



Università degli Studi di Cagliari

DOTTORATO DI RICERCA

BIOLOGIA E BIOCHIMICA DELL'UOMO E DELL'AMBIENTE
Ciclo XXVII

Caratterizzazione proteomica di fluidi e tessuti in diverse condizioni fisio-patologiche

Settore scientifico disciplinare di afferenza: BIO/10

Presentata da:

Dott.ssa Morena Arba

Coordinatore Dottorato:

Prof. Emanuele Sanna

Tutor/Relatore:

Prof.ssa M. Teresa Sanna

Esame finale anno accademico 2013 – 2014



La presente tesi è stata prodotta durante la frequenza del corso di dottorato in Biologia e Biochimica dell'Uomo e dell'Ambiente, Università degli Studi di Cagliari, a.a. 2013/2014 – XXVII ciclo con il supporto di una borsa di studio finanziata con le risorse del P.O.R. SARDEGNA F.S.E. 2007-2013 - Obiettivo competitività regionale e occupazione, Asse IV Capitale umano, Linea di attività 1.3.1 “Finanziamento di corsi di dottorato finalizzati alla formazione di capitale umano altamente specializzato, in particolare per i settori dell'ICT, delle nanotecnologie e delle biotecnologie, dell'energia e dello sviluppo sostenibile, dell'agroalimentare e dei materiali tradizionali.

Indice

•	Abstract	4
•	Introduzione:	
	Proteomica <i>versus</i> Genomica	6
	Spettrometria di massa	10
	Elettroforesi bidimensionale	14
	Piattaforme <i>top-down</i> e <i>bottom-up</i>	19
•	Figure	23
•	Parte I:	
	<i>Caratterizzazione proteomica della frazione insolubile in acido della saliva di neonati pretermine e confronto con la saliva di neonati a termine e adulti</i>	25
1.1	Introduzione: la Saliva	26
1.2	Materiali e Metodi	34
1.2.1	Raccolta dei campioni di saliva, estrazione peptidi e proteine solubili in acido e lavaggio della frazione insolubile in acido	34
1.2.2	Solubilizzazione della frazione insolubile in acido	35
1.2.3	Quantificazione delle proteine totali	36
1.2.4	Prima dimensione: IEF	38
1.2.5	Seconda dimensione: SDS-PAGE	40
1.2.6	Colorazione dei gel	41
1.2.7	Analisi delle immagini	43
1.2.8	Digestione triptica	45
1.3	Risultati e Discussione	48
1.4	Conclusioni	60
	Tabelle	62
	Figure	65

•	Parte II:	
	<i>Caratterizzazione proteomica della frazione insolubile in acido di tessuti tumorali cerebrali infantili di Medulloblastoma e Astrocitoma Pilocitico mediante analisi in elettroforesi bidimensionale accoppiata alla spettrometria di massa HPLC-ESI-MS/MS ad alta risoluzione</i>	71
2.1	Introduzione	72
2.1.1	Medulloblastoma	74
2.1.2	Astrocitoma Pilocitico	75
2.2	Materiali e Metodi	77
2.3	Risultati e Discussione	81
2.4	Conclusioni	88
	Tabelle	89
	Figure	92
•	Bibliografia	98
•	Articoli scientifici	109
•	Comunicazioni a congressi	109

Abstract

This thesis has been focused on the proteomic characterization of human saliva from donors of different ages, starting from birth up to adult age, and pediatric brain tumor tissues. The first study has been performed in order to compare the acid-insoluble fraction of saliva from preterm with at-term newborns and adults and establish if differences exist. In the second study medulloblastoma and pilocytic astrocytoma pediatric brain tumor extracts have been compared. In both studies 2-DE analysis was coupled with high resolution tandem mass spectrometry (MS/MS). The proteomic characterization of the acid-insoluble fractions of saliva from preterm newborns allowed to integrate data previously obtained on the acid-soluble fraction by HPLC-electrospray ionization (ESI)-mass spectrometry (MS), and to evidence several differences between preterm newborns, at-term newborns and adults.

Spots differentially expressed between the three groups, according to image analysis of the gels, were submitted to *in-gel* tryptic digestion and the peptide mixture analyzed by high performance HPLC-ESI-MS/MS for their characterization. By this strategy, we identified three over-expressed proteins in at-term newborns with respect to preterm newborns and adults (BPI fold-containing family A member 1, two proteoforms of annexin A1, and keratin type 1 cytoskeletal 13), and several over-expressed proteins in adults (fatty acid-binding protein, S100A6, S100A7, two proteoforms of S100A9, several proteoforms of prolactin-inducible protein, Ig kappa chain, two proteoforms of cystatin SN, one proteoform of cystatin S and several proteoforms of α -amylase 1). Moreover, for the first time, it was possible to assign by MS/MS four spots of human saliva 2-DE, already detected by other authors, to different proteoforms of S100A9. The strategy applied used a sequential staining protocol to the 2-DE gels, first with Pro-Q Diamond, that allows specific detection of phosphoproteins, and successively with total protein SYPRO Ruby stain.

In the second study, proteomic analysis of two pediatric brain tumor tissues pointed out differences between medulloblastoma, the prevalent malignant tumor in childhood, and pilocytic astrocytoma, the most common, that only rarely shows a malignant progression. Due to the limited availability of bioptic tissue, the study was performed on pooled tumor tissues, and was focused on acid-insoluble fraction to integrate the characterization performed by a group of colleagues in Rome on the acid-soluble fraction by high performance HPLC-ESI-MS/MS. The results

indicated that the two tumors exhibit different proteomic profiles and evidenced interesting differential expression of several proteins. Among them, peroxiredoxin-1, peptidyl-prolyl cis–trans isomerase A, heterogeneous nuclear ribonucleoproteins A2/B1, mitochondrial isoform of malate dehydrogenase, nucleoside diphosphate kinase A, glutathione S-transferase P and fructose biphosphate aldolase A resulted significantly over-expressed in medulloblastoma while glial fibrillary acidic protein, serotransferrin, α crystallin B chain, ferritin light chain, annexin A5, fatty acid-binding protein (brain), sorcin and apolipoprotein A-I resulted significantly over-expressed in pilocytic astrocytoma.

In conclusion, the work done allowed to evidence the usefulness of using an integrated *bottom-up/top-down* approach, based on 2-DE-MS analysis and high performance MS in order to obtain a complete characterization of the proteome under investigation, revealing and identifying, not only peptides and small proteins, but also proteins with higher MW, that often it is not possible to identify by using exclusively a *top-down* ESI-MS approach.

Keyword: proteomics, 2-dimensional gel electrophoresis, mass spectrometry, human saliva, brain tumors

Corresponding author: Morena Arba,
Department of Life and Environment Sciences,
Biomedical Section
University of Cagliari, 09042 Monserrato-Cagliari, Italy.
E-mail: morenarba@gmail.com

Introduzione

Proteomica versus Genomica

La proteomica è una scienza nata agli inizi degli anni '90 che si occupa di studiare il “proteoma” analizzando l'insieme delle proteine espresse da una cellula, un tessuto o un fluido biologico in un certo intervallo di tempo e sotto determinate condizioni ambientali. Il termine proteoma è stato utilizzato per la prima volta in tal senso nel 1994 da Mark Wilkins durante un congresso organizzato dall'Università di Siena. La proteomica è una disciplina estremamente vasta e complessa che permette non solo l'identificazione e la quantificazione di tutte le proteine e i peptidi di una cellula, di un tessuto o di un organismo, ma anche l'analisi di profili proteici complessi, la caratterizzazione di proteine ancora sconosciute, lo studio delle interazioni che ciascuna proteina stabilisce con altre molecole, la caratterizzazione strutturale e funzionale delle proteine identificate e la loro localizzazione (Wilkins, 1996).

A differenza della proteomica, la genomica si occupa dello studio della struttura, del contenuto, della funzione ed evoluzione del genoma degli organismi viventi. Mentre il contenuto proteico cellulare è altamente dinamico e si modifica in continuazione in relazione ai cambiamenti dell'ambiente extracellulare, alla specificità di azione del tipo di cellula esaminata, alle sue necessità metaboliche e allo stadio di sviluppo dell'organismo cui appartiene, il genoma è un sistema statico e identico in ogni cellula o organismo, deputato a garantire la conservazione dell'informazione genetica. Oltre alle variazioni del proteoma dovute a variazioni fisiologiche (età, sesso, etc.), dobbiamo tenere presente anche le variazioni che avvengono in particolari stati di stress, in presenza di patologie o sotto somministrazione di farmaci (Pandey, 2000).

Grazie al completamento del sequenziamento del genoma di molti organismi sono state ottenute un enorme numero di sequenze di DNA che hanno fornito innumerevoli informazioni. Tuttavia, è emerso chiaramente che le informazioni che possono essere ottenute con questo approccio non sono sempre sufficienti per spiegare le funzioni biologiche di una cellula. Di conseguenza, negli ultimi anni è diventato sempre più evidente che il genoma rappresenta solo il primo livello di complessità di un sistema biologico e che tale complessità aumenta analizzando il proteoma piuttosto che il genoma. Non vi è infatti sempre una corrispondenza univoca tra il numero dei geni e quello dei possibili prodotti proteici, in quanto il vecchio paradigma “un gene, una proteina” non riflette la realtà esistente all'interno

delle cellule. Il proteoma cellulare è estremamente complesso, articolato e mutabile nel tempo. Ad un singolo genoma corrisponde una molteplicità di proteomi il cui limite per ora non è definibile. Il proteoma rappresenta una collezione dinamica di proteine che differiscono da individuo ad individuo o anche da cellula a cellula, che è determinata dalla regolazione dei geni, degli RNA messaggeri (mRNA), dall'influenza di fattori extracellulari e dalle stesse proteine.

Oltre ai processi successivi alla trascrizione, come la maturazione dell'mRNA e/o i vari livelli di regolazione traduzionale, le eventuali soppressioni, amplificazioni e splicing alternativi che si possono verificare, le proteine possono andare incontro a modificazioni post-traduzionali (PTMs). Dopo la traduzione dell'mRNA, specifiche molecole (fosfati, acetati, carboidrati, lipidi e ubiquitina) possono reagire con la catena polipeptidica in formazione (che viene così fosforilata, acetilata, glicosilata, lipidata, ubiquitinilata), oppure la catena polipeptidica può subire la rimozione di un certo numero di residui amminoacidici dall'estremità N- o C-terminale. Si stima che siano almeno 250 le PTMs nelle cellule eucariote e queste modificazioni possono alterare significativamente le funzioni e le proprietà delle proteine.

Data l'impossibilità della genomica di evidenziare le PTM delle proteine, non tutti i processi biologici possono essere descritti in maniera esauriente unicamente tramite studi genomici.

Recentemente è stato proposto e introdotto il termine proteoforme per indicare tutte le differenti forme molecolari delle proteine prodotte da un singolo gene, incluse le modificazioni dovute a variazioni genetiche, splicing alternativi e PTMs (Smith, 2013). L'utilizzo di questo termine è preferibile rispetto al termine isoforma, che nella definizione IUPAC indica le proteine prodotte a partire da geni diversi, ma che spesso viene utilizzato per indicare proteine derivanti dallo stesso gene (per esempio da splicing alternativi o PTMs). In questa tesi il termine proteoforme è stato ampiamente utilizzato, vista la sua non ambiguità rispetto a termini alternativi precedentemente utilizzati.

I risultati che la proteomica e la genomica forniscono non sono quindi sovrapponibili ma complementari, dato che la proteomica evidenzia le proteine espresse da un organismo in un dato momento, mentre la genomica evidenzia tutti i geni di un determinato organismo, anche quelli silenti. I risultati ottenuti con i due approcci possono risultare profondamente differenti.

Negli ultimi anni gli studi proteomici sono stati utilizzati in modo sempre crescente in diversi campi della biologia e della medicina allo scopo di:

- caratterizzare le proteine espresse in un dato tessuto o linea cellulare,

-identificare le proteine differenzialmente espresse in un dato proteoma in due diverse condizioni sperimentali (ad esempio sano *vs* patologico), evidenziando potenziali biomarcatori (molecole diversamente espresse in un tessuto sano e in un tessuto patologico),

- caratterizzare eventuali PTMs presenti.

