di soglia che permette di discriminare i risultati in positivi e negativi, cioè quel valore assunto dalla variabile misurata nel test al di sopra della quale il soggetto viene dichiarato positivo e al di sotto del quale viene definito negativo. Un test diagnostico ideale vorrebbe che al 100% di sensibilità corrispondesse un 100% di specificità; questo sarebbe possibile solo nel caso in cui le due popolazioni messe a confronto avessero dei valori che ricadono in intervalli ben separati (Figura 25). Il test ideale dovrebbe consentire di discriminare completamente tra pazienti sani e malati. Nel caso in cui non ci sia alcuna sovrapposizione tra le due popolazioni, è immediato individuare sull'asse delle ascisse il valore di *cut-off* che permette di discriminare, in questo caso e solo in questo, con assoluta precisione tra pazienti sani e pazienti malati.

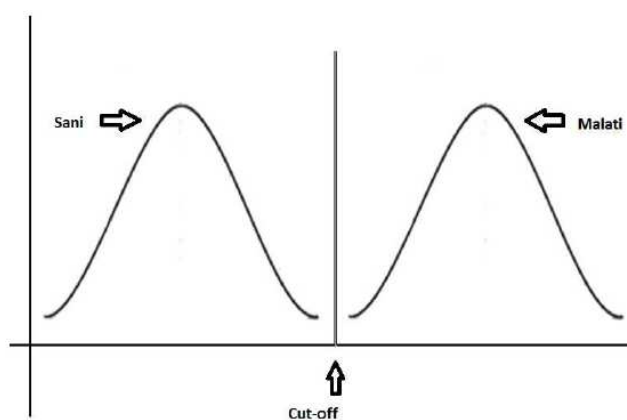


Figura 25. Valore del *cut-off* in un test diagnostico ideale.

Tuttavia nella pratica comune si verifica una sovrapposizione più o meno ampia delle due distribuzioni come mostrato in Figura 26.

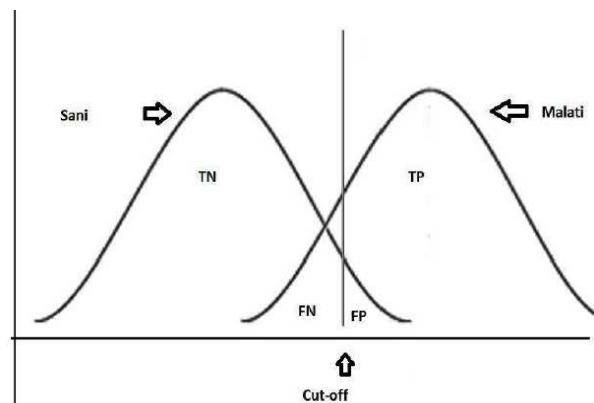


Figura 26. Situazione reale in cui si ha una sovrapposizione delle popolazioni in esame.

In questo caso si avrà sempre un certo numero di soggetti sani che risulteranno positivi al test (falsi positivi, FP) e un certo numero di malati che erroneamente verranno classificati come sani (falsi negativi, FN). Dunque nella realtà è impossibile individuare sull'asse delle x un valore di cut-off che consenta una classificazione perfetta, ossia da azzerare i falsi positivi e i falsi negativi. Il valore di *cut-off* deve essere fissato per decidere il livello di sensibilità e specificità migliore per un determinato test; per scegliere il valore di *cut-off* si utilizza la curva ROC (*Receiving Operating Characteristic* – caratteristiche operative del ricevente). La curva ROC offre una visione grafica circa la specificità e sensibilità di un test e ne fornisce una stima della capacità discriminante; altro non è che una rappresentazione grafica che mette in luce come al variare del *cut-off* corrisponda una variazione in termini di sensibilità e specificità. La curva ROC viene costruita considerando tutti i possibili valori del test e, per ognuno di questi, si calcola la proporzione di veri positivi (sensibilità) e la proporzione di falsi positivi (1-Specificità). La curva ROC fornisce anche un'indicazione dell'accuratezza del test: infatti attraverso la costruzione della curva ROC e del calcolo dell'area sottesa ad essa (AUC, *Area under the ROC curve*), è possibile stimare la probabilità di assegnare una unità statistica al suo reale gruppo di appartenenza (Figura 27).

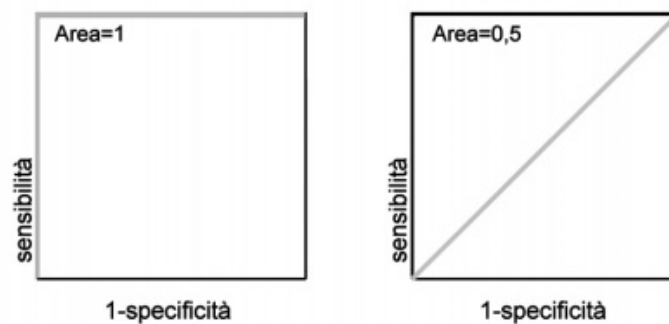


Figura 27. Esempio di test perfettamente performante (a sinistra) e di test completamente inutile (a destra).

Un valore AUC uguale a 0 indica una capacità discriminante del test nulla, mentre un valore AUC uguale a 1 un'ottima capacità discriminante (Tabella 7).

Tanto maggiore è l'area sotto la curva (cioè tanto più la curva si avvicina al vertice del grafico) tanto maggiore è il potere discriminante del test. Per l'interpretazione dei valori dell'area sottostante la curva ROC è possibile riferirsi alla classificazione proposta da Swets [229]:

AUC=0,5	Il test non è informativo
$0,5 < \text{AUC} \leq 0,7$	Il test è poco accurato
$0,7 < \text{AUC} \leq 0,9$	Il test è moderatamente accurato
$0,9 < \text{AUC} < 1,0$	Il test è altamente accurato
AUC=1	Test perfetto

Tabella 7. Interpretazione dei valori AUC ottenuti dall'analisi della curva ROC per la misura dell'accuratezza di un modello diagnostico.

1.4. LA SPETTROMETRIA DI MASSA

La spettrometria di massa consiste in un insieme di tecniche analitiche, particolarmente utilizzate in chimica organica, in grado di fornire informazioni riguardanti:

- composizione qualitativa e quantitativa di analiti sia organici che inorganici in miscele complesse;
- struttura di una vasta gamma di specie molecolari complesse;
- rapporti isotopici degli atomi nei campioni;
- struttura e composizione delle superfici solide.

Insieme all'elevata selettività, che le permette di riconoscere una determinata sostanza anche se presente in miscela con molte altre, essa è dotata di alta sensibilità (nell'ordine delle femto (10^{-15}) o atto (10^{-18}) moli), così che spesso è considerata la tecnica d'elezione per l'analisi di sostanze presenti in basse concentrazioni in matrici complesse, quali sono quelle di origine biologica [123].

Uno spettrometro di massa separa le particelle cariche in funzione del loro rapporto massa/carica (m/z). La capacità di uno spettrometro di massa di differenziare le masse degli ioni è generalmente espressa in termini di risoluzione R , definita come:

$$R = \frac{m}{\Delta m}$$

dove Δm è la differenza di massa tra due picchi adiacenti risolti e m è la massa nominale del primo picco. La massa di un ione è misurata in *dalton* (*Da*), che corrisponde a 1/12 della massa dell'isotopo carbonio-12 (^{12}C).

L'iniziale necessità di formare ioni del campione in fase gassosa per i processi di separazione e rivelazione, ha reso la spettrometria di massa inaccessibile per lungo tempo all'analisi di macromolecole come acidi nucleici e proteine; questo limite era dovuto al fatto che il processo di ionizzazione avveniva per riscaldamento e bombardamento con particelle, processi che causano nella maggior parte dei casi la distruzione del campione biologico [230]. Attualmente esistono tecniche di ionizzazione cosiddette *soft* ovvero che non causano la distruzione degli acidi nucleici e delle proteine e che rendono quindi possibile l'analisi di queste macromolecole biologiche in fase solida, liquida oppure gassosa.

Questo diagramma a blocchi mostra i componenti principali di uno spettrometro di massa (Figura 28).

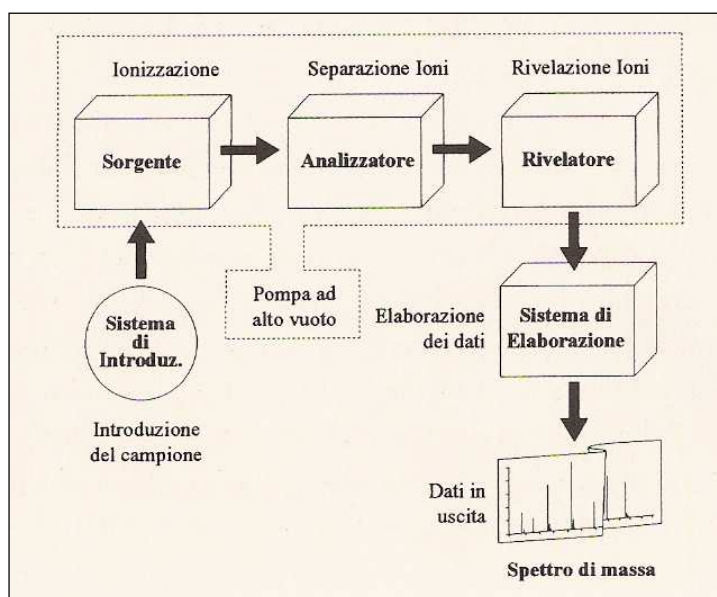


Figura 28. Schema a blocchi di uno spettrometro di massa.

Una caratteristica peculiare degli spettrometri di massa è la necessità di un complesso sistema ad alto vuoto che mantenga bassi valori di pressione (da 10^{-4} a 10^{-8} torr) in tutte le componenti dello strumento, eccetto l'elaboratore e il visualizzatore del segnale. Il sistema di alto vuoto è necessario, affinché non vi sia una perdita di ionizzazione per urto con i gas atmosferici.

Il sistema di introduzione del campione viene detto *on-line* nel caso in cui sia interfacciato direttamente con un sistema di separazione, come ad esempio un HPLC o un gascromatografo, mentre viene detto *off-line* se il campione viene introdotto nello strumento senza l'utilizzo dell'interfaccia con un sistema di separazione.

Tutti gli spettrometri sono, dunque, costituiti essenzialmente da tre parti:

- camera di ionizzazione o sorgente ionica;
- analizzatore;
- rivelatore.

1.4.1. SORGENTE IONICA

La ionizzazione è il processo a cui sono sottoposte tutte le sostanze introdotte nello spettrometro di massa prima della separazione degli ioni.

Il campione viene ionizzato in un'apposita camera di ionizzazione, in cui il fascio di elettroni è prodotto da una sorgente ionica, che varia a seconda della tecnica utilizzata. Le sorgenti si suddividono in due principali categorie: la prima è costituita dalle sorgenti in fase gassosa, nelle quali il campione viene dapprima volatilizzato e i componenti gassosi vengono quindi ionizzati in diverso modo; la seconda comprende le sorgenti a desorbimento, che non richiedono la vaporizzazione totale del campione e impiegano una sonda per il campione. Il sistema di ionizzazione è, quindi, molto importante in spettrometria di massa, poiché da esso dipendono il numero, la natura (ione positivo o negativo, ione monocarica, doppia carica, ecc.) e l'abbondanza dei frammenti molecolari che compaiono nello spettro di massa.

Esistono diverse tecniche di ionizzazione, tra queste le più diffuse sono:

- Ionizzazione per interazione elettronica (*Electron Impact ionization*, EI)
- Ionizzazione chimica (*Chemical Ionization*, CI)
- Ionizzazione chimica a pressione atmosferica (*Atmospheric Pressure Chemical Ionization*, APCI)
- Ionizzazione per desorbimento con laser assistita da matrice (*Matrix-Assisted Laser Desorption Ionization*, MALDI)
- Ionizzazione *elettrospray* (*ElectroSpray Ionization*, ESI).

La numerosità delle tecniche attualmente disponibili ha generato, quindi, particolari varianti della spettrometria di massa che vengono utilizzate in base alla natura del campione da analizzare. La principale caratteristica che differenzia le diverse sorgenti è l'energia con cui vengono generati gli ioni: esistono sorgenti *hard* che operano ad elevata energia e portano ad una frammentazione spinta del campione e sorgenti *soft* che invece operano a minore energia e producono una frammentazione molto più limitata del campione. La frammentazione del campione all'interno della sorgente ionica può essere uno strumento utile per ottenere informazioni sulla struttura del composto in analisi, soprattutto nel caso si tratti di molecole organiche di piccole dimensioni [231].

1.4.1.1. IONIZZAZIONE PER DESORBIMENTO LASER ASSISTITA DA MATRICE (MALDI)

La ionizzazione per desorbimento con laser assistita da matrice (MALDI) è una tecnica di ionizzazione *soft*, adatta per l'analisi di composti simili a quelli analizzabili con ESI. La tecnica MALDI è basata essenzialmente sulla ionizzazione per desorbimento con laser. Gli analiti sono co-cristallizzati con una matrice solida e il processo di allontanamento dello ione dalla matrice consiste in una irradiazione con laser, il quale permette ad un numero ristretto di molecole poste sulla superficie delle matrici di essere proiettato ad alte velocità nell'analizzatore [232]. Si possono analizzare sia molecole cariche positivamente che molecole cariche negativamente; in genere, si ha il trasferimento di un protone dalla matrice all'analita e,

dunque, il processo di osservazione in modalità positiva è quello più frequente. La ionizzazione mediante questa tecnica può avvenire anche a pressione atmosferica (*Atmospheric Pressure- Matrix Assisted Laser Desorption Ionization*, AP-MALDI).

1.4.1.2. IONIZZAZIONE ELETTROSPRAY (ESI)

La ionizzazione *electrospray* (ESI) è una tecnica di ionizzazione *soft* e utilizza un gas inerte (di solito azoto) per determinare un processo di nebulizzazione del campione che è in fase liquida o in soluzione (Figura 29).

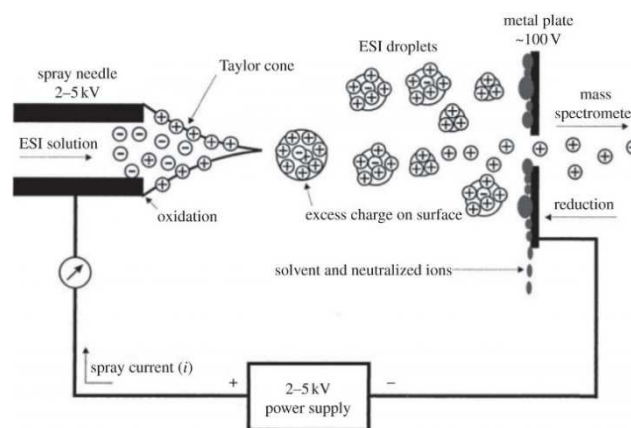


Figura 29. Meccanismo di ionizzazione della sorgente ESI.

Nell'ESI la ionizzazione si verifica a pressione atmosferica e sulla specie molecolare si possono depositare più di una carica, formando ioni multicarica. Questo processo avviene in soluzioni (generalmente di metanolo e acqua), che vengono nebulizzate in una camera a cui è applicato un campo elettrico (ottenuto applicando una elevata differenza di potenziale di diversi kV). La nebulizzazione causa la formazione di piccole goccioline che contengono le specie ionizzate (analita carico). Nella sorgente il solvente evapora, così che le goccioline diventano sempre più piccole sino a che la repulsione elettrica supera la tensione superficiale della goccia. A questo punto la gocciolina si rompe, creando una corrente di ioni che vengono poi indirizzati da un gradiente di campo verso l'analizzatore. La velocità alla quale vengono accelerati gli ioni dipende esclusivamente dalla loro carica complessiva e dalla forza di accelerazione. Di conseguenza, a parità di forza di accelerazione, maggiore è il numero di cariche, maggiore sarà anche la velocità raggiungibile. La ionizzazione *electrospray* generalmente produce solo ioni pseudomolecolari; in qualche caso è possibile osservare anche ioni frammento, ed in alcuni spettrometri è anche possibile favorire la frammentazione in sorgente agendo sulle condizioni sperimentali. Lo spettro ESI di un composto puro rappresenta una distribuzione di picchi (ioni multicarica) derivanti tutti dalla stessa molecola. Successivamente, lo spettro multicarica viene trasformato matematicamente (deconvoluzione) dal computer per generare un profilo spettrale, in cui appare il valore di peso molecolare (PM) della proteina con notevole accuratezza. Inoltre, dato che la ionizzazione *electrospray* risulta particolarmente compatibile

con metodiche di separazione in fase liquida, questa tecnica è notevolmente impiegata nell'analisi farmaceutica e biologica [233].

Fra i principali vantaggi della ionizzazione *electrospray* vi è il fatto che sia una sorgente adatta alla maggior parte dei composti, dalle piccole molecole organiche (anche non volatili) fino a macromolecole come proteine (purché esse abbiano disponibili molti siti ionizzabili). Solo composti estremamente apolari come gli idrocarburi non sono analizzabili. Inoltre, la sensibilità della ionizzazione *electrospray* non diminuisce sostanzialmente all'aumentare della massa della sostanza analizzata e i solventi volatili usati nell'ESI fanno sì che il rumore di fondo sia molto basso.

La presenza di metalli alcalini non è particolarmente fastidiosa per l'analisi di piccole molecole organiche; al contrario, è molto dannosa per ioni poliprotonati: infatti se uno o più Na^+ sostituiscono gli H^+ , si formano ioni aggiuntivi che complicano lo spettro e riducono la sensibilità. All'interno della sorgente, la frazione di elettroni che non urta contro le molecole è raccolta da una trappola per gli elettroni, le molecole che non sono ionizzate sono allontanate dalla pompa ad alto vuoto, mentre quelle ionizzate sono accelerate e convogliate verso l'analizzatore.

1.4.2. ANALIZZATORE

L'analizzatore consente di differenziare gli ioni generati in base al loro rapporto massa/carica e poiché nella maggioranza dei casi la carica degli ioni è +1, la separazione avviene sulla base delle rispettive masse ($m/1$). Ioni con differente m/z escono dall'analizzatore in tempi diversi e di conseguenza raggiungeranno il rivelatore in tempi diversi ed appariranno nello spettro a valori di m/z differenti. Tutti gli analizzatori lavorano sotto alto vuoto. Esistono diversi tipi di analizzatori, ognuno dei quali è caratterizzato da un differente potere risolutivo; tra questi, quelli attualmente più diffusi sono:

- analizzatori quadrupolari
- analizzatori a tempo di volo (*Time Of Flight*, TOF)
- analizzatori a trappola ionica
- analizzatori Qq-TOF
- analizzatori *ultra-high resolution time-of-flight* (UHR-TOF)
- analizzatori a risonanza ciclotronica ionica in trasformata di *Fourier* (FT-ICR).

1.4.2.1. ANALIZZATORI QUADRUPOLARI

Gli analizzatori quadrupolari hanno solitamente una risoluzione minore (circa 4000) dei sistemi a settore. Sono costituiti da quattro barre di metallo disposte parallelamente; alle barre opposte del quadrupolo è applicata una differenza di potenziale generata da una corrente continua e da una alternata. Solo gli ioni con una traiettoria stabile potranno passare attraverso le barre e raggiungere l'elettromoltiplicatore. A determinati valori di tensione applicata, solo ioni aventi un certo rapporto m/z usciranno dal quadrupolo; dunque, variando nel tempo la tensione, tutti gli ioni saranno messi in condizione di uscire a tempi diversi dal quadrupolo (Figura 30).

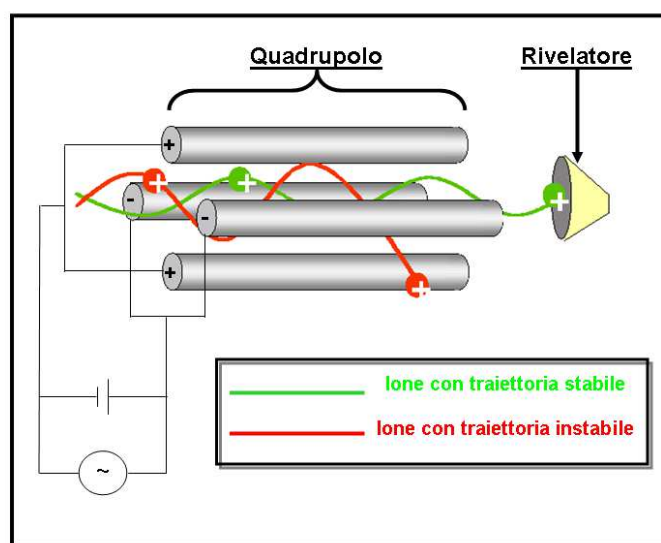


Figura 30. Schema di un analizzatore di massa quadrupolare.

1.4.2.2. ANALIZZATORI A TRAPPOLA IONICA (*ION TRAP*)

Questi tipi di analizzatori sono molto simili a quello quadrupolare, ma in essi il filtro quadrupolare è sferico e trattiene tutti gli ioni generati, i quali vengono poi rilasciati progressivamente verso il rivelatore. La trappola ionica utilizza campi elettrici alternati a radiofrequenza (RF) applicati agli elettrodi disposti in una geometria a “sandwich”, con un elettrodo anulare nel centro e due elettrodi “coperchi” sopra e sotto di esso [234]. Il dispositivo intrappola nello spazio tra gli elettrodi gli ioni aventi un dato intervallo di rapporti m/z , determinato dai potenziali applicati. Lo spettro di massa è prodotto variando il potenziale RF in modo da espellere sequenzialmente dalla trappola gli ioni, secondo un ordine di rapporto m/z crescente. Il potere risolutivo di questo tipo di analizzatore è compreso tra 10^3 e 10^4 .

1.4.2.3. ANALIZZATORI A TEMPO DI VOLO

Gli analizzatori a tempo di volo (TOF) si basano sul principio che ioni di differente valore m/z hanno uguale energia cinetica ($E=1/2 mv^2$), ma differente velocità dopo l'accelerazione subita nella camera di ionizzazione da un campo elettrico di 10^3 - 10^4 V. Il tempo che ciascun ione impiega ad attraversare l'analizzatore, che è costituito da un tubo (tubo di deriva) lungo pochi metri in cui è presente un alto vuoto, è diverso:

$$t = a (m/z)^{1/2} + b$$

Gli ioni che hanno il rapporto m/z minore giungono al rivelatore prima di quelli più pesanti. Quindi il tempo di volo è direttamente proporzionale alla massa. I tempi di volo, comunque, variano in genere da 1 a 30 μ s. Questo analizzatore riesce a coprire un'ampia regione spettrale ed ha un'alta sensibilità. Inoltre, il potere risolutivo può raggiungere anche valori nell'ordine di 6×10^4 (*reflectron* TOF).

Negli spettrometri a tempo di volo il rivelatore è generalmente un moltiplicatore di elettroni e la presentazione dello spettro di massa è praticamente istantanea.

1.4.2.4. ANALIZZATORI Qq-TOF

Questi tipi di analizzatori sono detti "ibridi" poiché combinano in un unico spettrometro due diversi settori analitici. Internamente, essi sono costituiti da tre quadrupoli, Q_0 , Q_1 e Q_2 , seguiti da un tubo di volo (TOF). Fra i quadrupoli, Q_1 e Q_2 operano grazie a radiofrequenze (*RF-mode*). Per analisi di spettrometria di massa convenzionali, il Q_1 funge da filtro per gli ioni mentre la separazione vera e propria degli analiti secondo il loro rapporto m/z viene eseguita dal tubo di volo. Il quadrupolo Q_2 può essere attivo come cella di frammentazione, grazie alla possibilità di far entrare gas di diversa natura per la collisione, o semplicemente per focalizzare e trasferire gli ioni al settore successivo nel caso di analisi di massa semplice.

1.4.2.5. ANALIZZATORI ULTRA-HIGH RESOLUTION TIME-OF-FLIGHT (UHR-TOF)

Questi tipi di analizzatori sono quadrupoli "ibridi" dotati di elevata accuratezza e alta risoluzione. Gli analizzatori UHR-TOF sono chiamati ibridi in quanto possiedono segmenti multipolari e quadrupolari prima del tubo di volo. Gli ioni generati in sorgente sono convogliati all'interno dell'analizzatore grazie ad una differenza di potenziale fra le estremità di un capillare in vetro; gli ioni subiscono poi un doppio passaggio attraverso un settore costituito da una serie di anelli di diametro via via minore, le cui funzioni principali sono quelle di aumentare l'efficienza di trasmissione degli analiti al settore successivo, con l'obiettivo di perdere ciò che è aspecifico. Dopo l'analizzatore è presente un multipolo che focalizza gli ioni in un fascio per favorire l'accumulo e il trasferimento degli ioni; il multipolo, più precisamente un esapolo, termina con una lente "gate" mediante la quale vengono minimizzati i ritardi di tempo fra analisi MS e MS/MS. Il terzo settore dopo il multipolo è il quadrupolo. Questo è il primo vero analizzatore negli spettrometri UHR-TOF. Esso è costituito da due tipi di segmenti quadrupolari: i segmenti centrali, che

fungono da “filtro” per gli analiti, e i segmenti esterni per il trasferimento degli stessi. Gli ioni che escono dal settore quadrupolare giungono nella cella di collisione in cui gas inerti come Argon o Azoto possono contribuire alla frammentazione degli analiti per le analisi MS/MS; inoltre, un dispositivo legato a questa cella raffredda gli ioni per aumentare l'efficacia di focalizzazione degli ioni e la loro trasmissione al tubo di volo (Figura 31) [235].

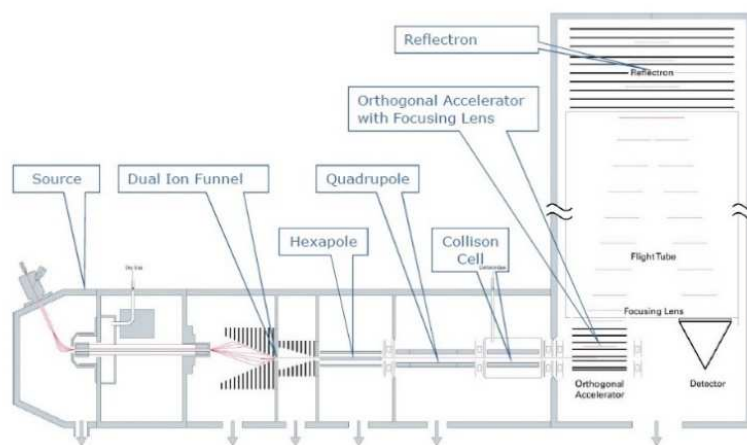


Figura 31. Immagine schematica dell'interno dello spettrometro di massa UHR-TOF [235].

La recente introduzione degli analizzatori *ultra-high resolution quadrupole time-of-flight* (UHR-QToF) e *tandem time-of-flight* (UHR-ToF/ToF) ha garantito agli spettrometri TOF di avvicinarsi molto alla risoluzione di analizzatori a risonanza ciclotronica ionica in trasformata di *Fourier* (FT-ICR). La possibilità di analisi rapide MS e MS/MS, anche nel caso di campioni complessi come quelli biologici era infatti prerogativa di sofisticate e costose apparecchiature come gli strumenti FT-ICR o, più recentemente, gli strumenti *Orbitrap* [236]. Gli strumenti UHR-TOF possono dunque essere efficacemente impiegati per studi proteomici di campioni biologici complessi; le caratteristiche di risoluzione, velocità e sensibilità di scansione sia a livello dei precursori che dei loro frammenti rendono inoltre gli spettrometri UHR-TOF compatibili con le tecniche di proteomica quantitativa come SILAC [237], ICAT [238], iTRAQ [205] o *Tandem Mass Tags* [239] o *label-free* [166].

1.4.3. RIVELATORE

La maggioranza dei rivelatori funzionano ad impatto ionico o per cattura ionica. Tutte le tipologie richiedono una superficie che raccolga gli ioni e dove la carica venga neutralizzata. Si ha quindi un trasferimento di elettroni ed un flusso di corrente, che può essere amplificato e convertito in un segnale processabile da un computer; quest'ultimo registra il segnale sottoforma di spettro (m/z). Il segnale generato è direttamente proporzionale al numero di ioni che raggiungono il *detector* e lo spettro di massa è un

diagramma che mette in relazione il rapporto massa/carica (asse delle ascisse) con l'abbondanza ionica o "intensità" del segnale degli ioni a diverso m/z (asse delle ordinate).

1.4.4. LA SPETTROMETRIA DI MASSA MALDI-TOF

La spettrometria di massa MALDI-TOF rappresenta la tecnica principale per l'analisi di miscele peptidiche [149, 240]. Questa tecnica è stata utilizzata per la prima volta nel 1976 [241] e prevedeva la produzione di peptidi in fase gassosa, allo scopo di analizzarli con uno spettrometro di massa; per riuscire a volatilizzarli si dovevano considerare due parametri fisici fondamentali, quali la lunghezza d'onda di assorbimento e la volatilità dei peptidi. Tuttavia questo metodo aveva il difetto di non consentire l'analisi di peptidi con peso superiore ai 1000 Da.

MALDI-TOF è un acronimo che indica la modalità di desorbimento/ionizzazione del campione co-cristallizzato con matrici organiche idonee, indotta da fasci di luce laser ed interfacciata con analizzatori a tempo di volo [242]. Tale tecnica è stata introdotta alla fine degli anni '80; Karas [243] e Tanaka [244] proposero, nello stesso anno, l'aggiunta al campione di una sovrabbondanza di matrice, in grado di assorbire alla lunghezza d'onda del laser in modo da preservare l'analita ed aiutarlo nel processo di ionizzazione. In seguito si sono ottenuti notevoli miglioramenti, soprattutto grazie agli esperimenti empirici e alle maggiori conoscenze chimico-fisiche riguardo a questa tipologia di ionizzazione.

Numerosi sono i vantaggi nell'utilizzo di questa tecnica:

- applicabilità ad una grande varietà di sostanze
- possibilità di misurare un ampio intervallo di masse contemporaneamente e di misurare molecole con masse molecolari maggiori di 100 kDa
- automatizzabilità
- elevata sensibilità (femtomoli) e risoluzione [245]
- rapidità e semplicità d'utilizzo
- notevole capacità di risolvere, in termini di massa, miscele peptidiche complesse; questo aspetto è dovuto alla produzione di spettri semplici contenenti ioni monovalenti $[M+H]^+$ [246].

Attualmente la spettrometria di massa MALDI-TOF ha assunto un ruolo fondamentale nel campo della proteomica, grazie alla sua elevata sensibilità e poiché permette analisi molto rapide, possedendo anche una buona tolleranza alla presenza nel campione di sali non volatili e tamponi comunemente utilizzati in biologia. Infatti essa viene spesso utilizzata nelle fasi iniziali dei progetti di proteomica, durante le quali è eseguito uno *screening* per l'identificazione di un gran numero di candidati proteici per le successive e più approfondite analisi; in questa fase risultano essere molto importanti la produttività ed il numero di proteine identificate positivamente. Le due caratteristiche che, tramite questa metodica, rendono automatizzabili le analisi per l'identificazione delle proteine sono il loro relativamente facile utilizzo e la facile interpretazione degli spettri ottenuti; mediante tecnica di ionizzazione MALDI si ottengono, infatti, quasi esclusivamente ioni monovalenti $[M+H]^+$ [246]. Nella modalità ionizzazione positiva gli ioni molecolari protonati $[M+H]^+$ sono in genere le specie più abbondanti. Questa ionizzazione è utilizzata per le analisi di proteine e peptidi.

Nel caso della modalità ionizzazione negativa le specie più abbondanti sono gli ioni molecolari deprotonati $[M+H]^-$. Questo tipo di ionizzazione viene utilizzata per l'analisi di oligonucleotidi e oligosaccaridi.

Il MALDI-TOF è utilizzato per lo studio di biomolecole con massa molecolare compresa tra 400 e 350×10^3 Da ed essendo un metodo molto sensibile, consente di analizzare quantità molto ridotte di campione con un'accuratezza di 0,1-0,01%.

La matrice utilizzata è solitamente un acido organico debole. Essa ha il compito di assorbire alla lunghezza d'onda del laser, proteggere ed evitare l'aggregazione dell'analita e desorbire quest'ultimo in maniera indipendente dalle sue caratteristiche intrinseche.

Esistono diverse tipologie di matrice adatte a specifiche esigenze di analisi:

- CHCA (Acido α -ciano-4-idrossicinnamico): questa matrice è soprattutto utilizzata negli studi di proteomica ed è utilizzata nell'analisi di proteine e peptidi con un valore di massa fino a 10 kDa, lipidi e carboidrati.
- DHB (Acido 2,5-diidrossi benzoico): questa matrice viene utilizzata nello studio di peptidi, carboidrati neutri e basici, glicolipidi carichi negativamente, polimeri sintetici polari e non polari, e piccole molecole. Negli studi di proteomica viene utilizzata soprattutto nell'analisi di peptidi con massa inferiore a 10 kDa, derivati dalla digestione di proteine glicosilate.
- SA (Acido sinapico): questa matrice è utilizzata nell'analisi di peptidi e proteine con massa molecolare maggiore di 10 kDa, glicoproteine e proteine di membrana. Nel campo della proteomica viene utilizzato per l'analisi di proteine intatte ed il suo impiego in MALDI-TOF è fondamentale per determinare, con elevata precisione, la massa molecolare di una proteina o per stabilire eventuali modificazioni post-traduzionali di proteine a massa molecolare nota.

Una volta selezionata la matrice opportuna, si deve procedere alla scelta del metodo di cristallizzazione da adottare, in quanto anch'esso varia a seconda dell'analita e può influenzare la riuscita dell'esperimento.

La ionizzazione avviene in due fasi distinte chiamate ionizzazione primaria e ionizzazione secondaria [247]; i processi di ionizzazione primaria sono dovuti ad una associazione tra un meccanismo di fotoionizzazione, imputabile ai protoni adsorbiti dalla matrice, ed un meccanismo termico. Il processo di ionizzazione primaria dà come risultato la formazione di molecole di matrice allo stato eccitato. Il processo di ionizzazione secondaria, invece, ha luogo nel plasma di espansione per un processo di trasferimento protonico in fase gassosa; durante questo tipo di trasferimento protonico possono avvenire due tipi di reazione, reazione matrice-matrice e reazione matrice-analita. Durante entrambi i processi di ionizzazione il parametro determinante per la ionizzazione dell'analita è l'affinità protonica; le proteine ed i peptidi hanno una affinità protonica di circa 240 kcal/mol, mentre la maggior parte delle matrici comunemente utilizzate ha valori di affinità protonica inferiori (massimo 223 kcal/mol). Essendo la reazione di protonazione di peptidi e proteine favorita dal punto di vista termodinamico, questa reazione risulta essere preponderante. Il processo è esotermico e l'energia liberata può causare la frammentazione dell'analita; questo permette inoltre di spiegare il processo di frammentazione degli ioni metastabili che viene descritto come una conseguenza dell'aumento di temperatura durante il processo di desorbimento [248]. Se al contrario viene

utilizzata una matrice con una affinità protonica uguale o superiore a quella dell'analita, si avrà una diminuzione della frammentazione.

L'analisi dei campioni, al MALDI-TOF, può essere eseguita in modalità lineare o in riflessione (*reflectron*) (Figura 32). L'analisi in lineare viene impiegata generalmente per campioni ad alto peso molecolare, mentre l'analisi in *reflectron* è prevista soprattutto per lo studio di proteine a basso peso molecolare o di digeriti triptici.

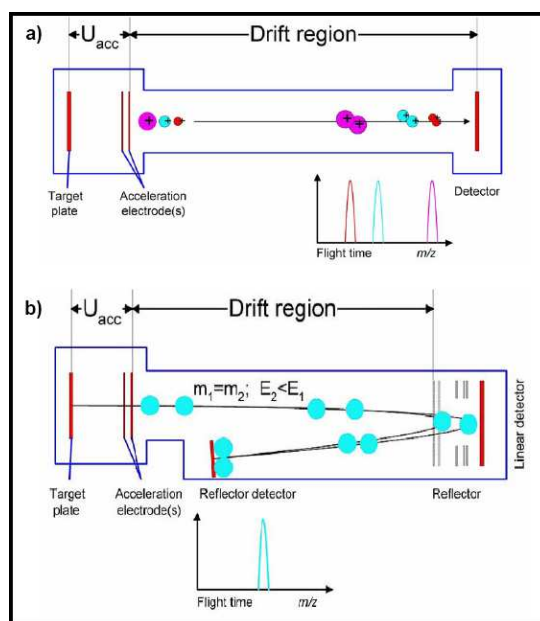


Figura 32. Schema di analizzatori a tempo di volo lineare (a) e *reflectron* (b).

L'analizzatore lineare è più semplice dal punto di vista costruttivo, ma ha minor potere risolutivo (1000), mentre il *reflectron* è più complesso, ma fornisce una risoluzione maggiore anche se non è utilizzabile per tutti gli analiti.

Nella modalità in riflessione, viene accoppiato al normale tubo di volo uno specchio elettrostatico con il compito di rallentare, riaccelerare e rifocalizzare gli ioni mediante un gradiente crescente di potenziali; questo permette quindi di aumentare lo spazio che gli ioni devono percorrere per raggiungere il rivelatore ed annullare le eventuali differenze di energia cinetica acquisite dagli ioni aventi uguale rapporto m/z , a causa di piccole disomogeneità del campo elettrico applicato. Esso viene utilizzato, infatti, per deviare il percorso degli ioni: gli ioni caratterizzati da una maggiore energia cinetica penetrano maggiormente nel campo elettrico creato da piastre metalliche, caricate ad una particolare differenza di potenziale, rispetto agli altri ioni con minore energia cinetica. Dunque, i primi risultano rallentati nel loro tragitto; questo permette agli ioni con diversa energia cinetica e velocità iniziali differenti, ma rapporto m/z uguale, di raggiungere nello stesso tempo il rivelatore. Questo sistema, ovviamente, consente una migliore risoluzione dei profili spettrali (può raggiungere un potere risolutivo pari a 20000) [249], però è adatto unicamente per molecole con massa inferiore a 10 kDa. Come sopra riportato, in seguito alla formazione degli ioni in sorgente, questi vengono accelerati verso il tubo di volo; questo processo non avviene in continuo, ma gli ioni vengono prodotti ed

accelerati a pacchetti discreti. Questi pacchetti vengono accelerati verso il tubo di volo mediante un potenziale decrescente (nel caso di ioni positivi), il quale conferisce a tutti gli ioni la medesima energia cinetica. Gli ioni però all'interno della sorgente non si formano tutti alla stessa distanza dall'inizio del tubo di volo a causa del ritardo che si ha fra la ionizzazione e l'estrazione, questo comporta una piccola differenza nei loro tempi di volo dell'ordine del centinaio di nanosecondi; a questa inoltre va sommata la differenza nel tempo di volo dovuta alla non perfetta omogeneità del campo elettrico applicato. Questi due fenomeni, sommati alle collisioni che possono avvenire fra gli ioni durante il processo di ionizzazione/desorbimento e che causano una dispersione dell'energia cinetica, fanno sì che l'analizzatore lineare abbia una risoluzione bassa e fornisca dei picchi allargati.

L'estrazione ritardata è una tecnica utilizzata per aumentare la risoluzione negli analizzatori TOF. Essa si basa sul principio secondo il quale ioni con velocità iniziali differenti possono essere focalizzati mediante l'uso di potenziali di accelerazione appropriati; l'introduzione di un ritardo comune a tutti gli ioni che si formano in sorgente ne permette la focalizzazione nello spazio correggendo la perdita di risoluzione. L'introduzione di lenti per l'estrazione ritardata ha portato così ad un aumento della risoluzione da 1000 a 5000 per gli analizzatori a tempo di volo lineari, ed ha portato ad un considerevole aumento della risoluzione anche per gli analizzatori *reflectron* che possono così arrivare ad oltre 10000. Un'altra differenza che si riscontra fra un analizzatore a tempo di volo lineare ed uno *reflectron*, è che il primo può essere utilizzato solamente per analisi MS, mentre con il secondo sono possibili anche analisi MS/MS, che consentono di ottenere informazioni parziali sulla sequenza di peptidi e proteine.

Il MALDI-TOF viene utilizzato preferibilmente rispetto ad altri spettrometri di massa, poiché è automatizzabile e consente di ottenere spettri in modo estremamente rapido. Tuttavia questa tecnica presenta alcuni svantaggi, tra cui la soppressione ionica, a causa della quale la copertura di sequenza delle proteine risulta minore rispetto a quella ottenuta con altri spettrometri. Tale fenomeno è dovuto alla differente capacità di ionizzare dei diversi peptidi in una miscela; alcuni di essi, ad esempio quelli particolarmente ricchi in residui acidi, ionizzano meglio di altri provocandone la soppressione nello spettro di massa finale; questo è anche il motivo per cui l'intensità del segnale nello spettro non può essere utilizzata come misura quantitativa dell'abbondanza di un peptide nel campione.

La maggior parte delle identificazioni proteiche viene effettuata con analisi nanoelectrospray-MS/MS [250] ma grazie agli sviluppi tecnologici degli ultimi anni, anche la tecnologia MALDI TOF/TOF permette di ottenere gli spettri MS/MS di peptidi e proteine con una elevata sensibilità (fino alle attomoli).

1.4.5. LA SPETTROMETRIA DI MASSA TANDEM (MS/MS)

Le informazioni strutturali delle molecole sono generalmente ottenute grazie all'utilizzo di spettrometri di massa con analizzatori multipli, noti come *tandem mass spectrometers* (MS/MS), che permettono l'analisi dei prodotti generati dalla frammentazione degli ioni del campione all'interno degli spettrometri di massa [251]. Questa tecnica è utile per determinare la struttura dei composti organici e per il sequenziamento di oligopeptidi, poiché consente di ottenere informazioni sulla composizione aminoacidica; inoltre, può essere

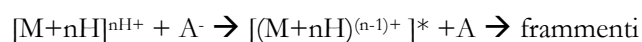
utilizzato anche per determinare la presenza di specifici composti in miscele complesse, basandosi su specifici e caratteristici meccanismi di frammentazione [252, 253]. Uno spettrometro “*tandem*”, nella forma più semplice, è costituito da due analizzatori disposti in serie. Il primo analizzatore (MS_1) ha la funzione di selezionare tra i vari ioni presenti in uno spettro di massa convenzionale lo ione desiderato. Lo ione selezionato viene successivamente fatto collidere con un opportuno gas di collisione (He, Ar) in una cella di collisione e i frammenti ottenuti dalla dissociazione dello ione molecolare vengono separati dal secondo analizzatore (MS_2) in base al rapporto m/z . Diverse tipologie di celle di collisione rompono i legami molecolari secondo schemi differenti.

- ***Collision-induced dissociation (CID)***

La cella di collisione più utilizzata è sicuramente la CID: in spettrometria di massa *tandem*, il CID è un meccanismo di frammentazione degli ioni in fase gassosa [254, 255]. Gli ioni generati in sorgente e accelerati in vuoto dal potenziale elettrico crescente entrano nella cella di frammentazione dove collidono con molecole neutre gassose (solitamente elio, azoto o argon); durante questa collisione, una parte dell'energia cinetica è convertita in energia interna e ciò fa sì che si rompano uno o più legami e lo ione sia frammentato. I tipi di frammenti che vengono prodotti nella cella CID sono soprattutto di tipo *b* [256] e *y*. I frammenti sono poi trasferiti verso l'analizzatore per l'analisi MS/MS.

- ***Electron transfer dissociation (ETD)***

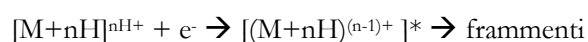
La dissociazione per trasferimento elettronico è un altro metodo per frammentare ioni [257, 258]. In modo analogo alla dissociazione per cattura elettronica, il processo ETD induce frammentazione di cationi trasferendogli elettroni; contrariamente al CID, l'ETD non usa elettroni liberi bensì radicali anionici come l'anione antracene:



dove A è l'anione [259]. La frammentazione ETD è casuale lungo lo scheletro peptidico (cosiddetti ioni *c* e *z*) mentre le catene laterali e le modificazioni post-traduzionali come la fosforilazione rimangono intatte. Questa tecnica di frammentazione è efficace per ioni con cariche alte ($z > 2$) in particolare per frammenti peptidici lunghi o proteine intere. Questo rende ETD una tecnica di frammentazione importante per la caratterizzazione delle modificazioni post-traduzionali.

- ***Electron capture dissociation (ECD)***

La dissociazione per cattura di elettroni è un metodo di frammentazione di ioni in fase gassosa. L'ECD è una tecnica di frammentazione MS/MS che si basa sull'introduzione diretta di elettroni a bassa energia catturati dagli ioni in fase gassosa [260, 261]. La dissociazione per cattura di elettroni tipicamente coinvolge una molecola multiprotonata M che interagisce con un elettrone libero per formare uno ione con un numero dispari di elettroni:



La liberazione del potenziale di energia elettrica provoca la frammentazione dello ione prodotto. L'ECD produce tipi di frammenti ionici significativamente diversi dagli altri metodi di frammentazione MS/MS, soprattutto di tipo c e z ma anche gli ioni b sono stati identificati in ECD [262]. Il sistema CID introduce l'energia interna vibrazionale in modo da causare la perdita di modifiche post-traduzionali durante la frammentazione. In ECD i frammenti mantengono le modifiche post-traduzionali come la fosforilazione [263]. La bassa efficienza di frammentazione che si ha utilizzando il sistema ECD ha limitato la sua diffusione.