Questo approccio ha permesso di ottenere numerose informazioni in stadi sempre più precoci delle patologie, aprendo la strada allo sviluppo di nuove strategie di diagnosi e di nuovi bersagli terapeutici, e in particolare il confronto tra il proteoma di un estratto tissutale di un soggetto in condizioni fisiologiche normali e di un soggetto affetto da una specifica patologia offre la possibilità di individuare potenziali biomarcatori della patologia rilevanti per la diagnosi e la prognosi di determinate patologie (Anagnostopoulos, 2011).

Il gran numero di dati ottenuti negli ultimi decenni grazie all'utilizzo di tecnologie avanzate nel campo della genomica e della proteomica hanno contribuito alla scoperta e all'identificazione di marcatori biologici di numerose patologie e molecole bersaglio che permettano una diagnosi e una prognosi precoce, sicura e affidabile.

Dato che i più importanti fattori coinvolti nel malfunzionamento delle cellule sono le variazioni dei livelli di concentrazioni delle proteine e delle loro interazioni, l'analisi proteomica risulta essere un potente strumento di ricerca in tal senso. A questo quadro va aggiunta l'addizionale eterogeneità e complessità derivante dalla presenza delle PTMs che giocano un importante ruolo anche sulla stabilità, sulla localizzazione e sulle interazioni che le proteine possono instaurare.

Un tipico studio proteomico comporta l'identificazione delle proteine presenti in una specifica matrice biologica. Nelle fasi iniziali della proteomica la principale tecnica utilizzata era l'elettroforesi bidimensionale (2-DE) che permette di rilevare le differenze nei profili proteici in diverse condizioni (O'Farrell, 1975). Successivamente, le strategie applicate in proteomica hanno utilizzato una ampia varietà di combinazioni di tecniche basate principalmente sull'accoppiamento di metodiche di separazione ad elevate capacità con metodiche di analisi analitiche (Messana, 2013). Tra queste ultime, la spettrometria di massa (MS) risulta la più importante per l'identificazione e la caratterizzazione delle proteine e dei peptidi all'interno di una miscela, ma a causa della notevole complessità e della grande variabilità nel contenuto proteico dei fluidi e dei tessuti biologici, i risultati ottimali vengono ottenuti dall'accoppiamento delle analisi di MS con metodologie di

separazione in cromatografia liquida (LC) o in elettroforesi su gel mono- e bi-dimensionale.

Spettrometria di massa

La spettrometria di massa è la tecnica analitica maggiormente utilizzata in proteomica per l'identificazione, la determinazione quantitativa e l'analisi delle proprietà strutturali e chimiche delle proteine. Grazie alle tecniche di MS è possibile determinare accuratamente le masse molecolari delle proteine e caratterizzare le PTMs presenti, utilizzando minime quantità di campione.

L'utilizzo della MS per lo studio delle macromolecole biologiche risale alla fine degli anni '30 ma è dai primi anni '80 che il suo impiego per ricerche di tipo biochimico è diventato più rilevante, grazie alla possibilità di produrre ioni in fase gassosa a partire da composti polari e carichi (pre-requisito fondamentale per l'analisi).

Uno spettrometro di massa è composto dai seguenti componenti (Figura 1):

- Sistema di introduzione del campione
- Sorgente ionica (o camera di ionizzazione)
- Analizzatore (separatore di ioni)
- Rivelatore e sistema di elaborazione dei dati.

Dato che i migliori risultati vengono ottenuti se gli analiti presenti in una miscela vengono separati prima dell'analisi in MS, questa tecnica viene generalmente accoppiata a tecniche analitiche in modo che i componenti della miscela, dopo essere stati separati, vengano indirizzati verso il sistema di ionizzazione dove vengono generati ioni di natura diversa e in numero variabile in funzione del tipo di sistema di ionizzazione impiegato.

In relazione alle energie utilizzate, le tecniche di ionizzazione vengono distinte in hard e soft. Le tecniche di ionizzazione hard provocano un'estesa frammentazione della molecola che rende quindi difficoltosa la determinazione del PM e della struttura di un determinato analita, di conseguenza questa tecnica non trova applicazioni nell'analisi delle macromolecole biologiche. Le tecniche di ionizzazione soft al contrario, operano a bassa energia e per questo motivo non sono distruttive. Tra le tecniche di ionizzazione soft rientrano le tecniche MALDI (*Matrix Assisted Laser Desorption Ionization*) ed ESI (*Electrospray ionization*), quest'ultima ampiamente utilizzata per gli esperimenti di MS descritti in questa tesi. Nella tecnica ESI, per favorire il processo di ionizzazione vengono utilizzate soluzioni acquose di metanolo e acetonitrile, in presenza di basse concentrazioni di acido acetico, formico o trifluoroacetico. Tale soluzioni vengono nebulizzate, utilizzando un gas inerte (di solito azoto), attraverso un capillare tenuto ad alto

potenziale elettrico (5000-6000 V). Il campo elettrostatico è sufficientemente forte da disperdere la soluzione emergente in una sottile nube di gocce di solvente contenenti gli ioni. L'elevata temperatura e la presenza di N₂ che attraversa la sorgente a bassa velocità favoriscono il processo di evaporazione del solvente. Man mano che le gocce rimpiccioliscono, perdendo la loro sfera di solvatazione, al loro interno aumenta la densità di carica e, quando la repulsione elettrostatica supera la tensione superficiale delle gocce, queste esplodono liberando i singoli ioni nudi in esse contenuti. In questo modo il campione è posto nelle condizioni ottimali per essere analizzato. Gli ioni fuoriescono dal capillare accelerati dal campo elettrico applicato e vengono indirizzati all'analizzatore dove, in condizioni di vuoto spinto (10⁻⁴-10⁻⁸ mm Hg), vengono separati in base al loro rapporto m/z.

Una tecnica molto efficace per la caratterizzazione della struttura primaria di un oligopeptide è la spettrometria di tandem massa (MS/MS) che può accoppiare fisicamente due analizzatori (*in space* MS/MS) oppure trattenere l'analita entro una trappola e analizzarlo/frammentarlo/ionizzarlo più volte sempre all'interno dello stesso analizzatore (*in time* MS/MS). Nel caso della ionizzazione *in space* il primo analizzatore ha la funzione di selezionare lo ione desiderato tra i vari ioni presenti. Questo ione "precursore" o "genitore" viene convogliato in una cella di collisione dove, utilizzando un opportuno gas di collisione (He, Ar), viene frammentato in "ioni figli" che vengono poi separati dal secondo analizzatore. Gli analizzatori utilizzati per le procedure *in time* sfruttano invece la possibilità di trattenere in una trappola ionica tutti gli ioni presenti nello spettro primario, di isolare successivamente lo ione desiderato (espellendo contemporaneamente gli altri ioni dalla trappola) e di frammentarlo. I frammenti così generati vengono analizzati all'interno della stessa trappola ionica.

In questo lavoro di tesi è stato utilizzato lo spettrometro di massa LTQ-Orbitrap (*Linear Trap Quadrupole*) per gli esperimenti di *in time* MS/MS dopo accoppiamento con un RP-HPLC (*Reversed-Phase-High-Performance Liquid Chromatography*). In Figura 2 sono schematizzati i componenti dell'Orbitrap.

L'LTQ-Orbitrap è un sistema ibrido in cui l'alta sensibilità della trappola lineare LTQ si coniuga all'elevata accuratezza e risoluzione analitica dell'analizzatore Orbitrap. In questo caso, gli ioni generati dalla sorgente vengono inviati alla trappola quadrupolare lineare (LTQ), che accumula gli ioni, li isola, li frammenta e li convoglia verso un quadrupolo a radiofrequenza (chiamato C-trap per la caratteristica forma a C) in cui gli ioni vengono compressi in una piccola nube. Da qui, grazie ad un repentino aumento del voltaggio sull'elettrodo posto al centro

dello strumento, gli ioni sono inviati all'Orbitrap, che funziona da secondo analizzatore. L'Orbitrap è infatti costituito da un elettrodo centrale fusiforme avvolto da un paio di elettrodi esterni a forma di campana.

Gli ioni si muovono con oscillazioni armoniche intorno all'elettrodo centrale producendo un segnale periodico sugli elettrodi esterni. Vengono rilevate solo le oscillazioni assiali ed il segnale è trasmesso come una immagine di corrente ionica. Dopo l'amplificazione, le immagini dei segnali di corrente ionica vengono convertite in uno spettro di frequenza usando un algoritmo a trasformata di Fourier. La frequenza di oscillazione di ciascuno ione è direttamente correlata al rapporto m/z e quindi lo spettro di frequenza è rapidamente trasformato in uno spettro di massa.

Nello spettrometro LTQ-Orbitrap i frammenti generati dalla prima frammentazione possono essere analizzati con un secondo spettro di massa, ed un frammento può essere a sua volta selezionato, frammentato ulteriormente ed i frammenti di seconda classe analizzati a loro volta. Questi eventi si realizzano nella cella di collisione HCD (*Higher Energy Collision Dissociation*) che consiste in un multipolo lineare alloggiato in un tubo metallico connesso in linea retta con la C-Trap, dalla quale è separato attraverso una lente responsabile della trasmissione degli ioni da e verso la trappola C dove è presente il gas di collisione (elio). Per gli esperimenti HCD, gli ioni una volta accumulati nella trappola lineare, vengono selezionati e trasferiti alla C-Trap, da cui sono successivamente inviati alla cella HCD, attraverso un sistema di lenti. Al suo interno gli ioni collidono con le molecole del gas e nell'urto acquisiscono energia portandosi nello stato attivato. L'energia acquisita viene dissipata con la frammentazione dello ione stesso. I frammenti ottenuti sono analizzati in seguito, quando trasferiti indietro nella C-trap, grazie a variazioni del potenziale, ed inviati all'analizzatore Orbitrap.

Comunemente un moltiplicatore elettronico, costituito da una serie di elettrodi in cascata, funge da collettore e rivelatore di ioni. Quando uno ione giunge al primo elettrodo, questo emette un fascio di elettroni che va a colpire il secondo elettrodo, il quale a sua volta emette una quantità maggiore di elettroni e così via.

Il risultato è una forte amplificazione del segnale che viene poi digitalizzato ed elaborato mediante un computer interfacciato allo spettrometro, per costruire lo spettro di massa. Nel caso dell'LTQ-Orbitrap in cui è applicato il metodo FTMS (spettrometro di massa a trasformata di Fourier), il sistema di rivelazione è un ricevitore di radiofrequenza, a sua volta collegato al calcolatore per l'analisi di Fourier dell'interferogramma da cui si ottiene lo spettro di massa. Oltre a

quest'ultimo, in uscita dallo strumento è possibile ottenere anche la corrente ionica totale (TIC) che rappresenta l'intensità della corrente prodotta da tutti gli ioni generati in uscita dalla colonna cromatografica e fornisce una quantificazione delle specie cariche che sono state intrappolate ed analizzate dallo spettrometro di massa. La trappola ionica in particolare effettua una scansione di un campo a radiofrequenza ad intervalli di tempo stabiliti, cui segue l'acquisizione dello spettro di massa. Questo processo discontinuo di riempimento e svuotamento della trappola viene ripetuto durante tutta la durata della corsa HPLC fornendo per ogni scansione contemporaneamente spettro di massa e TIC.

Si può decidere di rivelare i frammenti analizzati nel solo LTQ o nell'Orbitrap; nel primo caso si ottengono spettri di MS e MS/MS ad alta sensibilità, ma con una bassa risoluzione e accuratezza di massa, mentre gli spettri registrati nell'Orbitrap presentano un minor rumore di fondo poiché contengono meno ioni che interferiscono durante l'analisi ed hanno quindi una risoluzione maggiore.

Elettroforesi bidimensionale (2-DE)

Le tecniche elettroforetiche sfruttano la migrazione di molecole dotate di carica elettrica sotto l'influsso di un campo elettrico. Le tecniche elettroforetiche monodimensionali non riescono a separare un numero di proteine superiori al centinaio e quindi, per poter analizzare i componenti di una miscela proteica complessa, come le matrici biologiche, è stato messo a punto l'accoppiamento di due tecniche elettroforetiche indipendenti. La versione classica della 2-DE deriva dalla combinazione di una separazione in base alla carica come prima dimensione, l'isoelettrofocalizzazione (IEF), che separa le proteine in base al loro punto isoelettrico (pI), e di una separazione in base alla massa come seconda dimensione, l'elettroforesi su gel di poliacrilammide in presenza di sodio dodecil solfato (SDS-PAGE) che separa le proteine in base al peso molecolare (PM). Fin dalla sua introduzione a metà degli anni '70, la 2-DE si è rivelata una tecnica particolarmente utile ed efficace nella separazione delle proteine in miscele complesse. L'elevata risoluzione della tecnica bidimensionale deriva dal fatto che la direzione di migrazione delle proteine nella seconda dimensione risulta ortogonale a quella nella prima dimensione, e quindi il numero di proteine che possono essere separate contemporaneamente all'interno di un gel bidimensionale è notevolmente superiore rispetto a quello delle tecniche monodimensionali. Allo stato attuale la 2-DE presenta elevata capacità di separazione: nelle condizioni sperimentali ottimali è in grado di risolvere più di 5000 proteine contemporaneamente (2000 proteine nelle analisi di routine), ed è in grado di rilevare quantità <1 ng di proteina per spot (Görg, 2004). La 2-DE permette di ottenere una mappa delle proteine intatte che riflette i cambiamenti nel livello di espressione delle proteine, rivelando la presenza di isoforme e PTMs. Molte mappe di riferimento sono disponibili sul sito Swiss-Prot 2D (<http://world-2dpage.expasy.org/swiss-2dpage>), e su altre banche dati.

Nelle fasi iniziali della 2-DE l'IEF veniva effettuata su sottili cilindri di gel di poliacrilammide contenenti anfoliti carrier che, a seguito della pre-elettroforesi, si disponevano a creare il gradiente di pH all'interno del gel (O'Farrel, 1975). La riproducibilità ottenuta con questa strategia operativa era, però, scarsa e, conseguentemente, il confronto tra analisi effettuate, non solo tra laboratori differenti, ma anche nello stesso laboratorio in tempi successivi, era molto problematico.

Un grande contributo allo sviluppo della 2-DE è stato l'introduzione nel 1982 delle immobiline, derivati dei monomeri di acrilammide che hanno come sostituenti del gruppo amminico uno o più gruppi acidi o basici. Le immobiline vengono utilizzate

come monomeri per la polimerizzazione del gel, permettendo la creazione di un gradiente di pH immobilizzato (IPG), estremamente stabile nel tempo (Görg, 2000). In questo modo viene evitato il fenomeno del *drift* catodico, che si verifica nei gradienti di pH creati con gli anfolti quando sottoposti all'applicazione di campi elettrici molto prolungati. Tale fenomeno comporta una ridotta stabilità del gradiente a livello dei pH più basici a causa del lento spostamento nel tempo degli anfolti verso il catodo. Nelle migliori condizioni sperimentali l'IPG può consentire la separazione di proteine il cui pI differisce anche solo di 0.001 unità di pH.

I principali passaggi della 2-DE sono schematizzati in Figura 3. Tra questi, la preparazione del campione è sicuramente uno dei punti più critici dell'intera analisi, dato che è importante eliminare molecole non proteiche interferenti e, nei casi in cui le proteine più abbondanti risultino particolarmente concentrate (ad esempio albumina nel plasma) è richiesta la loro eliminazione per consentire la rilevazione delle proteine meno rappresentate. Non esiste un protocollo universale da utilizzare per la preparazione di tutti i campioni a causa della molteplicità ed eterogeneità delle proteine. Fattori quali solubilità, dimensioni, carica, e pI delle proteine entrano in gioco nella scelta del protocollo più vantaggioso per la preparazione del campione. Generalmente è utile ricorrere a procedure di precipitazione selettive, per separare le proteine presenti nel campione da specie contaminanti - come sali, detergenti, acidi nucleici e lipidi - che interferiscono con i risultati dell'analisi elettroforetica. Sono molto utilizzate la precipitazione in acetone, la precipitazione in acido tricloroacetico (TCA) e la precipitazione in acetone e TCA. In queste condizioni le proteine precipitano, mentre le sostanze interferenti rimangono in soluzione e possono essere così eliminate (Burgess, 2009).

Un protocollo della 2-DE applicato generalmente all'analisi di miscele proteiche complesse utilizza per la prima dimensione un intervallo di pH 3-10, ma se le proteine della miscela sono distribuite in un intervallo di pH inferiore, si preferisce restringere l'intervallo di pH della prima dimensione. Per la seconda dimensione tipicamente vengono utilizzati gel in gradiente di T che migliorano la risoluzione della separazione delle proteine con PM molto diversificati (8000- 200000 Da).

Tipicamente, il riconoscimento delle proteine separate mediante la 2-DE richiede la digestione proteolitica dei singoli spot (*in-gel digestion*) e l'analisi in MS dei peptidi estratti dal gel.

Per questo tipo di esperimenti è importante che l'agente proteolitico garantisca la produzione di peptidi di dimensioni medie di circa 15-20 amminoacidi con una composizione amminoacidica che ne favorisca la ionizzazione e quindi permetta la

successiva analisi in MS. L'enzima che più soddisfa questi requisiti è la tripsina, un enzima proteolitico ad alta specificità che catalizza l'idrolisi del legame peptidico in corrispondenza del gruppo carbossilico di residui di lisina (K) o arginina (R).

Per l'identificazione delle proteine separate nel gel sono applicabili due approcci: il *peptide mass fingerprinting* (PMF) e la spettrometria di massa tandem.

Il PMF si basa sulla determinazione delle masse dei peptidi prodotti dalla digestione che sono utilizzate come una "impronta" identificativa della proteina genitrice. L'identificazione avviene attraverso il confronto delle masse dei peptidi triptici sperimentali con le masse di peptidi "virtuali" generati attraverso l'uso di strumenti bioinformatici da sequenze proteiche note depositate in banche dati. Con questo approccio l'accurata identificazione della proteina genitrice dipende dall'accuratezza dei valori di massa misurati per i suoi peptidi proteolitici e dal numero di peptidi rivelati: più alto è il numero dei peptidi rivelati, maggiore sarà la copertura della sequenza della proteina genitrice. Questo approccio fornisce l'identificazione della proteina più probabile su basi statistiche. Per rendere rigorosa l'identificazione della proteina è richiesta la determinazione dei valori di massa di almeno 4 peptidi con massa corrispondente a quella dei peptidi virtuali generabili a partire dalla proteina presente nella banca dati con un errore massimo accettabile di 10 ppm.

L'utilizzo del PMF richiede l'isolamento delle proteine in esame a partire dalla miscela proteica o comunque un frazionamento del campione che semplifichi la composizione della miscela. Questo approccio non fornisce informazioni sulla sequenza amminoacidica dei peptidi riconosciuti, perciò per validare l'identificazione ottenuta spesso viene richiesto di confermare l'identificazione della proteina con altre metodiche. Molto utilizzato in tal senso è l'*immunoblotting*, che permette l'identificazione certa di una determinata proteina sfruttando la sua interazione all'interno del gel elettroforetico con anticorpi specifici.

Nella spettrometria di massa tandem l'identificazione delle proteine è ottenuta mediante esperimenti di frammentazione MS/MS eseguiti dopo la separazione dei peptidi proteolitici per cromatografia liquida. Il pattern di frammentazione MS/MS che si ottiene è specifico e permette di identificare la proteina in esame sulla base della sequenza amminoacidica e non solo sulla base dei valori della massa dei peptidi proteolitici ottenuti. Gli spettri di frammentazione MS/MS raccolti durante l'analisi possono essere analizzati in modo automatico mediante programmi bioinformatici, come Sequest o Mascot, che confrontano i dati sperimentali con le sequenze proteiche presenti nelle banche dati. L'identificazione della proteina

genitrice è univoca se i peptidi sequenziati mediante analisi MS/MS sono peptidi “unici” (detti proteotipici), ossia peptidi specifici che possono essere ottenuti solo da quella specifica proteina e, in particolare, per un'identificazione non ambigua di una proteina è richiesto il sequenziamento di un numero minimo di due peptidi unici. Anche in questo caso l'identificazione della proteina avviene su considerazioni statistiche, ma le informazioni ottenute dalle sequenze amminoacidiche hanno un peso statistico sicuramente superiore rispetto ai valori di massa utilizzati nel PMF.

Gli aspetti più interessanti della 2-DE si riferiscono alla sua eccellente capacità di separare proteine in spot proteici definiti, permettendo un'analisi comparativa per individuare proteine sovra- o sotto espresse confrontando la stessa tipologia di campione prelevato in diverse condizioni fisio-patologiche. Inoltre, grazie al suo accoppiamento con la MS, è possibile ottenere una notevole quantità di informazioni dall'identificazione della specifica proteina e delle proteoforme che da quella data proteina possono originare, ad esempio a seguito di splicing alternativi. Tra le limitazioni della 2-DE sono da riportare la sua ridotta capacità di separare sia le proteine di membrana che, essendo idrofobiche, sono particolarmente difficili da risolubilizzare nel tampone di reidratazione (Chevalier, 2010), sia le proteine ad alto PM e le proteine basiche, e di evidenziare le proteine presenti a basse concentrazioni che risultano mascherate dalle proteine più abbondanti (Greenough, 2004; Yamada, 2002). Un'altra limitazione della 2-DE è la ridotta riproducibilità di questa tecnica per l'esistenza di una variabilità dovuta alla preparazione del gel o del campione (Weist, 2008), a problemi nella migrazione e nella colorazione dovuti alle specifiche proprietà delle diverse proteoforme delle proteine. Proprio a causa dell'alta variabilità della tecnica, è necessario esaminare un numero significativo di campioni diversi (repliche biologiche) per evidenziare differenze statisticamente significative. Al fine di agevolare il confronto tra le diverse mappe 2-DE recentemente è stata sviluppata una versione avanzata della classica 2-DE, la 2-D DIGE (*Difference gel electrophoresis*, Unlü, 1997). Questa strategia prevede che fino a tre diversi campioni proteici possano essere analizzati nello stesso 2-DE gel dopo essere stati incubati in presenza di cromofori fluorescenti diversi (Cy3, Cy5 e Cy2) e miscelati in quantità uguali. Al termine dell'analisi, il gel viene scansionato in successione alle lunghezze d'onda di eccitazione di ciascun cromoforo, ottenendo la quantificazione delle diverse specie proteiche presenti in ciascun campione, eliminando in tal modo la variabilità inter-

gel dal momento che le proteine dei diversi campioni (ad esempio sano vs. malato) vengono analizzate nello stesso gel.

Per poter confrontare un numero di campioni superiore a 3, nella 2D-DIGE si può prevedere di includere uno standard interno che viene preparato miscelando alcuni o tutti i campioni. In questo modo si misura l'abbondanza di una proteina in ciascun campione rispetto allo standard interno. Dato che la quantità di ciascuna proteina nello standard interno è nota, in questo modo si riduce la variabilità inter-gel. Le eventuali differenze quantitative tra i campioni sono quindi evidenziate da uno sbilanciamento della fluorescenza dei diversi cromofori rilevabile a seguito della sovrapposizione delle immagini.

Data l'estrema complessità di un proteoma, che può contenere migliaia di proteine diverse, ad oggi non è ancora stata messa a punto una tecnica in grado di consentire da sola di fornire tutte le informazioni per una corretta identificazione di tutte le proteine presenti nel campione, comprese le eventuali le PTMs e i loro rapporti quantitativi, ma l'accoppiamento della 2-DE e della MS risulta estremamente vantaggioso in quanto permette di visualizzare una mappa delle proteine presenti in un dato campione biologico in quel preciso momento, separando, quantificando e identificando un gran numero di proteine.

Piattaforme top-down e bottom-up

Le strategie utilizzate in proteomica per la caratterizzazione del profilo proteico vengono classificate come piattaforme *top-down* o *bottom-up* (Tipton, 2011; Cui, 2011).

L'approccio *top-down* è l'analisi proteomica di proteine intere o in forma di grandi frammenti ottenuti durante la spettrometria di massa tandem. Le proteine vengono analizzate evitando il più possibile qualsiasi alterazione del campione, separate in un sistema analitico o preparativo e frammentate direttamente nello spettrometro di massa, che, oltre a fornire il dato del PM della proteina intera, ne permette la caratterizzazione attraverso l'identificazione della sequenza amminoacidica in modalità di frammentazione MS/MS o MSⁿ.