1.4.6. LO SPETTRO DI MASSA

Lo spettrometro di massa fornisce i risultati sotto forma di uno spettro di massa, che consiste in una serie di picchi di intensità variabile; la posizione di ogni picco corrisponde ad un determinato valore di m/z . In uno spettro di massa, l'asse X riporta i valori di rapporto m/z e l'asse Y i valori di abbondanza relativa degli ioni analizzati. Le intensità dei picchi sono espresse in percentuali del picco più intenso, il cosiddetto picco base, cui si assegna arbitrariamente il valore 100. Gli spettri prodotti sono normalizzati, abbastanza indipendenti dallo strumento impiegato e quindi sono comparabili direttamente. Se la risoluzione dello strumento è sufficientemente elevata, è possibile determinare la massa esatta dei singoli ioni, dalla quale si può dedurre la composizione elementare dello ione stesso. Dallo spettro MS/MS si può, infatti, risalire alla struttura di un composto sconosciuto, attribuendo ai singoli ioni una composizione elementare e ricostruendo i meccanismi di frammentazione.

Nell'interpretazione di uno spettro si segue generalmente una determinata procedura:

- Identificazione dello ione molecolare o molecola ionizzata
- Identificazione di ioni caratteristici
- Identificazione di processi di frammentazione caratteristici
- Ricostruzione della struttura della molecola sulla base della conoscenza dei meccanismi di frammentazione standard.

Le qualità che uno spettro di massa acquisito deve avere per poter essere utilizzato sono:

- un buon rapporto segnale/rumore di fondo
- l'intensità dei picchi non deve superare la soglia di acquisizione, poiché in questo caso risulterebbe difficile attribuire loro un valore esatto
- picchi ben risolti, ovvero alti e stretti.

Al termine dell'acquisizione si ottiene uno spettro di massa finale, che viene rielaborato da programmi specifici; quest'ultimi riducono e/o eliminano per sottrazione il rumore di fondo, evidenziano i picchi relativi alla matrice ed i possibili addotti con ioni metallici (principalmente sodio e potassio), infine selezionano i picchi più intensi associando ad essi un valore di m/z .

1.5. HIGH PERFORMANCE LIQUID CHROMATOGRAPHY (HPLC)

La cromatografia liquida ad alta prestazione è una forma di cromatografia liquida che, tramite l'utilizzo di colonne di dimensioni minori rispetto alla LC tradizionale (anche capillari), permette di separare volumi di campione anche molto ridotti in breve tempo; questo sistema necessita però di elevate pressioni. Una tipica strumentazione per HPLC è composta da contenitori con i solventi della fase mobile, un sistema di pompe, un iniettore, una colonna per la separazione ed un *detector*.

Il campione da analizzare è iniettato all'inizio della colonna cromatografica dove è trascinato attraverso la fase stazionaria dalla fase mobile, applicando pressioni dell'ordine delle centinaia di atmosfere. Per ottenere un'elevata efficienza nella separazione è necessario che le dimensioni delle particelle di riempimento siano molto ridotte ed è indispensabile applicare un'elevata pressione se si vuole mantenere una giusta velocità di flusso dell'eluente e quindi un tempo di analisi conveniente. La lunghezza delle colonne è solitamente compresa tra 10 e 30 cm; il diametro interno è compreso tra 4 e 10 mm, mentre il diametro delle particelle del riempimento tra 3 e 10 μm . I solventi devono essere privi di impurità e degasati per evitare la formazione di bolle nel circuito che possono inficiare la corretta separazione dei composti; le pompe provvedono a generare una pressione elevata, costante e possono essere programmate per variare la composizione del solvente durante la separazione. La misurazione del segnale al detector può essere basata su diversi parametri, come il cambiamento dell'indice di rifrazione, l'assorbimento nell'UV-visibile o l'emissione di fluorescenza dopo eccitazione a determinate lunghezze d'onda. La dimensione ridotta della colonna evita movimenti longitudinali e percorsi alternativi della fase mobile; inoltre la colonna può essere utilizzata per numerosi cicli di analisi. I tempi di lavoro risultano quindi molto brevi e si ha la possibilità di separare sostanze termolabili e sostanze non volatili. Sono necessari volumi molto piccoli di campione (ng) grazie alla elevata sensibilità raggiunta oggi dai rivelatori (ad esempio, gli spettrometri di massa); tutto questo fornisce una maggiore accuratezza e precisione d'analisi e rappresenta un vantaggio fondamentale nel campo delle ricerche biologiche e biochimiche dove la disponibilità di campione è spesso molto limitata. Per questi motivi l'HPLC ha riscontrato ultimamente notevoli applicazioni, quali la separazione di farmaci e loro metaboliti, vitamine, polifenoli e, in campo proteomico, peptidi.

1.5.1. NANO-CROMATOGRAFIA LIQUIDA (nano-LC)

Un incremento della sensibilità, in cromatografia liquida, è stato raggiunto con l'introduzione della nano-cromatografia liquida (nano-LC) [264] che riduce ulteriormente i volumi di flusso e il diametro della colonna utilizzata (sotto il millimetro). Nel campo della proteomica l'accoppiamento nanoLC-ESI-MS [265] è diventato la strategia analitica più utilizzata per la caratterizzazione e l'identificazione proteica; essa permette un'analisi riproducibile e molto sensibile, vantaggio derivante dall'alta capacità separativa dello strumento nano-LC. I sistemi nano-LC hanno quindi assunto il ruolo di sistema di introduzione preferenziale per la spettrometria ESI-MS/MS. La riduzione del diametro delle colonne presenti nei sistemi cromatografici

nano-LC consente di ottenere velocità di flusso che vanno dall'ordine dei nL./min a quello dei $\mu\text{L./min}$ così da essere compatibili con le sorgenti *electrospray* miniaturizzate che lavorano con flussi di nL./min (in genere 300-200 nL./min), riducendo così al massimo il consumo di campione (il volume di campione da iniettare è di solito nell'ordine dei microlitri); la riduzione del flusso è associata ad un aumento della sensibilità, in quanto aumentano il numero di molecole che arrivano al rilevatore. Dimezzando il diametro interno della colonna analitica la sensibilità del sistema aumenta di quattro volte (aumento quadratico della sensibilità): l'impiego di colonne per nano HPLC con diametro interno pari a $75\ \mu\text{m}$ consente di ottenere una sensibilità tremilaottocentovolte maggiore di una colonna tradizionale poiché la concentrazione di analiti altrettanto abbondanti nella fase mobile LC è proporzionale al quadrato del diametro interno della colonna (Figura 33).

Concentration Factor	
Concentration:	$(d\ \text{standard}/d\ \text{micro})^2$
Capillary:	$(4.6/0.3)^2 = 235\ \times$
Nano LC	$(4.6/0.75)^2 = 3800\ \times$

Figura 33. Legame tra diametro della colonna cromatografica e sensibilità.

Ulteriori vantaggi derivanti dell'impiego di una colonna per nano-LC sono:

- riduzione considerevole dell'impiego di solventi organici rispetto alle tecniche tradizionali;
- maggiore efficienza di separazione e dunque una maggiore separazione dei picchi cromatografici.

Solitamente prima di introdurre il campione in un sistema nano-LC è richiesta una fase di rimozione dei sali: infatti, puntali di estrazione C18 monouso possono essere usati per rimuovere le particelle insolubili e i sali, oltre che per concentrare il campione. Per lo stesso motivo i dispositivi per nano-LC sono dotati di precolonne dove il campione viene purificato, de salificato e concentrato prima di essere iniettato nella colonna analitica [266]. La separazione mediante nano-LC oltre a garantire eccellente sensibilità, fornisce anche elevata riproducibilità, efficienza di separazione ed è inoltre molto semplice da utilizzare; tuttavia in un sistema nano-LC non è possibile introdurre campioni troppo concentrati e la riduzione del diametro interno della colonna richiede l'utilizzo di pressioni maggiori rispetto a quelle convenzionali dei sistemi HPLC; quest'ultima limitazione è stata superata dallo sviluppo di sistemi di cromatografia liquida a prestazioni ultra-elevate.

1.5.2. CROMATOGRAFIA LIQUIDA A PRESTAZIONI ULTRA-ELEVATE (UHPLC)

Nell'ultimo decennio sono stati sviluppati sistemi ancora più performanti noti come UHPLC (Ultra High Performance Liquid Chromatography), con i quali la cromatografia ha fatto un ulteriore passo in avanti. L'UHPLC è il risultato di un'innovazione riguardante la tecnologia delle particelle, delle colonne, degli iniettori e delle pompe, allo scopo di ottenere una maggiore sensibilità, risoluzione e velocità di analisi. In particolare viene combinato l'uso di particelle ibride con granulometria inferiore ai 2 μm per la fase stazionaria alla capacità dell'LC di erogare fase mobile ad alte pressioni evitando dispersioni. Lo sviluppo di UHPLC è infatti andato di pari passo con la diffusione di colonne in silice porosa che permettono di migliorare la risoluzione e di ottenere corse cromatografiche più rapide: infatti, colonne corte con particelle piccole offrono la possibilità di ridurre drasticamente i tempi di analisi aumentando la velocità di flusso senza abbassare le prestazioni in termini di separazione. Nel caso invece della separazione di campioni complessi, colonne lunghe con particelle e diametro ridotti offrono un'efficienza superiore e una maggiore risoluzione dei picchi; di fatto una minore dispersione implica anche meno diluizione dei picchi dell'analita, con conseguente guadagno in termini di sensibilità. La pressione necessaria per guidare il solvente attraverso queste colonne aumenta però rapidamente al crescere della velocità di flusso al crescere della lunghezza della colonna e al diminuire del diametro della colonna e del particolato. Dunque per operare in tali condizioni, è stato necessario sviluppare dei sistemi LC che consentano pressioni superiori rispetto ai consueti 400 bar (fino circa 1000 bar). Il principio che sottende all'evoluzione in sistemi UHPLC, è governato dall'equazione di Deemter, formula empirica che descrive la relazione tra la velocità lineare e l'altezza del piatto teorico (HEPT o efficienza della colonna). La curva di Van Deemter viene usata per stabilire la velocità di flusso ottimale (punto minimo della curva) per ottenere la migliore efficienza di separazione da una colonna; l'efficienza è proporzionale alla lunghezza della colonna e inversamente proporzionale alle dimensioni delle particelle presenti in essa. In conclusione, alcuni dei vantaggi che si possono ottenere da sistemi di cromatografia liquida a prestazioni ultra-elevate sono:

- riduzione del tempo di analisi e aumento della sensibilità
- mantiene la selettività, la specificità e il range dinamico dei sistemi LC
- mantiene le performance risolutive
- riduzione del costo d'analisi
- Minore consumo di solvente.

Tuttavia questo sistema non è esente da svantaggi, tra cui la richiesta di una manutenzione più frequente e un consumo più rapido della colonna a causa delle alte pressioni a cui viene sottoposta [267].

1.5.3. LA CROMATOGRAFIA LIQUIDA ASSOCIATA ALLA SPETTROMETRIA DI MASSA

La tecnica LC-MS è uno strumento potente per l'analisi di peptidi e proteine; questo approccio combina l'efficiente separazione dei componenti del campione eterogeneo di partenza (per esempio una miscela di peptidi derivanti da digestione enzimatica) con l'identificazione sensibile di ogni singolo analita mediante spettrometria di massa e ricerca in banca dati. Nonostante ci siano numerosi metodi per interfacciare la cromatografia liquida alla spettrometria di massa, è stato l'accoppiamento con la ionizzazione *electrospray* (ESI) [268] a rendere la tecnica LC-MS una procedura abbastanza sensibile per analizzare peptidi e proteine in studi di proteomica.

In una sorgente ESI il campione viene introdotto nello spettrometro direttamente in soluzione e forzato a passare attraverso un capillare insieme ad un gas di nebulizzazione inerte, solitamente azoto (N₂); il potenziale applicato alla punta del capillare è tale da nebulizzare la soluzione in una miriade di goccioline cariche contenenti il campione. La nebulizzazione ed il processo di evaporazione del solvente vengono coadiuvati da due flussi di gas inerte, uno coassiale al flusso dell'analita (chiamato *nebulizer*) e l'altro, invece, controcorrente rispetto al flusso dell'analita e riscaldato (chiamato *drying gas*). Nella sorgente il gas (*drying gas*) fluisce a bassa velocità e a temperatura relativamente alta, facilitando l'evaporazione del solvente e permettendo di analizzare anche composti termicamente labili come proteine e peptidi in studi di proteomica. La differenza di potenziale applicata tra il capillare, all'interno del quale passa l'analita in soluzione, ed il controlettrodo, che si trova a qualche centimetro dalla sorgente, provoca la formazione di gocce di piccole dimensioni e cariche che all'estremità del capillare assumono la forma di un cono detto cono di *Taylor*. Le sorgenti *electrospray* miniaturizzate o *nano-Electrospray* (nano-ESI), lavorano con flussi nell'ordine di nanolitri al minuto [269], riducendo così al minimo il consumo di solvente, concentrando gli analiti in un volume di eluzione piccolo, aumentando la sensibilità analitica e permettendo quindi analisi con quantità minime di analita; questo le rende particolarmente adatte per il sequenziamento di peptidi o proteine. In generale il capillare metallico della sorgente nano-ESI è in silice fusa e ha un diametro interno di 10-15 µm; esso è rivestito esternamente da un sottile strato metallico in modo che possa condurre corrente. Le dimensioni delle gocce di analita che si vengono a formare sono dunque fortemente ridotte e con una densità di carica superiore ed un rapporto superficie/volume maggiore; ciò permette un desorbimento degli ioni più veloce. In questo caso, la ionizzazione non necessita dell'utilizzo dei gas di solvatazione coassiali. Inoltre la distanza fra il capillare ed il controlettrodo è minore e variabile e la differenza di potenziale applicata è minore, circa 1000 V, in modo da evitare scariche a corona all'interno della sorgente. Tipicamente, fra la sorgente e l'analizzatore di massa vi è una zona che costituisce l'interfaccia; questa assicura la transizione e la focalizzazione degli ioni nel passaggio dalla sorgente, che lavora a pressione atmosferica, e l'analizzatore di massa che si trova invece sotto alto vuoto. L'interfaccia è costituita inizialmente da un capillare di vetro, avute le estremità ricoperte da un materiale conduttore; a queste estremità è applicata una differenza di potenziale tale da favorire l'attraversamento del capillare da parte degli ioni. Lenti elettrostatiche vengono utilizzate per la focalizzazione del fascio di ioni, cosicché gli

ioni sono trasportati all'analizzatore sotto l'effetto dei gradienti di pressione e potenziale. Al processo di focalizzazione degli ioni contribuiscono fortemente anche altri elementi, ad esempio esapoli o quadrupoli, che permettono il convogliamento del fascio di ioni al centro delle barre, trasmettendo un fascio di ioni monocinetico verso l'analizzatore. Inoltre nell'interfaccia si possono trovare anche due dispositivi in serie chiamati *dual-funnel*, leggermente disallineati tra di loro, che hanno il compito di facilitare la definitiva eliminazione dei solventi eventualmente presenti in forma gassosa all'interno dello strumento aumentando la focalizzazione del fascio di ioni. La differenza di potenziale generata fra la sorgente e l'interfaccia accelera gli ioni, che accumulano quindi anche energia interna; questi possono frammentarsi in seguito alla collisione con molecole di gas residue all'interno dello strumento (*in source fragmentation*). Sfruttando questa caratteristica è possibile favorire o meno tale frammentazione, aumentando o diminuendo l'energia con la quale si indirizzano gli ioni verso l'analizzatore. Recentemente, sono state sviluppate sorgenti ottimizzate tra cui la *CaptiveSpray nanoBooster™* in cui si è cercato di ottimizzare le condizioni di ionizzazione (Figura 34).

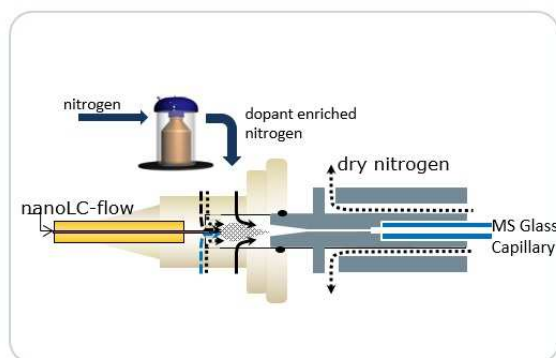


Figura 34. *CaptiveSpray nanoBooster*

Queste sorgenti presentano due importanti innovazioni. La prima innovazione è stata introdotta per favorire il processo di focalizzazione del cono di Taylor a livello dell'ingresso nello spettrometro di massa; infatti lo *sprayer* non è più perpendicolare ma è convogliato all'ingresso del capillare; inoltre non solo il *drying* gas è riscaldato ma anche il *nebulizer*. La seconda innovazione è la possibilità di queste sorgenti di fruire di gas di solvatazione arricchiti con dopanti, quali ad esempio l'acetonitrile, che aumentando o diminuendo il numero di cariche facilitano la ionizzazione di particolari composti. Queste strategie innovative mirano ad aumentare sempre di più il potere identificativo di questi strumenti; infatti permettono di ottenere un aumento dell'intensità del segnale, un aumento della capacità di ionizzazione con un effetto positivo sul potere identificativo dello strumento.

2. SCOPO

I tumori renali maligni rappresentano il 4% di tutte le neoplasie dell'uomo [14, 270] e tra questi, il carcinoma a cellule renali (RCC) costituisce circa l'85% dei tumori renali primari e la sua incidenza è in continua crescita [22, 23]. Il RCC comprende un gruppo di tumori eterogeneo a livello clinico-patologico con diversa prognosi, in cui la variante a cellule chiare (ccRCC) è la più frequente rappresentando circa il 60% dei tumori del rene [39]. Questa patologia ha una sintomatologia che si manifesta solo negli ultimi stadi della malattia. Inoltre, il tempo medio di sopravvivenza si attesta intorno ai 13 mesi, con meno del 10% dei pazienti che sopravvivono più di 5 anni [271]. Circa il 30% dei pazienti presentano metastasi già al momento della diagnosi e il 30-40% di soggetti aventi una massa localizzata andrà incontro a recidiva [272]. Nonostante l'utilizzo, sempre più frequente, di tecniche di diagnostica per immagini abbia migliorato e aumentato la rilevazione del RCC in fase asintomatica, tuttavia, la natura di alcune particolari lesioni renali, anche benigne, non può essere distinta con confidenza utilizzando le sole tecniche di imaging. Inoltre, ad oggi, la classificazione TNM (Tumor-Node-Metastasis), il grading secondo Fuhrman e la misurazione dei marcatori serici risultano essere ancora i principali fattori utilizzati per predire l'*outcome* del paziente. Per questi motivi, l'RCC rappresenta una sfida sia per quanto riguarda la diagnosi precoce ma anche per la gestione clinica dei pazienti.

Dato che le principali alterazioni dei processi regolatori che sono alla base della tumorigenesi sono a carico non solo dell'espressione genica ma anche delle proteine, le investigazioni proteomiche costituiscono un valido approccio per la scoperta di biomarcatori multipli e di pattern di alterazioni proteiche specifiche della patologia in esame. In particolare, la proteomica clinica è un campo di ricerca in continua espansione i cui obiettivi primari sono l'identificazione dei meccanismi fisiopatologici della malattia, il controllo dell'efficacia terapeutica e la diagnosi precoce. Negli ultimi anni le tecniche di *protein profiling* hanno subito un sostanziale miglioramento tecnologico; a tal proposito, sono stati di grande rilievo non solo i progressi nella strumentazione ma anche quelli relativi agli aspetti bioinformatici. In effetti, questi ultimi hanno consentito lo sviluppo di programmi in grado di elaborare a fini diagnostici o prognostici le centinaia di migliaia di informazioni ottenute dal processo analitico e hanno reso gli approcci proteomici compatibili con la ricerca clinica.

Il proteoma dei fluidi biologici fornisce informazioni sullo stato dell'organismo in quanto essi veicolano proteine secrete dai tessuti. In particolare, la ricerca nelle urine di proteine correlate ad un alterato stato fisiopatologico, rappresenta un valido approccio di *biomarker discovery*, in quanto questo fluido biologico può essere ottenuto in maniera semplice con metodi per nulla o poco invasivi. Inoltre, le urine risentono poco del range dinamico di concentrazione proteico [273], rispetto ad altri fluidi biologici come il siero, risultando sufficientemente stabili se raccolte con protocolli standardizzati [170]. Inoltre, il contenuto proteico/peptidico della matrice urinaria risulta fortemente influenzato dai cambiamenti indotti dagli organi coinvolti nell'escrezione quali ad esempio i reni. Attraverso l'analisi di tale fluido è possibile ottenere informazioni sulle isoforme, sulle modificazioni post-traduzionali e sul livello di degradazione delle proteine

e sui peptidi endogeni naturalmente presenti [157]. Infine, occorre considerare che le cause alla base di molte malattie tra cui anche il RCC sono spesso multifattoriali e la complessità biologica associata alla patologia spesso limita il processo di comprensione dei meccanismi molecolari implicati [215]. L'integrazione di informazioni ottenute da diversi approcci -omici o da diverse piattaforme tecnologiche potrebbe rappresentare una possibile soluzione a tale problema in quanto apporterebbe un forte beneficio e permetterebbe di focalizzare l'attenzione su target di interesse [274, 275].

Pertanto questo progetto di dottorato si è focalizzato sullo studio quali-quantitativo del peptidoma e del proteoma urinari di una ampia coorte di pazienti affetti da RCC e volontari sani, al fine di costruire modelli di potenziali biomarcatori in grado di discriminare le diverse classi in studio, di distinguere significativamente le lesioni maligne da quelle benigne e dai controlli e di individuare peptidi endogeni la cui espressione urinaria fosse alterata secondo lo stage, il grade e/o la dimensione tumorale. È da notare che, sebbene il significato diagnostico di questi segnali sia indipendente dalla determinazione della loro sequenza amminoacidica, la loro identificazione potrebbe chiarire il ruolo biologico incrementando le conoscenze a nostra disposizione sulle funzioni molecolari coinvolte nella malattia. Pertanto una volta individuati i segnali d'interesse appartenenti ai modelli discriminanti o variati in relazione alla progressione tumorale, si è provveduto all'identificazione della sequenza amminoacidica mediante la tecnica nLC-ESI-MS/MS. Infine, ulteriore obiettivo del lavoro di tesi è stato quello di integrare tutte le informazioni ottenute dai diversi approcci sia a livello proteomico che peptidomico in modo da ottenere una panoramica più ampia dei cambiamenti sottesi dalla trasformazione neoplastica e far così emergere eventuali proteine *target* da caratterizzare e validare.

3. MATERIALI E METODI

3.1. DISEGNO SPERIMENTALE

Il presente lavoro di tesi ha previsto l'analisi del proteoma e del peptidoma urinari per la valutazione delle firme molecolari tipiche della patologia tumorale in pazienti affetti da carcinoma renale (RCC) e volontari sani secondo il protocollo sperimentale illustrato in Figura 35.

Il processo analitico che è stato seguito ha previsto, per l'analisi peptidomica, un singolo passaggio di pre-frazionamento di campioni di urine, mediante l'utilizzo di biglie magnetiche a fase inversa di tipo C8. In seguito i campioni purificati sono stati analizzati mediante spettrometria di massa MALDI-TOF per acquisire i rispettivi profili spettrali e tramite nLC-ESI-MS/MS per la successiva fase identificativa, al fine di rilevare possibili firme molecolari di peptidi urinari in pazienti affetti da ccRCC, da altre tipologie di tumori renali e in soggetti controllo. La valutazione del proteoma urinario è stata invece effettuata mediante digestione enzimatica di due pools di urine e successiva analisi quali-quantitativa tramite analisi nLC-ESI MS/MS.

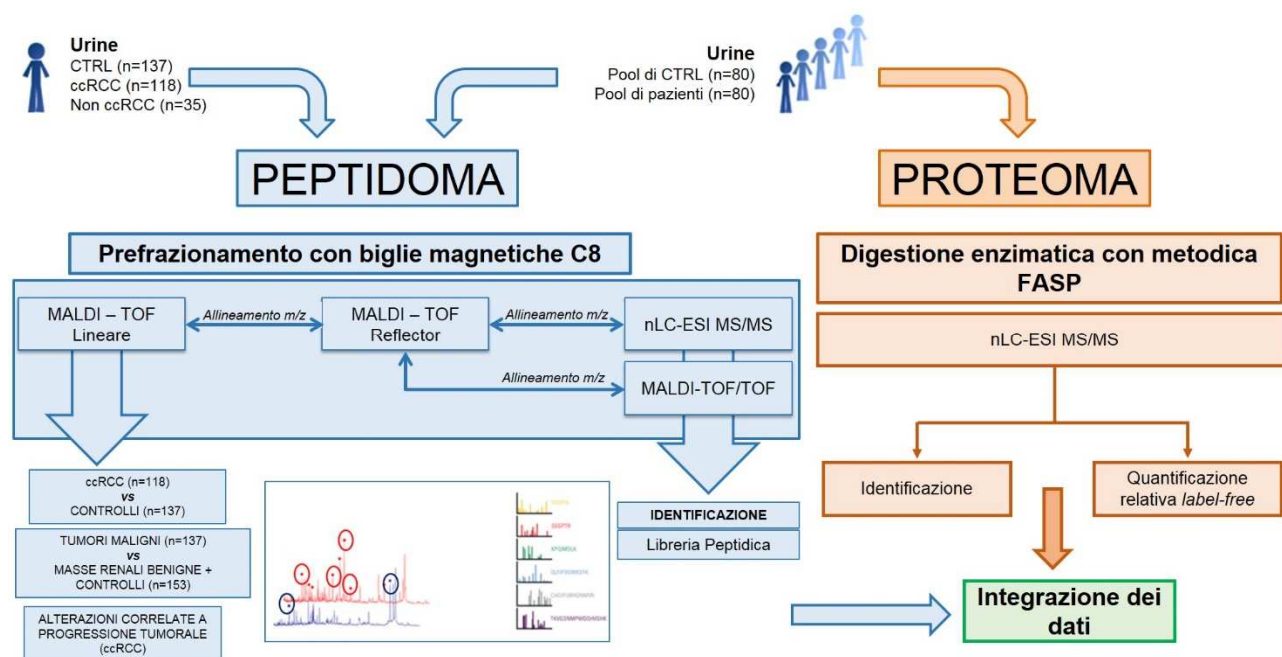


Figura 35. Schema riassuntivo del disegno sperimentale

3.2. RACCOLTA E PROCESSAMENTO DEI CAMPIONI DI URINE

I campioni di urine sono stati raccolti da volontari sani e da pazienti con tumore renale il giorno prima dell'intervento presso la fondazione "Ospedale Maggiore Policlinico" (Milano, Italia), presso l'Ospedale San Gerardo (Monza, Italia) e presso la struttura ospedaliera di Desio (Desio, Italia). Tutti i soggetti hanno firmato un consenso informato prima della donazione. I protocolli di raccolta e le procedure di preparazione dei campioni attuati sono stati approvati dal comitato etico locale e le analisi sono state effettuate in accordo con la *Dichiarazione di Helsinki*. Il protocollo di raccolta usato è coerente con quello europeo per analisi di proteomica proposto da *Eu Action EuroKUP (Urine and kidney Proteomics)*. Nel dettaglio, è stato raccolto il mitto intermedio della seconda minzione mattutina in provette sterili. Entro un'ora il campione è stato centrifugato a 4°C per 10 minuti a 1000 g per rimuovere i detriti cellulari. Il surnatante recuperato è stato quindi aliquotato in provette da 2 mL e conservato a -80°C fino al momento dell'analisi. Una provetta per ciascun campione è stata scongelata per la successiva procedura automatica di isolamento peptidico e per il pre-frazionamento manuale dei *pool* urinari.

3.3. ANALISI DEL PEPTIDOMA URINARIO

3.3.1. ESTRAZIONE DEL PEPTIDOMA URINARIO MEDIANTE BIGLIE MAGNETICHE

Al fine di ottenere l'estrazione specifica di proteine/peptidi dai campioni di urine, sono state utilizzate biglie magnetiche a fase inversa di tipo C8 (*Hydrophobic Interaction Chromatography*, HIC 8). Lo stesso *kit* di *profiling* è stato impiegato per purificare tutti i campioni in studio e controllato prima dell'uso mediante indagine al microscopio ottico, al fine di valutare l'ottimale funzionalità della sospensione di biglie. La procedura di estrazione è stata eseguita tramite un sistema automatizzato, che riduce la variabilità dovuta alla preparazione manuale dei campioni, basato sull'utilizzo di una piattaforma robotizzata a 8 canali, *ClinProt Robot™*.

Brevemente, 40 µL di urine sono stati aggiunti a 5 µL di biglie e 80 µL di *binding buffer*, miscelati e incubati a temperatura ambiente per un minuto. Dopo rimozione del surnatante, mediante un tampone di lavaggio vengono eliminate le proteine legate in maniera aspecifica, i sali e i contaminanti, che possono interferire quindi con la successiva analisi in spettrometria di massa; questa fase prevede una serie di tre lavaggi, due da 45 µL e uno da 30 µL di *washing solution* (fornita dal *kit*). Successivamente i peptidi/proteine sono stati eluiti con 10 µL di una soluzione 50% acetonitrile/H₂O per permettere la rottura del legame tra le proteine e le biglie. Due *pools* costituiti rispettivamente da urine di 80 pazienti e 80 soggetti sani sono stati ottenuti miscelando 40 µL di urina di ciascun soggetto. Tali *pools* sono stati purificati manualmente con biglie magnetiche HIC 8, utilizzando lo stesso protocollo sopra descritto per la procedura automatizzata, ma adattando i volumi al campione di partenza. Gli eluati sono stati conservati a -80°C per le successive analisi

in MALDI-TOF e nLC-ESI-MS/MS.

3.3.2. DEPOSITO SU TARGET ED ANALISI MALDI-TOF

Al fine di ottenere in spettrometria di massa MALDI-TOF i profili dei campioni di urine di tutti i pazienti studiati, gli eluati peptidici purificati con biglie sono stati depositati su una piastra metallica in maniera automatizzata sempre mediante utilizzo del Robot. In dettaglio, 4 μ L di ciascuna frazione eluita sono stati miscelati con 15 μ L della soluzione di matrice CHCA (acido α -ciano-4-idrossicinnamico) (0,3 g/L in etanolo/acetone 2:1) in una piastra da 96 pozzetti dedicata. Successivamente, 0,8 μ L della miscela sono stati depositati in quadruplicato direttamente su un target MALDI *AnchorChip*TM 600/384. Il target è stato lasciato asciugare qualche minuto per favorire l'evaporazione del solvente con co-cristallizzazione del campione in condizioni di temperatura e umidità controllate e quindi trasferito nella sorgente MALDI.

I campioni purificati sono stati analizzati con lo spettrometro di massa *UltrafleXtreme*TM MALDI-TOF/TOF e gli spettri di massa sono stati acquisiti automaticamente in modalità lineare positiva (LM). L'acquisizione è stata eseguita in un *range* di m/z da 1 a 12 kDa. Prima di procedere all'acquisizione dei profili proteici dei campioni, è stata eseguita una calibrazione esterna dello spettrometro di massa; esistono miscele di calibrazione diverse a seconda dell'intervallo di valori di m/z in cui ricade la massa molecolare del campione da analizzare. In questo studio è stato utilizzato lo standard *ProtMix I*, miscela di calibrazione adeguata per proteine a basso peso molecolare. Le analisi sono state eseguite con il metodo automatico *AutoXecute* (versione 3.3.108.0) del *software FlexControl*TM 3.3 utilizzando i seguenti parametri:

- *Ion Source 1* = 25,13 kV;
- *Ion Source 2* = 23,77 kV;
- *Lens* = 6,46 kV;
- *Pulsed ion extraction* = 300 ns;
- *High gating factor with matrix suppression up to* 900 Da.

La ionizzazione del campione è stata ottenuta con un laser *Smartbeam*^{TM-II solid-state} che opera ad una frequenza di 1000 Hz. L'acquisizione viene eseguita sommando spettri ottenuti da spari (bombardamenti con fotoni) successivi in posizioni diverse del pozzetto in cui è stato depositato il campione; in questo modo si riesce a migliorare il rapporto segnale/rumore di fondo dello spettro. Per ogni deposito, gli spettri sono stati ottenuti in 6 posizioni diverse (200 spari per ogni posizione) e sommati (1200 spari soddisfacenti). Poiché la matrice non cristallizza in modo uniforme è importante che venga analizzato tutto il pozzetto contenente il campione per poter ottenere il maggior numero di informazioni ed il più possibile riproducibili. L'acquisizione dei dati è stata eseguita impostando la potenza del laser al 60% della sua energia massima. La massa accurata dei peptidi è stata poi determinata in modalità *reflectron* positiva (RM) e gli spettri di massa sono stati calibrati esternamente utilizzando i peptidi standard della miscela *PepMix II*.

I parametri dello strumento erano i seguenti:

- *Ion Source 1* = 25,13 kV;
- *Ion Source 2* = 22,43 kV;

- $Lens = 7,45$ kV;
- *Deflection mode con matrix suppression up to 400 Da.*

Gli eluati ottenuti dai *pools* di urine prefrazionate con biglie magnetiche sia da pazienti che da controlli sono stati concentrati e desalificati come descritto nel paragrafo 3.3. Un'aliquota di ciascun eluato è stata utilizzata per l'analisi in MALDI-TOF, mentre il rimanente impiegato nelle successive indagini in nano-LC-ESI-MS/MS. Per quanto riguarda l'analisi di *profiling* in MALDI-TOF, 2 μL della frazione eluita sono stati miscelati con 4 μL di una soluzione di matrice CHCA (6.2 g/L CHCA in metanolo/acetonitrile/acqua 50:40:10). Quindi 0,8 μL di questa miscela sono stati depositati direttamente sul target MALDI-TOF MTP 384 *ground steel*. Il target è stato lasciato asciugare all'aria e immediatamente trasferito nello spettrometro di massa MALDI-TOF/TOF.

3.3.3. PROCESSAMENTO DEI PROFILI SPETTRALI

La visualizzazione degli spettri di massa, acquisiti con lo spettrometro di massa MALDI-TOF/TOF, ed il processamento iniziale dei dati sono stati eseguiti con il *software FlexAnalysis™* 3.3. Il pre-processamento degli spettri è necessario per ridurre la variabilità sperimentale all'interno del *dataset* prima dell'analisi statistica; il procedimento consiste in una correzione della linea di base e in un riallineamento degli spettri grezzi utilizzando la funzione di *Batch Process*. La sottrazione della linea di base è stata eseguita utilizzando l'algoritmo *TopHat*. Successivamente è stato individuato un insieme di picchi comuni alla maggior parte degli spettri in studio e i cui valori di m/z sono stati impiegati per l'allineamento. Relativamente ai dati acquisiti in modalità lineare, sono stati individuati sette picchi (valori medi assegnati ai picchi: m/z 1162, m/z 1511, m/z 1681, m/z 1895, m/z 1912, m/z 2236 e m/z 3373) a seguito di un'analisi manuale degli spettri, in modo tale da creare una lista di masse di riferimento avente una tolleranza di 1000 ppm.

È stata inoltre eseguita una calibrazione interna anche dei dati ottenuti in modalità *reflectron*, utilizzando quattro segnali comuni (a m/z 1680,93, m/z 1912,06, m/z 2040,17 e m/z 2659,32) e una tolleranza nell'assegnazione del picco di 100 ppm, ad eccezione del segnale a m/z 2659,32 la cui tolleranza è stata impostata a 200 ppm. Per l'allineamento è stato applicato un algoritmo di calibrazione quadratica. L'assegnazione dei valori m/z ai picchi dei profili spettrali è stata ottenuta impiegando l'algoritmo *SNAP*, con una soglia segnale/rumore di 2.

3.3.4. ANALISI STATISTICA DEI DATI SPERIMENTALI

▪ Scelta della lista di picchi

La comparazione degli spettri è stata eseguita mediante il *software ClinProTools™* 2.2. Prima del calcolo statistico dell'area dei picchi, sono state effettuate sugli spettri alcune operazioni per ottimizzare il numero e l'accuratezza dei segnali: una normalizzazione sulla corrente ionica totale (TIC) di tutti gli spettri in elaborazione e un nuovo allineamento in base ai picchi di maggiore intensità.

La selezione dei dati è stata ottenuta applicando i seguenti filtri:

- Risoluzione = 800;
- *Convex Hull* = 0,8;
- Esclusione degli spettri privi di segnali o non ricalibrabili.

Lo spettro medio di ciascun campione, dunque solo uno spettro sui quattro acquisiti per ogni soggetto mediante analisi MALDI-TOF, è stato poi usato per l'elaborazione statistica. L'individuazione dei picchi negli spettri è stata eseguita impostando un valore di rapporto S/N di 3 e le aree dei picchi sono state calcolate utilizzando un'integrazione di tipo *end-point level*.

In tal modo il *software* ha generato una lista di picchi (m/z) con ciascuno il proprio valore di area, che è stata utilizzata per la successiva analisi statistica.

▪ Individuazione delle differenze di abbondanza relativa tra profili spettrali

L'analisi statistica dei profili proteici ottenuti dopo purificazione dei campioni con biglie a fase inversa (C8) è stata eseguita con l'aiuto del Dipartimento di Informatica Sistemistica e Comunicazione dell'Università Milano-Bicocca.

Uno schema dei processi che sono stati effettuati è riportato in Figura 36.

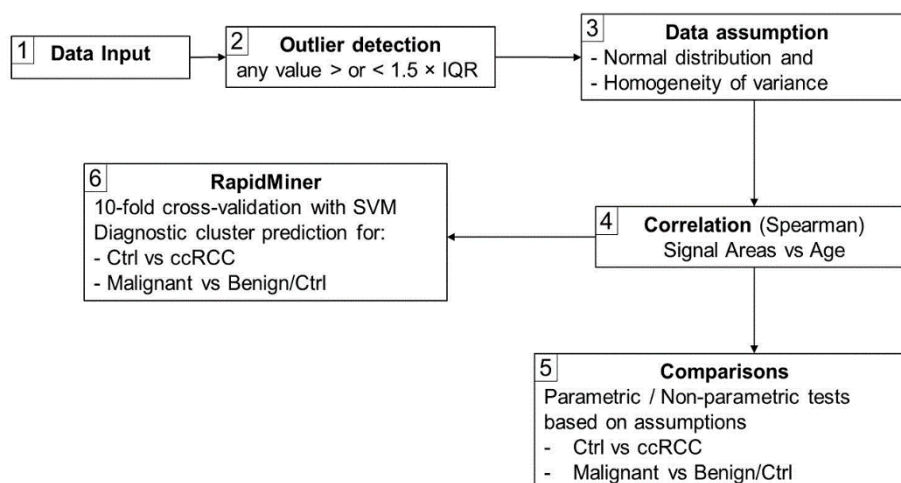


Figura 36. Schema generale delle analisi statistiche effettuate usando il *software* statistico *RapidMiner*.

Al fine di operare una corretta analisi statistica, in primo luogo è stata valutata l'ipotesi di una distribuzione normale dei dati e l'omogeneità della varianza; ad esempio è stato verificato per ciascun segnale presente negli spettri di pazienti/controlli se avesse una distribuzione normale e per ciascun segnale è stato verificato che sia i dati relativi ai controlli che quelli relativi ai pazienti avessero la stessa varianza (Figura 36, riquadro 3).

Quindi appropriati test parametrici e non parametrici sono stati utilizzati per il confronto e la correlazione tra i pazienti e i controlli. Per la verifica della normalità è stato applicato il test di *Shapiro-Wilk* (livello di significatività fissato a 0.05), mentre per verificare l'omogeneità delle varianze sono stati applicati il test di

Bartlett se la distribuzione dei dati era normale e il test di *Levene* per distribuzioni dei dati non normali (livello di significatività fissato a 0.05).

Lo scarto (o differenza) interquartile IQR è stato utilizzato per rilevare gli *outliers*; le successive analisi statistiche sono state effettuate sia considerando gli *outliers* che rimuovendoli dai dati (Figura 36, riquadro 2).

Dal momento che la maggior parte dei dati deviava significativamente dalle assunzioni parametriche è stato effettuato il test di *Spearman* per valutare la relazione tra le variabili in esame (Figura 36, riquadro 4). In questo caso, per ciascun segnale è stata considerata sia la forza di correlazione tra l'età del soggetto controllo e il valore dell'area associata al picco (coefficiente di correlazione di *Spearman*) sia il significato statistico di questo coefficiente, ovvero valori di *p-value* inferiori a 0.05 indicano una rilevanza statistica.

Per eseguire un'indagine statistica al fine di rilevare le differenze di espressione tra le classi in esame sono stati eseguiti i seguenti test: *t-test/Welch test* per insieme di dati parametrici con varianza uguale/disuguale rispettivamente, *Wilcoxon/Kolmogorov-Smirnov test* per dati non parametrici, con varianza uguale/disuguale rispettivamente (Figura 36, riquadro 5). Sulla base di questi test viene calcolato per ogni picco il valore di *p-value* indicativo della probabilità che la differenza di espressione trovata nelle popolazioni in esame non sia fondata su un evento casuale. Dunque questo indice rappresenta il livello di significatività effettiva ed è indipendente dall'altezza e dall'area del picco. Se il *p-value* è ≤ 0.05 il segnale è statisticamente significativo ai fini di una buona separazione, mentre se il *p-value* è > 0.05 il segnale non permette una discriminazione delle classi.

▪ Costruzione di *clusters* diagnostici

Successivamente la selezione dei *clusters* di segnali in grado di separare le popolazioni in studio è stata eseguita con il *software RapidMiner* (RaM). RaM è un *software* sviluppato per eseguire rapidamente i processi di *machine learning* e *knowledge discovery* (KD) [276, 277]. RaM è attualmente utilizzato per risolvere problemi di classificazione, clusterizzazione e anche di integrazione di grandi quantità di dati. L'utente viene supportato con semplici interfacce grafiche dove si possono combinare i diversi operatori messi a disposizione per effettuare le analisi di Data Mining e il flusso di lavoro viene costruito collegando gli operatori tra di loro.

Al fine di indagare le capacità diagnostiche dei picchi presenti nei profili spettrali delle popolazioni in esame e di individuare un sottogruppo di segnali capaci di discriminare pazienti affetti da diverse patologie è stato ottimizzato uno specifico *workflow* (Figura 37); fondamentalmente sono stati eseguiti i processi di *features selection*, di *cross-validation* con una *k-fold* = 10, e di *performance evaluation* basata sull'algoritmo SVM (*Support Vector Machine*), evitando l'*overfitting*.

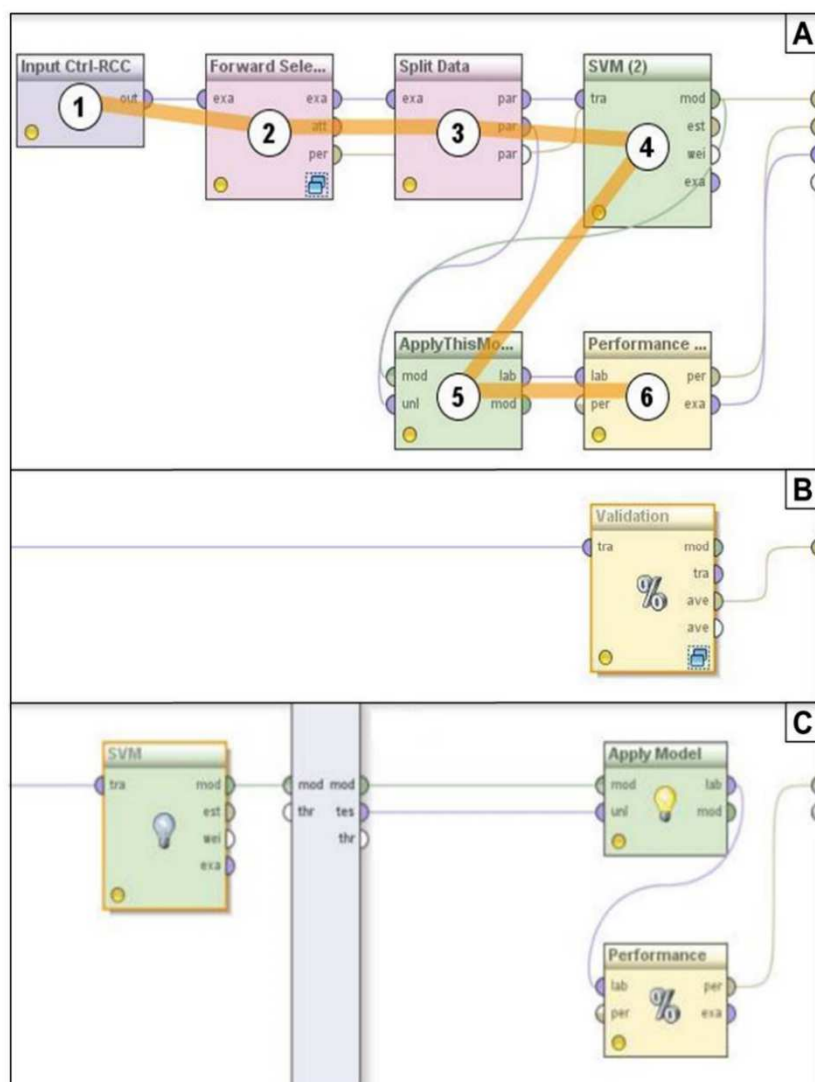


Figura 37. RapidMiner workflow. A) Il processo di *feature selection* e *K-fold cross-validation* con $K = 10$ viene eseguito sull'intera popolazione. In seguito l'operatore *Split data* suddivide la popolazione in due parti: la prima parte viene usata per la creazione del modello, la seconda per la sua validazione. Le performance che si otterranno saranno relative al processo di validazione. B) Validazione delle *feature selection* scelte tra le tante a disposizione. C) Performance del processo di *feature selection* e *cross-validation*.

L'indagine di possibili *pattern* proteici in grado di discriminare le popolazioni in esame è stata condotta tramite utilizzo di algoritmi matematici utili per la creazione di un modello diagnostico. Il *software RM* ha usato un algoritmo *Support Vector Machine* (SVM) per valutare la capacità di classificazione di un insieme di segnali.

Il SVM seleziona una combinazione di picchi che permettono di separare le popolazioni in esame attraverso l'utilizzo di un vettore. Si basa sul principio di "minimizzazione del rischio strutturale (SRM)", ovvero ha lo scopo di identificare la combinazione migliore di *features* e con il minor rischio di fallimento [278-280]. Esso utilizza le intensità delle *features* come coordinate in uno spazio N-dimensionale (nel nostro caso $N = 12$).

La selezione delle *features* è stata condotta utilizzando un approccio di tipo *wrapper* che utilizza il classificatore come una “scatola nera” al fine di identificare un sottoinsieme di *features* in base al loro potere predittivo. Dal momento che il numero di tutte le possibili combinazioni di *features* aumenta esponenzialmente con il numero di *features* considerate, la ricerca di un sottoinsieme di *features* che fornisca la più accurata precisione di classificazione è spesso una fase critica da realizzare. Per superare questo problema molti algoritmi matematici sono stati sviluppati per indirizzare la ricerca di un sottogruppo ottimale di *features*. Nel nostro caso, è stato applicato uno schema *forward selection*.

In particolare, tre sono gli algoritmi di SVM considerati in questo studio: *Dot*, *Radial* e *ANOVA* allo scopo di trovare il modello avente la migliore efficienza discriminativa. Inoltre sono state condotte una serie di analisi statistiche in cui sono stati impiegati diversi *Data Input* e diversi valori di suddivisione dei dati da destinare alla fase di costruzione del modello e a quella di validazione (9:1, 8:2 e 6:4).

I valori di sensibilità e specificità sono stati calcolati per ogni modello selezionato e la performance diagnostica dei segnali è stata valutata all'interno dell'intera casistica mediante un processo di *k-fold cross-validation*. Il processo di *cross-validation* è stato condotto mediante una validazione di tipo *k-fold* ($K = 10$). L'insieme di dati iniziale S viene diviso in sottoinsiemi (S_1, S_2, \dots, S_k). Prima, il modello di classificazione viene addestrato su $K-1$ sottoinsiemi e poi viene utilizzato per classificare i campioni dei rimanenti sottoinsiemi. Una stima dell'accuratezza (e delle altre misure di performance) del classificatore viene ottenuta attraverso l'operatore *performance*. Le performance ottenute sia nel processo di *feature selection* che di *cross-validation* sono state definite da indici standard normalmente impiegati nella valutazione della precisione e accuratezza di un'indagine inferenziale: sensibilità, specificità, valori predittivi negativi e positivi (VPN e VPP) [281]. L'analisi della curva ROC (*Receiver Operating Characteristic curve*) consente di valutare l'accuratezza di un particolare test diagnostico, ossia la sua capacità di distinguere i casi malati dai soggetti sani. La curva ROC offre una visione grafica circa la specificità e sensibilità di un test e fornisce una stima della abilità discriminante di un determinato modello. L'asse x riporta la specificità in termini di falsi positivi, mentre l'asse y la sensibilità. Il valore AUC (*Area Under Curve*), che rappresenta l'area sottesa alla curva ROC, indica la capacità del *cluster* di classificare correttamente i dati. L'analisi della curva ROC e dei rispettivi valori AUC è stata eseguita come descritto in un lavoro precedente [282].

▪ Individuazione di peptidi endogeni correlati con la progressione tumorale

L'analisi statistica per l'individuazione di alterazioni di peptidi endogeni correlate con la progressione tumorale è stata effettuata come sopra-descritto. Inizialmente sono state valutate la normalità e l'omogeneità della varianza mediante i test di Shapiro–Wilk, Barlett e Leven, seguiti dai relativi test parametrici o non parametrici per il confronto tra gruppi. Le variazioni di espressione dei peptidi urinari in base ai dati clinici sono state esaminate utilizzando la correlazione per ranghi di Spearman (*Spearman's rho*). Tutti i test sono stati applicati utilizzando un p-value di 0.05 come soglia di significatività statistica.

3.3.5. IDENTIFICAZIONE DEI PEPTIDI ENDOGENI MEDIANTE TECNICA nLC-ESI-MS/MS

I peptidi endogeni urinari presenti nelle frazioni ottenute dalla purificazione con biglie magnetiche HIC 8 di campioni di pazienti e controlli sono stati identificati mediante la tecnica *nano liquid chromatography* (nLC)-ESI-MS/MS. Il *pool* di eluati è stato in un primo momento concentrato in centrifuga *Heto Speed-Vac* e successivamente desalificato e privato delle biglie rimanenti utilizzando puntali da pipetta *ZipTip™* μ -C18 seguendo il protocollo standard fornito dalla ditta produttrice. Le frazioni desalificate sono state iniettate con un flusso di 1,8 μ L./min mediante il sistema *Dionex UltiMate 3000 rapid separation (RS) LC nano system* accoppiato *on-line* allo spettrometro *Impact HD™* UHR-qToF (Ultra High Resolution-qTOF).

Le analisi HPLC sono state eseguite con le seguenti condizioni operative con una configurazione della valvola del comparto colonna che prevedeva l'utilizzo di tutte le 10 vie:

- Pre-colonna per la desalificazione (*Dionex*, Acclaim PepMap 100 C18, cartridge, 300 μ m i.d. \times 5 mm, 5 μ m);
- Pre-colonna per i calibranti (*Dionex*, Acclaim PepMap 100 C18, nanoviper, 75 μ m i.d. \times 2 cm, 3 μ m);
- Colonna analitica: nano colonna da 50 cm (*Dionex*, 0.075 mm ID, Acclaim PepMap100, C18, 2 μ m);
- Fase mobile *loading pump*: H₂O + 2% CH₃CN + 0,1% Acido Trifluoroacetico;
- Fase mobile A *nano pump*: H₂O + 0,1% Acido Formico;
- Fase mobile B *nano pump*: H₂O + 80% CH₃CN + 0.08% Acido Formico;
- Miscela di calibrazione: MMI-L *Low concentration Tuning Mix*;
- Flusso della nanopompa: 300 nl/min;
- Flusso di caricamento del campione: 10 μ l/min;
- Flusso della pompa siringa per calibranti: 30 μ l/min;
- Temperatura: 40°;
- Gradiente della durata di 360 minuti (Tabella 8).

Minuti	% solv. B
0	4
0	4
10	4
18	6
189	24
228	29
260	35
308	66
310	98
323	98
323	4
360	4

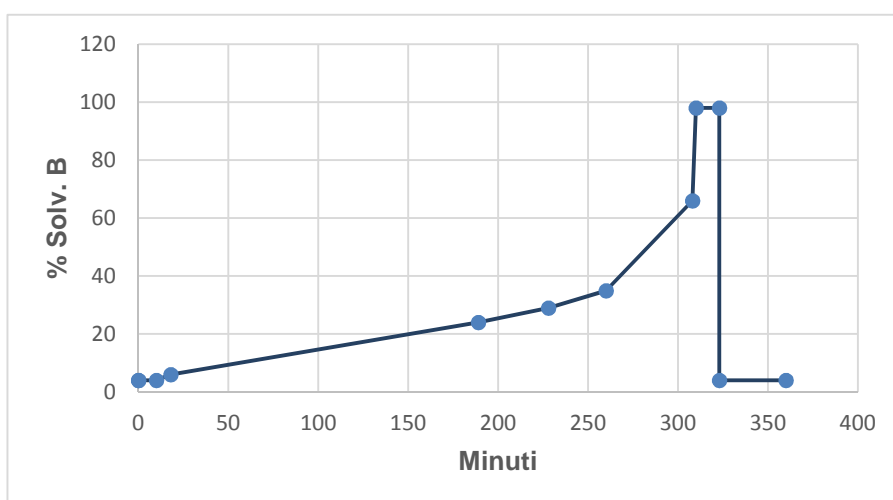


Tabella 8: Gradiente utilizzato per l'analisi HPLC. Solvente B H₂O + 80% CH₃CN + 0.08% Acido Formico.

La colonna era connessa ad una sorgente di tipo *CaptiveSpray*TM, funzionante in modalità *Enhanced*, ossia con l'ausilio di un gas di solvatazione arricchito con acetonitrile in modo da facilitare la ionizzazione dei campioni poco concentrati, aumentando lo stato di carica e riducendo di conseguenza il rapporto m/z .

Lo spettrometro di massa ha operato in modalità *data-dependent-acquisition* in modo tale da alternare automaticamente una scansione completa MS all'acquisizione degli spettri MS/MS.

Le analisi sono state eseguite utilizzando i seguenti parametri:

- *Funnel 1* RF = 400 V_{pp};
- *Funnel 2* RF = 600 V_{pp};
- *Hexapole* RF = 400 V_{pp};
- *Prepuls* = 10 μs.

Uno dei parametri ottimizzati è l'energia di collisione; essendo i peptidi endogeni di difficile frammentazione, tale energia è stata impostata a 7eV selezionando una rampa da 100% a 120%.

Il numero di ioni precursori è stato regolato automaticamente dal *software* al fine di mantenere costante la frequenza di campionamento sul picco cromatografico, con cicli fissi di 5 secondi.

Le funzionalità IDAS (*Intensity Dependent Acquisition Speed*) e RT2 (*RealTime Re-Think*) sono state applicate in una finestra di acquisizione MS/MS di 300-1221 e 1125-2000 m/z . Impostando la funzionalità IDAS è possibile fissare il *cycle time* ovvero il tempo che intercorre tra una scansione MS e quella successiva. Di conseguenza il numero di ioni precursori MS/MS in un singolo *cycle time* e il relativo tempo di acquisizione MS/MS varia automaticamente sulla base dell'intensità dello ione precursore, al fine di ottenere risultati indipendenti sia dalla quantità che dalla complessità del campione. La funzionalità RT2 permette di effettuare la frammentazione dello ione precursore in *cycle time* consecutivi solo se si verifica un determinato

aumento dell'intensità di quello ione, ottimizzando la qualità degli spettri MS/MS e il numero di identificazioni particolarmente nei campioni complessi.

La fase di elaborazione dei dati è stata effettuata in modo automatico eseguendo un appropriato *script* implementato nel *software Compass DataAnalysis™* v.4.1 Sp4. I dati MS/MS grezzi sono stati ricalibrati internamente su un iniziale segmento di analisi corrispondente all'uscita dei picchi della miscela dei calibranti e successivamente la curva di calibrazione corretta con la *lock-mass* (m/z 1221.9906) depositata su un apposito filtro del gas di solvatazione in entrata nella sorgente. I dati sono stati poi convertiti in una lista di picchi, deconvoluti e salvati in *file* di tipo .mgf e .xml. Per eseguire l'identificazione accurata dei frammenti, i dati MGF sono stati esaminati utilizzando il motore di ricerca *Mascot* (versione 2.4.1) attraverso lo strumento *Mascot Daemon*. La ricerca è stata ristretta alla sezione umana del *database Swissprot* (ultimo accesso Aprile 2014, 544996 sequenze; 193815432 residui); non sono stati impostati come parametri di ricerca gli enzimi di restrizione o modificazioni fisse. Le tolleranze di MS e MS/MS sono state in generale impostate rispettivamente a 20-5 ppm e 0.5-0.05 Da. L'acetilazione (N-terminale) è stata selezionata come eventuale modificazione variabile nei parametri di ricerca di *Mascot*. Sono state quindi considerate le soglie di *Mascot* per la definizione di omologia e di identità e il *DECOY database* come filtri per la determinazione della significatività dei peptidi (*False Discovery Rate* < 1%).

3.4. ANALISI DEL PROTEOMA URINARIO

Per l'analisi del proteoma urinario sono stati utilizzati due pools di urine costituiti da 80 soggetti controllo e 80 pazienti cRCC. Al fine di isolare la frazione contenente il proteoma, i campioni di urine sono stati sottoposti a centrifugazione per 15 minuti a 4000g su filtri aventi cut-off di 30kDa. Sulla frazione contenente le proteine (>30kDa) è stata quindi effettuata una quantificazione mediante approccio BCA al fine di valutarne la concentrazione proteica.

3.4.1. DIGESTIONE ENZIMATICA CON METODICA FASP

Per la digestione enzimatica dei pools di urine è stato seguito il protocollo pubblicato in letteratura da Mann *et al* [161].

Brevemente, 100ug di proteine sono andati incontro a denaturazione e riduzione dei ponti disolfuro mediante incubazione in thermomixer a 95°C per 5 minuti con un egual volume di soluzione di denaturazione (4% SDS, 0.1M DTT in 0.1M TRIS HCL pH 7.6). I campioni denaturati e ridotti sono stati quindi trasferiti su filtri aventi cut-off di 30kDa precedentemente lavati al fine di eliminare eventuali residui di glicerina che potrebbe interferire con l'analisi in spettrometria di massa. Si è quindi proceduto all'eliminazione dei detergenti utilizzati nella precedente fase mediante 2 lavaggi con una soluzione contenente urea (8M urea in TRIS HCL 0.1M pH 8.5) e centrifugazione da 15 minuti a 14100g. I ponti disolfuro precedentemente ridotti sono stati quindi alchilati mediante aggiunta di iodoacetamide, incubazione per 20 minuti al buio e successiva eliminazione del reagente per centrifugazione a 14100g per

10 minuti. Sono stati quindi effettuati altri due lavaggi con urea (8M urea in TRIS HCL 0.1M pH 7.9) e due con soluzione tampone (0.05 M NH_4HCO_3). Il campione è stato quindi digerito mediante aggiunta di 1 μg di tripsina (rapporto proteine:tripsina 1:100) ed incubazione overnight a 37°C in stufa umida. La raccolta dei peptidi derivanti dalla digestione sono stati raccolti in nuove provette di raccolta mediante centrifugazione e lavaggi sequenziali con soluzione tampone e soluzione 0.5M di NaCl. I filtrati vengono quindi acidificati con TFA (Acido TrifluoroAcetico) e quantificati utilizzando lo spettrofotometro NanoDrop. I digeriti triptici sono stati conservati a -30°C fino al momento dell'analisi.

3.4.2. ANALISI nLC-ESI-MS/MS

Al fine di effettuare un'analisi quali-quantitativa del proteoma urinario dei *pool* di soggetti controllo e pazienti, i peptidi triptici sono stati purificati e concentrati con puntali ZipTip™ C18 secondo il protocollo standard. Successivamente, 7 μl di campione sono stati iniettati con un flusso di 0.3 $\mu\text{L}/\text{min}$ mediante il sistema UHPLC accoppiato on-line alla trappola ionica nESI Ion Trap AmaZon ETD. La separazione LC è stata eseguita utilizzando una colonna Acclaim® PepMap RSLC (75 μm , 50cm, C18, 2 μm , 100 Å) con un gradiente di acetonitrile (AcN) della durata di 690 minuti (Tabella 9).

% solvente B	min
4	0
4	10
6	26
24	368
29	446
35	510
66	586
98	590
98	616
4	616
4	690

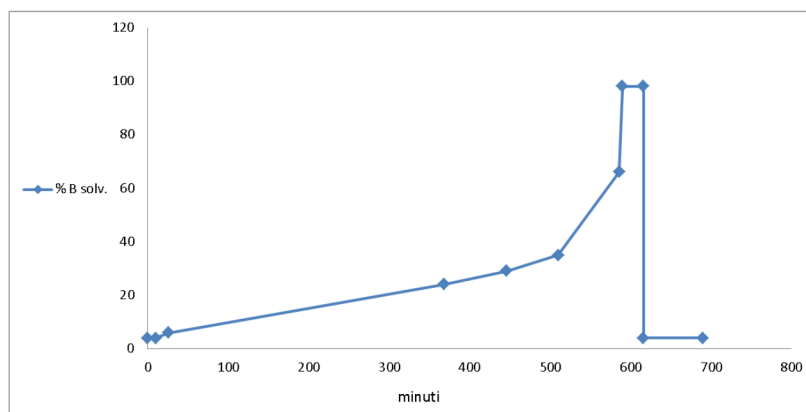


Tabella 9. Gradiente di acetonitrile: Solvente B 98% AcN / 2% H₂O / 0.1% FA (Acido Formico).

L'analisi nLC-MS/MS ha previsto l'iniezione in triplicato di circa 1 μg di peptidi (7 μL) derivanti da digestione proteolitica per ogni corsa eseguita. È stata quindi effettuata su ciascuna corsa un'analisi qualitativa al fine di identificare, mediante l'algoritmo Mascot (versione 2.4.01), le proteine intatte presenti in ciascun campione analizzato. I dati MS/MS grezzi sono stati deconvolti e convertiti in una lista di picchi mediante l'utilizzo del software Compass DataAnalysis™ v.4.0 Sp4 (Bruker Daltonics, Germania). Al fine di eseguire l'identificazione accurata dei frammenti, la ricerca è stata ristretta alla sezione umana del database Swissprot. In particolare, sono stati impostati i seguenti parametri di ricerca: tripsina come enzima proteolitico, modificazione fissa "carbamidometil (C)" (carbamidometilazione delle cisteine indotta da Iodoacetamide) e le tolleranze di MS e MS/MS sono state in generale impostate rispettivamente a 2Da e 0.8Da. Infine, il risultato è stato ricalcolato usando l'algoritmo "Percolator" per determinare una soglia di

significatività univoca per l'identificazione delle proteine (p value < 0.05) e migliorare la confidenza del risultato.

3.4.3. QUANTIFICAZIONE RELATIVA *LABEL-FREE*

Al fine di determinare i rapporti proteici nei campioni di urine ed individuare alterazioni proteiche tipiche del tumore renale, le analisi nLC-ESI MS/MS sono state ulteriormente elaborate mediante il software Progenesis LC-MS™ (versione 4.1). In particolare, l'utilizzo di tale software permette di ottenere una quantificazione relativa label-free grazie ad un'analisi dei dati che prevede i seguenti passaggi:

- ❖ *Import e Peak Modelling*: i dati derivati dall'analisi nLC-ESI-Ion Trap (.d) vengono importati, ridotti e successivamente rappresentati graficamente mediante una visualizzazione pseudo-gel bidimensionale.
- ❖ *Run alignment* (Figura 38): l'allineamento delle corse cromatografiche è un passaggio necessario per poter comparare i dati relativi all'analisi LC-MS con i profili proteici di ogni campione. Viene selezionata una corsa di riferimento e tutte le altre sono allineate in automatico dal software. Al termine dell'allineamento di default è possibile perfezionare i risultati manualmente per migliorare il valore di *score* di ogni corsa cromatografica.
- ❖ *Peak picking*: questa fase è utilizzata per determinare i picchi presenti nel cromatogramma e ridurre ulteriormente il dataset in quanto, per ogni spettro, vengono sostituite le coordinate di ciascun punto con le coordinate dell'apice del picco. I parametri impostati in questa fase dell'elaborazione sono:
 - Sensibilità: *fewer*
 - *Max peak width*: 0.5 minuti. Questo valore indica l'ampiezza minima dei picchi cromatografici in termini di tempo di ritenzione.
- ❖ *Filtering*: in questa fase è possibile filtrare i dati in termini di cariche, tempo di ritenzione e rapporto isotopico. Sono stati selezionati solo i peptidi che possiedono un numero di cariche massimo di 6.

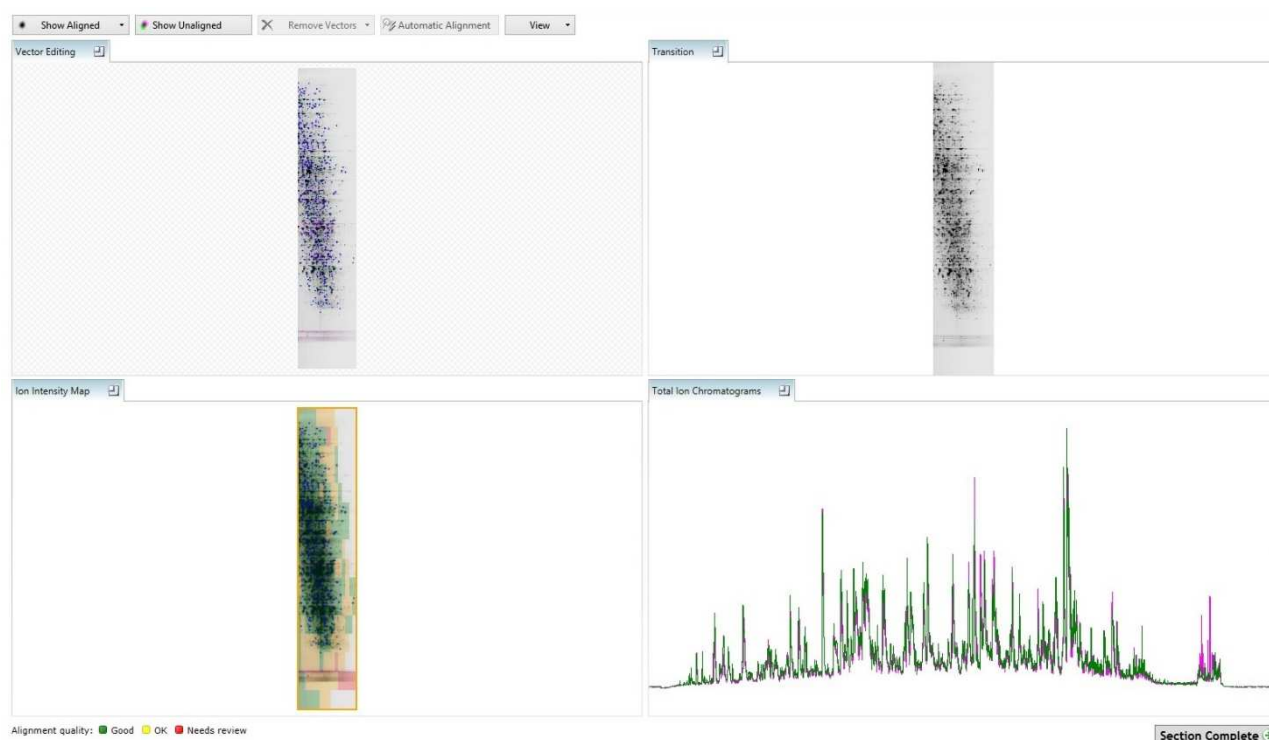


Figura 38. Rappresentazione grafica della fase di allineamento. In alto è possibile osservare la visualizzazione in pseudo-gel bidimensionale (destra) e l'allineamento automatico eseguito dal software (sinistra). Nella parte bassa dello schermo si nota il cromatogramma allineato rispetto alla corsa di riferimento (destra) e lo pseudo-gel allineato (sinistra) secondo una scala di colore coerente con l'efficienza dell'allineamento.

- ❖ *Normalization*: negli esperimenti di proteomica basati sull'uso di strumentazioni LC-MS, questo processo è richiesto per calibrare i dati provenienti da differenti corse cromatografiche ed eliminare la variabilità analitica (Figura 39). Per i campioni in studio, non essendo presente un calibrante interno, si utilizza il metodo di normalizzazione “*to all protein*”. Inizialmente il software seleziona una corsa di riferimento (può non essere la stessa scelta per la fase di allineamento) e successivamente crea i rapporti di abbondanza di ogni singolo ione. Ogni rapporto viene convertito in un fattore logaritmico per ottenere una distribuzione normale ed una migliore visualizzazione grafica. Sulla base della distribuzione di tutti i peptidi, viene calcolato il rapporto di normalizzazione. Il software è anche in grado di filtrare ed eliminare gli *outliers* calcolando varianza e deviazione standard di ogni ione rispetto ai valori presenti nella corsa di riferimento.

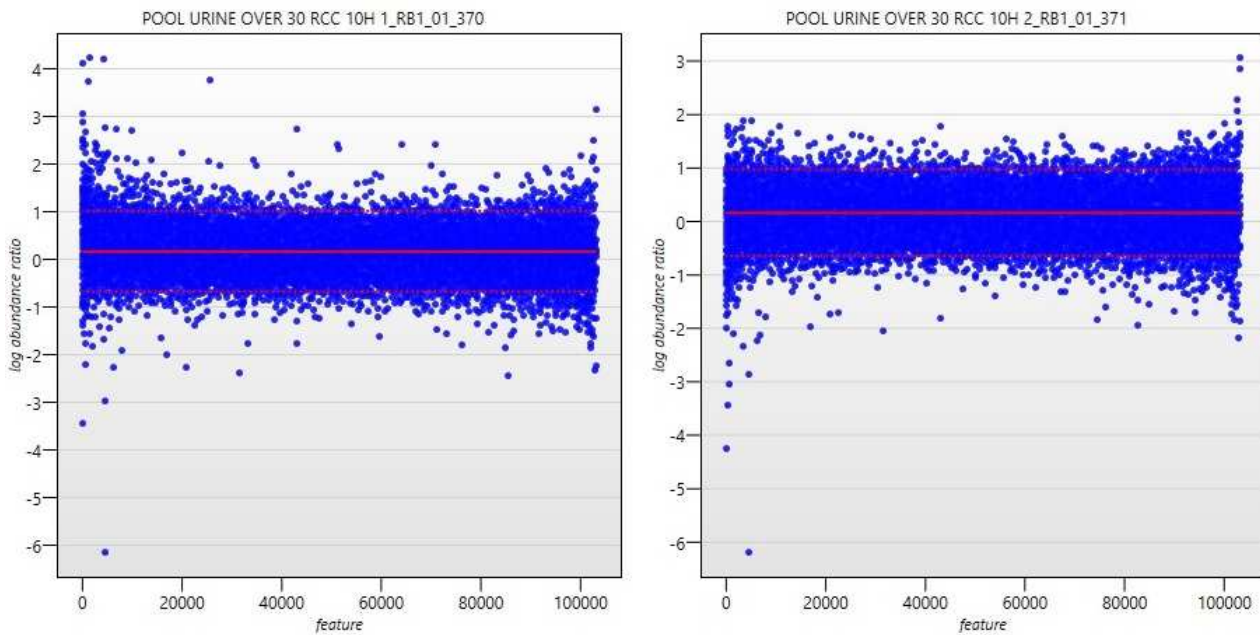


Figura 39. Rappresentazione grafica della normalizzazione dove ogni punto corrisponde al singolo peptide dell'analisi di riferimento. La linea rossa indica il punto zero calcolato in base alla corsa di riferimento.

- ❖ *Statistical analysis:* in questa sezione è possibile calcolare l'abbondanza relativa di ogni peptide all'interno di ogni singola analisi ed eseguire alcuni test statistici quali la *Principal Component Analysis* (PCA) e lo *student t-test* per valutare la significatività statistica del picco in esame all'interno dei due gruppi (campioni CTRL e ccRCC – Figura 40).



Figura 40. Rappresentazione grafica dell'abbondanza relativa di un peptide selezionato nelle diverse analisi. In azzurro le corse dei campioni controllo, in viola di quelli tumorali.

- ❖ *Identification:* vengono esportati gli spettri MS/MS sotto forma di file .mgf, viene eseguita una ricerca in database utilizzando Mascot, come descritto nel paragrafo 3.4.2 e i risultati ottenuti sono poi reimportati in Progenesis.

- ❖ *Review the protein of interest*: in questa fase viene scelta la tipologia di quantificazione da utilizzare. In particolare è stata utilizzata il metodo *Relative quantification using all peptides* che per la quantificazione considera tutti i peptidi appartenenti alla proteina in analisi.
- ❖ *Report*: in ultimo è possibile esportare le tabelle/grafici riassuntivi dei dati relativi all'identificazione e alla quantificazione.

3.5. VALIDAZIONE DI UNA PROTEINA DI INTERESSE MEDIANTE *WESTERN BLOT*

Al fine di validare i risultati ottenuti dall'integrazione delle informazioni relative alle variazioni del peptidoma e del proteoma urinario per una proteina *target*, è stato eseguito un saggio western blot.

È stata quindi effettuata una elettroforesi monodimensionale in cui sono stati caricati 40ug di proteine totali di 5 campioni CTRL e 5 campioni ccRCC. Tali proteine, sottoposte ad un campo elettrico, migrano secondo il proprio peso molecolare attraverso le maglie di un gel di poliacrilamide *pre-cast* in gradiente di concentrazione 4-12% usando il sistema NuPage (Life Technologies). Tale sistema lavora a pH neutro (pH 7.0), garantendo così alle proteine ed alla matrice la massima stabilità durante tutta la corsa. Il gradiente di concentrazione della poliacrilamide (4-12%) consente di ottenere una buona separazione di proteine a basso ed alto peso molecolare nel medesimo gel.

È inoltre possibile usare, in base alle esigenze, due diversi tamponi di corsa che permettono una migliore separazione di proteine ad alto (NuPAGE MOPS SDS Running Buffer) o a basso (NuPAGE MES SDS Running Buffer) peso molecolare.

<u>Running Buffer</u>	<u>NuPAGE MES SDS</u>	<u>NuPAGE MOPS SDS</u>
MES pH 7.2	50 mM	50 mM
Tris Base	50 mM	50 mM
SDS	0,1%	0,1%
EDTA	1 mM	1 mM

NuPAGE LDS Sample Buffer 4X

Glicerolo	10%
Tris Base	14 mM
Tris HCl	106 mM
LDS	2%
EDTA	0,51 mM
Serva Blue G250	0,22 mM
Phenol Red pH 8.5	0,175 mM

La solubilizzazione del campione ha previsto il trattamento con opportuni volumi di *NuPAGE LDS Sample Buffer (4X)* contenente: il detergente anionico litiiododecilsolfato (LDS), il glicerolo per appesantire il campione e facilitarne quindi il caricamento sul gel, un tampone adeguato e il Blu di Bromofenolo per permettere di visualizzare l'andamento della corsa elettroforetica. Il LDS denatura le proteine e conferisce loro la carica negativa che ne permette la migrazione verso il polo positivo. Ai campioni è stato poi aggiunto anche l'agente riducente, il ditiotreitolo (DTT 50 mM), che permette la rottura dei ponti disolfuro responsabili della struttura terziaria delle proteine. La denaturazione è necessaria per garantire che la migrazione avvenga solamente in base al peso molecolare specifico delle proteine e non in base alla loro conformazione nativa. I campioni così preparati sono stati quindi agitati ed incubati per 10 minuti a 70°C (Thermomixer, Eppendorf) per consentire la completa solubilizzazione delle proteine, raffreddati in ghiaccio per pochi secondi e, dopo essere stati spinnati brevemente (MiniSpin Plus, Eppendorf) per eliminare eventuale materiale insolubilizzato, caricati nei pozzetti del gel insieme agli standard a peso molecolare noto. Entrambe le camere di corsa, che si creano inserendo i gel nella vasca principale, sono state riempite completamente con i tamponi di corsa; l'antiossidante NuPAGE è stato aggiunto solo al tampone della camera interna dell'apparato elettroforetico (500 µl per 200 ml) al fine di impedire il processo di riossidazione durante la corsa elettroforetica. L'apparecchiatura è stata infine collegata ad un generatore di corrente applicando un voltaggio costante pari a 200V. Grazie alla presenza del colorante Blu di Bromofenolo nella soluzione in cui sono risospesi i campioni, è stato possibile seguire la corsa e fermarla prima che il fronte della corsa fuoriuscisse dal gel. Una volta spento il generatore di corrente si è proceduto all'estrazione del gel dalle armature e al successivo saggio *Western Blot*.

La tecnica del Western Blotting consente di trasferire le proteine, separate per mezzo della corsa elettroforetica, dal gel di poliacrilammide ad una membrana di nitrocellulosa rispecchiando il grado di risoluzione ottenuto nel gel. La membrana è poi sottoposta ad immunodecorazione con anticorpi specifici, permettendo così di rilevare e quantificare le proteine di interesse. Il trasferimento avviene mediante un campo elettrico diretto perpendicolarmente al gel e alla membrana, inseriti in una griglia a formare un "sandwich". La membrana di nitrocellulosa è stata tagliata in base alle dimensioni del gel, attivata in acqua bidistillata per 5 minuti e successivamente equilibrata per 10 minuti nel Transfer Buffer (MetOH 10%, CAPS pH11 10mM); anche il gel è stato fatto equilibrare per 5 minuti nel Transfer Buffer. È stato quindi assemblato il "sandwich" nella griglia disponendo le componenti nel modo seguente: una spugnetta, 2 fogli di carta assorbente 3M, membrana di nitrocellulosa, gel, altri 2 fogli di carta ed un'altra spugnetta (Figura 41).

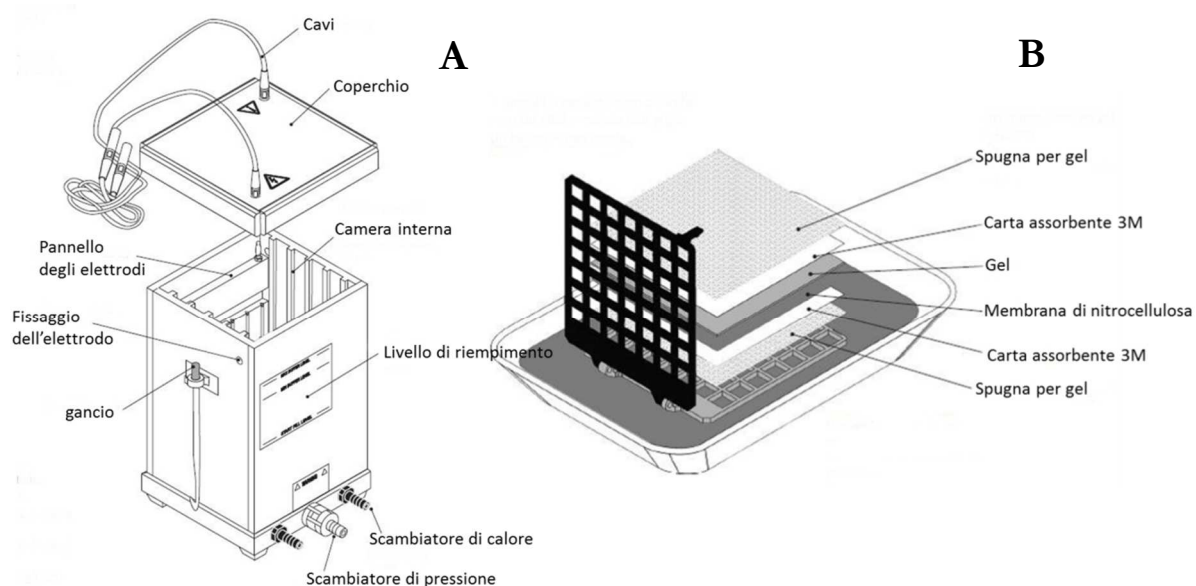


Figura 41. Schematizzazione dell'apparato per WB. A) Schema della vasca per il trasferimento Hoefler; B) schema di montaggio del sandwich.

Sia le spugnette che la carta sono state prima imbevute di Transfer Buffer. Il “sandwich” è stato quindi trasferito nella camera per il Western Blotting, facendo attenzione che la membrana fosse rivolta verso il polo positivo: in tal modo le proteine, che hanno acquisito una carica negativa durante la corsa elettroforetica, migrano dal gel verso la membrana, legandosi quindi ad essa. Il trasferimento è quindi avvenuto a 250 mA per 2 ore e 15 minuti in camera fredda (4°C). La durata del trasferimento stesso può variare in base alle dimensioni delle proteine che si desiderano studiare e allo spessore del gel. Ultimato il trasferimento, il “sandwich” viene smontato e la presenza e l'integrità delle proteine su nitrocellulosa vengono valutate mediante una colorazione reversibile delle proteine con Rosso Ponceau. Dopo un minuto di incubazione nel colorante, si decolora la membrana con dell'acqua bidistillata. È stata quindi acquisita l'immagine tramite CCD camera (ImageQuantTMLAS4000, GE Healthcare Life Sciences) e il Rosso Ponceau è stato eliminato mediante un lavaggio con PBS.

È stata quindi effettuata l'immunodecorazione della membrana, che consiste nell'analisi specifica delle proteine mediante anticorpi in grado di riconoscere uno specifico epitopo antigenico delle proteine di cui si vuole studiare il profilo di espressione. Al termine del Western Blotting la membrana è stata incubata per un'ora in agitazione (piastra basculante Heidolph, Polymax 1040) nella soluzione di *blocking* (Latte in polvere 5%, Tween 0,2%, PBS 10x), in modo che le proteine del latte andassero a saturare gli eventuali siti liberi della membrana che potrebbero avrebbero potuto interagire in modo non specifico con l'anticorpo primario. La membrana è stata poi incubata con l'anticorpo primario specifico, opportunamente diluito nella stessa soluzione di *blocking*, per 2 ore a temperatura ambiente (RT) o a 4°C overnight. Al termine

dell'incubazione, sono stati eseguiti 5 lavaggi da 10 minuti l'uno con la soluzione di blocking e si è quindi proceduto all'incubazione con i rispettivi anticorpi secondari per 1 ora e 30 minuti a RT o a 4°C *overnight*. Gli anticorpi secondari utilizzati sono coniugati con un enzima, la perossidasi di rafano (Horse radish peroxidase, HRP) e sono stati anch'essi diluiti nella soluzione di blocking (1:10.000). Dopo aver eliminato l'eccesso di anticorpo secondario sempre mediante 5 lavaggi da 10 minuti con la soluzione di blocking e un lavaggio terminale in PBS, è stata effettuata la rivelazione della proteina di interesse mediante un saggio di chemiluminescenza (ECL, Enhanced Chemiluminescence Assay); in presenza di una miscela di perossido di idrogeno (H₂O₂) e luminolo (substrato chemiluminescente), miscelati in quantità equimolari, la perossidasi coniugata all'anticorpo secondario ossida il luminolo producendo luce dove è presente la proteina d'interesse. La membrana è stata incubata con la miscela per 5 min e la luce emessa è stata visualizzata grazie ad una CCD camera (Image Quant LAS4000, GE Healthcare Life Sciences). Questa tecnica di rivelazione consente di catturare con precisione sia segnali forti che segnali deboli in un range di 4 ordini di magnitudo; inoltre, la temperatura di funzionamento della macchina (-25°C) consente di esporre la membrana per lunghi intervalli di tempo riducendo il rumore di fondo.

4. RISULTATI

4.1. DATI CLINICI

In questo studio, volto alla ricerca di firme molecolari del tumore renale, sono stati analizzati, a livello peptidomico tramite MALDI-MS, i profili di urine purificate con biglie magnetiche funzionalizzate provenienti da tre gruppi di soggetti: 137 soggetti controllo (Ctrl) (81 uomini e 56 donne), 118 soggetti affetti da tumore renale a cellule chiare (ccRCC) (73 uomini e 45 donne) e 35 soggetti affetti da tumori con differente sottotipo istologico (non-ccRCC) (22 uomini e 13 donne) di cui 16 soggetti affetti da tumore renale benigno e 19 soggetti affetti da tumore renale maligno non ccRCC.

L'età media dei controlli era di 49 anni in un intervallo di 24-79 anni, mentre per i pazienti l'età media era di 65 anni in un intervallo compreso tra 33 e 88 anni.

Le caratteristiche cliniche dei pazienti, classificati secondo il sistema TNM (Tumor-Node-Metastasis) del 2009, sono riassunte in Tabella 10 [283].

Le analisi istologiche eseguite sui pazienti hanno incluso il sistema di grading di Fuhrman, il riconoscimento del sottotipo istologico, il differenziamento sarcomatoide e cistico, la necrosi del tumore, l'infiltrazione tumorale microvascolare e della pelvi urinaria. I pazienti affetti da tumore sono stati sottoposti a exeresi della lesione.

		N° di pazienti
NUMERO TOTALE		153
Età media \pm SD alla diagnosi		64,53 \pm 10,97
Età media alla diagnosi (<i>range</i>)		33-88
GENERE		
	Maschi	95
	Femmine	58
ISTOLOGIA		
	RCC a cellule chiare	118
	RCC papillare	9
	RCC cromofobo	5
	Oncocitoma*	8
	Angiomiolipoma*	5
	Altri sottotipi	8
TIPO TUMORALE		
	Maligno	137
	Benigno	16
STADIAZIONE (solo ccRCC)		
Tumore Primario (T)		
	pT1a	45
	pT1b	39
	pT2a	23
	pT2b	1
	pT3a	9
	pT3b	0
GRADO (solo ccRCC)		
	1	6
	2	79
	3	20
	4	3
Dimensione massa tumorale in cm (solo ccRCC - <i>range</i>)		1,5 - 18

Tabella 10. Caratteristiche cliniche dei pazienti arruolati nello studio, classificati secondo il sistema di classificazione TNM (Tumor-Node-Metastasis) del 2009. * = masse renali benigne. Altri sottotipi = 5 sottotipi maligni + 3 masse renali benigne.

In considerazione della diversa percentuale di soggetti maschili rispetto a quelli femminili inclusi nello studio, è stato verificato che l'intensità dei segnali dei peptidi presenti negli spettri acquisiti non correlasse con il sesso dei pazienti mediante il test statistico di Fisher che non ha rilevato alcuna dipendenza con il sesso in questa coorte di pazienti.

4.2. ANALISI DEL PEPTIDOMA URINARIO

4.2.1. ANALISI DEI PROFILI PEPTIDICI E COSTRUZIONE DI *CLUSTER* DIAGNOSTICI

Al fine di ottenere clusters di potenziali biomarcatori in grado di discriminare le popolazioni in esame, sono stati acquisiti, allineati e valutati a livello statistico confrontando le diverse popolazioni, gli spettri MALDI-TOF del peptidoma urinario dei campioni appartenenti alla casistica sopradescritta (Tabella 10). La Figura 42 mostra i profili spettrali medi dei controlli, dei pazienti ccRCC, dei controlli addizionati ai pazienti con masse renali benigne e dei pazienti affetti da ccRCC più i pazienti affetti da tumori maligni, dopo purificazione con biglie magnetiche a fase inversa C8.

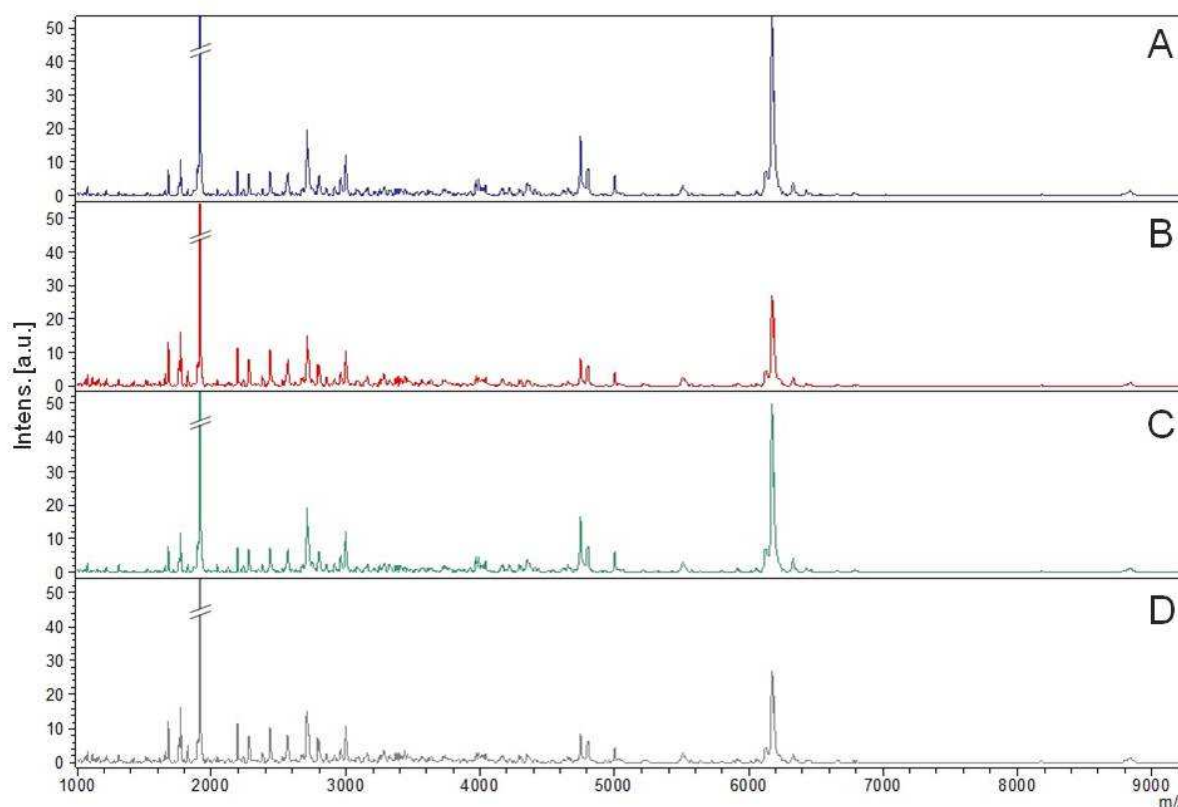


Figura 42. Profili spettrali medi di campioni di urine di controlli (A), pazienti affetti da ccRCC (B), controlli addizionati ai pazienti con masse renali benigne (C) e pazienti ccRCC più pazienti con lesioni maligne non-ccRCC (D), purificati con biglie magnetiche HIC 8.