La prima fase dell'analisi prevede quindi un opportuno pretrattamento del campione con lo scopo di rimuovere le proteine più abbondanti e di elevato PM, rendendo possibile la caratterizzazione delle proteine più piccole e meno rappresentate e dei peptidi. Generalmente l'analisi delle proteine in forma intera viene effettuata sulla frazione proteica e peptidica solubile del campione in esame, per esempio un fluido biologico o un tessuto trattati con soluzioni acide, da sole o in miscela con solventi organici, dopo miscelazione o omogeneizzazione (nel caso dei tessuti) e centrifugazione. In quest'ultima fase, unitamente alla frazione solubile, si ottiene, nella maggior parte dei casi, un precipitato insolubile, spesso visibile anche ad occhio nudo, contenente le specie non solubili nella miscela di estrazione utilizzata, in generale costituito dalle proteine più abbondanti e di elevato PM. La frazione solubile può essere analizzata con tecniche separative (tipicamente cromatografia liquida o elettroforesi capillare) accoppiate *on line* alla spettrometria di massa.

L'approccio *top-down* è generalmente l'approccio principalmente utilizzato per la caratterizzazione comparativa del contenuto proteico/peptidico di campioni biologici in termini qualitativi e quantitativi (ad esempio: campioni di soggetti sani da confrontare con quelli di soggetti affetti da una certa patologia), permettendo di individuare proteoforme che possono essere di fondamentale aiuto nell'individuare i meccanismi responsabili dell'insorgenza della patologia, strategie di diagnosi precoce, arrivando anche a individuare possibili biomarcatori della patologia in esame

L'approccio *top-down* presenta dei limiti di applicazione per campioni molto complessi, contenenti proteine ad elevato PM (> 50 kDa), con un contenuto proteico/peptidico caratterizzato da un ampio intervallo dinamico di concentrazioni, (McLafferty, 2007).

L'analisi dei dati ottenuti risulta sempre più complessa all'aumentare del PM della proteina oggetto di studio, in quanto i peptidi con masse superiori ai 3000-4000 Da di solito non forniscono spettri di frammentazione MS/MS completi, a causa della frammentazione parziale, e quelli ottenuti, sono difficilmente interpretati da parte dei programmi automatizzati utilizzati nell'analisi dei dati prodotti dalle strategie *bottom-up*. Per questo motivo gli spettri di frammentazione MS/MS di peptidi con masse superiori ai 3000-4000 Da devono essere analizzati manualmente.

Un grande vantaggio della strategia *top-down* è la possibilità di effettuare la quantificazione comparativa di proteine/peptidi evitando le reazioni con agenti derivatizzanti (*label-free*), senza alcuna limitazione nel numero delle proteine e peptidi oggetto di studio. Ad esempio in LC-ESI-MS, tutti gli ioni generati da una proteina alla sorgente ESI possono essere selettivamente estratti dal profilo di corrente ionica totale (TIC). Questa strategia, che prende il nome di corrente ionica estratta (XIC), permette di localizzare nel profilo HPLC il picco appartenente ad una specifica proteina o peptide, ricercando i valori m/z degli ioni generati da quella specifica proteina o peptide. L'area di questo picco è direttamente proporzionale alla quantità della proteina in esame in un vasto intervallo dinamico che va da circa 100 femtomoli a 100 picomoli (Messana, 2008).

Dato che la protonazione di biopolimeri ad alto PM non genera un'unica specie molecolare carica, ma più specie cariche, gli spettri ESI delle proteine ad elevato PM forniscono molti valori m/z , mentre le molecole più piccole, e quindi i piccoli peptidi, generano di solito pochi segnali che corrispondono generalmente ad uno o due valori m/z . Nelle procedure XIC è quindi richiesta una scelta accurata dei valori m/z da utilizzare al fine di evitare di utilizzare per la ricerca valori m/z condivisi con altre proteine che eluiscono in zone limitrofe.

Analogamente ad altre piattaforme, anche la strategia *top-down* presenta alcuni svantaggi specifici. Una limitazione è relativa all'impossibilità di analizzare con questa strategia le frazioni insolubili nelle soluzioni idro-organiche acide utilizzate per facilitare la protonazione delle proteine e il processo di *electrospray*. Di conseguenza la caratterizzazione proteomica di un certo fluido biologico non può essere effettuata esclusivamente con l'approccio *top-down* perché non è possibile analizzare tale frazione insolubile derivante dal pretrattamento del campione. In questa tesi è stato messo a punto un sistema di analisi delle frazioni insolubili che risultano dal pretrattamento in ambiente acido di campioni di saliva e di estratti di tessuto cerebrale al fine di integrare l'analisi proteomica *top-down* effettuata mediante analisi HPLC-ESI-MS.

La strategia *bottom-up* invece, si basa sulla possibilità di identificare le proteine a partire dall'analisi dei frammenti proteolitici che si generano a seguito della digestione della proteina utilizzando un enzima proteolitico, e collegando ogni frammento rilevato alla proteina genitore.

Infatti, considerato che ogni proteina è caratterizzata da una specifica sequenza amminoacidica, a seguito della digestione enzimatica si ottengono un certo numero di peptidi specifici e riconducibili solo alla proteina che li ha generati (prototipici).

Questa procedura si applica sia alle proteine separate su gel (SDS-PAGE o 2-DE), sia alle proteine in soluzione (metodologie *gel-free*). Nel caso in cui l'identificazione delle proteine presenti in miscele complesse preveda la digestione triptica del campione seguita da un'analisi in cromatografia mono o multidimensionale (o elettroforesi capillare) ad alta risoluzione accoppiata online con un'analisi ESI-MS (Castagnola, 2012) la strategia *bottom-up* applicata è indicata come *shotgun*.

Questo approccio è diventato uno dei più utilizzati per identificare le proteine nella maggior parte degli studi proteomici su larga scala, ma perché l'analisi possa avere successo è richiesta una notevole capacità risolutiva dei sistemi separativi. La natura peptide-centrica della proteomica *shotgun* complica l'analisi e l'interpretazione biologica dei dati, perché la stessa sequenza peptidica può essere presente in più proteine diverse e in organismi eucarioti superiori, a causa della presenza delle PTMs, in più proteoforme della stessa proteina. Tali peptidi condivisi possono portare a riconoscimenti e determinazioni ambigue dell'identità delle proteine presenti nel campione in esame (Nesvizhskii, 2005). Il limite dell'approccio *shotgun* è relativo soprattutto all'impossibilità di evidenziare i tagli proteolitici che molte proteine subiscono durante la loro maturazione, così come proteoforme della stessa proteina o derivati da splicing alternativi.

L'enzima proteolitico scelto può inoltre tagliare la proteina in siti diversi da quelli codificati o mancare dei siti di tagli, generando peptidi non attesi. Questo inconveniente nel processo di digestione può portare ad una mancata identificazione dei peptidi da parte dei software automatizzati. Inoltre, molti peptidi spesso co-eluiscono ed è più probabile che vengano selezionati i frammenti delle proteine più abbondanti rispetto a quelli derivanti da proteine meno abbondanti; quindi, spesso non si riesce ad ottenere la copertura completa nel sequenziamento di una proteina. Lo svantaggio più critico dell'approccio *shotgun* è quindi la mancanza di informazioni relative alla struttura del proteoma intatto.

La strategia *bottom-up* offre una copertura del proteoma maggiore rispetto a quella che si può ottenere con l'approccio *top-down* e per questo motivo viene ampiamente usata routinariamente nei laboratori di proteomica.

Tra le strategie *bottom-up* rientrano anche le piattaforme che utilizzano la 2-DE. In questi esperimenti, rispetto allo *shotgun*, c'è il vantaggio che la separazione avviene sulle proteine intatte, e solo successivamente le singole proteine vengono digerite, e identificate.

Anche se attualmente l'approccio *bottom-up* è la strategia più applicata e ampiamente standardizzata, l'analisi proteomica *top-down* viene sempre di più applicata per lo studio delle modifiche post-traduzionali delle proteine, l'identificazione di proteoforme, l'analisi di proteine in forma nativa, lo studio delle interazioni proteina-proteina e il “*de novo sequencing*”, anche grazie alla continua e rapida evoluzione e implementazione di risolutive tecniche analitiche separative e sofisticati spettrometri di massa. Tuttavia, l'approccio *top-down* soffre di alcune limitazioni, tra cui la necessità di operare con costosi strumenti ad alta risoluzione e la non soddisfacente disponibilità di algoritmi per l'analisi dei dati, da cui il vantaggio dell'integrazione delle metodologie *top-down* e *bottom-up*, che con le loro caratteristiche e i loro limiti risultano complementari e di supporto reciproco nel permettere la caratterizzazione del proteoma.

Attualmente è ragionevole ritenere che una piattaforma che possa essere considerata universale è ancora lontano dall'essere stata sviluppata (Castagnola, 2012).

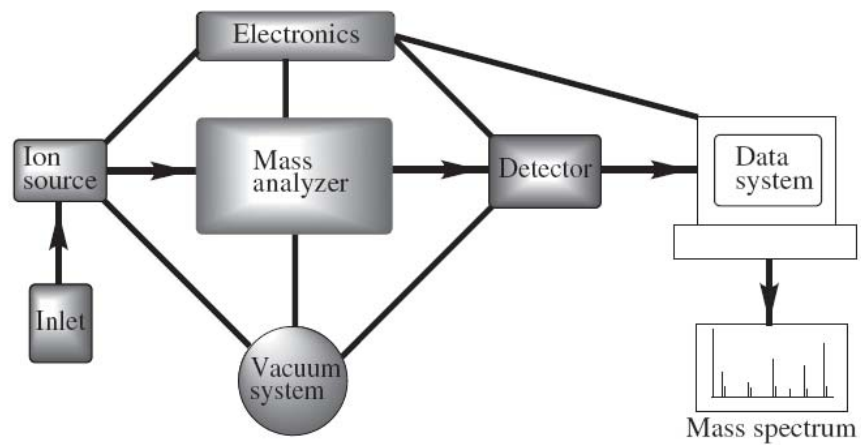


Figura 1: Schema dei componenti di uno spettrometro di massa

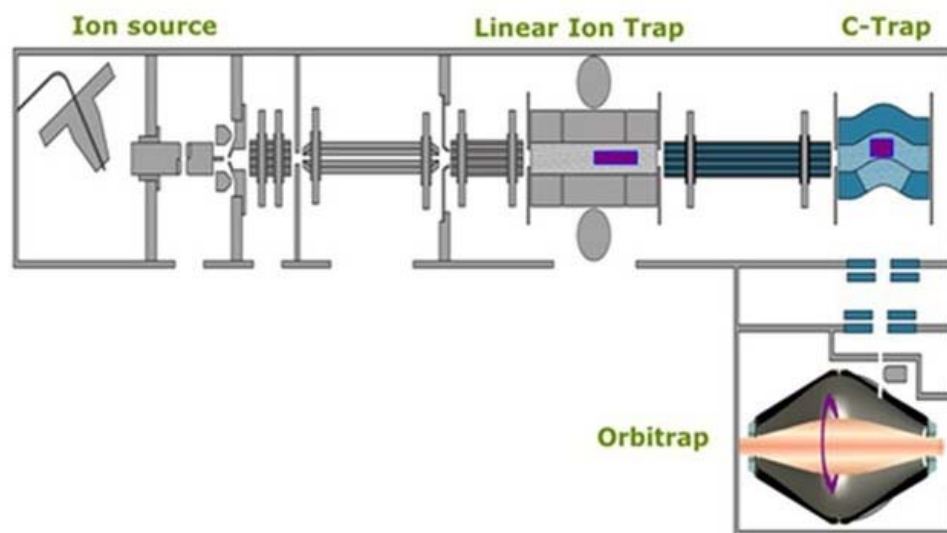


Figura 2: Schema dei componenti di uno spettrometro di massa LTQ-Orbitrap

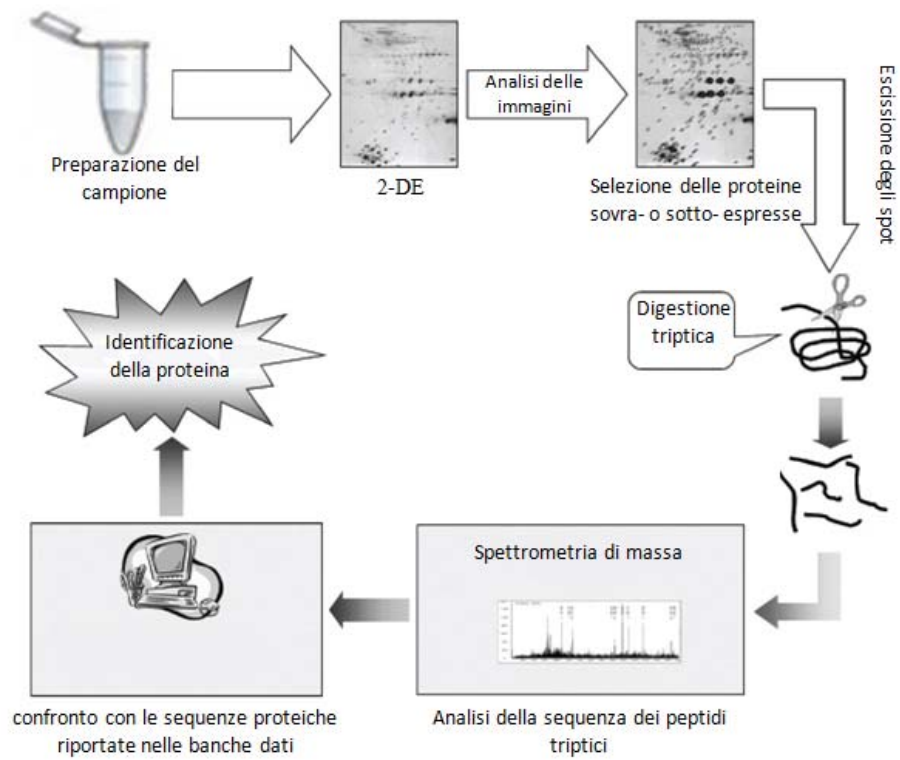


Figura 3: Schema per l'identificazione in spettrometria di massa delle proteine separate mediante elettroforesi bidimensionale

Parte I

Caratterizzazione proteomica della frazione insolubile in acido della saliva di neonati pretermine e confronto con la saliva di neonati a termine e adulti

1.1 Introduzione: la Saliva

La saliva è il fluido corporeo che idrata continuamente la mucosa della cavità orale, dell'orofaringe e della laringe. Il termine “saliva” viene generalmente utilizzato per indicare al prodotto della secrezione delle ghiandole salivari, mentre con il termine “saliva intera” si fa riferimento ad una soluzione complessa che comprende, oltre al secreto delle ghiandole salivari maggiori e minori, il fluido gengivale crevicolare, un essudato costituito da siero alterato che si trova nel solco gengivale, le secrezioni mucose della cavità nasale e faringea, la popolazione batterica orale non aderente, residui alimentari, cellule epiteliali desquamate e cellule del sangue (Humphrey, 2001).

La saliva svolge ruoli estremamente diversificati: ha la funzione di proteggere i tessuti del cavo orale mantenendoli umidi, fornire un secreto mucoso lubrificante, mantenere un ambiente con una elevata concentrazione di calcio e fosfato e elevata capacità di tamponare gli acidi. È molto importante nella masticazione, nella deglutizione, nella formazione del bolo alimentare (è anche responsabile della prima fase della digestione dell'amido), nella percezione del gusto, rende possibile la mineralizzazione del dente (alcune sue componenti controllano l'omeostasi del calcio orale e la sua mobilizzazione) e facilita la fonazione (Sreebny, 2000). Contiene sostanze antibatteriche che forniscono protezione contro i microorganismi che entrano nel tubo digerente attraverso la bocca, e possiede, oltre all'attività antibatterica, anche attività antifungina ed antivirale (Mandel, 1987), mediata da un considerevole numero di proteine e peptidi con obiettivi molecolari estremamente specifici.

La saliva è composta in massima parte di acqua (98-99%), contiene notevoli quantità di proteine, tra cui enzimi, come amilasi, lisozima, lipasi, fosfatasi acida, lattoperossidasi, superossido dismutasi, vari ormoni peptidici, glucidi (glicosamminoglicani costituenti principali delle secrezioni mucose), lipidi (ormoni come il testosterone e il progesterone) e ioni inorganici come sodio, cloruro, potassio, calcio, magnesio, bicarbonato e fosfato. Contiene inoltre cellule derivate dall'esfoliazione dell'epitelio del cavo orale, cellule batteriche (microflora commensale e/o patogena) e virus. Contiene inoltre composti derivati dal sangue come proteine plasmatiche, eritrociti e leucociti (Aps, 2005) e residui di cibo durante o immediatamente dopo la sua assunzione.

Le ghiandole salivari sono classificate, in base alle cellule secernenti che le caratterizzano ed in base al tipo di secreto prodotto, in:

- Sierose: il secreto è un fluido acquoso povero in mucine e ricco in ptialina;

- Mucose: il secreto è ricco in mucine;
- Miste.

La secrezione della saliva umana è dovuta in massima parte (> 90%) all'attività di tre paia di ghiandole salivari maggiori (Figura 4): le parotidi (a secrezione sierosa), le ghiandole sottolinguali (a secrezione prevalentemente mucosa) e le ghiandole sottomandibolari (a secrezione prevalentemente sierosa).

Le parotidi sono due ghiandole pari situate nella loggia parotide, lobulari tubulo-acinose, che producono un secreto principalmente sieroso. La loro secrezione basale (prima del pasto) contribuisce solo in minima parte (25%) alla secrezione totale della saliva, mentre il loro contributo aumenta notevolmente dopo il pasto, fino a raggiungere il 50% della saliva totale (dopo il pasto il flusso è di 6-7 mL/min). Le proteine e i peptidi di provenienza parotide, finora identificati, ammontano a circa un migliaio.

Le ghiandole sottomandibolari (pari) presentano un corpo a forma di mandorla, sono situate in una loggia fibrosa, chiamata loggia sottomandibolare, e sono delle ghiandole miste costituite per il 90% da acini a secrezione sierosa e per il 10% a secrezione mucosa.

Le ghiandole sottolinguali (pari) sono situate a ridosso della fossetta sottolinguale della mandibola e sono delle ghiandole miste che presentano un secreto principalmente mucoso. Il loro secreto è ricco in mucina e povero in ptialina, rispetto a quello delle altre ghiandole salivari maggiori.

Alla secrezione salivare contribuiscono in modo sostanziale anche un gran numero di ghiandole più piccole (da 450 a 750) dette minori o accessorie (labiali, vestibolari, palatine, glosso palatine), situate sulla superficie interna delle guance, sul palato duro, sul palato molle e sulla superficie della lingua. Con l'eccezione delle ghiandole linguali, il secreto delle ghiandole accessorie è principalmente mucoso, ricco di mucoproteine e la sua funzione principale è quella di lubrificante.

Le ghiandole salivari si originano tra la sesta e l'ottava settimana di gestazione e le prime a svilupparsi sono le parotidi. L'unità secretoria salivare classica è formata da una struttura ramificata di tipo tubulo-acinare.

Almeno l'85% delle proteine neosintetizzate nelle ghiandole salivari viene immagazzinato in granuli secretori e poi secreto attraverso un processo di esocitosi dei granuli. Le cellule acinari sono riccamente innervate da fibre nervose del simpatico e del parasimpatico, e i due tipi di stimolazione hanno differenti effetti: il parasimpatico evoca la secrezione di volumi di saliva maggiori ma con minor contenuto proteico, il simpatico invece stimola la produzione di minori volumi con

una grande concentrazione proteica (Castle, 1998). Nel corso della giornata la secrezione delle ghiandole minori è continua; anche la secrezione basale delle ghiandole maggiori è cospicua. Queste ultime rispondono in maniera più (parotide) o meno (sottomandibolare) marcata alla stimolazione parasimpatica derivante da stimoli fisici, chimici e psicologici.

Pertanto, la composizione della saliva varia in relazione alla particolare miscela di secreti prodotti dalle diverse ghiandole con diversi gradi di stimolazione. Nella saliva prodotta in assenza di stimoli prevale il contributo delle ghiandole sottomandibolari, mentre nella saliva prodotta sotto stimoli (ad esempio acidi) il contributo della ghiandola parotide è maggiore. È stato dimostrato che esiste un andamento circadiano nella produzione della saliva (Dawes, 1975; Ferguson, 1980), con un picco di produzione (acrofase) nelle ore post-prandiali. L'uomo adulto può produrre fino a 2 L di saliva al giorno (>1 mL/min).

La saliva è ipotonica rispetto al plasma: in generale la concentrazione di sodio e cloro salivare è inferiore rispetto a quella plasmatica, mentre quella di bicarbonato e potassio è superiore. Tra le componenti inorganiche citate, particolare rilievo hanno il bicarbonato (utile per tamponare il pH intorno alla neutralità tra 6,5 e 7,4), il calcio e il fosfato (che mantengono l'integrità del dente). Nella protezione della superficie dello smalto dei denti sono coinvolti anche il fluoro e il tiocianato, il primo grazie alla sua capacità di scambiarsi nel reticolo di idrossiapatite, mentre il secondo in virtù della sua azione antibatterica che si esplica quando viene convertito in ipotiocianato dalla lattoperossidasi salivare.

Nella saliva intera, il calcio ha tipicamente una concentrazione di 1,4 mmoli/L, solo il 50% circa è presente nella saliva in forma ionica, il 40% è presente in forma di complessi con altri ioni mentre il 10% circa si trova legato a proteine salivari. Nelle secrezioni delle singole ghiandole, la concentrazione del calcio aumenta all'aumentare del flusso, probabilmente a causa del maggior apporto di saliva parotidea in casi di flusso elevato.

Il fosfato ha una concentrazione di circa 6 mmoli/L ed è quasi tutto presente in forma ionica, ad eccezione di un 10% che si ritrova sotto forma di fosfato organico. La sua concentrazione tende a diminuire all'aumentare del flusso e la sua concentrazione è particolarmente bassa nel secreto delle ghiandole salivari minori.

Per quanto riguarda invece il bicarbonato, nella saliva di tutte le ghiandole in condizioni basali, la sua concentrazione è bassa, ma aumenta notevolmente in caso di flusso elevato. Il bicarbonato è il principale tampone della saliva, porta il pH della saliva prodotta dopo stimolazione ad un valore di circa 8, inoltre rappresenta

una difesa efficace contro i prodotti acidi dei batteri cariogeni. Deriva dall'anidride carbonica generata dall'attività metabolica delle ghiandole salivari.

In particolare, alcuni studi si sono concentrati sulla composizione salivare degli ioni nella saliva dei neonati e dei bambini, ed è risultato che le concentrazioni di elettroliti salivari (Na^+ , K^+ , Ca^{2+} e Mg^{2+}) nei bambini sani differiscono da quelle presenti negli adulti e inoltre che nei neonati è presente una grande variabilità. Non sono state riscontrate significative differenze nella composizione tra maschi e femmine. La concentrazione di Na^+ , Cl^- e K^+ è risultata più elevata alla nascita ed è stata riportata una loro diminuzione all'aumentare dell'età. Anche le concentrazioni di Ca^{2+} e Mg^{2+} sono risultate più alte durante il primo anno di età, mentre la concentrazione di fosfato è inferiore alla nascita ed aumenta con l'età (Ben-Aryeh, 1984).

Anche se tra i vari fluidi corporei la saliva risulta di grande interesse per la facilità e non invasività dei metodi di raccolta, il suo utilizzo come fluido diagnostico e prognostico è relativamente recente. Per poter comprendere il significato funzionale della saliva, è necessario definire la sua composizione proteica, individuare e caratterizzare le proteine salivari, conoscere le modificazioni subite dai peptidi e dalle proteine durante la secrezione e stabilire quali modificazioni ne determinano la forma biologicamente attiva. È necessario quindi caratterizzare la composizione proteica in soggetti sani, in modo da poterla poi confrontare con quella di soggetti affetti da specifiche patologie al fine di valutare le modificazioni che si osservano.

Le difficoltà che si oppongono ad una chiara ed esaustiva definizione del proteoma e del peptidoma salivare sono legate principalmente all'elevata variabilità di questo fluido biologico. Infatti la presenza di un'elevata variabilità fisiologica sia inter-individuale che dipendente da numerosi fattori, come i ritmi circadiani e l'età (Cabras, 2009), e di una importante variabilità genetica (polimorfismi nelle sequenze geniche codificanti per le proteine salivari, mutazioni individuali, fenomeni di splicing alternativo e PTMs) ha sicuramente giocato a sfavore dell'utilizzo clinico del fluido salivare per scopi diagnostici e prognostici.