In particolare l'analisi MALDI-TOF ha permesso di individuare circa 202 segnali urinari in comune alle popolazioni in studio; 121 di questi segnali risultano essere con un'area differente in modo statisticamente significativo tra controlli e ccRCC, ($p < 0.05$) (Tabella 11).

Essendo l'età media differente tra i gruppi di soggetti presi in esame (controlli = 48.7 e pazienti = 64.53), è stato valutato per quale di questi 202 ioni, comuni a tutte le popolazioni oggetto dello studio, l'età potesse costituire un fattore confondente. È stato effettuato un test statistico (test di *Spearman*) per individuare gli ioni la cui area correlasse con l'età nei soggetti controllo, in modo da poterli escludere dalla successiva elaborazione ($p < 0.05$).

ESPRESSIONE CONTROLLI <i>versus</i> pazienti ccRCC		
Ione (m/z)	p -value	Test statistico
1023,13*	< 0.001	<i>Kolmogorov-Smirnov test</i>
1051,87	n.s.	
1060,16*	< 0.001	<i>Kolmogorov-Smirnov test</i>
1065,76*	< 0.05	<i>Wilcoxon test</i>
1078,37	n.s.	
1105,88*	< 0.05	<i>Wilcoxon test</i>
1117,42*	< 0.001	<i>t-test</i>
1133,64*	< 0.001	<i>Kolmogorov-Smirnov test</i>
1149,2*	< 0.05	<i>Kolmogorov-Smirnov test</i>
1162,27*	< 0.001	<i>Kolmogorov-Smirnov test</i>
1196,3*	< 0.001	<i>Kolmogorov-Smirnov test</i>
1210,72*	< 0.001	<i>Kolmogorov-Smirnov test</i>
1219,12*	< 0.001	<i>Kolmogorov-Smirnov test</i>
1226,12*	< 0.001	<i>Kolmogorov-Smirnov test</i>
1229,44*	< 0.001	<i>Kolmogorov-Smirnov test</i>
1309,53*	< 0.001	<i>Kolmogorov-Smirnov test</i>
1332,43*	< 0.001	<i>Kolmogorov-Smirnov test</i>
1405,09*	< 0.001	<i>Kolmogorov-Smirnov test</i>
1425,52*	< 0.001	<i>Kolmogorov-Smirnov test</i>
1511,36*	< 0.001	<i>Kolmogorov-Smirnov test</i>
1524,97*	< 0.001	<i>Kolmogorov-Smirnov test</i>
1582,74*	< 0.001	<i>Kolmogorov-Smirnov test</i>
1613,15*	< 0.001	<i>Kolmogorov-Smirnov test</i>
1655,49*	< 0.001	<i>Kolmogorov-Smirnov test</i>
1670,1*	< 0.001	<i>Kolmogorov-Smirnov test</i>
1681,84*	< 0.001	<i>Wilcoxon test</i>
1718,39	n.s.	
1727,42*	< 0.001	<i>Kolmogorov-Smirnov test</i>
1757,01*	< 0.001	<i>Kolmogorov-Smirnov test</i>
1768,94*	< 0.001	<i>Kolmogorov-Smirnov test</i>
1799,84*	< 0.001	<i>Kolmogorov-Smirnov test</i>
1826,23*	< 0.001	<i>Kolmogorov-Smirnov test</i>
1870,17	n.s.	
1883,28*	< 0.05	<i>Kolmogorov-Smirnov test</i>
1896,13*	< 0.05	<i>Wilcoxon test</i>
1913,15*	< 0.05	<i>Wilcoxon test</i>
1934,87*	< 0.05	<i>Wilcoxon test</i>
1951,34*	< 0.001	<i>Wilcoxon test</i>
2041,6	n.s.	
2128,46	n.s.	
2192,74	n.s.	
2216,93*	< 0.05	<i>Kolmogorov-Smirnov test</i>
2237,29	n.s.	
2528,98*	< 0.001	<i>Kolmogorov-Smirnov test</i>
2571,48*	< 0.05	<i>Kolmogorov-Smirnov test</i>
2660,82*	< 0.001	<i>t-test</i>

ESPRESSIONE CONTROLLI <i>versus</i> pazienti ccRCC		
Ione (m/z)	<i>p</i> -value	Test statistico
2671,37*	< 0.05	<i>Wilcoxon test</i>
2682,71	n.s.	
2698,15	n.s.	
2714,32*	< 0.05	<i>Wilcoxon test</i>
2726,94	n.s.	
2755	n.s.	
2790,73	n.s.	
2857,03*	< 0.05	<i>Wilcoxon test</i>
2915,85	n.s.	
2961,05*	< 0.05	<i>Wilcoxon test</i>
3003,74	n.s.	
3151,64	n.s.	
3162,23*	< 0.001	<i>Kolmogorov-Smirnov test</i>
3168,37*	< 0.05	<i>Wilcoxon test</i>
3251,59	n.s.	
3267,76*	< 0.001	<i>Kolmogorov-Smirnov test</i>
3283,92*	< 0.001	<i>Kolmogorov-Smirnov test</i>
3291,39*	< 0.05	<i>Kolmogorov-Smirnov test</i>
3328,59	n.s.	
3371	n.s.	
3388,34*	< 0.001	<i>Wilcoxon test</i>
3427,1*	< 0.05	<i>Kolmogorov-Smirnov test</i>
3443,23*	< 0.001	<i>Kolmogorov-Smirnov test</i>
3484,49*	< 0.001	<i>Kolmogorov-Smirnov test</i>
3568,41*	< 0.05	<i>Unequal Variance t-test</i>
3616,81*	< 0.001	<i>Kolmogorov-Smirnov test</i>
3636,33*	< 0.001	<i>Wilcoxon test</i>
3723,84*	< 0.05	<i>Wilcoxon test</i>
3738,44	n.s.	
3752,71	n.s.	
3777,82	n.s.	
3930,04	n.s.	
3972,22*	< 0.001	<i>Kolmogorov-Smirnov test</i>
3989,97*	< 0.001	<i>Kolmogorov-Smirnov test</i>
4015,46*	< 0.05	<i>Kolmogorov-Smirnov test</i>
4028*	< 0.05	<i>Kolmogorov-Smirnov test</i>
4047,02*	< 0.05	<i>Kolmogorov-Smirnov test</i>
4163,54	n.s.	
4173,11	n.s.	
4220,67*	< 0.001	<i>Kolmogorov-Smirnov test</i>
4293,38	n.s.	
4309,08*	< 0.05	<i>Wilcoxon test</i>
4355,12*	< 0.001	<i>Kolmogorov-Smirnov test</i>
4366,87*	< 0.001	<i>Kolmogorov-Smirnov test</i>
4380,95*	< 0.05	<i>Kolmogorov-Smirnov test</i>
4409,64*	< 0.05	<i>Wilcoxon test</i>
4439,07*	< 0.001	<i>Kolmogorov-Smirnov test</i>
4537,99	n.s.	
4622,87*	< 0.001	<i>Kolmogorov-Smirnov test</i>
4638,54*	< 0.001	<i>Kolmogorov-Smirnov test</i>
4657,71*	< 0.05	<i>Wilcoxon test</i>
4675,57	n.s.	
4707,65	n.s.	
4722,16	n.s.	
4751,51*	< 0.001	<i>Kolmogorov-Smirnov test</i>
4773,82*	< 0.001	<i>Kolmogorov-Smirnov test</i>
4804,35*	< 0.001	<i>Kolmogorov-Smirnov test</i>
4834,51	n.s.	
4849,11*	< 0.05	<i>Kolmogorov-Smirnov test</i>
4866,17*	< 0.001	<i>Kolmogorov-Smirnov test</i>
4936,41	n.s.	
4962,79*	< 0.001	<i>Kolmogorov-Smirnov test</i>
5004,43*	< 0.001	<i>Wilcoxon test</i>

ESPRESSIONE CONTROLLI <i>versus</i> pazienti ccRCC		
Ione (m/z)	<i>p</i> -value	Test statistico
5032,21*	< 0.001	<i>Wilcoxon test</i>
5043,58*	< 0.05	<i>Wilcoxon test</i>
5068,82	n.s.	
5216,49	n.s.	
5231,47*	< 0.05	<i>Kolmogorov-Smirnov test</i>
5245,26	n.s.	
5303,87*	< 0.05	<i>Wilcoxon test</i>
5431,82*	< 0.001	<i>Wilcoxon test</i>
5514,25*	< 0.001	<i>Wilcoxon test</i>
5529,85*	< 0.001	<i>Kolmogorov-Smirnov test</i>
5578,23*	< 0.05	<i>Kolmogorov-Smirnov test</i>
5642,96*	< 0.05	<i>Kolmogorov-Smirnov test</i>
5803,17*	< 0.001	<i>Kolmogorov-Smirnov test</i>
5831,86	n.s.	
5863,96*	< 0.05	<i>Kolmogorov-Smirnov test</i>
5889,71*	< 0.001	<i>Kolmogorov-Smirnov test</i>
5917,77*	< 0.05	<i>Kolmogorov-Smirnov test</i>
5931,47*	< 0.001	<i>Wilcoxon test</i>
5962,02	n.s.	
6019,67	n.s.	
6044,4*	< 0.001	<i>Wilcoxon test</i>
6060,19*	< 0.001	<i>Wilcoxon test</i>
6075,82*	< 0.05	<i>Kolmogorov-Smirnov test</i>
6131,66*	< 0.001	<i>Kolmogorov-Smirnov test</i>
6175,03*	< 0.001	<i>Wilcoxon test</i>
6190,13*	< 0.001	<i>Kolmogorov-Smirnov test</i>
6229,1*	< 0.001	<i>Kolmogorov-Smirnov test</i>
6237,8*	< 0.05	<i>Wilcoxon test</i>
6261,41*	< 0.001	<i>Wilcoxon test</i>
6276,27	n.s.	
6305,2*	< 0.001	<i>Wilcoxon test</i>
6333,39*	< 0.001	<i>Wilcoxon test</i>
6430,63*	< 0.001	<i>Wilcoxon test</i>
6445,61*	< 0.001	<i>Wilcoxon test</i>
6463,69*	< 0.001	<i>Kolmogorov-Smirnov test</i>
6544,09*	< 0.05	<i>Wilcoxon test</i>
6565,49	n.s.	
6656,13*	< 0.05	<i>Kolmogorov-Smirnov test</i>
6691,14	n.s.	
6786,43*	< 0.001	<i>Kolmogorov-Smirnov test</i>
6806,67*	< 0.001	<i>Wilcoxon test</i>
6864,19*	< 0.05	<i>Wilcoxon test</i>
6949,1	n.s.	
7021,04*	< 0.05	<i>Wilcoxon test</i>
7056,11	n.s.	
7108,83	n.s.	
7243,71	n.s.	
7268,48	n.s.	
7357,7	n.s.	
7383,14	n.s.	
7513,65	n.s.	
7628,61	n.s.	
7659,38	n.s.	
7672,93	n.s.	
7910,16	n.s.	
8011,55	n.s.	
8115,84	n.s.	
8179,35*	< 0.05	<i>Kolmogorov-Smirnov test</i>
8292,5	n.s.	
8378,18	n.s.	
8449,94	n.s.	
8499,22	n.s.	
8560,45	n.s.	

ESPRESSIONE CONTROLLI <i>versus</i> pazienti ccRCC		
Ione (m/z)	<i>p-value</i>	Test statistico
8838,64*	< 0.001	<i>Kolmogorov-Smirnov test</i>
8915,93	n.s.	
9074,31	n.s.	
9177,25	n.s.	
9221,69	n.s.	
9317,23	n.s.	
9623,42	n.s.	
9747,52	n.s.	
9865,9*	< 0.05	<i>Kolmogorov-Smirnov test</i>
10013,36	n.s.	
10197	n.s.	
10327,51	n.s.	
10528,76	n.s.	
10556,05	n.s.	
10638,02	n.s.	
10654,4	n.s.	
10751,41	n.s.	
10996,05	n.s.	
11039,23	n.s.	
11351,06	n.s.	
11714,78	n.s.	
11858,18	n.s.	

Tabella 11. Elenco degli ioni differentemente espressi tra controlli e pazienti ccRCC. Viene indicato il *p-value* e il relativo test statistico applicato; valori di *p-value* < 0.05 indicano una differenza statisticamente significativa delle aree dei segnali peptidici tra le popolazioni in esame (n.s. = non significativo).

Data l'elevata complessità dei profili spettrali, l'analisi statistica e l'individuazione dei *patterns* discriminanti sono state effettuate mediante l'utilizzo dell'algoritmo *Support Vector Machine* (SVM) fornito dal *software RapidMiner* (RaM).

Allo scopo di valutare il modello avente la migliore efficienza discriminativa, sono state condotte una serie di elaborazioni statistiche in cui sono stati impiegati diversi *data Input*, diversi algoritmi di SVM (*ANOVA*, *Radial*, *Dot*) e diverse suddivisioni dell'intera popolazione in due gruppi da utilizzare per la costruzione del modello e per la loro validazione (*Splitting ratio* 9:1, 8:2 e 6:4).

In particolare, inizialmente sono stati utilizzati come *data Input* solo quei segnali con una differenza statisticamente significativa tra le classi in esame (controlli *vs* ccRCC) (Tabella 11 – valori con asterisco). I risultati ottenuti sono riportati in Tabella 12. In seguito sono stati utilizzati come *data Input* tutti i 194 segnali comuni alle due popolazioni al fine di comparare le performance di ciascuno dei due approcci (Tabella 13). Le capacità classificative sono risultate leggermente superiori usando tutti i segnali presenti nei profili proteici (Tabella 12 e 13). Tra i vari algoritmi valutati, *Dot* si è rilevato più efficace in termini di performance del classificatore, rispetto agli altri due, *ANOVA* e *Radial* (Tabella 13).

Infine ciascun modello, generato dall'algoritmo SVM, è stato valutato suddividendo la popolazione in diverse percentuali da destinare alla creazione del modello e al processo di validazione. Più precisamente nella suddivisione 9:1, il 90% dei dati è stato utilizzato per la creazione del modello e il 10% dei dati è stato utilizzato come popolazione indipendente su cui validare il modello stesso; ugualmente nella suddivisione 8:2 e 6:4, l'80% e il 60% sono stati destinati, rispettivamente, alla creazione del modello e il 20% e il 40%, rispettivamente, al processo di validazione (Tabella 13). Dall'osservazione degli indici di precisione e

accuratezza delle indagini inferenziali effettuate (Tabella 13), e dalla necessità di bilanciamento tra *overfitting* e generalizzazione, il miglior compromesso è risultato essere quello di destinare circa il 60% dei dati alla creazione del modello e circa il 40% alla fase di validazione.

Sulla base dei risultati mostrati, si è scelto pertanto di utilizzare queste condizioni di classificazione (tutti i dati di *input*, algoritmo *Dot*, *Splitting ratio* 6:4) sia per il confronto tra Ctrl e ccRCC che in quello tra pazienti con masse tumorali maligne e soggetti con masse renali benigne addizionati ai soggetti controllo.

Dot (9:1)					Radial (9:1)					ANOVA (9:1)					A
feature	Training				feature	Training				feature	Training				
A_1511_36		Veri negativi	Veri positivi	class precision	A_5728_63		Veri negativi	Veri positivi	class precision	A_3168_37		Veri negativi	Veri positivi	class precision	
A_6445_61	pred. Falsi	100	25	80.00%	A_6463_69	pred. Falsi	96	10	90.57%	A_1105_88	pred. Falsi	97	28	77.60%	
A_1117_42	pred. Veri	14	71	83.53%	A_6864_19	pred. Veri	18	86	82.69%	A_1582_74	pred. Veri	17	68	80.00%	
A_2216_93	Spec., Sens.	87.72%	73.96%		A_1149_20	Spec., Sens.	84.21%	89.58%		A_2216_93	Spec., Sens.	85.09%	70.83%		
A_5889_71	Model				A_2216_93	Model				A_5889_71	Model				
A_1670_10		Veri negativi	Veri positivi	class precision	A_6190_13		Veri negativi	Veri positivi	class precision	A_6190_13		Veri negativi	Veri positivi	class precision	
	pred. Falsi	8	5	61.54%	A_1425_52	pred. Falsi	10	1	90.91%	A_6949_10	pred. Falsi	10	1	90.91%	
	pred. Veri	0	8	100.00%	A_1670_10	pred. Veri	0	10	100.00%	A_1670_10	pred. Veri	2	8	80.00%	
	Spec., Sens.	100.00%	61.54%			Spec., Sens.	100.00%	90.91%		Spec., Sens.	83.33%	88.89%			

Dot (8:2)					Radial (8:2)					ANOVA (8:2)					B
feature	Training				feature	Training				feature	Training				
A_1511_36		Veri negativi	Veri positivi	class precision	A_5728_63		Veri negativi	Veri positivi	class precision	A_3168_37		Veri negativi	Veri positivi	class precision	
A_6445_61	pred. Falsi	100	25	80.00%	A_6463_69	pred. Falsi	96	10	90.57%	A_1105_88	pred. Falsi	97	28	77.60%	
A_1117_42	pred. Veri	14	71	83.53%	A_6864_19	pred. Veri	18	86	82.69%	A_1582_74	pred. Veri	17	68	80.00%	
A_2216_93	Spec., Sens.	87.72%	73.96%		A_1149_20	Spec., Sens.	84.21%	69.58%		A_2216_93	Spec., Sens.	85.09%	70.83%		
A_5889_71	Model				A_2216_93	Model				A_5889_71	Model				
A_1670_10		Veri negativi	Veri positivi	class precision	A_6190_13		Veri negativi	Veri positivi	class precision	A_6190_13		Veri negativi	Veri positivi	class precision	
	pred. Falsi	16	7	69.57%	A_1425_52	pred. Falsi	16	2	88.89%	A_6949_10	pred. Falsi	19	5	79.17%	
	pred. Veri	3	16	84.21%	A_1670_10	pred. Veri	5	19	79.17%	A_1670_10	pred. Veri	2	16	88.89%	
	Spec., Sens.	84.21%	69.57%			Spec., Sens.	76.19%	90.48%		Spec., Sens.	90.48%	76.19%			

Dot (6:4)					Radial (6:4)					ANOVA (6:4)					C
feature	Training				feature	Training				feature	Training				
A_1511_36		Veri negativi	Veri positivi	class precision	A_5728_63		Veri negativi	Veri positivi	class precision	A_3168_37		Veri negativi	Veri positivi	class precision	
A_6445_61	pred. Falsi	100	25	80.00%	A_6463_69	pred. Falsi	96	10	90.57%	A_1105_88	pred. Falsi	97	28	77.60%	
A_1117_42	pred. Veri	14	71	83.53%	A_6864_19	pred. Veri	18	86	82.69%	A_1582_74	pred. Veri	17	68	80.00%	
A_2216_93	Spec., Sens.	87.72%	73.96%		A_1149_20	class recall	84.21%	89.58%		A_2216_93	class recall	85.09%	70.83%		
A_5889_71	Model				A_2216_93	Model				A_5889_71	Model				
A_1670_10		Veri negativi	Veri positivi	class precision	A_6190_13		Veri negativi	Veri positivi	class precision	A_6190_13		Veri negativi	Veri positivi	class precision	
	pred. Falsi	38	12	76.00%	A_1425_52	pred. Falsi	33	3	91.67%	A_6949_10	pred. Falsi	39	10	79.59%	
	pred. Veri	5	29	85.29%	A_1670_10	pred. Veri	10	38	79.17%	A_1670_10	pred. Veri	8	27	77.14%	
	class recall	88.37%	70.73%			class recall	76.74%	92.68%		class recall	82.98%	72.97%			

Tabella 12. Performance classificative dei diversi *clusters* di peptidi selezionati per il confronto pazienti ccRCC *versus* controlli, utilizzando come *data Input* solo i segnali statisticamente differenti tra le due popolazioni. Training = Performance relative alla fase di *feature selection* utilizzando diversi algoritmi (*Dot*, *Radial* e *ANOVA*); una *cross-validation* di tipo *k-fold* è stata applicata per valutare la capacità predittiva di questi modelli. Model = Performance relative alla fase validazione su una casistica indipendente; la popolazione di partenza è stata divisa in diverse percentuali: A) 90% per la fase di costruzione del modello e 10% per la sua validazione. B) 80% per la fase di costruzione del modello e 20% per la sua validazione. C) 60% per la fase di costruzione del modello e 40% per la sua validazione.

Dot (9:1)					Radial (9:1)					ANOVA (9:1)					A																																																														
Training					Training					Training																																																																			
Feature A_1670_10 A_1727_42 A_2192_74 A_3003_74 A_3251_59 A_3636_33 A_4622_87 A_5431_82 A_5529_85 A_5962_02 A_6060_19 A_6175_03					<table border="1"> <thead> <tr> <th></th> <th>true false</th> <th>true true</th> <th>class precision</th> </tr> </thead> <tbody> <tr> <td>pred. false</td> <td>123</td> <td>21</td> <td>85.42%</td> </tr> <tr> <td>pred. true</td> <td>14</td> <td>97</td> <td>87.39%</td> </tr> <tr> <td>class recall</td> <td>89.78%</td> <td>82.20%</td> <td></td> </tr> </tbody> </table>						true false	true true	class precision	pred. false	123	21	85.42%	pred. true	14	97	87.39%	class recall	89.78%	82.20%		Feature A_1051_87 A_1117_42 A_1655_49 A_1670_10 A_1727_42 A_2192_74 A_2216_93 A_3162_23 A_3443_23 A_3616_81 A_4962_79 A_5529_85 A_5728_63					<table border="1"> <thead> <tr> <th></th> <th>Veri negativi</th> <th>Veri positivi</th> <th>class precision</th> </tr> </thead> <tbody> <tr> <td>pred. Falsi</td> <td>105</td> <td>3</td> <td>97.22%</td> </tr> <tr> <td>pred. Veri</td> <td>32</td> <td>115</td> <td>78.23%</td> </tr> <tr> <td>Spec., Sens.</td> <td>76.64%</td> <td>97.46%</td> <td></td> </tr> </tbody> </table>						Veri negativi	Veri positivi	class precision	pred. Falsi	105	3	97.22%	pred. Veri	32	115	78.23%	Spec., Sens.	76.64%	97.46%		Feature A_1670_10 A_2216_93 A_3162_23 A_5728_63 A_6175_03 A_8378_18 A_11351_06					<table border="1"> <thead> <tr> <th></th> <th>Veri negativi</th> <th>Veri positivi</th> <th>class precision</th> </tr> </thead> <tbody> <tr> <td>pred. Falsi</td> <td>116</td> <td>22</td> <td>84.06%</td> </tr> <tr> <td>pred. Veri</td> <td>21</td> <td>96</td> <td>82.05%</td> </tr> <tr> <td>Spec., Sens.</td> <td>84.67%</td> <td>81.36%</td> <td></td> </tr> </tbody> </table>						Veri negativi	Veri positivi	class precision	pred. Falsi	116	22	84.06%	pred. Veri	21	96	82.05%	Spec., Sens.	84.67%	81.36%	
	true false	true true	class precision																																																																										
pred. false	123	21	85.42%																																																																										
pred. true	14	97	87.39%																																																																										
class recall	89.78%	82.20%																																																																											
	Veri negativi	Veri positivi	class precision																																																																										
pred. Falsi	105	3	97.22%																																																																										
pred. Veri	32	115	78.23%																																																																										
Spec., Sens.	76.64%	97.46%																																																																											
	Veri negativi	Veri positivi	class precision																																																																										
pred. Falsi	116	22	84.06%																																																																										
pred. Veri	21	96	82.05%																																																																										
Spec., Sens.	84.67%	81.36%																																																																											
Model <table border="1"> <thead> <tr> <th></th> <th>Veri negativi</th> <th>Veri positivi</th> <th>class precision</th> </tr> </thead> <tbody> <tr> <td>pred. Falsi</td> <td>8</td> <td>2</td> <td>80.00%</td> </tr> <tr> <td>pred. Veri</td> <td>1</td> <td>14</td> <td>93.33%</td> </tr> <tr> <td>Spec., Sens.</td> <td>88.89%</td> <td>87.50%</td> <td></td> </tr> </tbody> </table>						Veri negativi	Veri positivi	class precision	pred. Falsi	8	2	80.00%	pred. Veri	1	14	93.33%	Spec., Sens.	88.89%	87.50%		Model <table border="1"> <thead> <tr> <th></th> <th>Veri negativi</th> <th>Veri positivi</th> <th>class precision</th> </tr> </thead> <tbody> <tr> <td>pred. Falsi</td> <td>9</td> <td>0</td> <td>100.00%</td> </tr> <tr> <td>pred. Veri</td> <td>6</td> <td>10</td> <td>62.50%</td> </tr> <tr> <td>Spec., Sens.</td> <td>60.00%</td> <td>100.00%</td> <td></td> </tr> </tbody> </table>						Veri negativi	Veri positivi	class precision	pred. Falsi	9	0	100.00%	pred. Veri	6	10	62.50%	Spec., Sens.	60.00%	100.00%		Model <table border="1"> <thead> <tr> <th></th> <th>Veri negativi</th> <th>Veri positivi</th> <th>class precision</th> </tr> </thead> <tbody> <tr> <td>pred. Falsi</td> <td>17</td> <td>2</td> <td>89.47%</td> </tr> <tr> <td>pred. Veri</td> <td>0</td> <td>6</td> <td>100.00%</td> </tr> <tr> <td>Spec., Sens.</td> <td>100.00%</td> <td>75.00%</td> <td></td> </tr> </tbody> </table>						Veri negativi	Veri positivi	class precision	pred. Falsi	17	2	89.47%	pred. Veri	0	6	100.00%	Spec., Sens.	100.00%	75.00%																
	Veri negativi	Veri positivi	class precision																																																																										
pred. Falsi	8	2	80.00%																																																																										
pred. Veri	1	14	93.33%																																																																										
Spec., Sens.	88.89%	87.50%																																																																											
	Veri negativi	Veri positivi	class precision																																																																										
pred. Falsi	9	0	100.00%																																																																										
pred. Veri	6	10	62.50%																																																																										
Spec., Sens.	60.00%	100.00%																																																																											
	Veri negativi	Veri positivi	class precision																																																																										
pred. Falsi	17	2	89.47%																																																																										
pred. Veri	0	6	100.00%																																																																										
Spec., Sens.	100.00%	75.00%																																																																											

Dot (8:2)					Radial (8:2)					ANOVA (8:2)					B																																																				
Training					Training					Training																																																									
Feature A_1670_10 A_1727_42 A_2192_74 A_3003_74 A_3251_59 A_3636_33 A_4622_87 A_5431_82 A_5529_85 A_5962_02 A_6060_19 A_6175_03					<table border="1"> <thead> <tr> <th></th> <th>Veri negativi</th> <th>Veri positivi</th> <th>class precision</th> </tr> </thead> <tbody> <tr> <td>pred. Falsi</td> <td>123</td> <td>21</td> <td>85.42%</td> </tr> <tr> <td>pred. Veri</td> <td>14</td> <td>97</td> <td>87.39%</td> </tr> <tr> <td>Spec., Sens.</td> <td>89.78%</td> <td>82.20%</td> <td></td> </tr> </tbody> </table>						Veri negativi	Veri positivi	class precision	pred. Falsi	123	21	85.42%	pred. Veri	14	97	87.39%	Spec., Sens.	89.78%	82.20%		<table border="1"> <thead> <tr> <th></th> <th>Veri negativi</th> <th>Veri positivi</th> <th>class precision</th> </tr> </thead> <tbody> <tr> <td>pred. Falsi</td> <td>105</td> <td>3</td> <td>97.22%</td> </tr> <tr> <td>pred. Veri</td> <td>32</td> <td>115</td> <td>78.23%</td> </tr> <tr> <td>Spec., Sens.</td> <td>76.64%</td> <td>97.46%</td> <td></td> </tr> </tbody> </table>						Veri negativi	Veri positivi	class precision	pred. Falsi	105	3	97.22%	pred. Veri	32	115	78.23%	Spec., Sens.	76.64%	97.46%		<table border="1"> <thead> <tr> <th></th> <th>Veri negativi</th> <th>Veri positivi</th> <th>class precision</th> </tr> </thead> <tbody> <tr> <td>pred. Falsi</td> <td>116</td> <td>22</td> <td>84.06%</td> </tr> <tr> <td>pred. Veri</td> <td>21</td> <td>96</td> <td>82.05%</td> </tr> <tr> <td>Spec., Sens.</td> <td>84.67%</td> <td>81.36%</td> <td></td> </tr> </tbody> </table>						Veri negativi	Veri positivi	class precision	pred. Falsi	116	22	84.06%	pred. Veri	21	96	82.05%	Spec., Sens.	84.67%	81.36%	
	Veri negativi	Veri positivi	class precision																																																																
pred. Falsi	123	21	85.42%																																																																
pred. Veri	14	97	87.39%																																																																
Spec., Sens.	89.78%	82.20%																																																																	
	Veri negativi	Veri positivi	class precision																																																																
pred. Falsi	105	3	97.22%																																																																
pred. Veri	32	115	78.23%																																																																
Spec., Sens.	76.64%	97.46%																																																																	
	Veri negativi	Veri positivi	class precision																																																																
pred. Falsi	116	22	84.06%																																																																
pred. Veri	21	96	82.05%																																																																
Spec., Sens.	84.67%	81.36%																																																																	
Model <table border="1"> <thead> <tr> <th></th> <th>Veri negativi</th> <th>Veri positivi</th> <th>class precision</th> </tr> </thead> <tbody> <tr> <td>pred. Falsi</td> <td>20</td> <td>3</td> <td>86.96%</td> </tr> <tr> <td>pred. Veri</td> <td>5</td> <td>23</td> <td>82.14%</td> </tr> <tr> <td>Spec., Sens.</td> <td>80.00%</td> <td>88.46%</td> <td></td> </tr> </tbody> </table>						Veri negativi	Veri positivi	class precision	pred. Falsi	20	3	86.96%	pred. Veri	5	23	82.14%	Spec., Sens.	80.00%	88.46%		Model <table border="1"> <thead> <tr> <th></th> <th>Veri negativi</th> <th>Veri positivi</th> <th>class precision</th> </tr> </thead> <tbody> <tr> <td>pred. Falsi</td> <td>21</td> <td>0</td> <td>100.00%</td> </tr> <tr> <td>pred. Veri</td> <td>6</td> <td>24</td> <td>80.00%</td> </tr> <tr> <td>Spec., Sens.</td> <td>77.78%</td> <td>100.00%</td> <td></td> </tr> </tbody> </table>						Veri negativi	Veri positivi	class precision	pred. Falsi	21	0	100.00%	pred. Veri	6	24	80.00%	Spec., Sens.	77.78%	100.00%		Model <table border="1"> <thead> <tr> <th></th> <th>Veri negativi</th> <th>Veri positivi</th> <th>class precision</th> </tr> </thead> <tbody> <tr> <td>pred. Falsi</td> <td>25</td> <td>2</td> <td>92.59%</td> </tr> <tr> <td>pred. Veri</td> <td>6</td> <td>18</td> <td>75.00%</td> </tr> <tr> <td>Spec., Sens.</td> <td>80.65%</td> <td>90.00%</td> <td></td> </tr> </tbody> </table>						Veri negativi	Veri positivi	class precision	pred. Falsi	25	2	92.59%	pred. Veri	6	18	75.00%	Spec., Sens.	80.65%	90.00%						
	Veri negativi	Veri positivi	class precision																																																																
pred. Falsi	20	3	86.96%																																																																
pred. Veri	5	23	82.14%																																																																
Spec., Sens.	80.00%	88.46%																																																																	
	Veri negativi	Veri positivi	class precision																																																																
pred. Falsi	21	0	100.00%																																																																
pred. Veri	6	24	80.00%																																																																
Spec., Sens.	77.78%	100.00%																																																																	
	Veri negativi	Veri positivi	class precision																																																																
pred. Falsi	25	2	92.59%																																																																
pred. Veri	6	18	75.00%																																																																
Spec., Sens.	80.65%	90.00%																																																																	

Dot (6:4)					Radial (6:4)					ANOVA (6:4)					C																																																				
Training					Training					Training																																																									
Feature A_1670_10 A_1727_42 A_2192_74 A_3003_74 A_3251_59 A_3636_33 A_4622_87 A_5431_82 A_5529_85 A_5962_02 A_6060_19 A_6175_03					<table border="1"> <thead> <tr> <th></th> <th>Veri negativi</th> <th>Veri positivi</th> <th>class precision</th> </tr> </thead> <tbody> <tr> <td>pred. Falsi</td> <td>123</td> <td>21</td> <td>85.42%</td> </tr> <tr> <td>pred. Veri</td> <td>14</td> <td>97</td> <td>87.39%</td> </tr> <tr> <td>Spec., Sens.</td> <td>89.78%</td> <td>82.20%</td> <td></td> </tr> </tbody> </table>						Veri negativi	Veri positivi	class precision	pred. Falsi	123	21	85.42%	pred. Veri	14	97	87.39%	Spec., Sens.	89.78%	82.20%		<table border="1"> <thead> <tr> <th></th> <th>Veri negativi</th> <th>Veri positivi</th> <th>class precision</th> </tr> </thead> <tbody> <tr> <td>pred. Falsi</td> <td>105</td> <td>3</td> <td>97.22%</td> </tr> <tr> <td>pred. Veri</td> <td>32</td> <td>115</td> <td>78.23%</td> </tr> <tr> <td>Spec., Sens.</td> <td>76.64%</td> <td>97.46%</td> <td></td> </tr> </tbody> </table>						Veri negativi	Veri positivi	class precision	pred. Falsi	105	3	97.22%	pred. Veri	32	115	78.23%	Spec., Sens.	76.64%	97.46%		<table border="1"> <thead> <tr> <th></th> <th>Veri negativi</th> <th>Veri positivi</th> <th>class precision</th> </tr> </thead> <tbody> <tr> <td>pred. Falsi</td> <td>116</td> <td>22</td> <td>84.06%</td> </tr> <tr> <td>pred. Veri</td> <td>21</td> <td>96</td> <td>82.05%</td> </tr> <tr> <td>Spec., Sens.</td> <td>84.67%</td> <td>81.36%</td> <td></td> </tr> </tbody> </table>						Veri negativi	Veri positivi	class precision	pred. Falsi	116	22	84.06%	pred. Veri	21	96	82.05%	Spec., Sens.	84.67%	81.36%	
	Veri negativi	Veri positivi	class precision																																																																
pred. Falsi	123	21	85.42%																																																																
pred. Veri	14	97	87.39%																																																																
Spec., Sens.	89.78%	82.20%																																																																	
	Veri negativi	Veri positivi	class precision																																																																
pred. Falsi	105	3	97.22%																																																																
pred. Veri	32	115	78.23%																																																																
Spec., Sens.	76.64%	97.46%																																																																	
	Veri negativi	Veri positivi	class precision																																																																
pred. Falsi	116	22	84.06%																																																																
pred. Veri	21	96	82.05%																																																																
Spec., Sens.	84.67%	81.36%																																																																	
Model <table border="1"> <thead> <tr> <th></th> <th>Veri negativi</th> <th>Veri positivi</th> <th>class precision</th> </tr> </thead> <tbody> <tr> <td>pred. Falsi</td> <td>48</td> <td>8</td> <td>85.71%</td> </tr> <tr> <td>pred. Veri</td> <td>5</td> <td>41</td> <td>89.13%</td> </tr> <tr> <td>Spec., Sens.</td> <td>90.57%</td> <td>83.67%</td> <td></td> </tr> </tbody> </table>						Veri negativi	Veri positivi	class precision	pred. Falsi	48	8	85.71%	pred. Veri	5	41	89.13%	Spec., Sens.	90.57%	83.67%		Model <table border="1"> <thead> <tr> <th></th> <th>Veri negativi</th> <th>Veri positivi</th> <th>class precision</th> </tr> </thead> <tbody> <tr> <td>pred. Falsi</td> <td>37</td> <td>1</td> <td>97.37%</td> </tr> <tr> <td>pred. Veri</td> <td>20</td> <td>44</td> <td>68.75%</td> </tr> <tr> <td>Spec., Sens.</td> <td>64.91%</td> <td>97.78%</td> <td></td> </tr> </tbody> </table>						Veri negativi	Veri positivi	class precision	pred. Falsi	37	1	97.37%	pred. Veri	20	44	68.75%	Spec., Sens.	64.91%	97.78%		Model <table border="1"> <thead> <tr> <th></th> <th>Veri negativi</th> <th>Veri positivi</th> <th>class precision</th> </tr> </thead> <tbody> <tr> <td>pred. Falsi</td> <td>48</td> <td>5</td> <td>90.57%</td> </tr> <tr> <td>pred. Veri</td> <td>9</td> <td>40</td> <td>81.63%</td> </tr> <tr> <td>Spec., Sens.</td> <td>84.21%</td> <td>88.89%</td> <td></td> </tr> </tbody> </table>						Veri negativi	Veri positivi	class precision	pred. Falsi	48	5	90.57%	pred. Veri	9	40	81.63%	Spec., Sens.	84.21%	88.89%						
	Veri negativi	Veri positivi	class precision																																																																
pred. Falsi	48	8	85.71%																																																																
pred. Veri	5	41	89.13%																																																																
Spec., Sens.	90.57%	83.67%																																																																	
	Veri negativi	Veri positivi	class precision																																																																
pred. Falsi	37	1	97.37%																																																																
pred. Veri	20	44	68.75%																																																																
Spec., Sens.	64.91%	97.78%																																																																	
	Veri negativi	Veri positivi	class precision																																																																
pred. Falsi	48	5	90.57%																																																																
pred. Veri	9	40	81.63%																																																																
Spec., Sens.	84.21%	88.89%																																																																	

Tabella 13. Performance classificative dei diversi *clusters* di peptidi selezionati per il confronto pazienti ccRCC *versus* controlli, utilizzando come *data Input* tutti i dati a disposizione. Training = Performance relative alla fase di *feature selection* utilizzando diversi algoritmi (*Dot*, *Radial* e *ANOVA*); una *cross-validation* di tipo *k-fold* è stata applicata per valutare la capacità predittiva di questi modelli. Model = Performance relative alla fase validazione su una casistica indipendente; la popolazione di partenza è stata divisa in diverse percentuali: A) 90% per la fase di costruzione del modello e 10% per la sua validazione. B) 80% per la fase di costruzione del modello e 20% per la sua validazione. C) 60% per la fase di costruzione del modello e 40% per la sua validazione.

4.2.1.1. CLUSTER DISCRIMINANTE TRA CTRL *vs* ccRCC

L'intera casistica è stata inizialmente suddivisa in pazienti affetti da ccRCC (n = 118) e in soggetti controllo (n = 137). Le due popolazioni sono state a loro volta, divise in due gruppi: il primo gruppo è stato utilizzato per la creazione del modello, mentre il secondo gruppo è stato adottato per la validazione dei modelli costruiti (*performance evaluation* o fase di *test*).

Confrontando i profili urinari di 137 soggetti controllo con 118 soggetti affetti da ccRCC, è stato individuato un *cluster* discriminante di dodici peptidi (m/z in modalità lineare 1670, 1727, 2192, 3005, 3252, 3636, 4623, 5432, 5532, 5964, 6062 e 6175) avente rispettivamente specificità e sensibilità del 90% e 82%, (Tabella 14A). Una *cross-validation* di tipo *k-fold* (k = 10) è stata applicata per valutare la capacità predittiva di questo modello.

	Veri Negativi	Veri positivi	VPN ^a , VPP ^b	A
pred. Negativi	123	21	85,42% ^a	
pred. Positivi	14	97	87,39% ^b	
Spec., Sens.	89.78%	82.20%		

	Veri Negativi	Veri positivi	VPN ^a , VPP ^b	B
pred. Negativi	48	8	85,71% ^a	
pred. Positivi	5	41	89,13% ^b	
Spec., Sens.	90.57%	83.67%		

Tabella 14. Performance classifiche del *cluster* di dodici peptidi per il confronto pazienti ccRCC *versus* controlli. A) Performance relative alla fase di *feature selection*; una *cross-validation* di tipo *k-fold* è stata applicata per valutare la capacità predittiva di questo modello. B) Performance relative alla fase di validazione su una casistica indipendente (circa il 40% dei soggetti) rispetto a quella utilizzata per la creazione del modello (circa il 60% dei soggetti). VPN = valore predittivo negativo, VPP = valore predittivo positivo.

Sulla base delle indicazioni e delle linee guida suggerite in letteratura [284], il modello, costruito utilizzando circa il 60% della popolazione, è stato successivamente validato, nella fase di test, su una casistica indipendente di 53 controlli e 49 pazienti ccRCC. Il *cluster* discriminante di questi 12 segnali è stato in grado di discriminare i soggetti controllo dai soggetti affetti da ccRCC con valori di specificità e sensibilità di 91% e 84%, rispettivamente (Tabella 14B). Il valore AUC del modello era di 0.96 che corrisponde ad un test altamente accurato (Figura 43) [229].

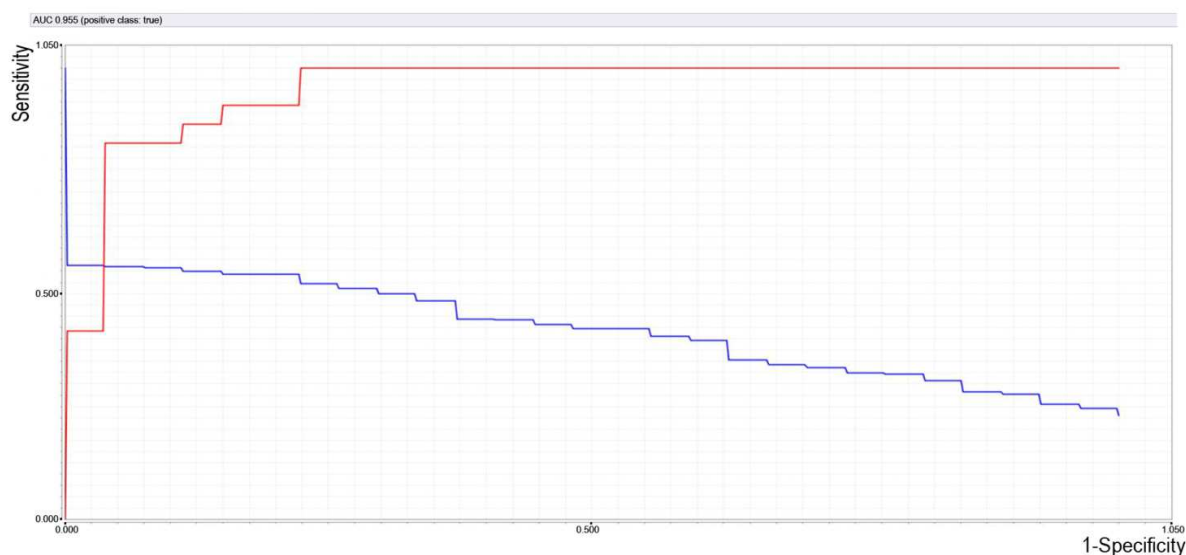


Figura 43. Curva ROC: la curva rossa mostra il potere diagnostico del *cluster* di 12 segnali in grado di discriminare i soggetti controllo dai ccRCC quando applicato a una coorte di soggetti indipendente.

Quattro di questi ioni, a m/z 4623, 5432, 6062 e 6175, erano presenti in minore concentrazione ($p < 0.05$) nei pazienti affetti da ccRCC, mentre altri quattro ioni, a m/z 1670, 1727, 3636 e 5532 erano sovra-rappresentati ($p < 0.05$) (Tabella 15 e Figura 44). I rimanenti 4 ioni (m/z 2192, 3005, 3252, 5964) sono risultati non significativamente differenti ($p > 0.05$) secondo *t-test/Welch test* per insieme di dati parametrici, e *Wilcoxon/Kolmogorov-Smirnov test* per dati non parametrici.

Ioni (m/z)	<i>p-value</i>	Espressione urinaria
1670	< 0.001	Up
1727	< 0.001	Up
2192	> 0.05	n.s.
3005	> 0.05	n.s.
3252	> 0.05	n.s.
3636	< 0.001	Up
4623	< 0.001	Down
5432	< 0.001	Down
5532	< 0.001	Up
5964	> 0.05	n.s.
6062	< 0.001	Down
6175	< 0.001	Down

Tabella 15. Valori di espressione urinaria e valori di *p-value* dei dodici ioni inclusi nel modello in grado di distinguere pazienti ccRCC da soggetti controllo (n.s.: non significativo). *P-value* calcolato con *t-test/Welch test* per insieme di dati parametrici, *Wilcoxon/Kolmogorov-Smirnov test* per dati non parametrici.