L'analisi completa del proteoma salivare umano e la caratterizzazione delle diverse proteoforme delle proteine presenti è un passo importante per poter ipotizzare l'uso della saliva a fini diagnostici e prognostici.

Negli ultimi anni numerosi studi proteomici hanno contribuito a definire la composizione del proteoma salivare e finora sono state caratterizzate più di 2400 proteine e più di 2000 peptidi (Amado, 2010; Bandhakavi, 2009) e a seguito di

questi risultati, la saliva sta assumendo una sempre crescente rilevanza come importante fluido diagnostico (Castagnola, 2011a).

Le proteine presenti nella saliva possono essere distinte in proteine di origine secretoria (circa 400 componenti) e non-secretoria (più di 2000, Cabras, 2014). Le proteine secrete rappresentano più del 90% in peso della saliva, mentre il restante 10% è rappresentato da numerosissime proteine di origine diversa, presenti a concentrazioni nettamente inferiori. La maggior parte di queste proteine derivano probabilmente da cellule di esfoliazione, da essudati del tessuto epiteliale (per esempio l'albumina sierica), dal fluido crevicolare gengivale (α -defensine e β -timosine), mentre altre sono di origine esogena, derivando dai contributi della microflora batterica orale.

Le più abbondanti proteine di origine secretoria appartengono alle seguenti famiglie:

- Proteine ricche in Prolina (PRP), ulteriormente suddivise in acide (aPRPs), basiche (bPRPs) e basiche glicosilate (gPRPs), costituiscono una classe di proteine poliedriche a causa del numero di varianti e di PTMs, come glicosilazioni, fosforilazioni e scissioni proteolitiche (Castagnola, 2012);

- α -amilasi salivare: una delle maggiori componenti della secrezione salivare e anche una delle proteine salivari con il più complesso pattern di proteoforme e PTMs. Infatti circa il 25% della α -amilasi esiste come forma glicosilata (Hirtz, 2005) e sono state identificate numerose proteoforme troncate;

- mucine salivari: vengono raggruppate in due sottofamiglie: MUC 5B (mucine polimeriche ad alto PM che tendono a formare una sorta di gel) e MUC7 (di PM inferiore e non polimeriche). Entrambe le sottofamiglie sono ampiamente O- e N-glicosilate.

- Cistatine salivari (tipo S): appartengono alla sottofamiglia delle cistatine di tipo 2, comprendente le cistatine S, SA, SN e C. Mentre la cistatina SA sembra essere espressa esclusivamente nella cavità orale, le cistatine S e SN sono state rilevate anche in altri fluidi corporei e in altri organi, mentre la cistatina C è ubiquitaria (Dickinson, 2002).

- Istatine: famiglia alla quale appartengono dei piccoli peptidi secreti dalle ghiandole salivari maggiori. Tutti i membri di questa famiglia derivano dai peptidi His-1 e His-3. L'His-1 è fosforilata sul secondo residuo di Ser ed è parzialmente solfatata sulla quarta Tyr C-terminale (nella sottomandibolare). L'His-3 invece può generare His-6 e His-5 in base al taglio proteolitico subito (Troxler, 1990). Questa

classe di proteine possiede attività antifungina e insieme allo ione HCO_3^- contribuisce al mantenimento del corretto pH della saliva (Holbrook, 1975).

- Staterina e peptide PB: la staterina è una piccola proteina fosforilata sul secondo e terzo residuo di Ser, è specifica della saliva ed è coinvolta nell'omeostasi orale dello ione calcio (Schlesinger, 1977). Il peptide PB appartiene invece alla famiglia delle bPRP, ma viene codificato da un gene differente (PROL3) localizzato nel cromosoma 4q13.3, molto vicino al locus genico dal quale deriva la staterina (Isemura, 2000). Inoltre, come la staterina, è secreto dalla ghiandole sottomandibolare e sublinguale.

Tra le più importanti famiglie di proteine di derivazione non secretoria ritroviamo le Timosine (β -4 e β -10), le defensine, le Cistatine A e B e le proteine S100 (A7, A8, A9, A11 e A12).

- Timosine β -4 e β -10: sono peptidi ubiquitari con funzioni intra- ed extra cellulari la cui via di secrezione non è stata ancora del tutto chiarita. La timosina β -4 è un potente iniziatore dell'angiogenesi così come la timosina β -10 (Huff, 2001);

- Defensine: Le defensine sono una famiglia di peptidi antimicrobici di origine granulocitaria non specifici del cavo orale che agiscono da regolatori del volume cellulare e della chemiotassi e da inibitori dell'attività delle cellule natural-killer. Sono state evidenziate nel plasma, nel tessuto infiammatorio in genere, nel midollo osseo, nella pelle e nelle cellule intestinali e nella saliva. Le α -defensine (HNP, *Human Neutrophil Peptides*) sono una classe di peptidi salivari basici ricchi di cisteina e tirosina, dotati di tre ponti disolfuro e secreti dai granulociti neutrofili. Un recente studio (Goebel, 2000) riporta che nella saliva umana sono presenti le α -defensine HNP-1, HNP-2, HNP-3, e le β -defensine 1 e 2. È dimostrato che la sorgente delle α -defensine salivari è il fluido crevicolare gengivale (GFC), di cui le α -defensine 1, 2, 3, e 4 sono le maggiori componenti, rappresentando il 40% del contenuto proteico.

- Cistatine A e B (Stefine): sono proteine che funzionano da inibitori intracellulari degli enzimi proteolitici quali la papaina e le catepsine L, S, H (lisosomiali); oltre che nella saliva, sono presenti anche in altri fluidi corporei;

- Proteine S100: sono piccole proteine con PM di ~12 kDa, presenti esclusivamente nei vertebrati e identificate per la prima volta nel 1965 (Moore, 1965). Esse possiedono la capacità di legare ioni calcio e di interagire tra di loro a formare omodimeri ed etero-dimeri; la loro presenza nella saliva esplica indirettamente un'attività antibatterica in quanto sottraggono dall'ambiente metalli di transizione

come manganese, ferro, zinco, utili alla proliferazione dei microrganismi patogeni (Heizmann, 1998).

Un importante contributo alla variabilità di un qualsiasi proteoma è dato dalle modificazioni fisiologiche che si verificano secondo la dimensione temporale. L'indagine dei cambiamenti proteomici in funzione del tempo è stata definita come "crono-proteomica" (Messana, 2015).

Ad oggi pochi lavori si sono focalizzati sulla determinazione della composizione proteomica della saliva in età infantile. Alcuni studi realizzati negli ultimi anni hanno dimostrato come le principali classi di proteine salivari seguano un'evoluzione nel soggetto in età pediatrica, fino ad assumere i caratteri tipici dell'adulto intorno al diciottesimo anno di età (Cabras, 2009), che la variabilità individuale del proteoma salivare è molto alta in età pediatrica, soprattutto nei neonati pretermine (Messana, 2015).

Uno studio pubblicato di recente ha esaminato la fascia di età 0-48 mesi, mediante una piattaforma top-down basata su esperimenti di RP-HPLC-ESI-MS (Manconi, 2013) da cui è emerso che durante il primo anno di vita il cavo orale dei bambini subisce importanti cambiamenti, e in particolare che il proteoma salivare dei bambini in questa fascia di età mostra importanti modifiche quantitative.

Un altro lavoro ha analizzato i cambiamenti nella composizione proteica salivare nella fascia di età 3-6 mesi mediante SDS-PAGE accoppiata alla MALDI-TOF/TOF-MS (Morzel, 2011), evidenziando anche in questo importanti cambiamenti nella composizione salivare.

Uno studio di Inzitari e coll. (2007) ha analizzato la variazione della concentrazione di diverse proteoforme di PRP acide nei neonati pretermine e a termine seguiti nel corso del primo anno di vita. Dallo studio emerge evidente che il campionamento della saliva può essere eseguito immediatamente dopo la nascita sia sul neonato prematuro che sul neonato a termine senza causare stress o dolore, evidenziando ulteriormente l'utilità del prelievo di saliva nel monitoraggio dello stato clinico di questa tipologia di pazienti particolarmente vulnerabili, che possono in questo modo evitare prelievi di sangue frequente.

Uno studio completo e mirato all'analisi della composizione proteica e peptidica della frazione solubile in acido della saliva di neonati prematuri mediante analisi HPLC-ESI-MS è stato pubblicato negli ultimi anni (Castagnola, 2011b). I risultati di questo lavoro hanno evidenziato che la composizione proteica e peptidica della frazione solubile in acido della saliva di neonati prematuri è molto diversa rispetto a quella dell'adulto, poiché nel bambino prematuro sono state rilevate 40 specie

proteiche/peptidiche i cui valori di massa non risultano presenti nell'adulto. Un certo numero di queste proteine e peptidi sono state identificate e caratterizzate utilizzando la tecnica HPLC-ESI-MS mediante l'approccio top-down (Castagnola, 2011b). Inoltre è stato possibile sottolineare una diminuzione della concentrazione di 25 di queste 40 proteine in funzione dell'aumento dell'età post-concezionale del neonato (PCA). La frazione insolubile in acido non può essere però analizzata con questo approccio proprio a causa della sua insolubilità nelle condizioni sperimentali utilizzate per l'analisi. E' possibile analizzare tale frazione insolubile mediante elettroforesi bidimensionale dopo solubilizzazione del campione in opportune condizioni sperimentali. Al fine di ottenere informazioni in grado di integrare i risultati ottenuti nel lavoro sopracitato, completando così il quadro proteomico della saliva dei neonati prematuri, in questo lavoro di tesi sono state analizzate le frazione insolubili in acido della saliva di neonati prematuri mediante elettroforesi bidimensionale, in modo da poter confrontare le mappe bidimensionali con quelle ottenute da neonati a termine e adulti.

1.2 Materiali e Metodi

1.2.1 Raccolta dei campioni di saliva, estrazione peptidi e proteine solubili in acido e lavaggio della frazione insolubile in acido.

I campioni utilizzati per questo studio sono stati raccolti da 7 bambini prematuri, 6 bambini a termine e 6 adulti, utilizzando una procedura di prelievo assolutamente non invasiva. La raccolta del campione è stata effettuata esclusivamente nei casi in cui la procedura di prelievo non causava alcun segno di stress al neonato.

Il protocollo e la scheda per la richiesta del consenso informato per la partecipazione alla ricerca descritta in questo studio sono stati approvati dal Comitato Etico del Dipartimento di Pediatria e dal Comitato Etico Medico della Facoltà di Medicina dell'Università Cattolica di Roma, in accordo con quanto previsto dalla Dichiarazione di Helsinki. Il consenso alla partecipazione è stato firmato dai genitori dei bambini che hanno partecipato allo studio.

Per i soggetti adulti il prelievo è avvenuto al mattino tra le 8 e le 10, a digiuno, in assenza di stimolazione, con un metodo diretto: il donatore è stato invitato a raccogliere la saliva nel cavo orale, nella parte anteriore della bocca davanti al frenulo linguale, e quindi a prelevarla per mezzo di una pipetta Pasteur monouso in plastica. La saliva così prelevata è stata trasferita all'interno di una provetta mantenuta in ghiaccio in modo da minimizzare il più possibile l'azione proteolitica delle proteasi presenti naturalmente nel campione, e immediatamente dopo il prelievo ciascun campione di saliva è stato miscelato in rapporto 1:1 (v/v) con una soluzione acquosa di acido trifluoroacetico (TFA) allo 0,2 % (v/v).

I campioni sono stati immediatamente centrifugati a 9000 g per 3 minuti a 4°C in modo da separare il supernatante, contenente la frazione proteica solubile in acido dal pellet, che contiene la frazione proteica insolubile in acido. I campioni così ottenuti sono stati conservati a -80°C, sino al momento dell'analisi.

I campioni di saliva dei neonati prematuri sono stati ottenuti da 7 soggetti aventi un'età post-concezionale (PCA) alla nascita compresa tra 24 e 28 settimane (PCA media $185,6 \pm 11,8$ giorni) ricoverati all'Unità di Cure Intensive Neonatali della Facoltà di Medicina e Chirurgia dell'Università Cattolica del Sacro Cuore. I neonati donatori prescelti non presentavano malformazioni o infezioni prenatali. I prelievi sono stati effettuati sullo stesso neonato ad intervalli più o meno regolari di (10-15 giorni) per diverse settimane dopo la nascita, durante le visite di controllo. Sono stati inoltre introdotti nello studio 6 neonati sani nati al termine di una gestazione

senza complicazioni (PCA media $282,8 \pm 12,8$ giorni) e per confronto è stata raccolta la saliva anche di 6 adulti sani non fumatori ($31,8 \pm 12,8$ anni).