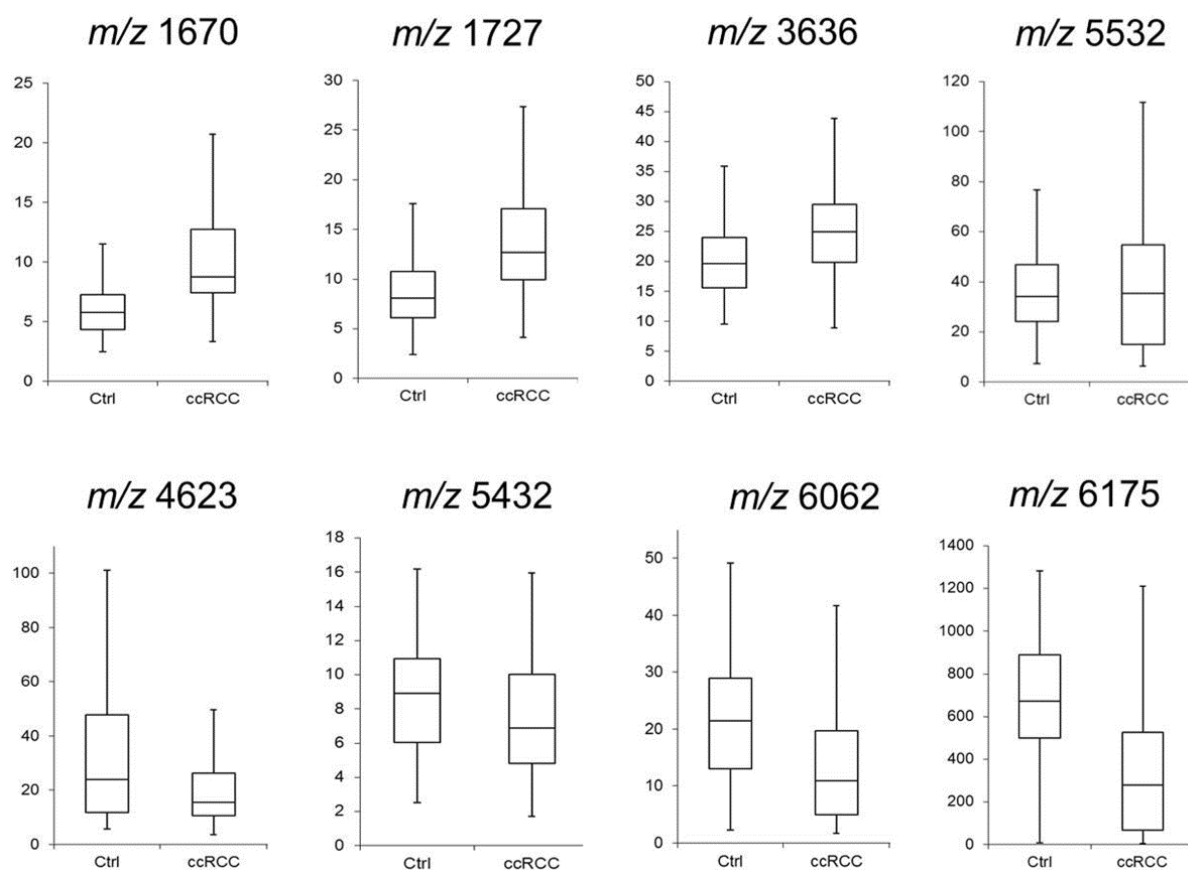


Figura 44. *Box-Plot* di otto ioni inclusi nel modello in grado di distinguere soggetti controllo da pazienti affetti da ccRCC e statisticamente differenti ($p < 0.05$) tra i due gruppi (vedi tabella 14).

4.2.1.2. CLUSTER DISCRIMINANTE TRA TUMORI RENALI MALIGNI VS MASSE RENALI BENIGNE E CONTROLLI

Tramite l'analisi interattiva, è stata effettuata un'ulteriore classificazione caso/controllo: soggetti con masse renali benigne addizionati ai soggetti controllo ($n = 153$) *versus* pazienti affetti da tumore renale maligno ($n = 137$). Le performance di classificazione relative sia al processo di *feature selection* e *cross-validation* che al *validation test* sono riportate nella Tabella 16.

	Veri Negativi	Veri positivi	VPN ^a , VPP ^b	A
pred. Negativi	120	17	87.6% ^a	
pred. Positivi	33	120	78.4% ^b	
Spec., Sens.	78.40%	87.60%		

	Veri Negativi	Veri positivi	VPN ^a , VPP ^b	B
pred. Negativi	53	13	80.3% ^a	
pred. Positivi	8	42	84.0% ^b	
Spec., Sens.	86.90%	76.40%		

Tabella 16. Performance del cluster di 12 peptidi nel classificare tumori maligni *versus* lesioni renali benigne e controlli. A) Performance relative alla fase di *feature selection*; una *cross-validation* di tipo *k-fold* è stata applicata per valutare la capacità predittiva di questo modello. B) Performance del modello, generato usando circa il 60% dei dati, nella fase di validazione effettuata su circa il 40% dei soggetti. VPN = valore predittivo negativo, VPP = valore predittivo positivo.

I processi di *feature selection* e di *cross-validation* sono stati eseguiti su tutti soggetti con masse renali benigne insieme ai soggetti controllo e i pazienti affetti da tumore maligno; mediante analisi statistica degli spettri MALDI-TOF è stato selezionato un *cluster* di dodici peptidi (m/z in modalità lineare 1116, 1670, 2216, 2528, 2661, 3162, 3443, 5032, 5532, 6130, 6786 e 10654). Le performance di classificazione del modello, hanno mostrato indici di specificità e sensibilità pari, rispettivamente, a 78% e 88% (Tabella 16A).

Anche in questa analisi le due popolazioni sono state, a loro volta, divise in due gruppi. La buona capacità diagnostica è stata quindi confermata mediante una validazione del modello utilizzando un secondo gruppo di dati indipendenti rispetto a quelli impiegati per la sua costruzione, 61 soggetti con masse renali benigne più soggetti controllo e 55 soggetti con tumore maligno; i risultati ottenuti hanno confermato i valori di specificità al 87% e sensibilità al 76% (Tabella 16B). Otto di questi ioni a m/z 1116, 1670, 2216, 2528, 2661, 3162, 3443 and 5532, erano sovra-rappresentati ($p < 0.05$) nei pazienti con tumori maligni rispetto ai soggetti controllo e soggetti con masse tumorali benigne, mentre tre ioni a m/z 5032, 6130 e 6786, erano sotto-rappresentati ($p < 0.05$) (Tabella 17 e Figura 45). Solo uno dei dodici ioni inclusi nel *cluster* discriminante (m/z 10654), è risultato avere un *p-value* superiore a 0.05.

Ioni (m/z)	p -value	Espressione urinaria
1116	< 0.001	Up
1670	< 0.001	Up
2216	< 0.05	Up
2528	< 0.001	Up
2661	< 0.001	Up
3162	< 0.001	Up
3443	< 0.001	Up
5032	< 0.001	Down
5532	< 0.05	Up
6130	< 0.001	Down
6786	< 0.001	Down
10654	> 0.05	n.s.

Tabella 17. Valori di espressione urinaria e valori di p -value dei 12 ioni inclusi nel modello in grado di distinguere pazienti con masse benigne e soggetti controllo da tumori maligni del rene (n.s.: non significativo). P -value calcolato con t -test/*Welch test* per insieme di dati parametrici, *Wilcoxon/Kolmogorov-Smirnov test* per dati non parametrici.

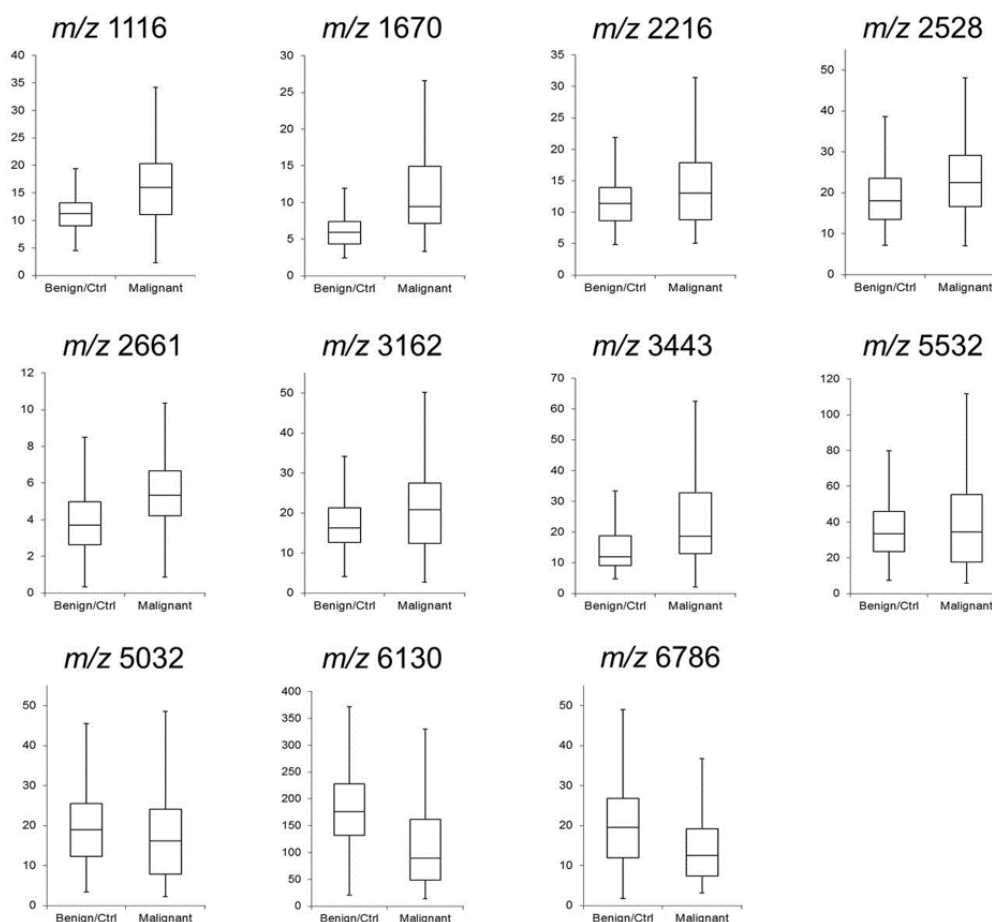


Figura 45. *Box-Plot* di undici ioni inclusi nel modello in grado di distinguere pazienti con masse benigne o controlli da pazienti con masse renali maligne e statisticamente differenti ($p < 0.05$) tra le due popolazioni (vedi Tabella 17).

Sulla base della curva ROC, i dodici ioni mostravano un buon potere diagnostico secondo i criteri suggeriti da Swets [229], con un valore di AUC pari a 0.89 (Figura 46).

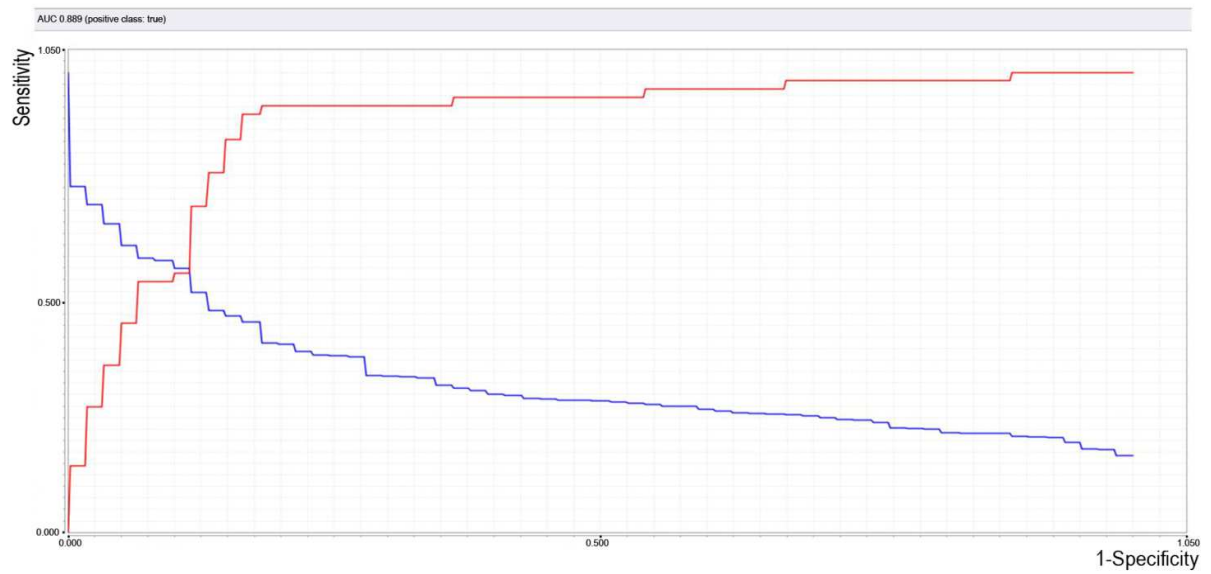


Figura 46. Curva ROC: la curva rossa mostra l'elevata accuratezza diagnostica del cluster di 12 segnali in grado di discriminare i soggetti controllo/benigni dai soggetti maligni quando applicato a una coorte di soggetti indipendente.

4.2.2. ANALISI DEI PROFILI PEPTIDICI E INDIVIDUAZIONE DI SEGNALI CORRELATI ALLA PROGRESSIONE TUMORALE

L'analisi statistica effettuata sui profili peptidici dei pazienti affetti da ccRCC ha permesso di valutare la possibile associazione tra le variazioni dei livelli urinari dei peptidi endogeni e i dati clinici (dimensione, pT e *grade*) dei soggetti arruolati nello studio.

4.2.2.1. VARIAZIONI PEPTIDICHE SECONDO LA DIMENSIONE TUMORALE

La Tabella 18 mostra i quindici peptidi che, a seguito dell'analisi statistica, presentano una variazione statisticamente significativa in base alla dimensione della massa tumorale; la maggior parte di essi mostra inoltre una sovra-rappresentazione rispetto ai soggetti controllo (Tabella 18 – seconda colonna).

MALDI-LM (m/z)	RCC vs CTRL P<0.05	Correlazione	<i>p-value</i>
1756	up	-0.307	0.0430
1912	down	-0.311	0.0380
1934	down	-0.312	0.0420
2661	up	0.514	0.0010
3724	down	0.320	0.0440
3990	down	0.345	0.0270
4410	down	0.357	0.0190
4538		0.363	0.0180
4627	down	0.331	0.0280
4752	down	0.363	0.0250
4835		0.481	0.0020
4849	up	0.310	0.0430
4866	up	0.315	0.0450
6261	down	0.352	0.0240
7268		0.342	0.0220

Tabella 18. Peptidi aventi valori di abbondanza urinaria (area) statisticamente significativi ($p < 0.05$) al variare della massa tumorale (cm). Up/Down si riferisce alla sovra/sotto rappresentazione del peptide nei pazienti ccRCC rispetto ai soggetti controllo.

È interessante notare come per tre di essi l'alterazione risulti essere negativa (da -0.307 a -0.312), mentre per i restanti 12 la variazione ha valore positivo (da 0.31 a 0.514).

4.2.2.2. VARIAZIONI PEPTIDICHE SECONDO LO STADIO TUMORALE

Ventisei peptidi mostrano una variazione statisticamente significativa rispetto allo stadio tumorale (tabella 19); quindici di essi risultano essere diversamente espressi nei pazienti ccRCC rispetto ai soggetti sani (Tabella 19 – seconda colonna). Solo cinque di essi mostrano un'alterazione negativa (da -0.202 a -0.326), mentre per gli altri ventuno essa risulta essere positiva (da 0.203 a 0.27).

MALDI-LM (<i>m/z</i>)	RCC vs CTRL P<0.05	Correlazione	<i>p-value</i>
1756	up	-0.251	0.0070
1894	down	-0.326	0.0000
1912	down	-0.273	0.0030
1934	down	-0.218	0.0210
2661	up	0.213	0.0260
2961	down	-0.202	0.0290
4027	down	0.253	0.0080
4639	down	0.240	0.0120
4752	down	0.212	0.0340
4835		0.259	0.0070
4936		0.237	0.0110
4963	up	0.211	0.0280
5232	up	0.223	0.0200
5578	down	0.203	0.0290
5832		0.205	0.0280
6261	down	0.270	0.0050
6305	down	0.221	0.0170
6949		0.207	0.0250
7109		0.203	0.0300
7269		0.257	0.0050
7358		0.226	0.0150
8182	down	0.241	0.0150
8293		0.207	0.0310
9074		0.221	0.0210
9748		0.233	0.0170
11858		0.214	0.0230

Tabella 19. Peptidi aventi valori di abbondanza urinaria (area) statisticamente significativi ($p < 0.05$) al variare del pT. Up/Down si riferisce alla sovra/sotto rappresentazione del peptide nei pazienti ccRCC rispetto ai soggetti controllo.

4.2.2.3. ALTERAZIONI PEPTIDICHE RISPETTO AL GRADO TUMORALE

Solo cinque peptidi mostrano un'alterazione statisticamente significativa rispetto al grado di differenziamento nucleare della massa tumorale (Tabella 20); quattro peptidi sono inoltre variati in maniera statisticamente significativa tra pazienti e soggetti controllo (Tabella 20 – Seconda colonna). Solo uno di essi mostra una variazione positiva, mentre per i restanti quattro l'alterazione presenta valori negativi di circa -0.2.

MALDI-LM (<i>m/z</i>)	RCC vs CTRL P<0.05	Correlazione	<i>p-value</i>
1024	up	-0.214	0.030
1052		-0.213	0.031
1194	up	-0.227	0.020
1826	up	-0.222	0.032
3724	down	0.251	0.011

Tabella 20. Peptidi aventi valori di abbondanza urinaria (area) statisticamente significativi ($p < 0.05$) al variare del grado tumorale. Up/Down si riferisce alla sovra/sotto rappresentazione del peptide nei pazienti ccRCC rispetto ai soggetti controllo.

4.2.2.4. PEPTIDI URINARI DIFFERENTEMENTE ESPRESSI RISPETTO AI DATI ISTOLOGICI TUMORALI

L'analisi statistica effettuata sui profili peptidici dei pazienti ccRCC divisi in quattro gruppi, ha permesso di investigare le possibili alterazioni a carico dei peptidi endogeni urinari rispetto al pT e al grado tumorale. In particolare, quindici peptidi hanno mostrato livelli urinari variati in maniera statisticamente significativa in pazienti affetti da tumori a diversi stadi (Tabella 21), e nove di essi risultano inoltre statisticamente alterati tra pazienti e controlli (Tabella 21 – seconda colonna).

MALDI-LM (<i>m/z</i>)	RCC vs CTRL <i>p</i> <0.05	pT
1756	up	1-2 1-3 1-4
1894	down	1-3 1-4
1912	down	1-3 1-4
1934	down	1-3 1-4
2192		1-2 2-3 2-4
2661	up	1-3 1-4 2-3 2-4
4355	down	1-2 1-4
4639	down	1-4
4658	down	1-4
4752	down	1-2 1-3 1-4
4835		1-4
4936		1-4
5832		1-3 1-4
7269		1-3-1-4
7358		1-3 1-4

Tabella 21. Peptidi aventi valori di abbondanza urinaria (area) statisticamente significativi ($p < 0.05$) nei soggetti affetti da ccRCC rispetto ai CTRL e a diversi valori di pT (1= pT1a; 2= pT1b; 3= pT2a and 4= pT > 2b). Up/Down si riferisce alla sovra/sotto rappresentazione del peptide nei pazienti ccRCC rispetto ai soggetti controllo.

Inoltre, considerando solo il confronto tra pazienti con tumori a stadi precoci (pT1a e pT1b), quattordici peptidi hanno mostrato una variazione statisticamente significativa (Tabella 22).

MALDI-LM (<i>m/z</i>)	<i>p</i> -value	pT1b vs pT1a
1219	0.007	down
1525	0.018	down
1756	0.014	down
2192	0.023	up
2282	0.027	down
2791	0.002	up
4355	0.014	up
4367	0.004	up
4439	0.018	up
4627	0.043	up
4752	0.004	up
5514	0.038	up
6238	0.003	up
6261	0.023	up

Tabella 22. Peptidi aventi valori di abbondanza urinaria (area) statisticamente significativi ($p < 0.05$) tra pT1a e pT1b. Up/Down si riferisce alla sovra/sotto rappresentazione del peptide nei pazienti pT1b rispetto ai soggetti pT1a.

Per quanto riguarda invece l'associazione tra l'alterazione di espressione urinaria e il grado tumorale, nove peptidi hanno mostrato una variazione statisticamente significativa nei soggetti affetti da tumori a diversi gradi (Tabella 23), con sette di essi alterati anche tra pazienti e soggetti sani (Tabella 23 – Seconda colonna).

MALDI-LM (<i>m/z</i>)	RCC vs CTRL P<0.05	Grade
3151		1-4 2-4 3-4
3571	up	1-4 2-4 3-4
4355	down	2-4
4752	down	1-2 2-3 2-4
5004	down	2-4
5028	down	2-4
5044	down	2-4
5068		2-4
5232	up	2-4

Tabella 23. Peptidi aventi valori di abbondanza urinaria (area) statisticamente significativi ($p < 0.05$) nei soggetti affetti da ccRCC rispetto ai CTRL e a diversi valori di *Grade* (1= pT1a; 2= pT1b; 3= pT2a and 4= pT> 2b). Up/Down si riferisce alla sovra/sotto rappresentazione del peptide nei pazienti ccRCC rispetto ai soggetti controllo.

4.2.3. IDENTIFICAZIONE DEI PEPTIDI ENDOGENI URINARI

L'analisi mediante nLC-ESI-MS/MS delle frazioni ottenute dopo pre-frazionamento con biglie magnetiche HIC 8 delle urine di un *pool* di 80 controlli e un *pool* di 80 pazienti ha permesso la costruzione di una libreria di peptidi endogeni urinari identificati. In particolare, essa include un totale di 2339 peptidi unici di cui 726 con uno *score* maggiore della soglia di identità peptidica, appartenenti a 251 diverse isoforme proteiche, e 1613 peptidi ottenuti utilizzando come *cut-off* la soglia di omologia (*false discovery rate* < 5%). Le 251 diverse proteine identificate sono state in seguito elaborate attraverso l'algoritmo *Panther* al fine di visualizzare i componenti cellulari di cui fanno parte; 219 dei 251 codici associati alle diverse isoforme proteiche a cui appartenevano i frammenti peptidici identificati nelle urine sono stati riconosciuti dall'algoritmo e mappati nei diversi database utilizzati nell'analisi funzionale (Figura 47).

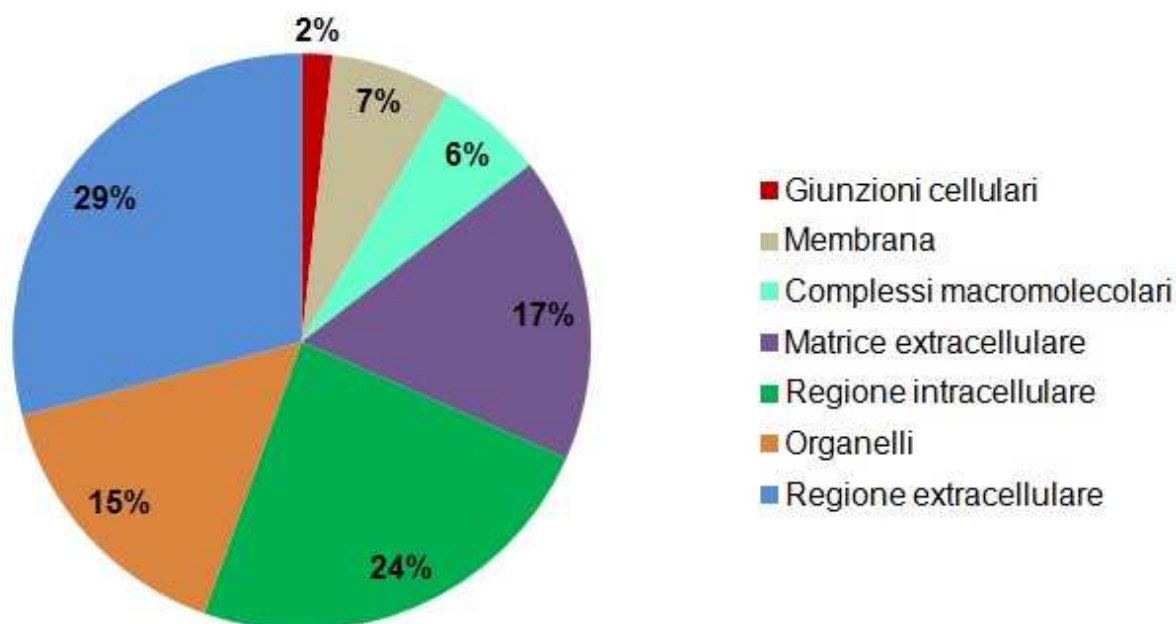


Figura 47: Localizzazione subcellulare delle proteine della libreria. I dati elaborati da *Panther* riguardano le isoforme proteiche a cui appartengono i 726 peptidi endogeni identificati con un punteggio superiore alla soglia di identità.

Più del 50% dei peptidi endogeni identificati appartengono a proteine localizzate a livello extracellulare, il 15% a proteine presenti negli organelli, il 24% alle altre porzioni intracellulari, il 7% a proteine di membrana mentre solo il 2% appartengono a proteine caratteristiche delle giunzioni cellulari.

4.2.3.1. IDENTIFICAZIONE DEI SEGNALI PRESENTI NEI *CLUSTER* DISCRIMINANTI

Al fine di associare l'identità dei peptidi in ESI ai segnali presenti nei profili ed inclusi nei modelli diagnostici, è stato necessario effettuare un accurato allineamento delle masse tra il valore di massa calcolato sulla sequenza amminoacidica identificata in ESI e il valore di massa $[M-H^+]$ dello spettro MALDI. I segnali acquisiti mediante lo spettrometro MALDI operante in modalità lineare (LM) sono stati perciò allineati con quelli ottenuti dall'analisi nLC-ESI. Per confermare tale allineamento è stata eseguita anche un'acquisizione MALDI-TOF in modalità *reflector* (RM).

Sette segnali peptidici osservati negli spettri MALDI-LM sono stati identificati come otto possibili frammenti proteici (Tabella 24).

Gli errori medi associati alle misurazioni delle masse molecolari (MME, *Mass Measurements Errors*) rilevate in MALDI-RM e MALDI-LM variavano da -53 a 72 ppm e -297 a 77 ppm, rispettivamente. Le differenze medie tra le masse molecolari rilevate in ESI e MALDI-RM variavano da -79 a 50 ppm. Al fine di ottenere masse più accurate e confermare la correttezza del *match* tra analisi nLC-ESI-MS/MS e MALDI-TOF sono stati acquisiti anche i valori m/z dei peptidi urinari di uno stesso campione depositato più volte su target metallico, oltre che da acquisizioni multiple su diversi soggetti.

MALDI			nLC ESI MS/MS										
Abbondanza relativa	LM m/z	RM m/z	Mr monoisotopica teorica + [H+]	Mr media teorica + [H+]	Mr sperimentale + [H+]	Cariche	Uniprot ID	MW (Da)	start	Sequenza	stop	PTM	Pep_score
↑ nei tumori maligni	1116	1115,60	1115,559	1116,209	1115,551	2	SCTM1_HUMAN	27039	69	RAHGQESAIF	78		58
↑ in ccRCC ↑ nei tumori maligni	1670	1668,84	1668,928	1669,899	1668,923	2	UROM_HUMAN	69761	590	GSVIDQSRVNLGPIT	605		32
↑ in ccRCC	1727	1725,92	1725,877	1726,865	1725,872	2	MEP1A_HUMAN	84419	636	EEALPVLSLQGGPSRQ	651		88
n.s.	2192	2190,89	2190,771	2192,179	2190,765	4	GP162_HUMAN	63930 (Isoforma 1)	344	IMSEEDGDDGGCDDYAEGR	363	N-acetilazione	26
								33062 (Isoforma 2)					
			2190,913	2192,253	2190,778	5	KPB1_HUMAN	137312	1021	AESQSPGTSMTSPSSGFSAYD	1042		22
↑ nei tumori maligni	2216	2215,39	2215,231	2216,523	2215,214	4	UROM_HUMAN	69761	587	FRSGVIDQSRVNLGPITR	606		54
↑ nei tumori maligni	2528	2527,26	2527,231	2528,706	2527,217	4	OSTP_HUMAN	33017-35423	19	VKQADSGSSEEKQLYNKYPDAVA	41		55
↑ nei tumori maligni	2661	2659,44	2659,257	2660,751	2659,252	5	FIBA_HUMAN	94973 (Isoforma 1)	605	DEAGSEADHEGTHSTKRGHAKSRPV	629		39
								69757 (Isoforma 2)					

Tabella 24. Elenco dei sette ioni inclusi nei *clusters* discriminanti che sono stati identificati con un punteggio superiore almeno alla soglia di omologia (FDR < 1%).

Peso molecolare della relativa proteina intatta (MW), abbondanze relative di ciascun peptide ($p < 0.05$) nel confronto tra ccCCR ($n = 118$) *versus* controlli ($n = 137$) e/o nel confronto tra tumori maligni ($n = 137$) e masse renali benigne più soggetti sani ($n = 153$), allineamento tra m/z determinato con la metodica MALDI-LM e MALDI-RM e Mr calcolato tramite metodica nLC-ESI-MS/MS dei peptidi identificati negli eluati dopo purificazione con biglie magnetiche RP-C8 di urine di pazienti e controlli. PTM = modificazioni post-traduzionali. ↑ = over-rappresentato; ↓ = sotto-rappresentato; n.s. = non significativo; LM = *linear mode*; RM = *reflector mode*.

L'identità del peptide corrispondente al segnale a m/z 2192 determinato in MALDI-TOF in modalità lineare (MALDI-LM) e incluso nel modello diagnostico, non può essere attribuita con certezza ad una sola sequenza amminoacidica (Tabella 24). In effetti la ridotta differenza di massa molecolare tra i due diversi peptidi identificati mediante nLC-ESI (m/z 2190,765 per GP162 e m/z 2190,778 per KPB1), e il potere risolutivo del MALDI operante in modalità *reflector* (MALDI-RM), non hanno permesso di discriminare lo specifico contributo di ciascuna sequenza al segnale osservato in MALDI-LM (Tabella 24). Nelle Figure 48 e 49 è mostrata la schermata, ottenuta da *Mascot*, relativa all'identificazione del frammento della GP162 e del frammento della KPB1, rispettivamente. Inoltre il frammento peptidico di sequenza ${}_{344}\text{IMSEEDGDDDGGCDDYAEGR}_{363}$ associato alla GP162 contiene una ipotetica N-acetilazione come modificazione post-traduzionale.

Mascot Search Results

Peptide View

MS/MS Fragmentation of **IMSEEDGDDDGGCDDYAEGR**

Found in **GP162_HUMAN** in **SwissProt**, Probable G-protein coupled receptor 162 OS=Homo sapiens GN=GPR162 PE=2 SV=1

Match to Query 18719: 2189.758416 from(548.446880,4+) intensity(51144.0000) index(16439)

Title: Cmpd 1143, +MSn(548.4469), 106.3 min

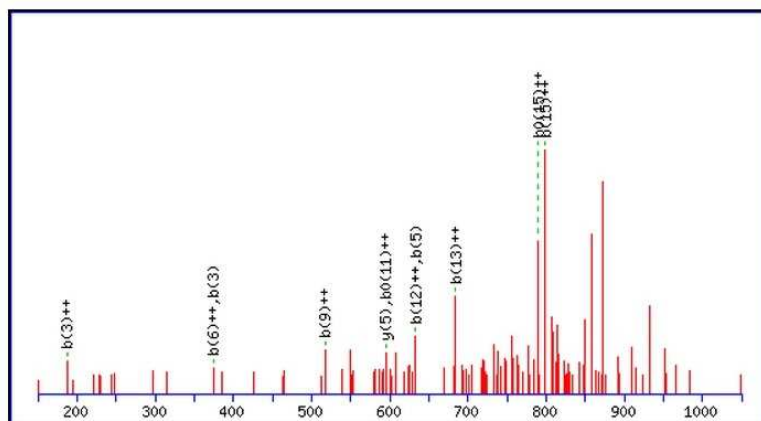
Data file C:\Program Files (x86)\Matrix Science\Mascot Daemon\mgf\19 C8 urine tutti file acetilate 10ppm\mascot_daemon_merge.mgf

Click mouse within plot area to zoom in by factor of two about that point

Or, Plot from to Da

Label all possible matches Label matches used for scoring

Show Y-axis



Monoisotopic mass of neutral peptide $M_r(\text{calc})$: 2189.7637

Variable modifications:

N-term : Acetyl (N-term)

Ions Score: 26 Expect: 0.0085

Matches : 11/182 fragment ions using 18 most intense peaks ([help](#))

Figura 48. Schermata di *Mascot* relativa all'identificazione del frammento peptidico di sequenza IMSEEDGDDDGGCDDYAEGR associato alla proteina GP162.

(MATRIX)
(SCIENCE) **Mascot Search Results**

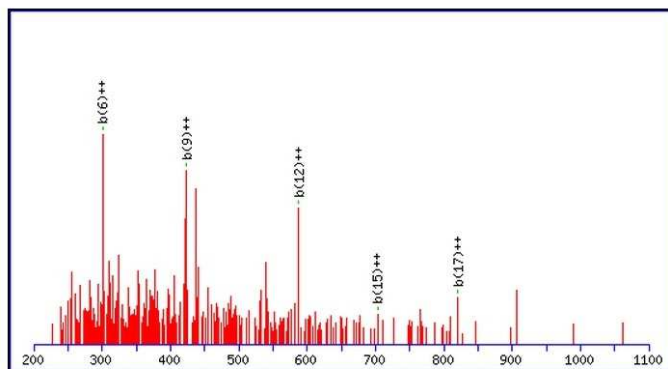
Peptide ViewMS/MS Fragmentation of **AESQSPGTSMTTPSSGSFPSAYD**Found in **KPB1_HUMAN** in **SwissProt**, Phosphorylase b kinase regulatory subunit alpha, skeletal muscle isoform OS=Homo sapiens GN=PHKA1 PE=1 SV=2

Match to Query 75792: 2189.771220 from(438.961520,5+) intensity(65302.0000) rtinseconds(7831.05) scans(MS: 33772 MSMS: 33785) index(32655)

Title: Cmpd 32656, AutoMSn(438.9615), 130.5 min

Data file C:\DAT\IMPACT HD\14-03-20\CLINPROT C8 RCC URINE pool 1_RB1_01_101.d\ProteinAnalysisResults.mgf

Click mouse within plot area to zoom in by factor of two about that point

Or, Plot from to Da Label all possible matches Label matches used for scoring Show Y-axis 

Monoisotopic mass of neutral peptide Mr(calc): 2189.9059

Ions Score: 22 Expect: 0.022

Matches : 5/204 fragment ions using 9 most intense peaks [\(help\)](#)

Figura 49. Schermata di *Mascot* relativa all'identificazione del frammento peptidico di sequenza AESQSPGTSMTTPSSGSFPSAYD associato alla proteina KPB1.

L'identità del peptide corrispondente al segnale m/z 2661 determinato in MALDI-LM è stata ulteriormente confermata dall'analisi MALDI-TOF/TOF (punteggio di Mascot maggiore della soglia di identità = 162 (p -value < 0.5)) (Figura 50).

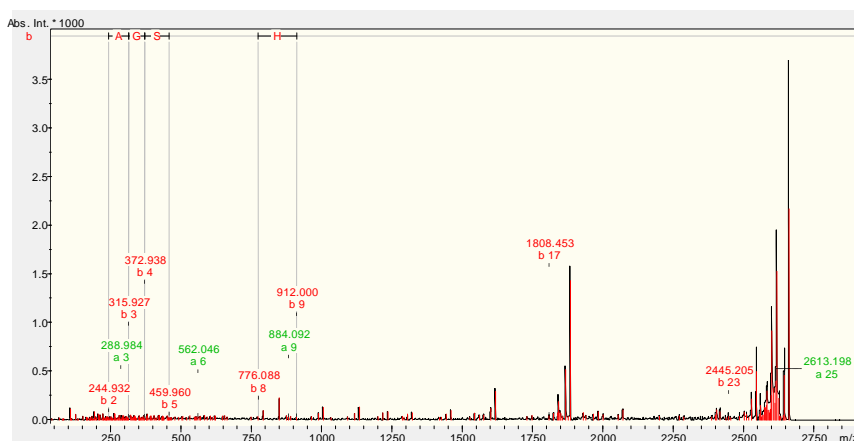


Figura 50. Spettro MS/MS dello ione a m/z 2661 determinato in MALDI-LM.

Il segnale a m/z 2659,44 in MALDI-RM e a m/z 2661 in MALDI-LM è stato identificato come frammento della catena alfa del fibrinogeno (FIBA) con sequenza amminoacidica ${}_{605}\text{DEAGSEADHEGTHSTKRGHAKSRPV}$, così come assegnato anche tramite analisi in nLC-ESI-MS/MS (Figura 51).

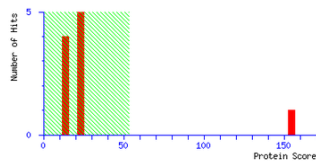
(MATRIX) Mascot Search Results

User : narta
 Email : n.oggiotti4@campus.unimib.it
 Search title : LIFT 2661
 MS data file : DRTA.TXT
 Database : SwissProt 2014_06 (545536 sequences; 194023197 residues)
 Taxonomy : Homo sapiens (human) (20214 sequences)
 Timestamp : 2 Sep 2014 at 15:18:41 GMT
 Protein hits : [FIBA_HUMAN](#) Fibrinogen alpha chain OS=Homo sapiens GN=FGA PE=1 SV=2

	SwissProt	Decoy	False discovery rate
Peptide matches above identity threshold	1	0	0.00 %
Peptide matches above homology or identity threshold	1	0	0.00 %

Mascot Score Histogram

Ions score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event.
 Individual ions scores > 53 indicate peptides with significant homology.
 Individual ions scores > 53 indicate identity or extensive homology ($p < 0.05$).
 Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As	Peptide Summary	Help
Significance threshold p<	0.05	Max. number of hits
Standard scoring	<input checked="" type="radio"/> MudPIT scoring <input type="radio"/> Ions score or expect cut-off	0
Show pop-ups	<input checked="" type="radio"/> Suppress pop-ups <input type="radio"/> Sort unassigned	Decreasing Score
Preferred taxonomy	All entries	Require bold red

Overview Table

Click on column header to jump to entry in results list.
 Move mouse over any indicator to highlight identical peptides.
 Click on an indicator to see details of individual match.
 Use check boxes to select sub-set of queries for new search.

Mouse over:

Hit: **1**
 [2659_3687](#) (1+)

Error tolerant

- [FIBA_HUMAN](#) Mass: 94914 Score: 154 Matches: 1(1) Sequences: 1(1)
 Fibrinogen alpha chain OS=Homo sapiens GN=FGA PE=1 SV=2
 Check to include this hit in error tolerant search or archive report

Query	Observed	Mr (exp)	Mr (calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
1	2659.3687	2658.3614	2658.2495	42.1	0	162	7.4e-013	1	0	A.DEAGSEADHEGTHSTKRGHAKSRPV.R

Search Parameters

Type of search : MS/MS Ion Search
 Enzyme : None
 Mass values : Monoisotopic
 Protein Mass : Unrestricted
 Peptide Mass Tolerance : ± 50 ppm
 Fragment Mass Tolerance : ± 1 Da
 Max Missed Cleavages : 1
 Instrument type : MALDI-TOF-TOF
 Number of queries : 1

Mascot: <http://www.matrixscience.com/>

Figura 51. Schermata di Mascot relativa all'identificazione del frammento peptidico di sequenza DEAGSEADHEGTHSTKRGHAKSRPV associato alla proteina FIBA.

I segnali a m/z 1668,84 (Mr calc. 1667,92) e a m/z 2215,39 (Mr calc. 2214,22) in MALDI-RM sono stati identificati come frammenti dell'uromodulina (UROM) mediante nLC-ESI-MS/MS (Figure 52 e 53).

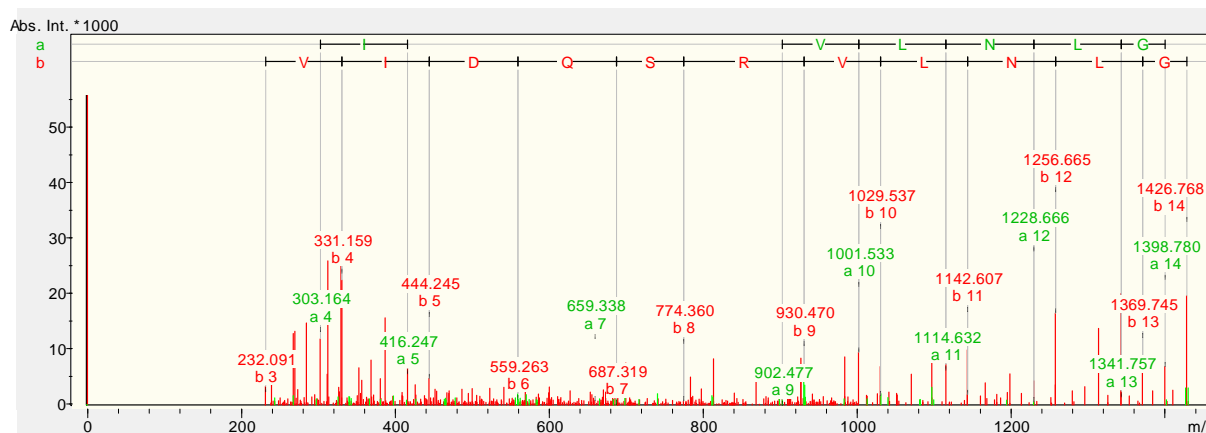


Figura 52. Spettro MS/MS del segnale a m/z 1668,84 identificato come frammento peptidico associato alla proteina UROM.

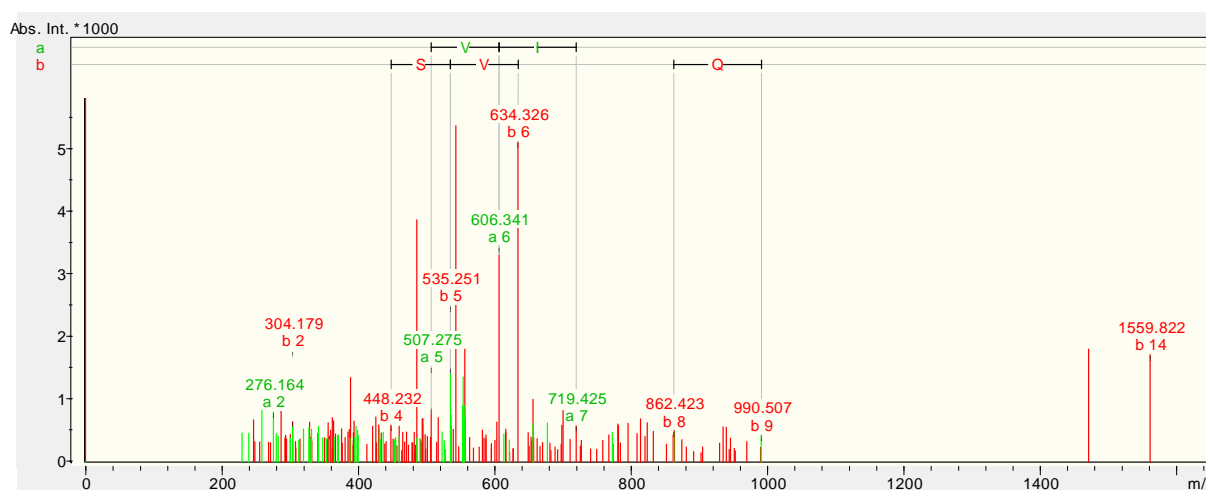


Figura 53. Spettro MS/MS del segnale a m/z 2215,39 identificato come frammento peptidico associato alla proteina UROM.

Inoltre, il segnale a m/z 1115,6 in MALDI-RM (Mr calc. 1114,55) è stato identificato come un frammento di sequenza ${}_{69}\text{RAHGQESAIF}_{78}$ (Mascot score = 58) appartenente alla SCTM1 (*Secreted and transmembrane protein 1*) (Figura 54).