In un primo set di esperimenti i campioni sono stati trattati direttamente con la soluzione di risolubilizzazione (DeStreak Rehydration Solution, GE Healthcare) senza alcun tipo di pretrattamento, ma solo neutralizzando l'ambiente acido con l'aggiunta di un piccolo volume di 0,1 M idrossido di sodio (NaOH).

In un secondo set di esperimenti, è stata valutata l'importanza di ridurre la forza ionica dei campioni, ed evitare problematiche nella separazione in IEF, si è scelto di rimuovere i sali presenti nelle frazioni insolubili in acido effettuando dei lavaggi dei pellet con dei concentratori con un limite di esclusione di 3000 Da in modo da rimuovere tutte le molecole a PM inferiore a 3000 Da e quindi principalmente sali, trattenendo però tutte le proteine e i peptidi. Sono stati perciò utilizzati filtri da centrifuga Amicon Ultra (Millipore) da 0,5 mL con un limite di esclusione di 3000 Da e sono stati effettuati 4 lavaggi utilizzando 0,5 mL di acqua ultrapura in ogni lavaggio per un totale di 2 mL.

In un terzo set di esperimenti, mirati ad evidenziare le proteoforme fosforilate della frazione insolubile in acido della saliva di un soggetto adulto, mediante il colorante fluorescente Pro-Q Diamond, il pellet è stato emulsionato manualmente in 1,5 mL di una soluzione al 10% (p/v) di acido tricloroacetico (TCA), 20 mM ditiotreitolo (DTT) in acetone per mezzo di un pestello in plastica utilizzando provette da 1,5 mL (Argos Technologies, VWR). Il campione così trattato è stato incubato a -20 °C per una notte al fine di ottenere la precipitazione delle proteine. Il giorno seguente è stato centrifugato a 24000 g per 30 minuti a 4 °C. Al termine della centrifugazione, il supernatante è stato eliminato, ed il pellet è stato sottoposto a 3 lavaggi consecutivi con una soluzione 20 mM DTT in acetone freddo (-20 °C), al fine di rimuovere il TCA dal campione, centrifugando ogni volta a 24000 g per 30 min a 4°C ed eliminando il supernatante al termine di ogni lavaggio. Il precipitato ottenuto è stato posizionato sotto cappa aspirante a temperatura ambiente per circa 3 ore per permettere l'evaporazione dei residui di acetone.

1.2.2 Solubilizzazione della frazione insolubile in acido (TFA e TCA)

I pellet, preparati come descritto nella sezione precedente, sono stati neutralizzati con l'aggiunta di 5 µL di NaOH 0,1M vorticando per 2 min, immediatamente prima di essere addizionati con la soluzione di risolubilizzazione (DeStreak). Considerato che, a seconda della lunghezza della strip impiegata per l'analisi in isoelettrofocalizzazione, il volume richiesto per la reidratazione della strip della

prima dimensione è variabile, e in particolare è pari a 125 μL per strip della lunghezza di 7 cm e 200 μL per strip della lunghezza di 11 cm, tutte le frazioni insolubili in acido destinate ad essere analizzate su una strip da 11 cm sono state risolubilizzate con 230 μL di DeStreak, al fine di disporre di un piccolo quantitativo di campione per la determinazione delle proteine totali.

I campioni sono stati incubati 3 ore a temperatura ambiente, vorticando occasionalmente. Al termine dell'incubazione, i campioni sono stati centrifugati a 24000 g per 10 min a 4 °C al fine di rimuovere il materiale che non è stato possibile solubilizzare nelle condizioni sperimentali utilizzate. Il supernatante ottenuto è stato prelevato, trasferito in una nuova provetta, misurato e portato ad un volume finale di 125 μL per le strip da 7 cm e 230 μL per le strip da 11 cm con l'aggiunta di un opportuno volume di soluzione di risolubilizzazione. In questo ultimo caso 200 μL sono stati utilizzati per la reidratazione passiva della strip, mentre 30 μL sono stati utilizzati per la quantificazione delle proteine totali.

1.2.3 Quantificazione delle proteine totali

La determinazione della concentrazione proteica totale dei campioni è stata effettuata utilizzando il 2D-Quant Protein Kit (GE Healthcare), che permette di determinare accuratamente la concentrazione proteica totale anche in presenza di reagenti che interferiscono con i metodi più utilizzati per il dosaggio delle proteine, come ad esempio detergenti, agenti riducenti, agenti caotropici. Nelle condizioni sperimentali da noi utilizzate i reagenti che possono interferire con i normali kit di determinazione quantitativa delle proteine sono l'urea, la tiourea, il CHAPS, e l'IPG Buffer, che devono essere necessariamente utilizzati nella preparazione di campioni da sottoporre a tecniche elettroforetiche e in particolare alla 2-DE. La compatibilità delle procedura utilizzata verso questi reagenti è realizzata facendo precipitare tutte le proteine e lasciando in soluzione gli agenti interferenti, che in questo modo vengono allontanati. Il saggio sfrutta il legame specifico di ioni rame alle proteine: le proteine precipitate vengono risospese in una soluzione contenente ioni rame che vanno a formare dei complessi rame-proteina. La quantificazione degli ioni rame liberi viene effettuata alla lunghezza d'onda (λ) di 480 nm attraverso l'utilizzo di un agente colorimetrico e la loro concentrazione risulta inversamente proporzionale alla concentrazione proteica, perciò l'Assorbimento (A) a 480 nm sarà inversamente proporzionale alla concentrazione delle proteine. Il test mostra una risposta lineare nell'intervallo 0-50 μg di proteine. Le concentrazioni proteiche dei campioni in esame sono state determinate costruendo una retta di calibrazione misurando il

valore di A di soluzioni contenenti diversi quantitativi proteici (0, 10, 20, 30, 40, 50 μg) ottenuti a partire da una soluzione 2 mg/mL di siero albumina bovina (BSA) fornita con il kit. Il kit utilizzato contiene diverse soluzioni coperte da brevetto (soluzione precipitante, co-precipitante, soluzione contenente rame, reagente A e reagente B) di cui non è riportata la loro composizione.

Ciascuna determinazione delle proteine totali è stata realizzata in doppio: 500 μL di soluzione precipitante sono stati aggiunti a 15 μL di ciascun campione proteico (sia per i campioni a concentrazione incognita che per le soluzioni di albumina a concentrazione nota). I campioni sono stati quindi vorticati ed incubati per 3 min a temperatura ambiente. Successivamente ad ogni provetta sono stati aggiunti 500 μL di soluzione co-precipitante e, dopo agitazione mediante vortex, i campioni sono stati centrifugati a 14000 g per 10 min. I supernatanti sono stati velocemente decantati ed i campioni sono stati ancora centrifugati per 1 min a 14000 g. Il supernatante rimanente è stato accuratamente rimosso, e al pellet è stato addizionato 100 μL di soluzione contenente rame e 400 μL di H_2O ultrapura. I campioni sono stati quindi vorticati per dissolvere le proteine precipitate e successivamente è stato aggiunto ad ogni provetta 1 mL di “*Working Color Reagent*” (WCR), soluzione colorante preparata miscelando 100 parti del reagente A con 1 parte del reagente B. Questa soluzione causa il viraggio di colore delle soluzioni in presenza di ioni rame permettendo di misurare gli ioni rame liberi presenti. I campioni sono stati rapidamente agitati per inversione immediatamente dopo l'aggiunta del WCR per assicurare una rapida miscelazione. Le provette sono state quindi incubate per 15 min a temperatura ambiente, prima di procedere con le misurazioni di $A_{480\text{ nm}}$ utilizzando lo spettrofotometro Cary 50 (Varian).

In Figura 5 è riportato il grafico ottenuto riportando nell'asse x i quantitativi proteici delle cinque diverse soluzioni di BSA utilizzate e nell'asse y i valori di $A_{480\text{ nm}}$ misurati, per ricavare il quantitativo proteico dei campioni a concentrazione incognita esaminati.

Le frazioni insolubili in acido hanno mostrato un contenuto proteico molto variabile: in particolare i soggetti adulti hanno mostrato un contenuto proteico tra circa 3 μg (adulto 6, Figura 5) e circa 40 μg (adulto 4, Figura 5) nei 15 μL di campione utilizzati per il saggio, perciò il quantitativo proteico totale nei 200 μL complessivi di campione è risultato compreso tra un minimo di 35.5 μg ed un massimo di 540 μg . Per quanto riguarda i neonati a termine il quantitativo proteico totale nei 200 μL complessivi di campione è risultato compreso tra 40 e 90 μg , mentre i campioni dei neonati pretermine hanno mostrato un quantitativo proteico

totale nei 200 μ L complessivi di campione compreso tra 40 e 185 μ g. Queste differenze possono essere spiegate considerando che i volumi dei campioni di saliva prelevati dai neonati sono stati sicuramente inferiori a quello prelevati dagli adulti, per ovvie difficoltà tecniche.

1.2.4 *Prima dimensione: IEF*

La caratterizzazione proteomica della frazione insolubile in acido della saliva dei neonati prematuri è stata effettuata dopo risolubilizzazione della frazione insolubile in acido in 125 μ L di DeStreak in presenza di anfoliti che coprono l'intervallo di pH 3-10 (0,5%), e reidratazione passiva di strisce di gel di poliacrilammide contenenti un gradiente di pH immobilizzato 3-10 non lineare della lunghezza di 7 cm (ReadyStrips IPG, Bio-Rad). Il gradiente di pH 3-10 non lineare è stato scelto in quanto consente una migliore separazione delle proteine con un pI compreso tra 5 e 7, che in genere sono le più abbondanti, rispetto alle proteine con pI più acidi o più basici.

Al fine di ottenere una maggiore risoluzione nella separazione bidimensionale delle proteine salivari della frazione insolubile in acido dei neonati pretermine, a termine e degli adulti, gli esperimenti di elettroforesi bidimensionale sono stati effettuati utilizzando IPG-strip più lunghe, e in particolare sono state scelte quelle di 11 cm (ReadyStrips IPG, Bio-Rad). Conseguentemente le frazioni insolubili in acido sono state risolubilizzate in 230 μ L di DeStreak.

L'alta variabilità del contenuto proteico nelle diverse frazioni di saliva insolubili in acido (20-90 μ g per i neonati pretermine, 20-45 μ g per i neonati a termine, 35,5-50 μ g per i soggetti adulti) non ci ha consentito di caricare lo stesso quantitativo proteico in ciascuna analisi bidimensionale con finalità analitiche, ma il confronto tra i livelli di espressione delle proteine nei diversi gel è stato possibile dopo aver effettuato l'operazione di normalizzazione nella fase di analisi quantitativa mediante il software SameSpots.

Tutte le suindicate analisi in elettroforesi bidimensionale sono state eseguite con finalità analitiche. Oltre queste analisi sono state eseguite analisi con finalità preparativa al fine di procedere alla digestione triptica degli spot e identificare la proteina di interesse. Nei gel preparativi sono stati caricati 150 μ g di proteine totali. Sia nella modalità analitica, sia in quella preparativa, l'esecuzione della prima e della seconda dimensione della 2-DE sono state eseguite seguendo lo stesso protocollo. Ciascun campione è stato quindi trasferito all'interno di un apposito

canale della vaschetta di reidratazione, evitando la formazione di bolle, e sopra di esso è stata adagiata la strip con il gel rivolto verso il basso in modo da permettere la sua reidratazione. Dopo 10 min dal caricamento, su ciascun canale del vassoio porta-strip sono stati depositati alcuni mL di olio minerale al fine di evitare la disidratazione della strip durante la reidratazione passiva che procede per tutta la notte.

La mattina successiva, ciascuna strip è stata risciacquata con H₂O ultrapura (utilizzata anche per preparare tutte le soluzioni utilizzate per gli esperimenti in modo da ridurre al minimo la presenza di contaminanti) e posizionata in uno dei canali dell'isoelettrofocalizzatore ETTAN IPGPhorII (Amersham Biosciences) questa volta con il gel rivolto verso l'alto. Su ognuna delle due estremità di ciascuna strip è stato posizionato un quadratino (0,5 cm²) di carta da filtro inumidito con 75 µL di H₂O ultrapura e al di sopra sono stati posizionati gli elettrodi per l'applicazione del campo elettrico.