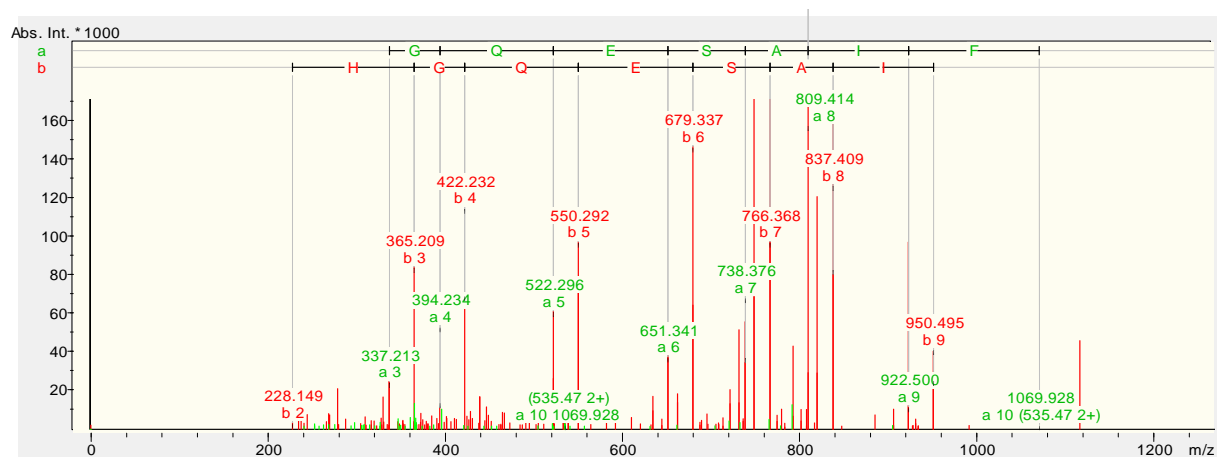


Figura 54. Spettro MS/MS del segnale a m/z 1115,6 identificato come frammento peptidico di sequenza RAHGQESAIF associato alla proteina SCTM1.

Altri due ioni a m/z 1725,92 (M_r calc. 1724,87) e a m/z 2527,26 (M_r calc. 2526,22) sono stati identificati rispettivamente come peptidi di sequenza ${}_{636}\text{EEALPVLSLQGGQPSRQ}_{651}$ (Mascot *score* = 88) e ${}_{19}\text{VKQADSGSSEEKQLYNKYPDAVA}_{41}$ (Mascot *score* = 55); il primo peptide è stato attribuito alla meprina A (MEP1A), mentre il secondo alla proteina osteopontina (OSTP) (Figure 55 e 56, rispettivamente).

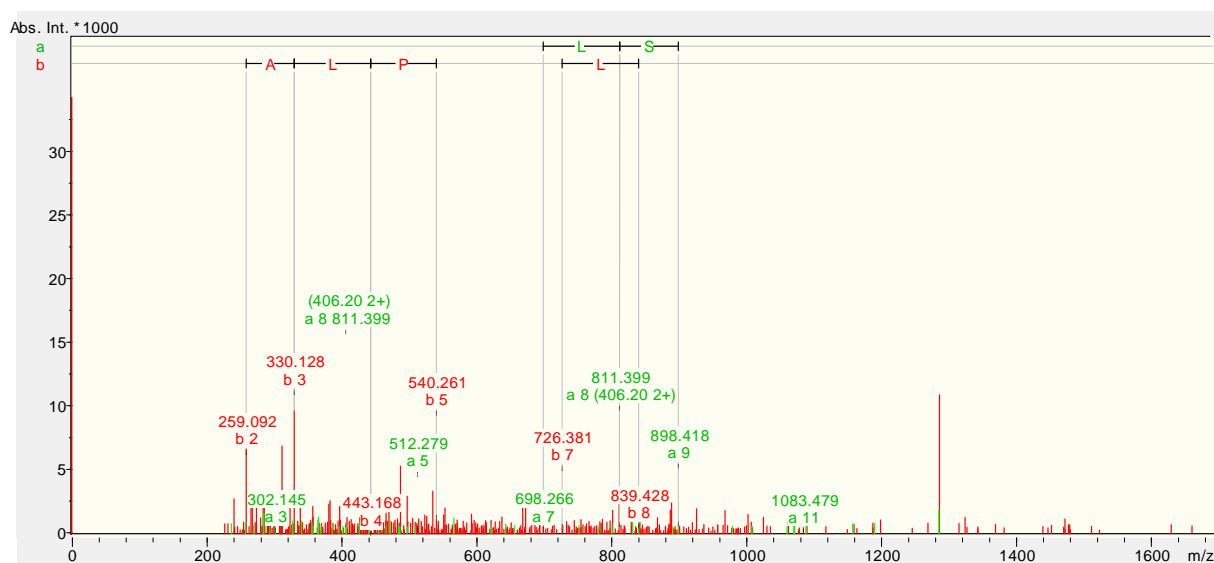


Figura 55. Spettro MS/MS del segnale a m/z 1725,92 identificato come frammento peptidico di sequenza EEALPVLSLQGGQPSRQ associato alla proteina MEP1A.

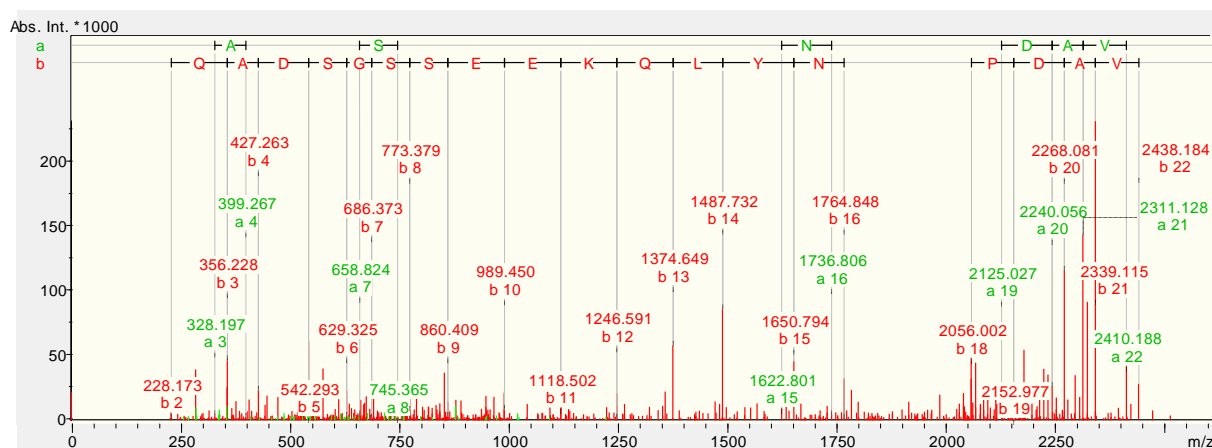


Figura 56 Spettro MS/MS del segnale a m/z 2527,26 identificato come frammento peptidico di sequenza VKQADSGSSEEKQLYNKYDPAVA associato alla proteina OSTP.

4.2.3.2. IDENTIFICAZIONE DEI PEPTIDI CORRELATI ALLA PROGRESSIONE TUMORALE

Per quanti riguarda i peptidi endogeni variati in maniera statisticamente significativa rispetto ai dati clinici di progressione tumorale, l'analisi nLC-MS/MS ha permesso l'identificazione di quindici di essi (Tabella 25). Sei ioni presenti nei profili peptidici acquisiti in modalità lineare utilizzando lo spettrometro di massa MALDI-TOF e non presenti nei profili acquisiti in modalità *reflector* sono stati assegnati a sequenze amminoacidiche identificate in ESI-MS/MS (errore medio associati alle misurazioni delle masse molecolari in modalità lineare: 222 ± 411.1 ppm): lo ione a m/z 3571.1 è stato riconosciuto come il frammento $_{362}\text{SQPPPLPPPPPAQETQEPSPILDSEETRAAAPQ}_{395}$ della proteina DREB; lo ione a m/z 3723.8 come il frammento $_{51}\text{STHGKFGHTVKAENGLVINGNPITIFQERDPSK}_{84}$ originato dalla proteina G3P; lo ione a m/z 4355.1 come la forma acetilata del peptide $_{386}\text{EEKDIKPIIKDEKGRVGSGRNLWVSGLSST}^{\text{T}}\text{TRATDLK}_{424}$ appartenente alla proteina SAFB2; lo ione a m/z 4626.9 come il frammento $_{868}\text{HEVKTIDMRFRHCQEARISPMHILNAKELVLNINKLE}_{906}$ della proteina CC168. I restanti nove segnali sono stati rilevati anche nei profili acquisiti in modalità *reflector*, rafforzando pertanto l'accuratezza del processo identificativo (errore medio associato alle misurazioni delle masse molecolari in modalità *reflector*: 36.7 ± 81.3 ppm). In particolare, lo ione a m/z 1755.8 è stato identificato come un peptide derivante dalla proteina A1AGx, che è comune ad entrambe le isoforme, A1AG1 e A1AG2. Lo ione a m/z 1934.2 potrebbe derivare invece da due peptidi differenti, per i quali non è stato possibile assegnare un'identità in maniera inequivocabile attraverso l'osservazione del cluster isotopico degli spettri acquisiti in modalità *reflector*: il frammento $_{1152}\text{ERLEEAGGATSAQIEMNK}_{1169}$ derivante dalla proteina MYH1_HUMAN o MYH4_HUMAN e/o dal frammento $_{40}\text{LGAPAVPAWSAAQGDVAALGL}_{60}$ originato dalla proteina MMP23. In maniera del tutto simile, lo ione a m/z 1825.6 potrebbe derivare dalla sequenza amminoacidica $_{59}\text{FRYNSKDRKSQPMGL}_{73}$ della proteina ZA2G_HUMAN e/o dal peptide $_{4283}\text{VSEDPINDGEWHRVTA}_{4298}$ di PGBM_HUMAN. Per due dei

tre dici segnali (m/z 2192.1 e m/z 3151.1) non è stato possibile attribuire con certezza ad una sola sequenza amminoacidica. In effetti, la ridotta differenza di massa molecolare tra i diversi peptidi identificati e il potere risolutivo del MALDI operante in modalità *reflector* non hanno permesso di discriminare lo specifico contributo di ciascuna sequenza al segnale osservato in MALDI-LM. Tre segnali sono stati assegnati a sequenze già identificate e presenti nei clusters diagnostici.

MALDI		nLC ESI MS/MS									
LM m/z	RM m/z	Mr monoisotopica teorica + [H+]	Mr sperimentale + [H+]	Carica	Uniprot ID	Gene Name	start	Sequenza	stop	PTM	Pep score
1024	1023,40	1023,562	1023,553	2	C1RL_HUMAN	C1RL	79	VRLVFQDF	86		53
1525		1525,689	1525,695	2	PTGDS_HUMAN	PTGDS, PDS	107	YRSPHWGSTYSVS	119		77
1756	1755,90	1755,843	1755,839	2	A1AG1_HUMAN	ORM1, AGP	145	YADKPETTKEQLGEF	159		80
		1755,843	1755,839	2	A1AG2_HUMAN	ORM2, AGP2	145	YADKPETTKEQLGEF	159		80
1826	1827,08	1826,933	1826,928	2	ZA2G_HUMAN	AZGP1, ZAG, ZNGP1	59	FRYNSKDRKSQPMGL	73		48
	1825,09	1824,851	1824,846	2	PGBM_HUMAN	HSPG2	4283	VSEDPINDGEWHRVTA	4298		94
1912	1912,15	1912,061	1912,048	3	UROD_HUMAN	UMOD	589	SGSVIDQSRVNLGPITR	606		83
1934	1935,10	1935,033	1935,027	2	MMP23_HUMAN	MMP23A, MMP21, MMP23B, MMP22	40	LGAPAVPAWSAAQGDVAALGL	60		27
	1934,10	1933,928	1933,918	3	MYH1_HUMAN	MYH1	1152	ERLEEAGGATSAQIEMNK	1169		27
		1933,928	1933,918	3	MYH4_HUMAN	MYH4	1152	ERLEEAGGATSAQIEMNK	1169		27
2192	2190,90	2190,771	2190,765	4	GP162_HUMAN	GPR162, GRCA	344	IMSEEDGDDGGCDDYAEGR	363	N-Acetilazione	26
		2190,913	2190,778	5	KPB1_HUMAN	PHKA1, PHKA	1021	AESQSPGTSMTSPSSGFSAYD	1042		22
2661	2659,44	2659,257	2659,252	5	FIBA_HUMAN	FGA	605	DEAGSEADHEGTHSTKRGHAKSRPV	629		39
2791		2791,481	2791,491	4	HBA_HUMAN	HBA1, HBA2	111	AAHLPAEFTPAVHASLDKFLASVSTVL	137		25
3151	3149,55	3149,245	3149,268	4	ADA19_HUMAN	ADAM19	330	MDHSENAIGVAATMAHEMGNFMTHDSA	358	Ossidazione (HW), Ossidazione (M)	17
		3149,276	3149,278	4	RSPH3_HUMAN	RSPH3, RSHL2, RSP3	472	EHGEDTHQSPPEPEDEPGGPGAMTESLEASE	501		17
3571		3569,775	3569,761	3	DREB_HUMAN	DBN1, D0S117E	362	SQPPPLPPPPPAQETQEPSPILDSEETRAAPQ	395		63
3724		3719,961	3719,932	3	G3P_HUMAN	GAPDH, GAPD	51	STHGKFGHTVKAENGLVINGNPTIFQERDPSK	84		24
4027	4027,07	4027,598	4027,580	4	NOTC2_HUMAN	NOTCH2	257	CERNIDDCPNHRCQNGGVCVGVNTYNCRCPPQWTG	292	Deamidazione (NQ)	21
4355		4354,300	4354,288	5	SAFB2_HUMAN	SAFB2, KIAA0138	386	EKDIKPIKDEKGRVSGSGRNLWVSLSSITRATDLK	424	Acetilazione (K)	25
4627		4626,458	4626,880	3	CC168_HUMAN	CCDC168, C13orf40	868	HEVKTIDMRFRHCQEARISPMSHLNKELVLNINKLE	906		22

Tabella 25. Identificazione dei segnali MALDI variati secondo la dimensione tumorale, il pT o il *Grade*. LM=*linear mode*, RM=*reflector mode*, PTM=possibili modificazioni post-traduzionali.

4.3. ANALISI DEL PROTEOMA URINARIO

L'analisi nLC-ESI-MS/MS dei pools di urine di pazienti e soggetti controllo digerite con protocollo FASP ha permesso di ottenere profili cromatografici paragonabili per intensità, numerosità di segnali e per tempo di ritenzione (shift massimo <3%), confermando quindi la riproducibilità del metodo (Figura 57).

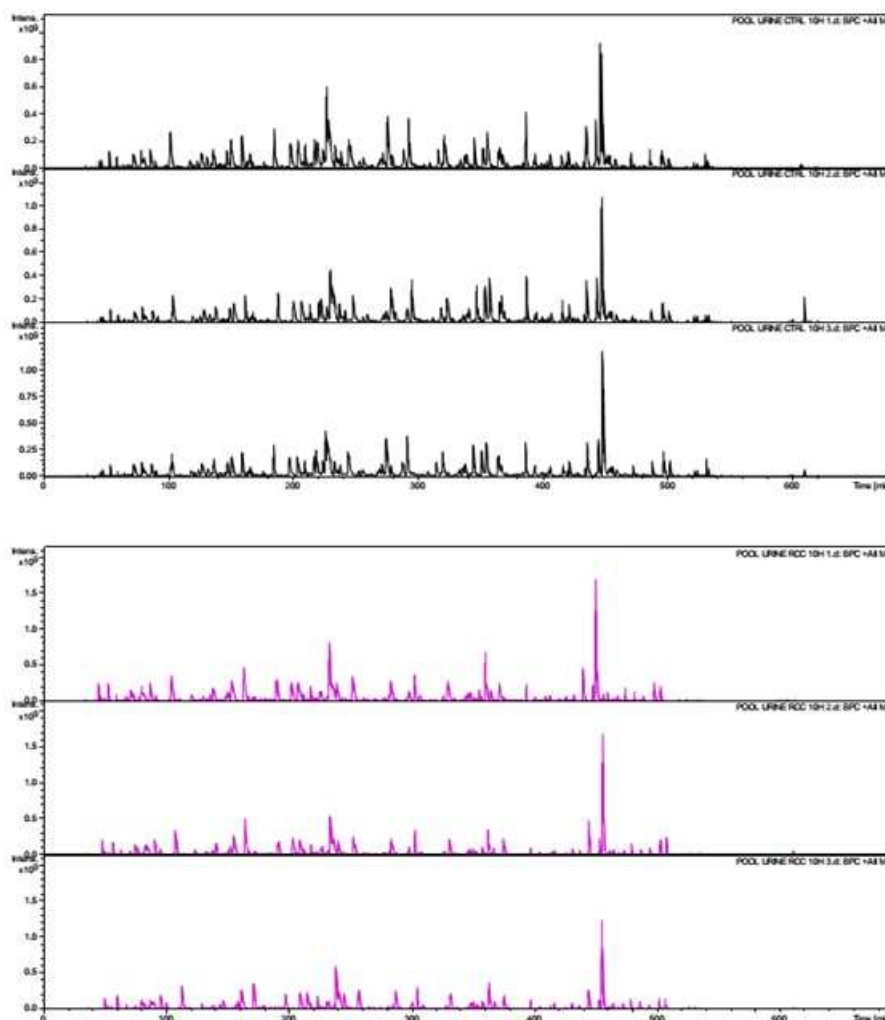


Figura 57. Comatogrammi BPC (*Base Peak Chromatogram*) derivanti dalle analisi nLC-MS/MS dei campioni di urine digerite dei soggetti controllo (in nero) e dei pazienti affetti da ccRCC (in rosa).

4.3.1. INDIVIDUAZIONE DI ALTERAZIONI PROTEOMICHE

L'elaborazione mediante il software DataAnalysis e la successiva ricerca in banca dati ha portato all'identificazione di 459 proteine nei pools di urine di soggetti controllo e pazienti affetti da ccRCC. La ricerca delle alterazioni del proteoma tra le due classi di soggetti è stata effettuata mediante il *software* Progenesis LC-MS™ considerando variata l'espressione proteica quando il rapporto tra le abbondanze relative normalizzate dei segnali appartenenti ad una specifica proteina è aumentato del 50% nei pazienti ccRCC (ratio>1.5) o nei controlli (ratio< 0.67) [213, 285].

In particolare, nei campioni di urine sono state identificate e quantificate 518 proteine di cui 309 sono risultate alterate tra controlli e ccRCC.

In prima analisi, le proteine urinarie sovra- (n=127) o sotto-espresse (n=182) nei pazienti sono state analizzate da un punto di vista funzionale in modo da chiarire quali processi risultassero arricchiti nella patogenesi tumorale. Il sistema di classificazione PANTHER ha permesso di chiarire in particolare il loro coinvolgimento nei processi biologici (Figura 58) e di determinarne la classificazione in base alla funzione molecolare svolta (Figura 59).

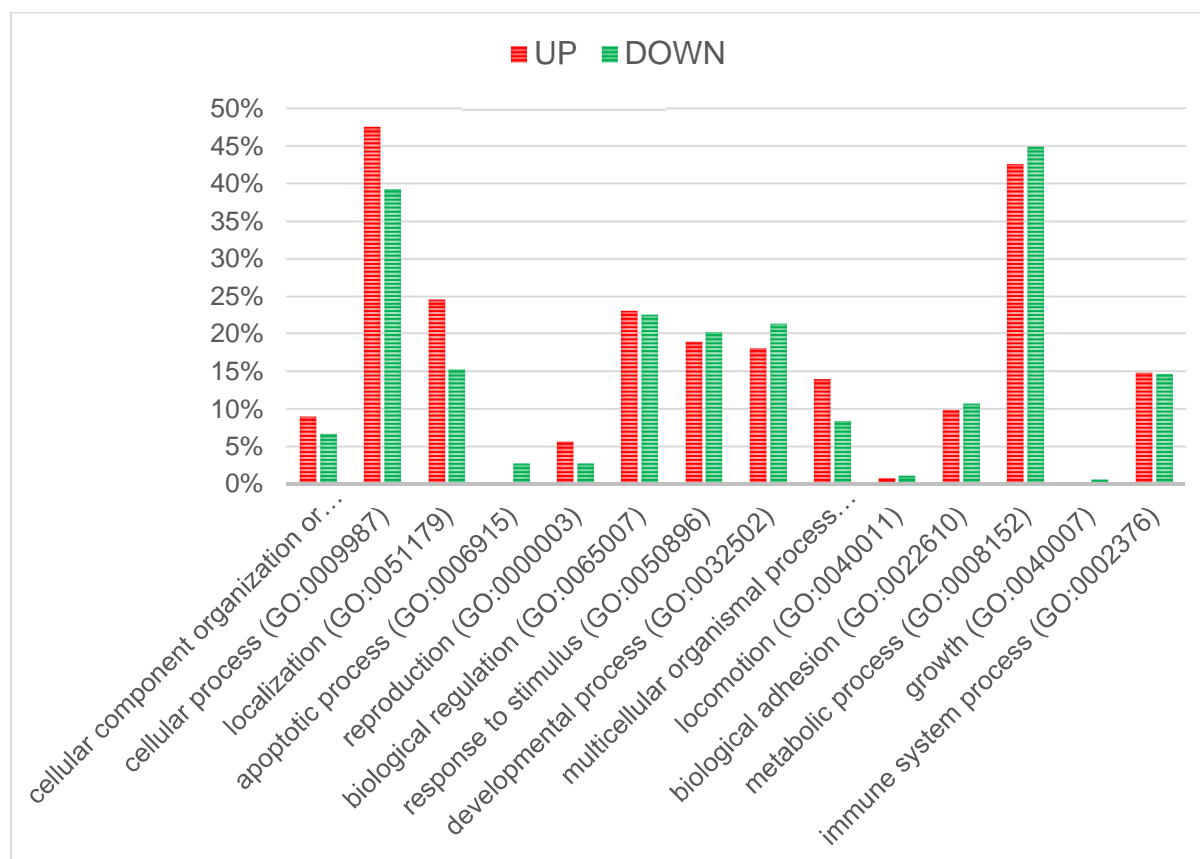


Figura 58. Grafico dei processi biologici in cui risultano coinvolte le proteine sovra (in rosso) o sotto (in verde) rappresentate nelle urine dei soggetti ccRCC; le percentuali indicate sull'asse delle ordinate si riferiscono al rapporto fra il numero di proteine alterate appartenenti ad una specifica classe di processi biologici, e il totale delle proteine alterate.

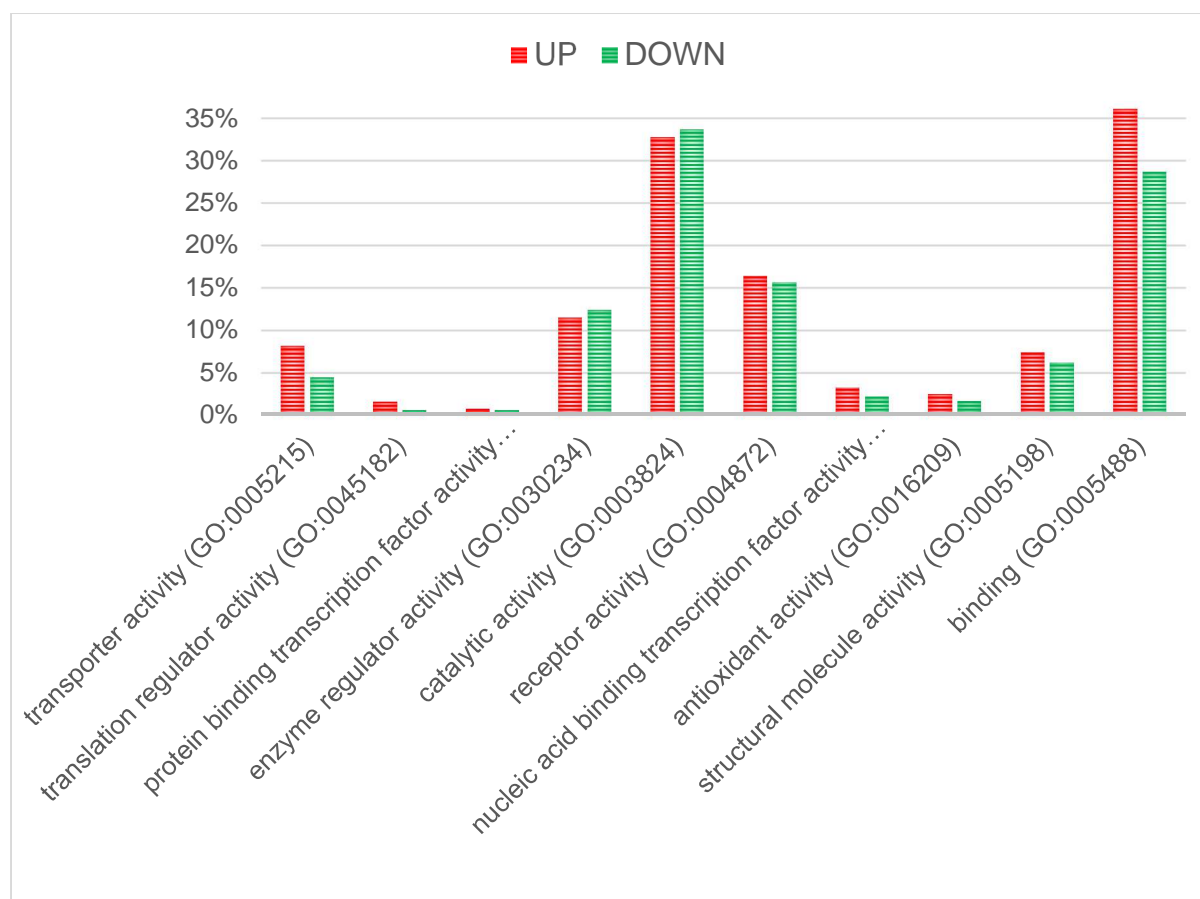


Figura 59. Grafico delle funzioni molecolari in cui risultano coinvolte le proteine sovra (in rosso) o sotto (in verde) rappresentate nelle urine dei soggetti ccRCC; le percentuali indicate sull'asse delle ordinate si riferiscono al rapporto fra il numero di proteine alterate appartenenti ad una specifica classe di processi biologici, e il totale delle proteine alterate.

È stato quindi possibile osservare che le principali alterazioni nelle urine risultano essere a carico di proteine coinvolte in processi metabolici e cellulari, in processi regolatori, di risposta a stimoli e di sviluppo. Per quanto riguarda invece l'analisi dal punto di vista funzionale, le funzioni molecolari a cui afferiscono le proteine alterate in maniera significativa nei pazienti affetti da ccRCC riguardano principalmente l'attività catalitica, di binding e le attività recettoriali e di regolazione dell'attività enzimatica.

4.4. INTEGRAZIONE DELLE INFORMAZIONI PEPTIDOMICHE E PROTEOMICHE

Al fine di valutare le firme molecolari tipiche del tumore renale, sono state inizialmente selezionate le 21 proteine identificate i cui peptidi erano presenti nei cluster diagnostici e/o presentavano variazioni statisticamente significative in relazione ai dati clinici del tumore (dimensione, pT, grado). Un accurato allineamento delle masse tra il valore di massa calcolato sulla sequenza amminoacidica identificata in ESI e il valore di massa [M-H⁺] dello spettro MALDI ha permesso di associare a ciascuna proteina *target* i rispettivi valori di m/z dei peptidi endogeni presenti nei profili urinari. Grazie all'analisi statistica condotta sui dati di peptidomica è stato inoltre possibile determinare quali dei peptidi *matchati* mostravano una variazione statisticamente significativa tra pazienti affetti da ccRCC e volontari sani (Tabella 26).

UNIPROT ID	LM m/z	UNIPROT ID	LM m/z
A1AG1_HUMAN	1756	OSTP_HUMAN	2529
A1AG2_HUMAN	1756		2961
ADA19_HUMAN	1211	PGBM_HUMAN	1134
C1RL_HUMAN	3152		1826
CC168_HUMAN	1024		2529
DREB_HUMAN	4623		2755
	1332	PTGDS_HUMAN	1525
2916	1951		
FIBA_HUMAN	3568		3252
	1883	RSPH3_HUMAN	3152
	2661	SAFB2_HUMAN	4355
	3268	SCTM1_HUMAN	1117
4439	UROM_HUMAN	1226	
G3P_HUMAN		3724	1229
GP162_HUMAN		2193	1310
HBA_HUMAN		1219	1426
		1426	1511
		1583	1583
		1769	1613
		2698	1655
		2727	1670
2857		1682	
3329		1896	
3427	1913		
KPB1_HUMAN	2193	2042	
MEP1A_HUMAN	1727	2217	
MMP23_HUMAN	1935	1826	
NOTC2_HUMAN	4028	2237	
		2714	

Tabella 26. Associazione tra le proteine target presenti nei cluster diagnostici o i cui peptidi endogeni risultavano correlati con la progressione tumorale e i rispettivi peptidi presenti nei profili urinari. I peptidi sopra rappresentati nelle urine di pazienti ccRCC

w CTRL sono indicati in rosso, quelli sotto rappresentati in verde e quelli invariati in giallo.

Inoltre, sono riscontrabili diversi livelli di coerenza a seguito dell'integrazione dei dati peptidomici: in alcuni casi i segnali dei peptidi appartenenti alla stessa proteina mostrano un trend coerente (la maggior parte sovra- o sotto-rappresentati), in altri casi questo livello di coerenza viene meno, a causa probabilmente di fattori biologici.

I dati peptidomici così ottenuti sono stati quindi *matchati* con i risultati ottenuti dalle analisi del proteoma urinario relativi alle proteine *target*. È stato quindi possibile definire l'andamento dell'espressione per 12 delle 22 proteine di interesse (Tabella 27).

	Peptidoma	Proteoma
UNIPROT ID	LM m/z	Espressione nei campioni di urina
A1AG1_HUMAN	1756	
A1AG2_HUMAN	1756	
	1211	
ADA19_HUMAN	3152	N.D.
C1RL_HUMAN	1023	
CC168_HUMAN	4623	
DREB_HUMAN	1332	N.D.
	2916	
	3668	
FIBA_HUMAN	1883	
	2661	
	3268	
	4439	
G3P_HUMAN	3724	
GP162_HUMAN	2193	N.D.
HBA_HUMAN	1219	
	1426	
	1583	
	1769	
	2698	
	2727	
	2867	
	3329	
	3427	
KPB1_HUMAN	2193	
MEP1A_HUMAN	1727	N.D.
MMP23_HUMAN	1935	N.D.
NOTC2_HUMAN	4028	N.D.
	2529	
OSTP_HUMAN	2961	
PGBM_HUMAN	1134	
	1826	
	2529	
	2755	
PTGDS_HUMAN	1625	
	1951	
	3252	
RSPH3_HUMAN	3152	N.D.
SAFB2_HUMAN	4355	N.D.
SCTM1_HUMAN	1117	N.D.
UROM_HUMAN	1226	
	1229	
	1310	
	1426	
	1511	
	1583	
	1613	
	1655	
	1670	
	1682	
	1896	
	1913	
	2042	
	2217	
ZA2G_HUMAN	1826	
	2237	
	2714	

Tabella 27. Integrazione dei dati ottenuti dalle analisi peptidomiche e proteomiche. I peptidi/proteine over rappresentati nelle urine di pazienti ccRCC vs CTRL sono indicati in rosso, quelli sotto rappresentati in verde e quelli invariati in giallo.

Anche in questo caso si possono notare delle incongruenze, in quanto alcune proteine presentano lo stesso *trend* di espressione dei rispettivi peptidi endogeni, mentre in altri casi il loro livello di espressione risulta essere opposto.

4.5. VALIDAZIONE DEI RISULTATI OTTENUTI MEDIANTE WESTERN BLOT

L'attenzione è stata quindi focalizzata sulle proteine per le quali vi fosse un trend di espressione coerente all'interno dei peptidi endogeni e la cui alterazione potesse avere un significato biologico nell'ambito del tumore renale a cellule chiare, in particolare sono state considerate le proteine A1AG1/A1AG2, PTGDS e PGBM. Per due di esse i dati di abbondanza relativi alle proteine intatte presentano trend non coerenti rispetto ai livelli di espressione dei peptidi endogeni. Per una delle proteine di interesse, la Prostaglandina-H2 D-isomerasi (PTGDS_HUMAN), è stato possibile valutarne il contenuto mediante immunodecorazione nelle urine di pazienti ccRCC e controlli al fine di confermare i dati ottenuti a livello proteomico e chiarire la mancata coerenza tra le analisi peptidomiche e proteomiche.

Come si può osservare dalla Figura 60 (blocco A), è stato possibile rilevare una banda a circa 30kDa che mostra come PTGDS risulti essere maggiormente presente nei campioni di urina dei soggetti affetti da tumore renale rispetto ai volontari sani. Tramite analisi statistica, effettuata con normalizzazione rispetto alla creatinina urinaria (uCr), è stata confermata la presenza di una differenza statisticamente significativa (Figura 60 – Blocco B).

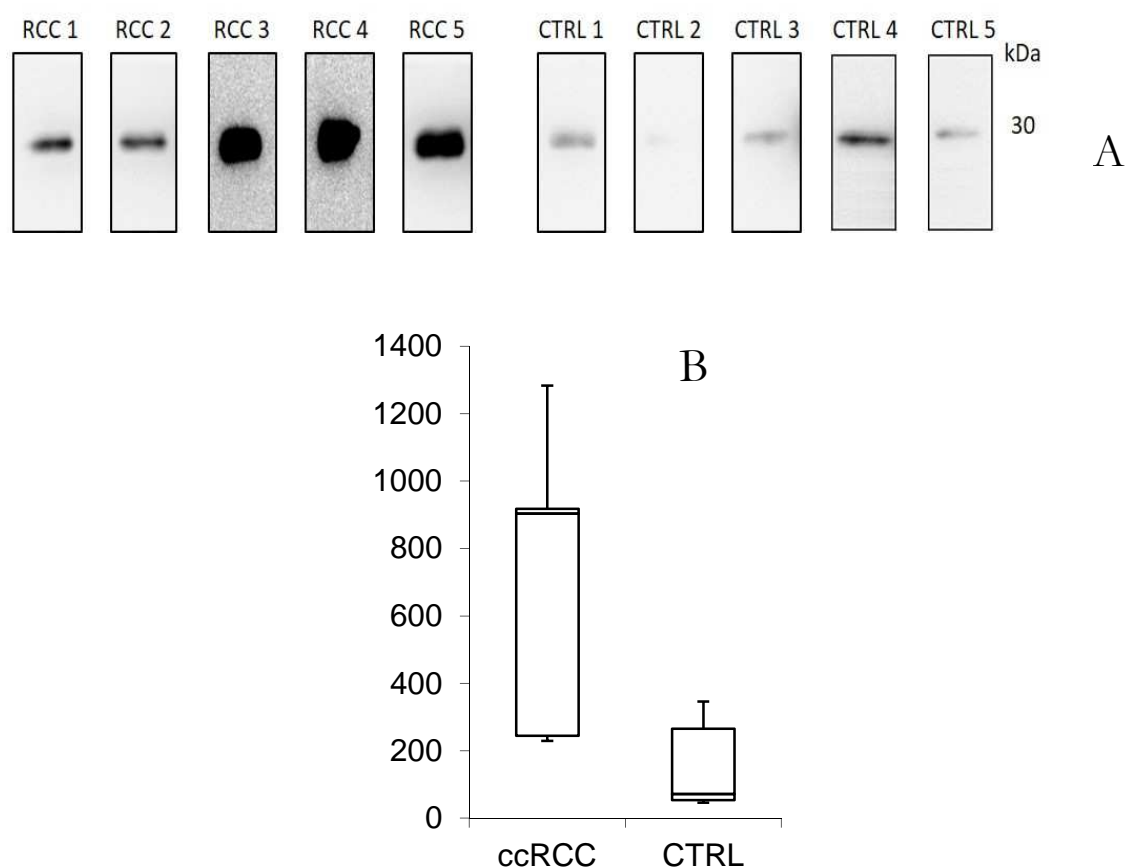


Figura 60. Contenuto di PTGDS nei campioni di urina di pazienti affetti da ccRCC e volontari sani. A) Immunodecorazione con anticorpo anti-PTGDS B) Analisi densitometrica delle bande di PTGDS nelle urine di controlli e pazienti normalizzate rispetto alla concentrazione di creatinina urinaria (uCr).

5. DISCUSSIONE

La ricerca di biomarcatori capaci di caratterizzare e predire una malattia multifattoriale come il cancro è un importante obiettivo delle scienze cosiddette “-omiche”, tra cui la proteomica e la peptidomica.

I tumori renali maligni rappresentano il 4% di tutti i cancri nell'uomo e il carcinoma a cellule renali chiare costituisce circa il 60% di tutti i tumori renali [39]. Il carcinoma renale è un tumore frequente nei paesi occidentali ed ogni anno in Europa vengono diagnosticati circa 86.000 nuovi casi [21]. Inoltre, a differenza di altri tumori urologici, il carcinoma renale ha un maggior potenziale metastatico ed è sia chemio che radio-resistente. Tuttora, l'escissione chirurgica appare l'unica vera terapia radicale per il carcinoma renale [286]. Dunque, una diagnosi precoce e una migliore conoscenza dei meccanismi molecolari alla base dell'insorgenza e della progressione della malattia risultano essere obiettivi molto importanti, e potrebbero consentire un significativo miglioramento nel trattamento di questo tipo di tumore. Negli ultimi decenni la progressiva diffusione della diagnostica per immagini ha determinato l'aumento delle diagnosi precoci di lesioni renali asintomatiche; tuttavia in alcuni casi può essere difficile distinguere una lesione benigna da una maligna: sono state infatti osservate parecchie cisti di morfologia ambigua, così come masse solide renali di aspetto dubbio. È inoltre possibile che all'interno di una diagnosi di RCC, il singolo tumore presenti delle caratteristiche molecolari eterogenee. Pertanto, alla luce della sua estesa variabilità fenotipica, una migliore caratterizzazione avrebbe delle potenziali ricadute positive per lo *screening*, la diagnosi precoce e la prognosi. Numerosi sforzi sono stati compiuti nel campo della ricerca al fine di sviluppare nuove metodiche adeguate per una diagnosi precoce del RCC o per un suo monitoraggio continuo. Tuttavia, ad oggi, non sono stati individuati biomarcatori specifici utilizzabili nella clinica che permettano, attraverso un'analisi poco invasiva come quella dei fluidi biologici, il riconoscimento di questa patologia né in stadio precoce né già avanzato.

Attualmente, la spettrometria di massa è sempre più utilizzata come tecnologia d'elezione nel campo della proteomica per l'identificazione di proteine e peptidi differentemente espressi in condizioni sia fisiologiche che patologiche. Infatti, tali *target* molecolari sono molto utili nella ricerca clinica perché possono essere utilizzati come potenziali biomarcatori a fini diagnostici, prognostici e di *screening* [287]. Ad oggi, migliaia di proteine e peptidi presenti nelle urine umane sono stati sequenziati, fornendo in tal modo indicazioni sempre più complete sul contenuto endogeno urinario in condizioni fisiologiche ed evidenziando nel contempo la necessità di ulteriori studi sulla controparte derivante da soggetti affetti da patologie al fine di trasferire tali informazioni nella pratica clinica. In letteratura, numerosi studi di *biomarker discovery* su campioni urinari sono già stati riportati per il cancro della prostata [288], la nefropatia diabetica [148], l'insufficienza renale cronica [289] e anche per il carcinoma a cellule renali [147, 175, 177]. La maggior parte di essi ha come fine ultimo quello di individuare marcatori per la diagnosi precoce, la prognosi e la predizione della risposta terapeutica dei pazienti. In particolare, campioni di urina prelevati da pazienti affetti da carcinoma a cellule renali sono stati comparati, in diversi studi condotti mediante *Western-blot*, a campioni ottenuti da soggetti sani ed hanno evidenziato la presenza di alterazioni nella concentrazione

urinaria di diverse proteine che potrebbero avere una potenziale capacità prognostica e diagnostica per questo tumore. A tal proposito, la proteina di matrice nucleare 22 (NMP22) è stata trovata overrappresentata nelle urine di 23 dei 35 pazienti affetti da RCC quando comparati a pazienti con calcoli e cisti renali, utilizzati come popolazione di controllo [290]. Anche la proteina 14-3-3 β/α è risultata essere maggiormente concentrata nelle urine di pazienti affetti da RCC rispetto a campioni prelevati da volontari sani [176], mostrando una capacità diagnostica espressa con un valore di AUC pari a 0.88. Sono stati inoltre descritti aumenti dei livelli urinari di altre due proteine, l'aquaporina 1 (AQP1) e la perilipina 2 (PLIN2), in 63 pazienti affetti da RCC confrontati con 43 soggetti sani [183]; queste due proteine hanno mostrato valori di sensibilità e specificità compresi in un *range* da 90% a 100%. Inoltre, 2-3 settimane dopo la rimozione chirurgica del tumore la concentrazione urinaria di queste due proteine ritornava a livelli simili a quelli trovati nei soggetti controllo, fornendo un'ulteriore prova della loro specificità nell'individuare la presenza di carcinoma a cellule renali. È stato anche dimostrato dagli autori del lavoro che i livelli urinari di AQP1 e PLIN2 sono correlati con la dimensione e lo stadio tumorale ma non con il *grade* [183]. La proteina KIM-1 (*Kidney Injury Molecule-1*), biomarcatore potenzialmente utile nel rilevare la presenza di un danno di natura ischemica o tossica a carico delle cellule epiteliali renali del tubulo prossimale nell'uomo, è stata trovata in maggiore concentrazione nelle urine di pazienti affetti da RCC [287]. La comparazione di tessuti derivanti da pazienti con RCC ha permesso l'osservazione di molte proteine correlate al pT [291]. Recentemente è stata inoltre riportata un'associazione significativa tra l'espressione tissutale di TGFBI e il tumore renale a diversi stadi e con il diametro della massa [292].

Tuttavia occorre notare che nessuno dei risultati appena descritti è stato validato su una casistica indipendente di pazienti né è stato in grado di predire l'aggressività del RCC e, ad oggi, la dimensione e il tasso di crescita tumorale risultano essere i fattori prognostici più utilizzati.

L'interesse scientifico ultimamente si è rivolto verso l'utilizzo di tecniche di *protein profiling* altamente sensibili e versatili, basate su processi cromatografici di frazionamento prima dell'analisi in spettrometria di massa, e che meglio rispettano i requisiti richiesti in ambito clinico, ossia rapidità d'analisi e *high-throughput* [293]. Attualmente, è possibile ottenere profili proteici di fluidi biologici da controlli e pazienti, mediante i quali è possibile selezionare *patterns* di segnali caratterizzanti uno specifico stato patologico [294]. In diverse patologie umane, comprese anche le malattie renali [147], è stata ampiamente descritta l'identificazione di un profilo peptidico discriminante, mediante *profiling* MALDI-TOF di urine pre-frazionate prima dell'analisi in spettrometria di massa con biglie dotate di superfici arricchite [295], oppure attraverso particelle di silice mesoporosa per l'arricchimento selettivo di proteine a basso peso molecolare [296, 297], oppure con biglie magnetiche aventi superfici variamente funzionalizzate [164]. Numerosi lavori finalizzati allo studio dei profili proteici urinari sono stati realizzati anche mediante la tecnologia SELDI allo scopo di rilevare *clusters* diagnostici del carcinoma a cellule renali. Rogers *et al.*, in particolare, hanno descritto un modello algoritmico avente sensibilità e specificità molto elevate tra il 98.3% e il 100%, valori che tuttavia crollavano al 41-77% quando il modello veniva verificato su una casistica indipendente di dati [189]. Successivamente anche Wu *et al.* hanno identificato un *cluster* di quattro peptidi differentemente rappresentati tra pazienti affetti da RCC

e controlli [190]. Tale modello mostrava una sensibilità e specificità dell'80% e dell'89.6%, rispettivamente, nella fase di *training*, e del 67.8% e 81.5%, rispettivamente [190], nella fase di validazione. In nessuno di questi studi è stato possibile definire l'identità dei biomarcatori inclusi nei diversi modelli diagnostici. Solo recentemente Alves *et al.* sono stati in grado sia di discriminare mediante un approccio SELDI-TOF i soggetti controllo dai pazienti affetti da RCC, sia di fornire un'identità mediante la tecnologia LC-MS abbinata alla separazione SDS-PAGE di alcuni di questi segnali proteici, senza però fornire in questo caso alcuna informazione sulla capacità diagnostica dei risultati [191].

Lo scopo del presente lavoro di tesi è stato dunque quello di ricercare potenziali firme molecolari tipiche del tumore renale attraverso l'applicazione di un approccio multi-omico e l'integrazione dei dati ottenuti al fine di far luce su potenziali target molecolari alla base della disregolazione indotta dalla neoplasia.

A questo fine, sono stati analizzati un ampio numero di pazienti affetti da ccRCC ($n = 118$) e soggetti controllo ($n = 137$) estendendo l'analisi anche a 35 pazienti affetti da tumori non-ccRCC. In questo studio la ricerca è stata focalizzata sia sul peptidoma urinario endogeno, ottenuto dal pre-frazionamento dei campioni con biglie magnetiche a fase inversa (RP-C8) seguito dall'analisi MALDI-TOF [147], sia sul proteoma urinario, ed analizzato quali- e quantitativamente mediante nLC-ESI MS/MS. Le proteine e i peptidi contenuti nei liquidi biologici riflettono lo stato fisiologico o patologico di un organismo e possono infatti rappresentare una valida fonte di potenziali biomarcatori [298]. In particolare l'urina è anche un importante fluido biologico che può essere ottenuto in maniera semplice con metodi per nulla o poco invasivi, e che a differenza del siero risente meno delle difficoltà legate all'ampiezza del *range* dinamico di concentrazione proteica [273] mostrandosi sufficientemente stabile se raccolta con protocolli standardizzati [170]. Inoltre, la matrice urinaria risulta fortemente influenzata dai cambiamenti indotti dagli organi coinvolti nell'escrezione quali ad esempio il rene: questo aspetto è degno di nota se si considerano i benefici relativi alla possibilità di effettuare una diagnosi precoce, di seguire l'andamento di tumori localizzati o di monitorare i risultati dei trattamenti terapeutici semplicemente analizzando marcatori tumore-specifici in un fluido facilmente accessibile come il prodotto di escrezione del rene [178, 299].

Secondo quanto riportato in letteratura da Zurbig *et al.* [300], la concentrazione urinaria di alcuni peptidi correla con l'aumentare dell'età, e quindi con la conseguente ridotta funzionalità renale. L'invecchiamento induce cambiamenti morfologici del rene e riduce la sua funzionalità. La malattia renale è spesso aggravata negli anziani [301, 302] e il declino della funzionalità renale con l'età è spesso senza significato clinico [303]. La velocità di filtrazione glomerulare diminuisce del 20-25% da 40 a 80 anni di età. Inoltre, la capacità del rene di concentrare l'urina diminuisce progressivamente [304]. Le principali osservazioni istologiche che accompagnano la caduta del tasso di filtrazione glomerulare nell'invecchiamento della popolazione sono la sclerosi glomerulare e aumenti di tessuto interstiziale [303, 305, 306] ovvero l'espansione della matrice extracellulare (ECM). In condizioni fisiologiche, l'ECM è mantenuta da un equilibrio di sintesi e degradazione che viene modificato durante la malattia cronica renale [307]. È importante comprendere e rilevare cambiamenti nella struttura renale e nella sua funzione che si verificano con l'aumentare dell'età, e distinguere queste modifiche da alterazioni che avvengono con la malattia e se queste correlano con l'età.

In questo studio, la casistica dei pazienti, a nostra disposizione, mostrava una media di età (64.53 anni) decisamente superiore rispetto alla media di età dei soggetti controllo (48.7 anni). Pertanto è stato verificato che i livelli urinari non variassero come conseguenza dell'età e che quindi la loro abbondanza fosse un processo indipendente rispetto all'invecchiamento. Per verificare questo, è stato applicato un test statistico (*test di Spearman*) al fine di individuare gli ioni segnale correlati con l'età nei soggetti controllo, i quali sono stati esclusi dalla successiva elaborazione in modo tale da considerare solo quei peptidi la cui variazione di espressione fosse un processo indipendente rispetto all'invecchiamento e specifico per la presenza della massa tumorale.

L'analisi statistica si è poi focalizzata sulla possibilità di individuare un *cluster* di marcatori peptidici urinari in grado di discriminare i pazienti affetti da ccRCC dai soggetti controlli e tumori maligni da masse renali benigne addizionate a soggetti sani. Nell'ambito della scoperta e classificazione di nuovi biomarcatori, la dimensione, ovvero il numero di variabili inclusi nel modello costruito, è un aspetto cruciale [308]. Il modello dovrebbe essere costruito evitando l'*overfitting* ma consentendo la generalizzazione; un compromesso tra questi due parametri permetterebbe di classificare correttamente una casistica indipendente da quella utilizzata per la costruzione del modello. Per questi motivi, in questo studio i *clusters* di segnali sono stati costruiti con non più di 20 *features*, come suggerito da dati presenti in letteratura [308]. Allo scopo di valutare il modello avente la migliore efficienza discriminativa, sono state condotte una serie di elaborazioni statistiche in cui sono stati impiegati diversi *data Input*, diversi algoritmi di SVM (*ANOVA*, *Radial*, *Dot*) e diverse suddivisioni dell'intera popolazione in due gruppi da utilizzare per la costruzione del modello e per la loro validazione (*Splitting ratio* 9:1, 8:2 e 6:4). Sulla base dei risultati ottenuti, si è scelto pertanto di utilizzare queste condizioni di classificazione (tutti i dati di *input*, algoritmo *Dot*, *Splitting ratio* 6:4) sia per il confronto tra Ctrl e ccRCC che in quello tra pazienti con masse tumorali maligne e soggetti con masse renali benigne addizionate ai soggetti controllo, in quanto questi parametri hanno permesso di ottenere le migliori performance.

Inizialmente al fine di discriminare i pazienti affetti da ccRCC dai soggetti sani, è stato individuato un *cluster* di dodici peptidi dotato di una buona capacità di classificazione (specificità del 90% e sensibilità del 82%, Tabella 14A), con un valore di AUC pari a 0.96 (Figura 43). Tale modello è stato confermato anche su una casistica indipendente di soggetti, mostrando valori di specificità e sensibilità di 91% e 84%, rispettivamente (Tabella 14B). Successivamente, è stata affrontata un'altra importante problematica clinica che riguardava la natura delle masse renali. In particolare si è cercato di distinguere sulla base del profilo peptidico urinario le masse renali benigne dalle lesioni renali maligne. È stato quindi costruito un *pattern* discriminante di dodici segnali peptidici con una specificità dell'78% e una sensibilità dell'88% (Tabella 16A). Questa elevata capacità di classificazione è stata anche confermata nella fase di validazione su una casistica indipendente di soggetti (valori di specificità al 87% e sensibilità al 76%, Tabella 16B). Sulla base della curva ROC, i dodici ioni mostravano un buon potere diagnostico secondo i criteri suggeriti da Swets [229], con un valore di AUC pari a 0.89 (Figura 46).

Successivamente, l'analisi statistica è stata indirizzata all'individuazione di possibili segnali peptidici che mostravano una variazione in relazione alle caratteristiche tumorali: dimensione, *stage* e *grade*. Al fine di preservare l'omogeneità della classe analizzata, sono stati inclusi nella casistica solo i tumori renali a cellule chiare. Attraverso questo approccio è stato possibile osservare una dipendenza dell'abbondanza urinaria di 15 e 26 peptidi rispettivamente con la dimensione e lo stadio tumorale (Tabella 18 e Tabella 19), e tra di essi otto erano in comune e mostravano un trend di correlazione coerente. Tuttavia, solo pochi segnali mostravano un'alterazione significativa in relazione al Grade secondo Fuhrman (Tabella 20). Anche se il grading secondo Fuhrman è considerato tra i fattori prognostici istologici più accettati [309], è necessario sottolineare che esistono degli aspetti controversi relativi alla bassa accuratezza del sistema, principalmente a causa dell'eterogeneità intrinseca al tumore, e alla variabilità inter- e intra osservatore nell'assegnazione del *grade* alla massa neoplastica. Inoltre, molti di questi peptidi erano sovra- o sotto-rappresentati ($p < 0.05$) nei campioni di urine dei pazienti ccRCC rispetto ai soggetti controllo (Tabelle 18, 19 e 20) e alcuni di essi presentavano variazioni nelle concentrazioni urinarie statisticamente significative al variare del pT o del *grade* (Tabella 21 e Tabella 23) e anche a stadi precoci come nel caso di neoplasie aventi pT1a e pT1b (Tabella 22).

Uno dei principali vantaggi della strategia di profiling MALDI è che i segnali rilevati non necessitano alcuna conoscenza pregressa circa la loro identità per poter essere utilizzati come biomarcatori. Essi potrebbero infatti rappresentare uno strumento diagnostico promettente, anche indipendentemente dall'identità delle loro strutture [147]. Occorre tuttavia considerare che la mancata assegnazione ai segnali di una sequenza amminoacidica comporta da un lato l'impossibilità di creare anticorpi specifici e dunque di sviluppare saggi immunologici e dall'altro la perdita di informazioni che, anche se legate a prodotti di degradazione, potrebbero migliorare la conoscenza dei processi biologici implicati nella patologia e delle funzioni molecolari e regolazioni di tali molecole bioattive [155]. Il peso molecolare dei segnali discriminanti le popolazioni in studio collocandosi sotto i 20-30 kDa [158] si riferisce essenzialmente al peptidoma [159]. Dato che l'identificazione dei peptidi mediante isolamento del precursore e frammentazione in MALDI-TOF/TOF risulta spesso difficoltosa, generalmente poco sensibile e informativa, soprattutto a causa della singola carica dello ione precursore [310], gli approcci peptidomici moderni si basano su un esteso utilizzo della tecnologia LC-ESI-MS/MS. Tale tecnica infatti è complementare a quella MALDI-TOF ed è caratterizzata da un più elevato potere identificativo e sensibilità riconducibili alle caratteristiche di ionizzazione (ioni multi carica) e frammentazione proprie della strumentazione. In questo studio è stato applicato un metodo ottimizzato in nLC-ESI-MS/MS che ha consentito di costruire una libreria di peptidi endogeni identificati nelle frazioni arricchite ottenute dal pre-frazionamento di urine di pazienti e controlli con biglie magnetiche a fase inversa di tipo C8. L'indagine di una matrice biologica così semplificata a livello peptidomico ha permesso l'identificazione di 2339 peptidi unici di cui 726 con uno *score* maggiore della soglia di identità peptidica, appartenenti a 251 diverse isoforme proteiche, e 1613 peptidi ottenuti utilizzando come *cut-off* la soglia di omologia (*false discovery rate* $< 5\%$). Alcuni di questi peptidi identificati hanno consentito di associare sequenze peptidiche specifiche ai segnali inclusi nei modelli diagnostici o che

presentavano una variazione significativa in relazione alla dimensione, allo stadio o al grado della massa tumorale, aumentando così le informazioni inerenti la patogenesi e la progressione tumorale e i meccanismi biologici coinvolti (Tabella 24 e Tabella 25). L'assegnazione dell'identità dei segnali MALDI tramite l'utilizzo della libreria di peptidi endogeni identificati in ESI è stata possibile solo mediante un accurato allineamento di massa. In effetti, il *match* tra le identificazioni ottenute in nLC-ESI-MS/MS e i segnali osservati negli spettri MALDI è stato confermato dal basso valore in ppm del MME dell'allineamento delle masse sia rispetto alla modalità lineare che *reflector*, come precedentemente descritto (Paragrafo 4.4.1).

Tra i segnali inclusi nei modelli discriminanti e quelli che correlavano con i dati clinici del tumore, venti sono stati identificati (Tabella 24 e Tabella 25). Recentemente Frantzi *et al.* hanno descritto un modello di 86 segnali rilevati nelle urine di soggetti sani e pazienti con carcinoma a cellule renali, mediante un approccio CE-LC, che raggiungeva valori di sensibilità e specificità, rispettivamente, dell'80% e dell'87% [193]. I suddetti autori sono riusciti ad identificare 40 di questi biomarcatori e la maggior parte di essi erano frammenti di collagene, del fibrinogeno e della pompa Na⁺/K⁺ ATP dipendente; essi hanno quindi suggerito che questi marcatori possano essere il risultato dell'attività proteolitica di enzimi presenti nel microambiente del tumore renale e associati al danno renale. È interessante osservare che la maggior parte dei peptidi identificati nel nostro studio erano diversi da quelli rilevati da Frantzi *et al.* [193], ma soprattutto erano correlati alla presenza di una massa tumorale (Tabella 24 e Tabella 25). Questo risultato ottenuto non deve sorprendere dal momento che i dati acquisiti dalle analisi del peptidoma urinario mediante diversi approcci di pre-frazionamento e diverse separazioni cromatografiche forniscono risultati tra loro complementari [311]. Dunque, alterazioni del proteoma/peptidoma urinario potrebbero spiegare la risposta dell'organismo ad una determinata patologia e potrebbero essere strettamente correlati alla proteolisi aumentata, alla disfunzione renale, alle alterazioni dei rapporti di produzione o a tutte queste implicazioni [312]. La maggioranza dei peptidi endogeni è infatti rappresentata da prodotti di degradazione originanti da proteine molto abbondanti, come il collagene, l'albumina, la beta 2-macroglobulina e l'uromodulina. Sebbene questi peptidi endogeni possono essere presenti nelle urine come risultato di una lesione generalizzata del rene, è anche vero che sono soprattutto dovuti a cambiamenti nell'attività di specifiche proteasi dei tumori o di altre patologie [157, 193].

Al fine di completare ulteriormente il quadro molecolare delle alterazioni indotte dall'insorgenza o dalla progressione della neoplasia renale, si è proceduto ad effettuare un'analisi di tipo proteomico su pools di campioni di urine raccolte da 80 volontari sani e 80 pazienti affetti da ccRCC. La digestione triptica e la successiva analisi *shotgun* mediante nLC-ESI MS/MS hanno permesso di elaborare dati ottenuti sia da un punto di vista qualitativo che quantitativo, mediante approccio *label-free*. In particolare, lo studio qualitativo ha permesso, a seguito di opportuna elaborazione dei dati, di identificare 459 proteine nei pools di urine di soggetti controllo e pazienti affetti da ccRCC. La successiva analisi quantitativa ha permesso di indagare le possibili alterazioni a carico delle proteine urinarie in presenza di un tumore renale a cellule chiare; in particolare, considerando variata l'espressione proteica quando il rapporto tra le abbondanze relative normalizzate dei segnali appartenenti ad una specifica proteina è aumentato del 50% (>1.5) nei pazienti

ccRCC o nei controlli (< 0.67) [213, 285], è stato possibile individuare 309 proteine alterate tra volontari sani e pazienti. Tra di esse, 127 proteine sono risultate sovra-esprese e 182 sotto-esprese nei soggetti affetti da ccRCC. L'analisi funzionale effettuata sulle proteine risultate alterate ha permesso di mettere in luce i processi biologici e le funzioni molecolari modificate in presenza di una massa renale (Figura 58 e Figura 59). È stato quindi possibile osservare che le principali alterazioni risultano essere a carico di processi metabolici e cellulari, processi regolatori, di risposta a stimoli e di sviluppo, l'attività catalitica, di binding e le attività recettoriali e di regolazione dell'attività enzimatica.

Dato che il rene in condizioni fisiologiche contribuisce al mantenimento dell'omeostasi dell'ambiente interno regolando il livello e/o le emivite delle proteine insieme con la perdita di molecole nelle urine, l'integrazione delle informazioni proteomiche e peptidomiche ottenute potrebbe essere d'aiuto nell'identificazione di modificazioni chiave e di firme molecolari caratterizzanti la trasformazione maligna e la sua progressione. Per questo motivo si è proceduto all'integrazione delle informazioni ricavate dall'approccio multi-omico applicato in modo da aumentare il potenziale informativo delle analisi prese singolarmente. Inizialmente sono state selezionate le 22 proteine identificate e presenti nei modelli discriminanti o correlate ai dati istologici delle masse renali oggetto di studio e a ciascuna proteina sono stati associati i relativi peptidi endogeni presenti nei profili urinari analizzati insieme con i loro livelli di rappresentazione (Tabella 26) e i dati ottenuti dalle indagini proteomiche. Il confronto ha mostrato che i peptidi appartenenti a ciascuna proteina presentavano diversi livelli di coerenza: infatti in alcuni casi tutti o la maggior parte dei peptidi che derivano dalla stessa proteina mostrano un trend di rappresentazione coerente tra loro suggerendo la possibilità che questo andamento possa rispecchiare quello relativo al livello di espressione della proteina intatta o, in generale, al grado di proteolisi della specifica proteina. Al contrario, per altre proteine si osservano differenti livelli di coerenza a livello peptidico, a causa probabilmente di fattori biologici. Inoltre, anche a livello proteomico sono state riscontrate possibili anomalie se si considera la concordanza di espressione tra la proteina intatta e i relativi peptidi endogeni, mentre in altri casi è stato possibile osservare coerenza di andamento. In particolare, è interessante osservare che, come atteso, l'albumina (HBA_HUMAN) e la maggior parte dei peptidi endogeni da essa derivanti sono risultati sovra-espressi nelle urine dei pazienti affetti da ccRCC. Infatti, è stato dimostrato che spesso i pazienti affetti da questa patologia presentano ematuria causata da danno renale come conseguenza della presenza della massa tumorale [229, 313]. Mentre, si può supporre che la presenza di trend non coerente possa essere il risultato dell'attività proteolitica di enzimi, come le proteasi extracellulari, che possono riflettere l'attività di specifiche malattie o della loro progressione [314]. Se ne deduce che un'attività proteolitica cancro-specifica differente da quella operante in condizioni fisiologiche potrebbe portare ad una modifica del pattern dei frammenti generati e la relativa alterazione dei livelli di espressione delle proteine intatte. I cambiamenti complessi nell'attività proteolitica possono essere studiati, oltre che attraverso la diretta valutazione dell'attività di una specifica proteasi, anche attraverso l'identificazione di un pattern di frammenti proteolitici generati, attraverso tecniche *high throughput* come la spettrometria di massa [315]. Le alterazioni dei peptidi e delle proteine identificati nelle urine possono essere attribuite a differenze nel rate di produzione, aumento

dell'organizzazione in filamenti, aumento della proteolisi, in disfunzioni renali, o una combinazione di questi. Inoltre è stato anche riportato in letteratura che i cambiamenti di espressione dei peptidi urinari possono essere attribuiti ad un *turnover* alterato della matrice extracellulare (ECM): degradazione della ECM dovuta ad un'invasione tumorale, cambiamenti nella composizione della stessa ECM e alterazioni dell'endotelio. E' stato anche descritto che la ridotta secrezione di frammenti peptidici da collagene di tipo I è correlata all'attività malattia-specifica di metalloproteasi di matrice (MMP) in risposta ad un danno tubulare e legata ad un aumentato rischio a lungo termine di sviluppare un'insufficienza renale cronica in pazienti con danno acuto renale [312, 316].

Quattro proteine, tra le 22 identificate, che presentavano sia dati di tipo peptidomico che di tipo proteomico, sono state osservate avere peptidi che mostravano un trend di rappresentazione coerente.

In particolare, una di queste proteine è risultata essere la glicoproteina umana uromodulina (UMOD/THP), una delle proteine più abbondanti nelle urine, e la maggioranza dei suoi peptidi sono stati osservati in maggiore concentrazione nei pazienti affetti da ccRCC (Tabella 27).

Tuttavia un interesse particolare è stato indirizzato, soprattutto, verso quelle proteine che mostravano una forte associazione con lo sviluppo e la progressione della massa tumorale: A1AG1/A1AG2, PGBM e PTGDS.

In particolare, la proteina A1AGx, anche conosciuta come *a1-acid glycoprotein* (AGP), è il membro principale della famiglia APP, e i suoi livelli sierici aumentano durante le reazioni di fase acuta [317]. Inoltre, un aumento della risposta di APP è stato osservato essere associato a tassi di sopravvivenza ridotti in casi di carcinomi polmonari, pancreatici, renali e coloretali e di linfoma, indipendentemente dallo stadio della patologia maligna [317]. In un altro studio, i livelli sierici di AGP in un gruppo di pazienti affetti da diverse forme tumorali (carcinoma dell'esofago, gastrico, del colon retto, del polmone, epatico e pancreatico) non hanno mostrato differenze statisticamente significative in relazione alla dimensione, allo stadio e alle informazioni cliniche [318]. Ciononostante, i pazienti che presentavano la malattia in stadio avanzato o una recidiva avevano livelli di AGP aumentati [318]. In maniera del tutto simile, con il nostro approccio peptidomico abbiamo rilevato un aumento dei livelli dei frammenti di AGP nei pazienti ccRCC rispetto ai controlli, una over-rappresentazione nello stadio precoce pT1a, che diminuisce progressivamente dal pT1 al pT3 in maniera del tutto compatibile con la sua funzione di proteina della fase acuta. Alterazioni dei livelli della forma non glicosilata della proteina in relazione al pT sono state riportate nel tessuto tumorale dei pazienti affetti da RCC [291]. Gli autori di questo studio hanno infatti analizzato 9 tessuti tumorali a pT1, pT2 e pT3 e hanno osservato una down-espressione della proteina nei soggetti a pT1 e pT2 ma non a pT3. A livello proteomico la proteina intatta risulta avere un livello di espressione coerente con i suoi peptidi, che potrebbe far ipotizzare, a seguito di saggi immunologici, verso una maggior produzione della proteina in presenza di tumore renale rispetto ai soggetti sani.

La proteina PGBM (*Basement membrane-specific heparan sulfate proteoglycan core protein*), anche conosciuta come perlecano o HSPG2, è un componente integrale delle membrane basali, in particolare delle membrane basali glomerulari, dove risulta essere responsabile della carica elettrostatica negativa e fornisce una barriera

selettiva sia a livello di dimensione che di carica. Essa gioca un ruolo essenziale nella vascolarizzazione e nella regolazione della risposta vascolare alle ferite. Interagendo con il suo ligando, l'integrina alfa2/beta1, funge anche da agente anti-tumorale nella neo-vascolarizzazione. Essa viene clivata dando origine a due catene, la cui sequenza risulta essere parzialmente sovrapposta: l'endorepellina e il peptide LG3 (*C-terminal laminin-like globular domain*). In particolare, l'endorepellina sembra inibire alcuni aspetti dell'angiogenesi: migrazione delle cellule endoteliali, morfogenesi del tubo endoteliale indotta da collagene e crescita dei vasi sanguigni [319]. L'endorepellina potrebbe quindi ritardare la neovascolarizzazione tipica delle neoplasie e quindi la crescita tumorale (Endorepellin, a novel inhibitor of angiogenesis derived from the C terminus of perlecan). La sua azione sembra inoltre derivare dall'interruzione del citoscheletro di actina e delle adesione focali (Crystal Structure of the LG3 Domain of Endorepellin, an Angiogenesis Inhibitor). Il peptide LG3, invece, costituisce la porzione C-terminale sia di PGBM che dell'endorepellina. Tale peptide sembra essere responsabile di gran parte dell'attività angiogenica dell'endorepellina e tale attività sembra essere possibile grazie al legame del peptide con gli ioni calcio [320]. Numerosi studi hanno dimostrato un'implicazione di tale proteina in vari stati patologici, incluso il tumore renale a cellule chiare; in particolare, la sua forma glicosilata sembra essere over-espressa nel plasma di soggetti affetti da ccRCC prima di essere sottoposti a nefrectomia [321]. Inoltre, l'analisi di tessuti congelati e fissati in formalina e inclusi in paraffina (FFPE) ha permesso di individuare una diminuzione statisticamente significativa della forma intatta della proteina all'aumentare del grado tumorale. Dal nostro studio è emersa una correlazione negativa di un peptide a m/z 1826 con il *grade*, una over rappresentazione della maggior parte dei peptidi endogeni nei pazienti affetti da ccRCC e una down-espressione della proteina intatta nelle urine dei pazienti. La maggior parte dei peptidi identificati e utilizzati per la quantificazione label-free della proteina risultano appartenere al peptide LG3 e presentano una sotto-rappresentazione, suggerendone una potenziale down-espressione nei pazienti ccRCC e una conseguente riduzione della sua attività angiogenica. Tale ipotesi dovrà essere ulteriormente valutata al fine di validare i risultati ottenuti, studiare lo stato di glicosilazione della proteina intatta e spiegare la presenza di un'incoerenza tra i risultati ottenuti negli studi di proteomica e quelli di peptidomica.

La Prostaglandina H2 D-isomerasi (PTGDS_HUMAN) è una proteina di circa 21kDa che catalizza la conversione di PGH2 in PGD2, una prostaglandina coinvolta nella contrazione e rilassamento della muscolatura liscia e un potente inibitore dell'aggregazione piastrinica. Essa è coinvolta in numerose funzioni del sistema nervoso centrale, come la sedazione, il sonno NREM e l'allodinia indotta da PGE2 e sembrerebbe avere un ruolo anti-apoptotico negli oligodendrociti. PTGDS lega molecole lipofile, come la biliverdina, la bilirubina, l'acido retinoico e l'ormone tiroideo e sembra che agisca da scavenger per le molecole idrofobiche dannose e da trasportatore per l'acido retinoico e l'ormone tiroideo. Una sua alterazione è stata determinata in numerose neoplasie. In particolare, è stata determinata e validata una diminuzione dei livelli di PTGDS nel liquido cerebrospinale di soggetti affetti da meningioma rispetto a soggetti affetti da altri tumori [322]. Inoltre, Hirawa *et al.* hanno determinato un aumento dei livelli di PTGDS escreta nelle urine di pazienti affetti da diabete mellito rispetto a soggetti sani [323]. È stato inoltre dimostrato in un ulteriore studio l'aumento dei livelli serici e urinari in soggetti con ipertensione, con

ulteriore aumento nel caso di peggioramento della funzione renale, suggerendo un potenziale ruolo della proteina nelle lesioni renali indotte da ipertensione [324]. PTGDS risulta inoltre una proteina altamente glicosilata ed è stata dimostrata in campioni di urine la presenza di un sito di glicosilazione sull'amminoacido 78 [325]. In particolare, tale forma glicosilata, dal peso molecolare di circa 30kDa [326], è stata rilevata come over-espressa nel plasma dei soggetti affetti da ccRCC rispetto a volontari sani. Dalle analisi effettuate nell'ambito del progetto di dottorato è emersa la presenza di un peptide endogeno (m/z 1525) che presentava una over-rappresentazione nei soggetti affetti da ccRCC a pT1a, e quindi nelle fasi più precoci della malattia rispetto a soggetti con pT1b, una over-rappresentazione di due peptidi endogeni presenti nei profili proteici urinari nei pazienti rispetto ai volontari sani e una down-espressione della proteina intatta nelle urine. L'espressione della proteina intatta è stata inoltre verificata mediante saggio *western blot* che ha evidenziato la presenza di una banda a 30kDa e una netta over-espressione della proteina nelle urine di pazienti ccRCC. Le discordanze verificate nell'approccio multi-omico possono essere spiegate con l'aumento di degradazione della proteina intatta che porta ad un'augmentata escrezione urinaria dei suoi peptidi, rilevabile come aumento dei peptidi endogeni, e ad una conseguente diminuzione della proteina intatta. Per quanto riguarda l'anomalia riscontrata tra l'espressione della proteina intatta misurata con l'approccio *label-free* e quella determinata con l'immunodecorazione, essa può essere spiegata il fatto che la banda rilevata a 30kDa e over-espressa nei pazienti è ascrivibile alla forma glicosilata della proteina, mentre i peptidi utilizzati per la quantificazione relativa *label-free* della proteina intatta sono tutti down-rappresentati e non possiedono possibili siti glicosilabili descritti per PTGDS. Pertanto, ipotizziamo che i pazienti ccRCC presentano un aumento nelle urine della forma glicosilata della proteina a discapito della forma non glicosilata.

6. CONCLUSIONI

I risultati ottenuti hanno mostrato la possibilità di costruire modelli discriminativi, sulla base di *clusters* di peptidi urinari, dotati di elevata accuratezza, che permettono di distinguere non solo i pazienti affetti da ccRCC dai controlli ma soprattutto i tumori renali maligni dalle masse renali benigne e dai soggetti sani. Inoltre, attraverso questo progetto è stato possibile evidenziare la presenza di peptidi endogeni urinari la cui espressione risulta alterata in relazione alla progressione tumorale, in particolare in base a dimensione, *stage* e *grade* della neoplasia. Le analisi di proteomica hanno inoltre messo in luce le alterazioni proteiche nelle urine dei pazienti affetti da ccRCC rispetto a volontari sani. Infine, l'integrazione dei dati ottenuti da differenti approcci e piattaforme tecnologiche ha permesso di caratterizzare alterazioni nell'espressione proteica potenzialmente implicate nella patogenesi e nella progressione del ccRCC e che potrebbero essere utili in campo diagnostico o prognostico. In particolare, è stato possibile individuare tre possibili target molecolari per il ccRCC (A1AGx, PGBM e PTGDS), alcuni dei quali già noti per essere implicati nella patogenesi di RCC e di altri tipi tumorali. Tali risultati forniscono infatti nuove chiavi di lettura per migliorare la comprensione dell'intricato sistema di meccanismi biologici ed effettori che stanno alla base del RCC.

Questi risultati potrebbero avere molteplici risvolti: potrebbero essere pertanto utili nel distinguere le masse renali di incerta diagnosi che richiedono una conferma mediante biopsia. In pazienti già particolarmente compromessi dove la biopsia renale potrebbe essere una scelta tecnicamente difficile oltre che pericolosa; e potrebbero trarre beneficio dalle scoperte qui descritte, evitando qualsiasi tipo di procedura invasiva, al fine di ottenere una corretta diagnosi. Questi risultati costituiscono il punto di partenza per studi futuri volti alla validazione dei loro livelli urinari e la possibilità per alcuni di essi di essere inclusi in algoritmi per la stratificazione del rischio dei pazienti ccRCC. Inoltre, le loro variazioni nelle urine potrebbero essere utilizzate negli approcci “watch-and-wait” per il monitoraggio delle piccole masse renali e potrebbero avere un possibile ruolo nella gestione delle strategie di follow-up.

La conoscenza di efficaci biomarcatori che permettano il controllo continuo dell'evoluzione del tumore e che permettano di seguire il *follow-up* dei pazienti attraverso approcci non invasivi, è infatti un'importante prerogativa clinica. I biomarcatori presenti nelle urine potrebbero dunque aiutare nel rilevare precocemente possibili recidive soprattutto dopo parziale nefrectomia del rene. Esse, difatti, spesso risultano di difficile valutazione clinica, dal momento che la sola diagnostica per immagini è incapace, ad oggi, di distinguere facilmente a seguito dall'operazione chirurgica, una alterazione del parenchima, originata da una cicatrice disomogenea, da una vera recidiva della malattia.

Inoltre i biomarcatori urinari potrebbero permettere una facile valutazione della vitalità residua del tumore dopo terapie non invasive come ad esempio l'ablazione percutanea mediante radiofrequenze, la crioterapia, le microonde e gli ultrasuoni focalizzati ad alta intensità (HIFU) che risultano essere trattamenti indicati per pazienti con una predisposizione genetica a tumori multipli, con un solo rene, con tumori bilaterali o anziani.

Sarebbe utile, inoltre, effettuare un'ulteriore analisi interattiva: controlli *versus* ccRCC pT1a. Quest'ultima discriminazione potrebbe portare all'identificazione di un *pattern* di segnali in grado di classificare, con una maggiore efficacia, i pazienti con una massa renale inferiore ai 4 cm rispetto ai controlli.

7. BIBLIOGRAFIA

1. [Http://Globocan.Iarc.Fr/Pages/Online.AspX](http://Globocan.Iarc.Fr/Pages/Online.AspX).
2. <http://seer.cancer.gov/statfacts/html/kidrp.html>.
3. Lopez-Beltran, A., et al., 2004 *WHO classification of the renal tumors of the adults*. Eur Urol, 2006. **49**(5): p. 798-805.
4. Zbar, B. and M. Lerman, *Inherited carcinomas of the kidney*. Adv Cancer Res, 1998. **75**: p. 163-201.
5. Thoenes, W., S. Storkel, and H.J. Rumpelt, *Human chromophobe cell renal carcinoma*. Virchows Arch B Cell Pathol Incl Mol Pathol, 1985. **48**(3): p. 207-17.
6. Amin, M.B., et al., *Renal oncocytoma: a reappraisal of morphologic features with clinicopathologic findings in 80 cases*. Am J Surg Pathol, 1997. **21**(1): p. 1-12.
7. Sauter G, E.J., Epstein JI, Sesterhenn IA, *World Health Organization Classification of Tumours: Pathology and Genetics of Tumours of the Urinary System and Male Genital Organs*. IARC Press. Lyons, 2004.
8. Geramizadeh, B., M. Ravanshad, and M. Rahsaz, *Useful markers for differential diagnosis of oncocytoma, chromophobe renal cell carcinoma and conventional renal cell carcinoma*. Indian J Pathol Microbiol, 2008. **51**(2): p. 167-71.
9. *Renal Oncocytoma*.
10. Eble, J.N., *Angiomyolipoma of kidney*. Semin Diagn Pathol, 1998. **15**(1): p. 21-40.
11. [Http://Library.Med.Utah.Edu/Webpath/Renabtml/Renal115.Html](http://Library.Med.Utah.Edu/Webpath/Renabtml/Renal115.Html).
12. Amin, M., *Diagnostic Pathology: Genitourinary*. 2010.
13. Martignoni, G., et al., *Renal angiomyolipoma with epithelioid sarcomatous transformation and metastases: demonstration of the same genetic defects in the primary and metastatic lesions*. Am J Surg Pathol, 2000. **24**(6): p. 889-94.
14. McLaughlin, J.K. and L. Lipworth, *Epidemiologic aspects of renal cell cancer*. Semin Oncol, 2000. **27**(2): p. 115-23.
15. Mathew, A., et al., *Global increases in kidney cancer incidence, 1973-1992*. Eur J Cancer Prev, 2002. **11**(2): p. 171-8.
16. Moch, H., *An overview of renal cell cancer: pathology and genetics*. Semin Cancer Biol, 2013. **23**(1): p. 3-9.
17. Bennington J.L., B.J., *Tumors of the Kidney, Renal Pelvis, and Ureter, Atlas of Tumor Pathology*. 1975.
18. Abbas A.K., K.V., Aster J.C., *Robbins. Fondamenti di patologia e fisiopatologia*. 2013.
19. Gupta, K., et al., *Epidemiologic and socioeconomic burden of metastatic renal cell carcinoma (mRCC): a literature review*. Cancer Treat Rev, 2008. **34**(3): p. 193-205.
20. Mueller-Lisse, U.G. and U.L. Mueller-Lisse, *Imaging of advanced renal cell carcinoma*. World J Urol, 2010. **28**(3): p. 253-61.
21. Bray F., F.J., Pisani P., Parkin D.M., *Globocan 2012 Cancer Incidence, Mortality and Prevalence Worldwide*. 2004.
22. Figure., A.w.g.I.C., *Incidence, Mortality and Estimates. Epidemiologia E Prevenzione 2006*. 2006.

23. *Survival. Epidemiologia E Prevenzione* 2008.
24. Lipworth, L., R.E. Tarone, and J.K. McLaughlin, *The epidemiology of renal cell carcinoma*. J Urol, 2006. **176**(6 Pt 1): p. 2353-8.
25. Dhote, R., et al., *Risk factors for adult renal cell carcinoma*. Urol Clin North Am, 2004. **31**(2): p. 237-47.
26. Kuper, H., P. Boffetta, and H.O. Adami, *Tobacco use and cancer causation: association by tumour type*. J Intern Med, 2002. **252**(3): p. 206-24.
27. La Vecchia, C., et al., *Smoking and renal cell carcinoma*. Cancer Res, 1990. **50**(17): p. 5231-3.
28. D., P., *Carcinoma De Cellulas Renales. Modelo Predictivo De Progresión*. 2008, University of Zaragoza.
29. Calle, E.E., et al., *Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults*. N Engl J Med, 2003. **348**(17): p. 1625-38.
30. Research, W.C.R.F.a.A.I.f.C., *Food, Nutrition, Physical Activity, and Prevention of Cancer: A Global Perspective*. 2007.
31. Kolonel, L.N., *Association of cadmium with renal cancer*. Cancer, 1976. **37**(4): p. 1782-7.
32. Baynham, S.A., H.P. Katner, and K.B. Cleveland, *Increased prevalence of renal cell carcinoma in patients with HIV infection*. AIDS Patient Care STDS, 1997. **11**(3): p. 161-5.
33. Enterline, P.E., J. Hartley, and V. Henderson, *Asbestos and cancer: a cohort followed up to death*. Br J Ind Med, 1987. **44**(6): p. 396-401.
34. Weng, C.J., et al., *Long-term online hemodiafiltration does not reduce the frequency and severity of acquired cystic kidney disease in hemodialysis patients*. Ren Fail, 2009. **31**(7): p. 555-61.
35. Bonsib, S.M., *Renal cystic diseases and renal neoplasms: a mini-review*. Clin J Am Soc Nephrol, 2009. **4**(12): p. 1998-2007.
36. Stewart, J.H., et al., *The pattern of excess cancer in dialysis and transplantation*. Nephrol Dial Transplant, 2009. **24**(10): p. 3225-31.
37. Ishikawa, I., et al., *Ten-year prospective study on the development of renal cell carcinoma in dialysis patients*. Am J Kidney Dis, 1990. **16**(5): p. 452-8.
38. Neuzillet, Y., et al., *De novo renal cell carcinoma of native kidney in renal transplant recipients*. Cancer, 2005. **103**(2): p. 251-7.
39. Ross, H., G. Martignoni, and P. Argani, *Renal cell carcinoma with clear cell and papillary features*. Arch Pathol Lab Med, 2012. **136**(4): p. 391-9.
40. Eble, J.N. and S.M. Bonsib, *Extensively cystic renal neoplasms: cystic nephroma, cystic partially differentiated nephroblastoma, multilocular cystic renal cell carcinoma, and cystic hamartoma of renal pelvis*. Semin Diagn Pathol, 1998. **15**(1): p. 2-20.
41. de Peralta-Venturina, M., et al., *Sarcomatoid differentiation in renal cell carcinoma: a study of 101 cases*. Am J Surg Pathol, 2001. **25**(3): p. 275-84.
42. Moch, H., et al., *Prognostic utility of the recently recommended histologic classification and revised TNM staging system of renal cell carcinoma: a Swiss experience with 588 tumors*. Cancer, 2000. **89**(3): p. 604-14.

43. Sauter G, E.J., Epstein JI, Sesterhenn IA, *Tumours of the Kidney. Tumours of the Urinary System and Male Genital Organs*, in *Who Classification of Tumours*. 2004, IARC Press.
44. Sconocchia, G., et al., *Defective infiltration of natural killer cells in MICA/B-positive renal cell carcinoma involves beta(2)-integrin-mediated interaction*. *Neoplasia*, 2009. **11**(7): p. 662-71.
45. Morra, L., et al., *Relevance of periostin splice variants in renal cell carcinoma*. *Am J Pathol*, 2011. **179**(3): p. 1513-21.
46. Kovacs, G., *Papillary renal cell carcinoma. A morphologic and cytogenetic study of 11 cases*. *Am J Pathol*, 1989. **134**(1): p. 27-34.
47. Mancilla-Jimenez, R., R.J. Stanley, and R.A. Blath, *Papillary renal cell carcinoma: a clinical, radiologic, and pathologic study of 34 cases*. *Cancer*, 1976. **38**(6): p. 2469-80.
48. Amin, M.B., et al., *Papillary (chromophil) renal cell carcinoma: histomorphologic characteristics and evaluation of conventional pathologic prognostic parameters in 62 cases*. *Am J Surg Pathol*, 1997. **21**(6): p. 621-35.
49. Delahunt, B. and J.N. Eble, *Papillary renal cell carcinoma: a clinicopathologic and immunohistochemical study of 105 tumors*. *Mod Pathol*, 1997. **10**(6): p. 537-44.
50. Hes, O., et al., *Oncocytic papillary renal cell carcinoma: a clinicopathologic, immunohistochemical, ultrastructural, and interphase cytogenetic study of 12 cases*. *Ann Diagn Pathol*, 2006. **10**(3): p. 133-9.
51. Storkel, S., et al., *The human chromophobe cell renal carcinoma: its probable relation to intercalated cells of the collecting duct*. *Virchows Arch B Cell Pathol Incl Mol Pathol*, 1989. **56**(4): p. 237-45.
52. Thoenes, W., et al., *Chromophobe cell renal carcinoma and its variants--a report on 32 cases*. *J Pathol*, 1988. **155**(4): p. 277-87.
53. Akhtar, M., et al., *Sarcomatoid renal cell carcinoma: the chromophobe connection*. *Am J Surg Pathol*, 1997. **21**(10): p. 1188-95.
54. Amin, M.B., et al., *Chromophobe renal cell carcinoma: histomorphologic characteristics and evaluation of conventional pathologic prognostic parameters in 145 cases*. *Am J Surg Pathol*, 2008. **32**(12): p. 1822-34.
55. Prasad, S.R., et al., *Common and uncommon histologic subtypes of renal cell carcinoma: imaging spectrum with pathologic correlation*. *Radiographics*, 2006. **26**(6): p. 1795-806; discussion 1806-10.
56. Kovacs, G., et al., *The Heidelberg classification of renal cell tumours*. *J Pathol*, 1997. **183**(2): p. 131-3.
57. Latif, F., et al., *Identification of the von Hippel-Lindau disease tumor suppressor gene*. *Science*, 1993. **260**(5112): p. 1317-20.
58. Neumann, H.P. and B. Zbar, *Renal cysts, renal cancer and von Hippel-Lindau disease*. *Kidney Int*, 1997. **51**(1): p. 16-26.
59. Gnarr, J.R., et al., *Mutations of the VHL tumour suppressor gene in renal carcinoma*. *Nat Genet*, 1994. **7**(1): p. 85-90.
60. Linehan, W.M., M.I. Lerman, and B. Zbar, *Identification of the von Hippel-Lindau (VHL) gene. Its role in renal cancer*. *JAMA*, 1995. **273**(7): p. 564-70.
61. Kaelin, W.G., Jr. and E.R. Maher, *The VHL tumour-suppressor gene paradigm*. *Trends Genet*, 1998. **14**(10): p. 423-6.

62. Yao, M., et al., *Enhanced expression of c-myc and epidermal growth factor receptor (C-erbB-1) genes in primary human renal cancer*. *Cancer Res*, 1988. **48**(23): p. 6753-7.
63. Weidner, U., et al., *Inverse relationship of epidermal growth factor receptor and HER2/neu gene expression in human renal cell carcinoma*. *Cancer Res*, 1990. **50**(15): p. 4504-9.
64. Martinez-Salamanca, J.I., et al., *Prognostic impact of the 2009 UICC/AJCC TNM staging system for renal cell carcinoma with venous extension*. *Eur Urol*, 2011. **59**(1): p. 120-7.
65. Schmidt, L., et al., *Novel mutations of the MET proto-oncogene in papillary renal carcinomas*. *Oncogene*, 1999. **18**(14): p. 2343-50.
66. Edge, S.B. and C.C. Compton, *The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM*. *Ann Surg Oncol*, 2010. **17**(6): p. 1471-4.
67. Schmidt, L., et al., *Germline and somatic mutations in the tyrosine kinase domain of the MET proto-oncogene in papillary renal carcinomas*. *Nat Genet*, 1997. **16**(1): p. 68-73.
68. Sidhar, S.K., et al., *The t(X;1)(p11.2;q21.2) translocation in papillary renal cell carcinoma fuses a novel gene PRCC to the TFE3 transcription factor gene*. *Hum Mol Genet*, 1996. **5**(9): p. 1333-8.
69. Haitel, A., et al., *Biologic behavior of and p53 overexpression in multifocal renal cell carcinoma of clear cell type: an immunohistochemical study correlating grading, staging, and proliferation markers*. *Cancer*, 1999. **85**(7): p. 1593-8.
70. Kugler, A., et al., *Expression of metalloproteinase 2 and 9 and their inhibitors in renal cell carcinoma*. *J Urol*, 1998. **160**(5): p. 1914-8.
71. Yoshida, M.A., et al., *Rearrangement of chromosome 3 in renal cell carcinoma*. *Cancer Genet Cytogenet*, 1986. **19**(3-4): p. 351-4.
72. Presti, J.C., Jr., et al., *Comparative genomic hybridization for genetic analysis of renal oncocytomas*. *Genes Chromosomes Cancer*, 1996. **17**(4): p. 199-204.
73. Speicher, M.R., et al., *Specific loss of chromosomes 1, 2, 6, 10, 13, 17, and 21 in chromophobe renal cell carcinomas revealed by comparative genomic hybridization*. *Am J Pathol*, 1994. **145**(2): p. 356-64.
74. Belldegrun, A., *Renal Cell Carcinoma: Prognostic Factors and Patient Selection*. *eur urol suppl*, 2007. **8**: p. 554-58.
75. Ficarra, V., et al., *TNM staging system for renal-cell carcinoma: current status and future perspectives*. *Lancet Oncol*, 2007. **8**(6): p. 554-8.
76. Fuhrman, S.A., L.C. Lasky, and C. Limas, *Prognostic significance of morphologic parameters in renal cell carcinoma*. *Am J Surg Pathol*, 1982. **6**(7): p. 655-63.
77. Sorbellini, M., et al., *A postoperative prognostic nomogram predicting recurrence for patients with conventional clear cell renal cell carcinoma*. *J Urol*, 2005. **173**(1): p. 48-51.
78. Zisman, A., et al., *Improved prognostication of renal cell carcinoma using an integrated staging system*. *J Clin Oncol*, 2001. **19**(6): p. 1649-57.

79. Frank, I., et al., *An outcome prediction model for patients with clear cell renal cell carcinoma treated with radical nephrectomy based on tumor stage, size, grade and necrosis: the SSIGN score.* J Urol, 2002. **168**(6): p. 2395-400.
80. Scoll, B.J., et al., *Age, tumor size and relative survival of patients with localized renal cell carcinoma: a surveillance, epidemiology and end results analysis.* J Urol, 2009. **181**(2): p. 506-11.
81. deKernion, J.B. and D. Berry, *The diagnosis and treatment of renal cell carcinoma.* Cancer, 1980. **45**(7 Suppl): p. 1947-56.
82. Bukowski, R.M., *Prognostic factors for survival in metastatic renal cell carcinoma: update 2008.* Cancer, 2009. **115**(10 Suppl): p. 2273-81.
83. Motzer, R.J. and P. Russo, *Systemic therapy for renal cell carcinoma.* J Urol, 2000. **163**(2): p. 408-17.
84. Gomella, L.G., et al., *Expression of transforming growth factor alpha in normal human adult kidney and enhanced expression of transforming growth factors alpha and beta 1 in renal cell carcinoma.* Cancer Res, 1989. **49**(24 Pt 1): p. 6972-5.
85. Sargent, E.R., et al., *Expression of mRNA for transforming growth factors-alpha and -beta and secretion of transforming growth factor-beta by renal cell carcinoma cell lines.* Cancer Commun, 1989. **1**(5): p. 317-22.
86. Kenneth, N.S., et al., *SWI/SNF regulates the cellular response to hypoxia.* J Biol Chem, 2009. **284**(7): p. 4123-31.
87. Lemon, B., et al., *Selectivity of chromatin-remodelling cofactors for ligand-activated transcription.* Nature, 2001. **414**(6866): p. 924-8.
88. Reisman, D., S. Glaros, and E.A. Thompson, *The SWI/SNF complex and cancer.* Oncogene, 2009. **28**(14): p. 1653-68.
89. Thompson, M., *Polybromo-1: the chromatin targeting subunit of the PBAF complex.* Biochimie, 2009. **91**(3): p. 309-19.
90. Dalglish, G.L., et al., *Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes.* Nature, 2010. **463**(7279): p. 360-3.
91. van Haafden, G., et al., *Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer.* Nat Genet, 2009. **41**(5): p. 521-3.
92. Niu, X., et al., *The von Hippel-Lindau tumor suppressor protein regulates gene expression and tumor growth through histone demethylase JARID1C.* Oncogene, 2012. **31**(6): p. 776-86.
93. Grosfeld, A., et al., *Interaction of hydroxylated collagen IV with the von hippel-lindau tumor suppressor.* J Biol Chem, 2007. **282**(18): p. 13264-9.
94. Kurban, G., et al., *Collagen matrix assembly is driven by the interaction of von Hippel-Lindau tumor suppressor protein with hydroxylated collagen IV alpha 2.* Oncogene, 2008. **27**(7): p. 1004-12.
95. Kurban, G., et al., *Characterization of a von Hippel Lindau pathway involved in extracellular matrix remodeling, cell invasion, and angiogenesis.* Cancer Res, 2006. **66**(3): p. 1313-9.
96. Ohh, M., et al., *The von Hippel-Lindau tumor suppressor protein is required for proper assembly of an extracellular fibronectin matrix.* Mol Cell, 1998. **1**(7): p. 959-68.

97. Clifford, S.C., et al., *Contrasting effects on HIF-1alpha regulation by disease-causing pVHL mutations correlate with patterns of tumourigenesis in von Hippel-Lindau disease*. Hum Mol Genet, 2001. **10**(10): p. 1029-38.
98. Hoffman, M.A., et al., *von Hippel-Lindau protein mutants linked to type 2C VHL disease preserve the ability to downregulate HIF*. Hum Mol Genet, 2001. **10**(10): p. 1019-27.
99. Harten, S.K., et al., *Regulation of renal epithelial tight junctions by the von Hippel-Lindau tumor suppressor gene involves occludin and claudin 1 and is independent of E-cadherin*. Mol Biol Cell, 2009. **20**(3): p. 1089-101.
100. Jonasch, E., et al., *State of the science: an update on renal cell carcinoma*. Mol Cancer Res, 2012. **10**(7): p. 859-80.
101. Xu, J., et al., *Proteolytic exposure of a cryptic site within collagen type IV is required for angiogenesis and tumor growth in vivo*. J Cell Biol, 2001. **154**(5): p. 1069-79.
102. Eng, C., et al., *A role for mitochondrial enzymes in inherited neoplasia and beyond*. Nat Rev Cancer, 2003. **3**(3): p. 193-202.
103. Gottlieb, E. and I.P. Tomlinson, *Mitochondrial tumour suppressors: a genetic and biochemical update*. Nat Rev Cancer, 2005. **5**(11): p. 857-66.
104. Vanharanta, S., et al., *Early-onset renal cell carcinoma as a novel extraparaganglial component of SDHB-associated heritable paraganglioma*. Am J Hum Genet, 2004. **74**(1): p. 153-9.
105. Isaacs, J.S., et al., *HIF overexpression correlates with biallelic loss of fumarate hydratase in renal cancer: novel role of fumarate in regulation of HIF stability*. Cancer Cell, 2005. **8**(2): p. 143-53.
106. Pollard, P.J., et al., *Accumulation of Krebs cycle intermediates and over-expression of HIF1alpha in tumours which result from germline FH and SDH mutations*. Hum Mol Genet, 2005. **14**(15): p. 2231-9.
107. Pollard, P.J., et al., *Targeted inactivation of fb1 causes proliferative renal cyst development and activation of the hypoxia pathway*. Cancer Cell, 2007. **11**(4): p. 311-9.
108. Yang, Y., et al., *UOK 262 cell line, fumarate hydratase deficient (FH-/FH-) hereditary leiomyomatosis renal cell carcinoma: in vitro and in vivo model of an aberrant energy metabolic pathway in human cancer*. Cancer Genet Cytogenet, 2010. **196**(1): p. 45-55.
109. Pfaffenroth, E.C. and W.M. Linehan, *Genetic basis for kidney cancer: opportunity for disease-specific approaches to therapy*. Expert Opin Biol Ther, 2008. **8**(6): p. 779-90.
110. Inoki, K., T. Zhu, and K.L. Guan, *TSC2 mediates cellular energy response to control cell growth and survival*. Cell, 2003. **115**(5): p. 577-90.
111. Brugarolas, J., et al., *Regulation of mTOR function in response to hypoxia by REDD1 and the TSC1/TSC2 tumor suppressor complex*. Genes Dev, 2004. **18**(23): p. 2893-904.
112. Huang, J. and B.D. Manning, *The TSC1-TSC2 complex: a molecular switchboard controlling cell growth*. Biochem J, 2008. **412**(2): p. 179-90.
113. Kucejova, B., et al., *Interplay between pVHL and mTORC1 pathways in clear-cell renal cell carcinoma*. Mol Cancer Res, 2011. **9**(9): p. 1255-65.
114. Brugarolas, J.B., et al., *TSC2 regulates VEGF through mTOR-dependent and -independent pathways*. Cancer Cell, 2003. **4**(2): p. 147-58.

115. Jiang, B.H., et al., *Phosphatidylinositol 3-kinase signaling controls levels of hypoxia-inducible factor 1*. Cell Growth Differ, 2001. **12**(7): p. 363-9.
116. Zhong, H., et al., *Modulation of hypoxia-inducible factor 1alpha expression by the epidermal growth factor/phosphatidylinositol 3-kinase/PTEN/AKT/FRAP pathway in human prostate cancer cells: implications for tumor angiogenesis and therapeutics*. Cancer Res, 2000. **60**(6): p. 1541-5.
117. Hortin, G.L., et al., *Proteomics: a new diagnostic frontier*. Clin Chem, 2006. **52**(7): p. 1218-22.
118. Tolson, J.P., et al., *Differential detection of S100A8 in transitional cell carcinoma of the bladder by pair wise tissue proteomic and immunohistochemical analysis*. Proteomics, 2006. **6**(2): p. 697-708.
119. Solassol, J., et al., *Clinical proteomics and mass spectrometry profiling for cancer detection*. Expert Rev Proteomics, 2006. **3**(3): p. 311-20.
120. Theodorescu, D., et al., *Discovery and validation of new protein biomarkers for urothelial cancer: a prospective analysis*. Lancet Oncol, 2006. **7**(3): p. 230-40.
121. de Noo, M.E., et al., *Reliability of human serum protein profiles generated with C8 magnetic beads assisted MALDI-TOF mass spectrometry*. Anal Chem, 2005. **77**(22): p. 7232-41.
122. International Human Genome Sequencing, C., *Finishing the euchromatic sequence of the human genome*. Nature, 2004. **431**(7011): p. 931-45.
123. Diamandis, E.P., *Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations*. Mol Cell Proteomics, 2004. **3**(4): p. 367-78.
124. Mayne, J., et al., *Fine tuning of proteomic technologies to improve biological findings: advancements in 2011-2013*. Anal Chem, 2014. **86**(1): p. 176-95.
125. Mischak, H., et al., *Clinical proteomics: A need to define the field and to begin to set adequate standards*. Proteomics Clin Appl, 2007. **1**(2): p. 148-56.
126. Megger, D.A., et al., *Proteomic differences between hepatocellular carcinoma and nontumorous liver tissue investigated by a combined gel-based and label-free quantitative proteomics study*. Mol Cell Proteomics, 2013. **12**(7): p. 2006-20.
127. Boja, E., et al., *Evolution of clinical proteomics and its role in medicine*. J Proteome Res, 2011. **10**(1): p. 66-84.
128. Strebhardt, K. and A. Ullrich, *Paul Ehrlich's magic bullet concept: 100 years of progress*. Nat Rev Cancer, 2008. **8**(6): p. 473-80.
129. Dijkers, E.C., et al., *Development and characterization of clinical-grade 89Zr-trastuzumab for HER2/neu immunoPET imaging*. J Nucl Med, 2009. **50**(6): p. 974-81.
130. Perk, L.R., et al., *Quantitative PET imaging of Met-expressing human cancer xenografts with 89Zr-labelled monoclonal antibody DN30*. Eur J Nucl Med Mol Imaging, 2008. **35**(10): p. 1857-67.
131. Petricoin, E., et al., *Clinical proteomics: revolutionizing disease detection and patient tailoring therapy*. J Proteome Res, 2004. **3**(2): p. 209-17.

132. Baltaci, S., et al., *Use of percent free prostate-specific antigen density to improve the specificity for detecting prostate cancer in patients with normal rectal examinations and intermediate prostate-specific antigen levels*. Urol Int, 2003. **70**(1): p. 36-41.
133. Landers, K.A., et al., *Use of multiple biomarkers for a molecular diagnosis of prostate cancer*. Int J Cancer, 2005. **114**(6): p. 950-6.
134. Wulfkuhle, J.D., et al., *Technology insight: pharmacoproteomics for cancer--promises of patient-tailored medicine using protein microarrays*. Nat Clin Pract Oncol, 2006. **3**(5): p. 256-68.
135. Li, J., et al., *Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer*. Clin Chem, 2002. **48**(8): p. 1296-304.
136. Weissinger, E.M., et al., *Proteomic patterns established with capillary electrophoresis and mass spectrometry for diagnostic purposes*. Kidney Int, 2004. **65**(6): p. 2426-34.
137. Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics*. Nature, 2003. **422**(6928): p. 198-207.
138. Nyman, T.A., *The role of mass spectrometry in proteome studies*. Biomol Eng, 2001. **18**(5): p. 221-7.
139. Pusch, W. and M. Kostrzewa, *Application of MALDI-TOF mass spectrometry in screening and diagnostic research*. Curr Pharm Des, 2005. **11**(20): p. 2577-91.
140. Purohit, S., et al., *Assessing the utility of SELDI-TOF and model averaging for serum proteomic biomarker discovery*. Proteomics, 2006. **6**(24): p. 6405-15.
141. Ahmed, N., et al., *An approach to remove albumin for the proteomic analysis of low abundance biomarkers in human serum*. Proteomics, 2003. **3**(10): p. 1980-7.
142. Petricoin, E.F. and L.A. Liotta, *SELDI-TOF-based serum proteomic pattern diagnostics for early detection of cancer*. Curr Opin Biotechnol, 2004. **15**(1): p. 24-30.
143. Skoog, W., Holler, *Fondamenti di chimica analitica*.
144. Liu, C., *The application of SELDI-TOF-MS in clinical diagnosis of cancers*. J Biomed Biotechnol, 2011. **2011**: p. 245821.
145. Mischak, H., A. Vlahou, and J.P. Ioannidis, *Technical aspects and inter-laboratory variability in native peptide profiling: the CE-MS experience*. Clin Biochem, 2013. **46**(6): p. 432-43.
146. Diamandis, E.P., *Point: Proteomic patterns in biological fluids: do they represent the future of cancer diagnostics?* Clin Chem, 2003. **49**(8): p. 1272-5.
147. Bosso, N., et al., *Human urine biomarkers of renal cell carcinoma evaluated by ClinProt*. Proteomics Clin Appl, 2008. **2**(7-8): p. 1036-46.
148. Gianazza, E., et al., *Different expression of fibrinopeptide A and related fragments in serum of type 1 diabetic patients with nephropathy*. J Proteomics, 2010. **73**(3): p. 593-601.
149. Zhang, X., et al., *Evaluation of a novel, integrated approach using functionalized magnetic beads, bench-top MALDI-TOF-MS with prestructured sample supports, and pattern recognition software for profiling potential biomarkers in human plasma*. J Biomol Tech, 2004. **15**(3): p. 167-75.
150. Wulfkuhle, J.D., L.A. Liotta, and E.F. Petricoin, *Proteomic applications for the early detection of cancer*. Nat Rev Cancer, 2003. **3**(4): p. 267-75.

151. Malik, G., et al., *Serum levels of an isoform of apolipoprotein A-II as a potential marker for prostate cancer*. Clin Cancer Res, 2005. **11**(3): p. 1073-85.
152. Boguski, M.S. and M.W. McIntosh, *Biomedical informatics for proteomics*. Nature, 2003. **422**(6928): p. 233-7.
153. Gast, M.C., et al., *Postoperative serum proteomic profiles may predict recurrence-free survival in high-risk primary breast cancer*. J Cancer Res Clin Oncol, 2011. **137**(12): p. 1773-83.
154. Schrader, M. and H. Selle, *The process chain for peptidomic biomarker discovery*. Dis Markers, 2006. **22**(1-2): p. 27-37.
155. Tinoco, A.D. and A. Saghatelian, *Investigating endogenous peptides and peptidases using peptidomics*. Biochemistry, 2011. **50**(35): p. 7447-61.
156. Lone, A.M., Y.G. Kim, and A. Saghatelian, *Peptidomics methods for the identification of peptidase-substrate interactions*. Curr Opin Chem Biol, 2013. **17**(1): p. 83-9.
157. Coon, J.J., et al., *CE-MS analysis of the human urinary proteome for biomarker discovery and disease diagnostics*. Proteomics Clin Appl, 2008. **2**(7-8): p. 964.
158. Zheng, Z.G., et al., *Comprehensive profiling of the low molecular weight proteins and peptides in weak cation exchange beads human serum retentate*. Protein Pept Lett, 2011. **18**(5): p. 498-506.
159. Svensson, M., et al., *Peptidomics-based discovery of novel neuropeptides*. J Proteome Res, 2003. **2**(2): p. 213-9.
160. Di Lorenzo, G., R. Autorino, and C.N. Sternberg, *Metastatic renal cell carcinoma: recent advances in the targeted therapy era*. Eur Urol, 2009. **56**(6): p. 959-71.
161. Wisniewski, J.R., et al., *Universal sample preparation method for proteome analysis*. Nat Methods, 2009. **6**(5): p. 359-62.
162. Baumann, S., et al., *Standardized approach to proteome profiling of human serum based on magnetic bead separation and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry*. Clin Chem, 2005. **51**(6): p. 973-80.
163. Siegel, R., et al., *Cancer statistics, 2011: the impact of eliminating socioeconomic and racial disparities on premature cancer deaths*. CA Cancer J Clin, 2011. **61**(4): p. 212-36.
164. Mainini, V., et al., *Modulation of urinary peptidome in humans exposed to high altitude hypoxia*. Mol Biosyst, 2012. **8**(4): p. 959-66.
165. Zhou, Y., J. Dong, and R.W. Vachet, *Electron transfer dissociation of modified peptides and proteins*. Curr Pharm Biotechnol, 2011. **12**(10): p. 1558-67.
166. Gianazza, E., et al., *Alterations of the serum peptidome in renal cell carcinoma discriminating benign and malignant kidney tumors*. J Proteomics, 2012. **76 Spec No.**: p. 125-40.
167. Strong, K.J., T.M. Osicka, and W.D. Comper, *Urinary-peptide excretion by patients with and volunteers without diabetes*. J Lab Clin Med, 2005. **145**(5): p. 239-46.
168. Thongboonkerd, V. and P. Malasit, *Renal and urinary proteomics: current applications and challenges*. Proteomics, 2005. **5**(4): p. 1033-42.

169. Sauer, S., et al., *Expression of the oncofetal ED-B-containing fibronectin isoform in hematologic tumors enables ED-B-targeted 131I-L19SIP radioimmunotherapy in Hodgkin lymphoma patients*. *Blood*, 2009. **113**(10): p. 2265-74.
170. [Http://Eurokup.Org](http://Eurokup.Org).
171. Johnston, P.B., C. Bondly, and I.N. Micallef, *Ibritumomab tiuxetan for non-Hodgkin's lymphoma*. *Expert Rev Anticancer Ther*, 2006. **6**(6): p. 861-9.
172. Smith, S. and J.W. Sweetenham, *Iodine 131 tositumomab in the treatment of non-Hodgkin's lymphoma*. *Future Oncol*, 2007. **3**(3): p. 255-62.
173. Mesrobian, H.G., et al., *Candidate urinary biomarker discovery in ureteropelvic junction obstruction: a proteomic approach*. *J Urol*, 2010. **184**(2): p. 709-14.
174. Kim, S.C., E.K. Page, and S.J. Knechtle, *Urine proteomics in kidney transplantation*. *Transplant Rev (Orlando)*, 2014. **28**(1): p. 15-20.
175. Ganti, S. and R.H. Weiss, *Urine metabolomics for kidney cancer detection and biomarker discovery*. *Urol Oncol*, 2011. **29**(5): p. 551-7.
176. Minamida, S., et al., *14-3-3 protein beta/alpha as a urinary biomarker for renal cell carcinoma: proteomic analysis of cyst fluid*. *Anal Bioanal Chem*, 2011. **401**(1): p. 245-52.
177. Del Boccio, P., et al., *A hyphenated microLC-Q-TOF-MS platform for exosomal lipidomics investigations: application to RCC urinary exosomes*. *Electrophoresis*, 2012. **33**(4): p. 689-96.
178. Raimondo, F., et al., *The urinary proteome and peptidome of renal cell carcinoma patients: a comparison of different techniques*. *Expert Rev Proteomics*, 2014. **11**(4): p. 503-14.
179. Pieper, R., et al., *Characterization of the human urinary proteome: a method for high-resolution display of urinary proteins on two-dimensional electrophoresis gels with a yield of nearly 1400 distinct protein spots*. *Proteomics*, 2004. **4**(4): p. 1159-74.
180. Vasudev, N.S., et al., *Pre-operative urinary cathepsin D is associated with survival in patients with renal cell carcinoma*. *Br J Cancer*, 2009. **101**(7): p. 1175-82.
181. Morrissey, J.J., et al., *Urinary biomarkers for the early diagnosis of kidney cancer*. *Mayo Clin Proc*, 2010. **85**(5): p. 413-21.
182. Morrissey, J.J. and E.D. Kharasch, *The specificity of urinary aquaporin 1 and perilipin 2 to screen for renal cell carcinoma*. *J Urol*, 2013. **189**(5): p. 1913-20.
183. Morrissey, J.J., et al., *Urinary concentrations of aquaporin-1 and perilipin-2 in patients with renal cell carcinoma correlate with tumor size and stage but not grade*. *Urology*, 2014. **83**(1): p. 256 e9-14.
184. Di Carlo, A., *Matrix metalloproteinase-2 and -9 in the sera and in the urine of human oncocyoma and renal cell carcinoma*. *Oncol Rep*, 2012. **28**(3): p. 1051-6.
185. A, D.I.C., *Evaluation of neutrophil gelatinase-associated lipocalin (NGAL), matrix metalloproteinase-9 (MMP-9) and their complex MMP-9/NGAL in sera and urine of patients with kidney tumors*. *Oncol Lett*, 2013. **5**(5): p. 1677-1681.

186. A, D.I.C., *Matrix metalloproteinase-2 and -9 and tissue inhibitor of metalloproteinase-1 and -2 in sera and urine of patients with renal carcinoma*. *Oncol Lett*, 2014. **7**(3): p. 621-626.
187. Zhang, P.L., et al., *Urine kidney injury molecule-1: a potential non-invasive biomarker for patients with renal cell carcinoma*. *Int Urol Nephrol*, 2014. **46**(2): p. 379-88.
188. Sandim, V., et al., *Proteomic analysis reveals differentially secreted proteins in the urine from patients with clear cell renal cell carcinoma*. *Urol Oncol*, 2016. **34**(1): p. 5 e11-25.
189. Rogers, M.A., et al., *Proteomic profiling of urinary proteins in renal cancer by surface enhanced laser desorption ionization and neural-network analysis: identification of key issues affecting potential clinical utility*. *Cancer Res*, 2003. **63**(20): p. 6971-83.
190. Wu, D.L., et al., *Proteomic evaluation of urine from renal cell carcinoma using SELDI-TOF-MS and tree analysis pattern*. *Technol Cancer Res Treat*, 2008. **7**(3): p. 155-60.
191. Alves, G., et al., *Urine screening by Seldi-Tof, followed by biomarker identification, in a Brazilian cohort of patients with renal cell carcinoma (RCC)*. *Int Braz J Urol*, 2013. **39**(2): p. 228-39.
192. Sim, S.H., et al., *Changes in the urinary proteome post-operatively in renal cancer patients - a reflection of tumour or kidney removal?* *Proteomics Clin Appl*, 2009. **3**(9): p. 1112-22.
193. Frantzi, M., et al., *Discovery and validation of urinary biomarkers for detection of renal cell carcinoma*. *J Proteomics*, 2014. **98**: p. 44-58.
194. Guerrero, I.C. and O. Kleiner, *Application of mass spectrometry in proteomics*. *Biosci Rep*, 2005. **25**(1-2): p. 71-93.
195. Ferguson, P.L. and R.D. Smith, *Proteome analysis by mass spectrometry*. *Annu Rev Biophys Biomol Struct*, 2003. **32**: p. 399-424.
196. Vaudel, M., A. Sickmann, and L. Martens, *Peptide and protein quantification: a map of the minefield*. *Proteomics*, 2010. **10**(4): p. 650-70.
197. Oda, Y., et al., *Accurate quantitation of protein expression and site-specific phosphorylation*. *Proc Natl Acad Sci U S A*, 1999. **96**(12): p. 6591-6.
198. Bottari, P., et al., *Design and synthesis of visible isotope-coded affinity tags for the absolute quantification of specific proteins in complex mixtures*. *Bioconjug Chem*, 2004. **15**(2): p. 380-8.
199. Brun, V., et al., *Isotope dilution strategies for absolute quantitative proteomics*. *J Proteomics*, 2009. **72**(5): p. 740-9.
200. Villanueva, J., M. Carrascal, and J. Abian, *Isotope dilution mass spectrometry for absolute quantification in proteomics: concepts and strategies*. *J Proteomics*, 2014. **96**: p. 184-99.
201. Harsha, H.C., H. Molina, and A. Pandey, *Quantitative proteomics using stable isotope labeling with amino acids in cell culture*. *Nat Protoc*, 2008. **3**(3): p. 505-16.
202. Asara, J.M., et al., *A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen*. *Proteomics*, 2008. **8**(5): p. 994-9.
203. Bantscheff, M., et al., *Quantitative mass spectrometry in proteomics: a critical review*. *Anal Bioanal Chem*, 2007. **389**(4): p. 1017-31.

204. Gafken, P.R. and P.D. Lampe, *Methodologies for characterizing phosphoproteins by mass spectrometry*. Cell Commun Adhes, 2006. **13**(5-6): p. 249-62.
205. Ross, P.L., et al., *Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents*. Mol Cell Proteomics, 2004. **3**(12): p. 1154-69.
206. Freund, D.M. and J.E. Prenni, *Improved detection of quantitative differences using a combination of spectral counting and MS/MS total ion current*. J Proteome Res, 2013. **12**(4): p. 1996-2004.
207. Grossmann, J., et al., *Implementation and evaluation of relative and absolute quantification in shotgun proteomics with label-free methods*. J Proteomics, 2010. **73**(9): p. 1740-6.
208. Old, W.M., et al., *Comparison of label-free methods for quantifying human proteins by shotgun proteomics*. Mol Cell Proteomics, 2005. **4**(10): p. 1487-502.
209. Braisted, J.C., et al., *The APEX Quantitative Proteomics Tool: generating protein quantitation estimates from LC-MS/MS proteomics results*. BMC Bioinformatics, 2008. **9**: p. 529.
210. Sechi, S., *Quantitative Proteomics by Mass Spectrometry*. 2007: Totowa.
211. Fatima, N., et al., *Label-free global serum proteomic profiling reveals novel celecoxib-modulated proteins in familial adenomatous polyposis patients*. Cancer Genomics Proteomics, 2009. **6**(1): p. 41-9.
212. Huang, S.K., et al., *LC/MS-based quantitative proteomic analysis of paraffin-embedded archival melanomas reveals potential proteomic biomarkers associated with metastasis*. PLoS One, 2009. **4**(2): p. e4430.
213. Levin, Y., et al., *Label-free LC-MS/MS quantitative proteomics for large-scale biomarker discovery in complex samples*. J Sep Sci, 2007. **30**(14): p. 2198-203.
214. Becnel, L.B. and N.J. McKenna, *Minireview: progress and challenges in proteomics data management, sharing, and integration*. Mol Endocrinol, 2012. **26**(10): p. 1660-74.
215. Altelaar, A.F., J. Munoz, and A.J. Heck, *Next-generation proteomics: towards an integrative view of proteome dynamics*. Nat Rev Genet, 2013. **14**(1): p. 35-48.
216. Chen, J.I., et al., *Proteomic characterization of midproliferative and midsecretory human endometrium*. J Proteome Res, 2009. **8**(4): p. 2032-44.
217. Lee, M.J., et al., *Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks*. Cell, 2012. **149**(4): p. 780-94.
218. Jia, L., et al., *An attempt to understand kidney's protein handling function by comparing plasma and urine proteomes*. PLoS One, 2009. **4**(4): p. e5146.
219. Farrah, T., et al., *State of the human proteome in 2013 as viewed through PeptideAtlas: comparing the kidney, urine, and plasma proteomes for the biology- and disease-driven Human Proteome Project*. J Proteome Res, 2014. **13**(1): p. 60-75.
220. Chinello, C., et al., *Urinary signatures of Renal Cell Carcinoma investigated by peptidomic approaches*. PLoS One, 2014. **9**(9): p. e106684.
221. Cole, L.M., et al., *MALDI-MSI and label-free LC-ESI-MS/MS shotgun proteomics to investigate protein induction in a murine fibrosarcoma model following treatment with a vascular disrupting agent*. Proteomics, 2014. **14**(7-8): p. 890-903.

222. Sorace, J.M. and M. Zhan, *A data review and re-assessment of ovarian cancer serum proteomic profiling*. BMC Bioinformatics, 2003. **4**: p. 24.
223. Baggerly, K.A., et al., *High-resolution serum proteomic patterns for ovarian cancer detection*. Endocr Relat Cancer, 2004. **11**(4): p. 583-4; author reply 585-7.
224. Baggerly, K.A., J.S. Morris, and K.R. Coombes, *Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments*. Bioinformatics, 2004. **20**(5): p. 777-85.
225. Baker, E.S., et al., *Mass spectrometry for translational proteomics: progress and clinical implications*. Genome Med, 2012. **4**(8): p. 63.
226. Japkowicz, N.C., N. Kotcz, A., *Editorial: Special Issue on Learning from Imbalanced Data Sets*. SIGKDD Explorations, 2004. **6**: p. 1-6.
227. Dziuda, D., *Data Mining for Genomics and Proteomics*. 2010.
228. Hara, T., et al., *Identification of 2 serum biomarkers of renal cell carcinoma by surface enhanced laser desorption/ionization mass spectrometry*. J Urol, 2005. **174**(4 Pt 1): p. 1213-7.
229. Swets, J.A., *Measuring the accuracy of diagnostic systems*. Science, 1988. **240**(4857): p. 1285-93.
230. Yates, J.R., 3rd, *Mass spectrometry and the age of the proteome*. J Mass Spectrom, 1998. **33**(1): p. 1-19.
231. Katta, V., D.T. Chow, and M.F. Rohde, *Applications of in-source fragmentation of protein ions for direct sequence analysis by delayed extraction MALDI-TOF mass spectrometry*. Anal Chem, 1998. **70**(20): p. 4410-6.
232. Hillenkamp, F., et al., *Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers*. Anal Chem, 1991. **63**(24): p. 1193A-1203A.
233. Bakhtiar, R. and R.W. Nelson, *Electrospray ionization and matrix-assisted laser desorption ionization mass spectrometry. Emerging technologies in biomedical sciences*. Biochem Pharmacol, 2000. **59**(8): p. 891-905.
234. Jonscher, K.R. and J.R. Yates, 3rd, *The quadrupole ion trap mass spectrometer--a small solution to a big challenge*. Anal Biochem, 1997. **244**(1): p. 1-15.
235. *Impact Hd*. 2013.
236. Makarov, A., *Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis*. Anal Chem, 2000. **72**(6): p. 1156-62.
237. Ong, S.E., et al., *Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics*. Mol Cell Proteomics, 2002. **1**(5): p. 376-86.
238. Gygi, S.P., et al., *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags*. Nat Biotechnol, 1999. **17**(10): p. 994-9.
239. Thompson, A., et al., *Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS*. Anal Chem, 2003. **75**(8): p. 1895-904.
240. Ye, B., et al., *Proteomic-based discovery and characterization of glycosylated eosinophil-derived neurotoxin and COOH-terminal osteopontin fragments for ovarian cancer in urine*. Clin Cancer Res, 2006. **12**(2): p. 432-41.
241. M. A. Posthumus, P.G.K., H. L. C. Meuzelaar, and M. C. Tennoeverdebrauw, *Laser Desorption-Mass Spectrometry of Polar Non-Volatile Bio-Organic Molecules*. Anal Chem, 1978. **50**: p. 985-91.

242. Marvin, L.F., M.A. Roberts, and L.B. Fay, *Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in clinical chemistry*. Clin Chim Acta, 2003. **337**(1-2): p. 11-21.
243. Karas, M. and F. Hillenkamp, *Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons*. Anal Chem, 1988. **60**(20): p. 2299-301.
244. Waki, H.T.K., Ido Y, et al., *Protein and Polymer Analyses up to M/Z 100 000 by Laser Ionization Time-of-Flight Mass Spectrometry*. Rapid Commun Mass Spectrom, 1988. **2**: p. 151-53.
245. Whittal, R.M. and L. Li, *High-resolution matrix-assisted laser desorption/ionization in a linear time-of-flight mass spectrometer*. Anal Chem, 1995. **67**(13): p. 1950-4.
246. Beavis, R.C. and B.T. Chait, *High-accuracy molecular mass determination of proteins using matrix-assisted laser desorption mass spectrometry*. Anal Chem, 1990. **62**(17): p. 1836-40.
247. Zenobi, R., Knochenmuss, R., *Ion Formation in Maldi Mass Spectrometry*. Mass Spectrom Rev, 1998. **17**: p. 337-66.
248. Karas, M., et al., *Matrix Dependence of Metastable Fragmentation of Glycoproteins in Maldi Tof Mass-Spectrometry*. Anal Chem, 1995. **67**: p. 675-79.
249. Cotter, R.J., *Time-of-flight mass spectrometry: an increasing role in the life sciences*. Biomed Environ Mass Spectrom, 1989. **18**(8): p. 513-32.
250. Pandey, A. and M. Mann, *Proteomics to study genes and genomes*. Nature, 2000. **405**(6788): p. 837-46.
251. Dooley, K.C., *Tandem mass spectrometry in the clinical chemistry laboratory*. Clin Biochem, 2003. **36**(6): p. 471-81.
252. Peng, J. and S.P. Gygi, *Proteomics: the move to mixtures*. J Mass Spectrom, 2001. **36**(10): p. 1083-91.
253. Yates, J.R., 3rd, A.L. McCormack, and J. Eng, *Mining genomes with MS*. Anal Chem, 1996. **68**(17): p. 534A-540A.
254. Sleno, L. and D.A. Volmer, *Ion activation methods for tandem mass spectrometry*. J Mass Spectrom, 2004. **39**(10): p. 1091-112.
255. Wells, J.M. and S.A. McLuckey, *Collision-induced dissociation (CID) of peptides and proteins*. Methods Enzymol, 2005. **402**: p. 148-85.
256. Harrison, A.G., *To b or not to b: the ongoing saga of peptide b ions*. Mass Spectrom Rev, 2009. **28**(4): p. 640-54.
257. Mikesch, L.M., et al., *The utility of ETD mass spectrometry in proteomic analysis*. Biochim Biophys Acta, 2006. **1764**(12): p. 1811-22.
258. Syka, J.E., et al., *Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry*. Proc Natl Acad Sci U S A, 2004. **101**(26): p. 9528-33.
259. McLuckey, S.A. and J.L. Stephenson, Jr., *Ion/ion chemistry of high-mass multiply charged ions*. Mass Spectrom Rev, 1998. **17**(6): p. 369-407.
260. McLafferty, F.W., et al., *Electron capture dissociation of gaseous multiply charged ions by Fourier-transform ion cyclotron resonance*. J Am Soc Mass Spectrom, 2001. **12**(3): p. 245-9.

261. Zubarev, R.A., et al., *Electron capture dissociation for structural characterization of multiply charged protein cations*. *Anal Chem*, 2000. **72**(3): p. 563-73.
262. Liu, H.C. and Hakansson, *Abundant b-type ions produced in electron capture dissociation of peptides without basic amino acids*. *Jasms*, 2007.
263. Creese, A.J. and H.J. Cooper, *The effect of phosphorylation on the electron capture dissociation of peptide ions*. *J. Am. Soc. Mass Spectrom.* *J Am Soc Mass Spectrom*, 2008.
264. Chervet, J.P., M. Ursem, and J.P. Salzmänn, *Instrumental requirements for nanoscale liquid chromatography*. *Anal Chem*, 1996. **68**(9): p. 1507-12.
265. Oosterkamp, A.J., E. Gelpi, and J. Abian, *Quantitative peptide bioanalysis using column-switching nano liquid chromatography/mass spectrometry*. *J Mass Spectrom*, 1998. **33**(10): p. 976-83.
266. Rappsilber, J., Y. Ishihama, and M. Mann, *Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics*. *Anal Chem*, 2003. **75**(3): p. 663-70.
267. Patil, V.P., et al., *Ultra Performance Liquid Chromatography: A Review*. *International Research Journal of Pharmacy*, 2011.
268. Lee, E.D., et al., *On-line capillary zone electrophoresis-ion spray tandem mass spectrometry for the determination of dynorphins*. *J Chromatogr*, 1988. **458**: p. 313-21.
269. Wilm, M. and M. Mann, *Analytical properties of the nanoelectrospray ion source*. *Anal Chem*, 1996. **68**(1): p. 1-8.
270. Vogelzang, N.J. and W.M. Stadler, *Kidney cancer*. *Lancet*, 1998. **352**(9141): p. 1691-6.
271. Hosoya, N., et al., *Proteomics identified nuclear N-myc downstream-regulated gene 1 as a prognostic tissue biomarker candidate in renal cell carcinoma*. *Biochim Biophys Acta*, 2013. **1834**(12): p. 2630-9.
272. Husillos Alonso, A., M. Carbonero Garcia, and C. Gonzalez Enguita, *Is there a role for systemic targeted therapy after surgical treatment for metastases of renal cell carcinoma?* *World J Nephrol*, 2015. **4**(2): p. 254-62.
273. Turtoi, A., E. De Pauw, and V. Castronovo, *Innovative proteomics for the discovery of systemically accessible cancer biomarkers suitable for imaging and targeted therapies*. *Am J Pathol*, 2011. **178**(1): p. 12-8.
274. Craven, R.A., N.S. Vasudev, and R.E. Banks, *Proteomics and the search for biomarkers for renal cancer*. *Clin Biochem*, 2013. **46**(6): p. 456-65.
275. Gibbs, D.L., et al., *Multi-omic network signatures of disease*. *Front Genet*, 2014. **4**: p. 309.
276. Zoppis, I., et al., *Mutual information optimization for mass spectra data alignment*. *IEEE/ACM Trans Comput Biol Bioinform*, 2012. **9**(3): p. 934-9.
277. Mierswa, I., et al., *Yale: Rapid prototyping for complex data mining tasks*, in *12th ACM SIGKDD int. conf. on know. disc. and data mining*. 2006. p. 935-40.
278. Schawe-Taylor, N. and J. Cristianini, *An introduction to support vector machines and other kernel based learning methods*. 2000.

279. Scholkopf, B. and A.J. Smola, *Learning with kernels: support vector machines, regularization, optimization and beyond* 2001.
280. Cava, C., et al., *Copy-number alterations for tumor progression inference*, in *14th conference on artificial intelligence in medicine, AIME*. 2013, Springer: Murcia, Spain.
281. Saeys, Y., I. Inza, and P. Larranaga, *A review of feature selection techniques in bioinformatics*. *Bioinformatics*, 2007. **23**(19): p. 2507-17.
282. Chinello, C., et al., *Serum biomarkers of renal cell carcinoma assessed using a protein profiling approach based on ClinProt technique*. *Urology*, 2010. **75**(4): p. 842-7.
283. Byrd, D.R., et al., *Kidney*, in *Ajcc cancer staging manual*. 2010, Springer: New York. p. 479-89.
284. Mischak, H., et al., *Recommendations for biomarker identification and qualification in clinical proteomics*. *Sci Transl Med*, 2010. **2**(46): p. 46ps42.
285. Siu, K.W., et al., *Differential protein expressions in renal cell carcinoma: new biomarker discovery by mass spectrometry*. *J Proteome Res*, 2009. **8**(8): p. 3797-807.
286. Lewis, G. and A.P. Maxwell, *Early diagnosis improves survival in kidney cancer*. *Practitioner*, 2012. **256**(1748): p. 13-6, 2.
287. Magni, F., et al., *Biomarkers discovery by peptide and protein profiling in biological fluids based on functionalized magnetic beads purification and mass spectrometry*. *Blood Transfus*, 2010. **8 Suppl 3**: p. s92-7.
288. M'Koma, A.E., et al., *Detection of pre-neoplastic and neoplastic prostate disease by MALDI profiling of urine*. *Biochem Biophys Res Commun*, 2007. **353**(3): p. 829-34.
289. Molin, L., et al., *A comparison between MALDI-MS and CE-MS data for biomarker assessment in chronic kidney diseases*. *J Proteomics*, 2012. **75**(18): p. 5888-97.
290. Kaya, K., et al., *Urinary nuclear matrix protein 22 for diagnosis of renal cell carcinoma*. *Scand J Urol Nephrol*, 2005. **39**(1): p. 25-9.
291. Junker, H., et al., *Stage-related alterations in renal cell carcinoma--comprehensive quantitative analysis by 2D-DIGE and protein network analysis*. *PLoS One*, 2011. **6**(7): p. e21867.
292. Lebdaï, S., et al., *Identification and validation of TGFBI as a promising prognosis marker of clear cell renal cell carcinoma*. *Urol Oncol*, 2015. **33**(2): p. 69 e11-8.
293. de Noo, M.E., et al., *Detection of colorectal cancer using MALDI-TOF serum protein profiling*. *Eur J Cancer*, 2006. **42**(8): p. 1068-76.
294. Latterich, M., M. Abramovitz, and B. Leyland-Jones, *Proteomics: new technologies and clinical applications*. *Eur J Cancer*, 2008. **44**(18): p. 2737-41.
295. Santucci, L., et al., *Combinatorial peptide ligand libraries for the analysis of low-expression proteins: Validation for normal urine and definition of a first protein MAP*. *Proteomics*, 2012. **12**(4-5): p. 509-15.
296. Terracciano, R., et al., *Derivatized mesoporous silica beads for MALDI-TOF MS profiling of human plasma and urine*. *Bioconjug Chem*, 2009. **20**(5): p. 913-23.
297. Savino, R. and R. Terracciano, *Mesopore-assisted profiling strategies in clinical proteomics for drug/target discovery*. *Drug Discov Today*, 2012. **17**(3-4): p. 143-52.

298. Albalat, A., H. Mischak, and W. Mullen, *Clinical application of urinary proteomics/peptidomics*. Expert Rev Proteomics, 2011. **8**(5): p. 615-29.
299. Pastore, A.L., et al., *Serum and urine biomarkers for human renal cell carcinoma*. Dis Markers, 2015. **2015**: p. 251403.
300. Zurbig, P., et al., *The human urinary proteome reveals high similarity between kidney aging and chronic kidney disease*. Proteomics, 2009. **9**(8): p. 2108-17.
301. Baylis, C. and B. Corman, *The aging kidney: insights from experimental studies*. J Am Soc Nephrol, 1998. **9**(4): p. 699-709.
302. Epstein, M., *Aging and the kidney*. J Am Soc Nephrol, 1996. **7**(8): p. 1106-22.
303. Martin, J.E. and M.T. Sheaff, *Renal ageing*. J Pathol, 2007. **211**(2): p. 198-205.
304. Hoang, K., et al., *Determinants of glomerular hypofiltration in aging humans*. Kidney Int, 2003. **64**(4): p. 1417-24.
305. Nyengaard, J.R. and T.F. Bendtsen, *Glomerular number and size in relation to age, kidney weight, and body surface in normal man*. Anat Rec, 1992. **232**(2): p. 194-201.
306. Kasiske, B.L., *Relationship between vascular disease and age-associated changes in the human kidney*. Kidney Int, 1987. **31**(5): p. 1153-9.
307. Liu, Y., *Renal fibrosis: new insights into the pathogenesis and therapeutics*. Kidney Int, 2006. **69**(2): p. 213-7.
308. Dziuda, D., *Data mining for genomics and proteomics: analysis of gene and protein expression data*, in *Wiley series on methods and applications in data mining*. 2010, Wiley: Hoboken, N.J. p. 319.
309. Volpe, A. and J.J. Patard, *Prognostic factors in renal cell carcinoma*. World J Urol, 2010. **28**(3): p. 319-27.
310. Tiss, A., et al., *A well-characterised peak identification list of MALDI MS profile peaks for human blood serum*. Proteomics, 2010. **10**(18): p. 3388-92.
311. Klein, J., et al., *Comparison of CE-MS/MS and LC-MS/MS sequencing demonstrates significant complementarity in natural peptide identification in human urine*. Electrophoresis, 2014. **35**(7): p. 1060-4.
312. Metzger, J., et al., *Urinary excretion of twenty peptides forms an early and accurate diagnostic pattern of acute kidney injury*. Kidney Int, 2010. **78**(12): p. 1252-62.
313. Vaglio, A., et al., *Prognostic significance of albuminuria in patients with renal cell cancer*. J Urol, 2003. **170**(4 Pt 1): p. 1135-7.
314. Candiano, G., et al., *Repetitive fragmentation products of albumin and alpha1-antitrypsin in glomerular diseases associated with nephrotic syndrome*. J Am Soc Nephrol, 2006. **17**(11): p. 3139-48.
315. Villanueva, J., et al., *Differential exoprotease activities confer tumor-specific serum peptidome patterns*. J Clin Invest, 2006. **116**(1): p. 271-84.
316. Lo, L.J., et al., *Dialysis-requiring acute renal failure increases the risk of progressive chronic kidney disease*. Kidney Int, 2009. **76**(8): p. 893-9.
317. Yildirim, A., et al., *Relationship between serum levels of some acute-phase proteins and stage of disease and performance status in patients with lung cancer*. Med Sci Monit, 2007. **13**(4): p. CR195-200.

318. Hashimoto, S., et al., *alpha1-acid glycoprotein fucosylation as a marker of carcinoma progression and prognosis*. Cancer, 2004. **101**(12): p. 2825-36.
319. Mongiat, M., et al., *Endorepellin, a novel inhibitor of angiogenesis derived from the C terminus of perlecan*. J Biol Chem, 2003. **278**(6): p. 4238-49.
320. Le, B.V., et al., *Crystal structure of the LG3 domain of endorepellin, an angiogenesis inhibitor*. J Mol Biol, 2011. **414**(2): p. 231-42.
321. Gbormittah, F.O., et al., *Comparative studies of the proteome, glycoproteome, and N-glycome of clear cell renal cell carcinoma plasma before and after curative nephrectomy*. J Proteome Res, 2014. **13**(11): p. 4889-900.
322. Kim, J.H., et al., *Proteome analysis of human cerebrospinal fluid as a diagnostic biomarker in patients with meningioma*. Med Sci Monit, 2012. **18**(11): p. BR450-60.
323. Hirawa, N., et al., *Urinary prostaglandin D synthase (beta-trace) excretion increases in the early stage of diabetes mellitus*. Nephron, 2001. **87**(4): p. 321-7.
324. Hirawa, N., et al., *Lipocalin-type prostaglandin d synthase in essential hypertension*. Hypertension, 2002. **39**(2 Pt 2): p. 449-54.
325. Halim, A., et al., *Human urinary glycoproteomics; attachment site specific analysis of N- and O-linked glycosylations by CID and ECD*. Mol Cell Proteomics, 2012. **11**(4): p. M111 013649.
326. Saso, L., et al., *Quantification of prostaglandin D synthetase in cerebrospinal fluid: a potential marker for brain tumor*. Biochem Mol Biol Int, 1998. **46**(4): p. 643-56.