Sia per le strip da 7 che da 11 cm le condizioni elettriche applicate sono state articolate in quattro fasi, impostando un limite di intensità di corrente massima applicabile per strip pari a 50 µA.

Per le strip da 7 cm il tempo totale della IEF è stato di circa 3,5 ore. Nella prima fase è stato applicato un voltaggio di 300 V per 60 min o fino al raggiungimento di 200 kV/h. In questa prima fase il campione contiene al suo interno un certo quantitativo di sali che limitano il raggiungimento della differenza di potenziale applicata. Di conseguenza, nella prima parte di questa fase il voltaggio aumenta gradualmente, mano a mano che diminuisce l'intensità della corrente. In alcuni casi è stato quindi necessario prolungare la durata di questa fase per raggiungere i 200 kV/h richiesti. La seconda e la terza fase sono state effettuate con una modalità in gradiente, e specificatamente variando il voltaggio applicato da 300 V a 1000 V in 30 min per la seconda fase e da 1000 V a 5000 V in 90 min per la terza fase. Nella quarta fase sono stati mantenuti 5000 V costanti per 30 min o fino al raggiungimento di 8000 kV/h totali.

Per le strip da 11 cm il tempo totale della IEF è stato di circa 5 ore. Nella prima fase è stato applicato un voltaggio di 500 V per 90 min o fino al raggiungimento di 500 kV/h. In alcuni casi è stato necessario prolungare la durata di questa fase per raggiungere i 500 kV/h richiesti. La seconda e la terza fase sono state effettuate con una modalità in gradiente, e specificatamente variando il voltaggio applicato da 500 V a 1000 V in 1 ora per la seconda fase e da 1000 V a 6000 V in due ore per la

terza fase. Nella quarta fase sono stati mantenuti 6000 V costanti per 30 min o fino al raggiungimento di 11000 kV/h totali.

1.2.5 Seconda dimensione: SDS-PAGE

Al termine dell' IEF, le proteine, separate nella strip in base al loro pI, sono state fatte migrare in un gel di poliacrilammide in presenza di sodio dodecil solfato (SDS), al fine di separarle in funzione delle loro dimensioni.

Prima di procedere con la seconda dimensione (SDS-PAGE) è necessario che le proteine presenti nell'IPG-strip vengano denaturate in presenza di SDS e vengano formati i complessi SDS-proteine. A tale scopo, ciascuna strip è stata equilibrata per 15 minuti in una soluzione di riequilibrio contenente una soluzione tampone 50 mM Tris-HCl pH 8,8 in presenza di 6 M urea, 30% glicerolo (v/v), 2% SDS (v/v), 0,01% blu di bromofenolo. Si esegue un primo passaggio in presenza dell'agente riducente (1% DTT) per rompere i ponti disolfuro e successivamente un secondo passaggio in presenza dell'agente alchilante iodoacetamide (IAM, 2,5%) in grado di alchilare sia i gruppi -SH precedentemente ridotti che quelli liberi in modo da evitare la formazione di ponti disolfuro.

Per effettuare la seconda dimensione a partire da strip da 7 cm sono stati utilizzati gel di poliacrilammide *home-made* (Laemmli, 1970) con un parametro T uguale al 12,5%, mentre per le strip da 11 cm sono stati utilizzati gel precast Criterion (Bio-Rad) con parametro T in gradiente dal 10 al 20% in modo che le dimensioni delle maglie del gel di poliacrilammide siano sempre più strette man mano che si passa dalla sommità del gel fino alla sua estremità inferiore. In questo modo è stato possibile separare efficacemente sia le proteine ad alto che quelle a basso peso molecolare consentendo di ottenere una separazione con aumentata risoluzione.

Per procedere con la seconda dimensione la strip è stata adagiata sul gel per SDS-PAGE provvisto di un pozzetto unico di lunghezza corretta e di un pozzetto per il caricamento della miscela di proteine a PM noto. Tra i due gel dovrà essere garantita la continuità per permettere che, sotto applicazione del campo elettrico, le proteine passino dall'IPG-strip nel gel per la seconda dimensione. Per realizzare questa continuità, dopo aver disposto con attenzione la strip sul top del gel per l'SDS-PAGE, è stato versato un piccolo volume di soluzione di agarosio allo 0,5% solubilizzata per riscaldamento in soluzione tampone 25 mM Tris, 0,192 M glicina pH 8,6 in presenza di 1% SDS. In breve tempo, a temperatura ambiente, la soluzione raffreddando gelifica e va a riempire tutti gli spazi vuoti rimasti tra i due supporti, garantendo così una perfetta continuità tra i due gel. Il gel è stato quindi

immerso nel tampone di corsa (25 mM Tris, 0,192 M glicina e 1% SDS). Nel pozzetto singolo sono stati caricati 3 μ L di una miscela di proteine a PM noto (SDS-PAGE Molecular Weight Standards, Low Range – Bio-Rad) contenente: glicogeno fosforilasi (97,4 kDa), albumina serica (66,2 kDa), ovoalbumina (45 kDa), anidrasi carbonica (31 kDa), inibitore della tripsina (21,5 kDa) e lisozima (14,4 kDa).

Per le corsa elettroforetiche dei gel *home-made* è stata utilizzata una camera elettroforetica Mini-PROTEAN Tetra System (Bio-Rad), e per i gel Criterion precast una camera elettroforetica Criterion Cell (Bio-Rad). E' stata applicata un'intensità di corrente di 15 mA mantenuta costante per i primi 20 minuti e successivamente una differenza di potenziale di 150 V costante fino alla fuoriuscita dal gel del tracciante blu di bromofenolo. Le analisi sono state effettuate in duplicato per aumentare l'affidabilità e la riproducibilità dei risultati ottenuti.

1.2.6. Colorazione dei gel

Al termine della corsa elettroforetica, si è proceduto alla fase di colorazione. I gel analitici sono stati colorati con il colorante in fluorescenza SYPRO Ruby (Molecular Probes). Prima della colorazione vera e propria si procede alla fase di fissaggio delle proteine al gel incubando il gel in 100 mL di una soluzione contenente 50% metanolo (MetOH), 7% acido acetico (CH_3COOH) in H_2O ultrapura per 30 min. Tutte le fasi di fissaggio, colorazione e decolorazione sono state effettuate mantenendo i gel in costante agitazione orbitale a 40 rpm. È raccomandato l'utilizzo di contenitori in polipropilene per minimizzare l'assorbimento del colorante, e il loro lavaggio accurato con etanolo prima dell'utilizzo. La soluzione di fissaggio viene poi sostituita con altri 100 mL di soluzione fresca, e l'incubazione prosegue per altri 30 min. Una volta eliminata la soluzione di fissaggio sono stati aggiunti ad ogni gel 100 mL di colorante SYPRO Ruby e l'incubazione è proseguita al buio per tutta la notte. La mattina successiva è stata eseguita la fase di decolorazione trasferendo i gel in contenitori puliti in cui sono stati versati 100 mL della soluzione decolorante contenente 10% MetOH, 7% CH_3COOH per 30 min. Prima di procedere con l'acquisizione delle immagini, sono stati eseguiti 3 lavaggi di 5 min con H_2O ultrapura.

La colorazione per evidenziare le proteine fosforilate è stata realizzata utilizzando il colorante in fluorescenza Pro-Q Diamond, che permette di rivelare direttamente nei gel di poliacrilammide la presenza del gruppo fosfato legato ai residui di serina o treonina delle proteine.

Il colorante Pro-Q Diamond consente di rilevare 1-16 ng di fosfoproteine per banda o spot a seconda dello stato di fosforilazione della proteina. Il suo utilizzo è particolarmente consigliato in combinazione con il colorante in fluorescenza per le proteine totali SYPRO Ruby. La determinazione del rapporto tra l'intensità del segnale ottenuto con la colorazione Pro-Q Diamond e quello ottenuto con il colorante SYPRO Ruby per ogni banda o spot fornisce una misura del livello di fosforilazione normalizzato rispetto alla quantità totale di proteine. Effettuando le due colorazioni in successione è possibile distinguere una proteina leggermente fosforilata da proteine fortemente fosforilate, distinguendo anche i falsi positivi (rappresentati da proteine presenti ad elevata concentrazione e debolmente reattive al colorante per le fosfoproteine).

Per i gel che devono essere colorati con il Pro-Q Diamond, dopo aver adagiato l'IPG strip sul gel per l' SDS-PAGE sono stati caricati, oltre ai 3 μ L della miscela di proteine a PM noto (SDS-PAGE Molecular Weight Standards, Low Range – Bio-Rad), 7 μ L di una miscela di proteine fosforilate e non fosforilate (PeppermintStick Phosphoprotein Molecular Weight Standards) contenente β -galattosidasi (116,25 kDa), sieralbumina bovina (66,2 kDa), ovalbumina (45 kDa), β -caseina (23,6 kDa), avidina (18 kDa) e lisozima (14,4 kDa) su un quadratino di carta assorbente posizionato all'estremità opposta del gel su cui è stata caricata la miscela di proteine a PM noto. A seguito della separazione elettroforetica e della colorazione con il colorante specifico per le fosfoproteine verranno visualizzate due bande corrispondenti all'ovalbumina e alla β -caseina che risultano fosforilate. Eseguendo la successiva colorazione con il colorante per le proteine totali SYPRO Ruby verranno evidenziate le 6 bande corrispondenti alle proteine non fosforilate e fosforilate. Questi standard forniscono un controllo positivo e uno negativo sul metodo di rilevazione delle proteine fosforilate utilizzato.

Per la colorazione con il Pro-Q Diamond, al termine della corsa elettroforetica si è proceduto alla fase di fissaggio delle proteine incubando il gel in 100 mL di una soluzione contenente 50% MetOH, 10%, CH₃COOH in H₂O ultrapura per 30 min utilizzando anche in questo caso contenitori in polipropilene accuratamente puliti con etanolo prima dell'utilizzo. La soluzione di fissaggio viene poi sostituita con altri 100 mL di soluzione fresca, e l'incubazione è proseguita per tutta la notte. Il giorno seguente, una volta eliminata la soluzione di fissaggio i gel sono stati incubati con 100 mL di H₂O ultrapura per 10 min, ripetendo questo passaggio per un totale di tre lavaggi in modo da rimuovere completamente il MetOH e il CH₃COOH dal gel in quanto eventuali residui potrebbero interferire con il

colorante. Al termine dei lavaggi, i gel sono stati incubati con 100 mL del colorante Pro-Q Diamond e incubati al buio per 90 min. Al termine dell'incubazione si è proceduto con la fase di decolorazione lavando i gel con 100 mL di una soluzione 20% acetonitrile (ACN), 50 mM acetato di sodio pH 4 per 30 min al buio. Questo passaggio è stato ripetuto altre due volte per un totale di tre lavaggi e di 90 minuti di decolorazione. Prima di procedere con l'acquisizione delle immagini, sono stati eseguiti 2 ulteriori lavaggi di 5 min con H₂O ultrapura.

Dopo l'acquisizione dell'immagine del gel colorato con il Pro-Q Diamond è opportuno colorare il gel con un colorante per le proteine totali, e il SYPRO Ruby è particolarmente indicato per la sua alta sensibilità. La colorazione per le proteine totali fornisce informazioni preziose sul campione, perché grazie all'allineamento delle due immagini è possibile riconoscere le proteine fosforilate nel quadro complessivo di tutte le proteine. La colorazione dello stesso gel con il colorante per le proteine totali SYPRO Ruby è stata effettuata secondo le modalità precedentemente descritte incubando il gel per tutta la notte in 100 mL di SYPRO Ruby.

I gel preparativi sono stati invece colorati con il colorante Coomassie colloidale EZBlue (Sigma-Aldrich). Al termine della corsa elettroforetica i gel sono stati lavati per 3 volte in H₂O ultrapura per 5 min al fine di rimuovere i residui di SDS. Successivamente sono stati immersi in 100 mL della soluzione colorante per 1 ora, e quindi lavati in H₂O ultrapura fino al mattino seguente. E' importante osservare che utilizzando questo colorante il fissaggio delle proteine al gel è più labile di quanto non avvenga con gli altri metodi di colorazione (in cui il fissaggio avviene in MetOH e CH₃COOH), e questo permette di ottenere una resa maggiore nell'estrazione dei peptidi triptici dalle maglie del gel.

1.2.7 Analisi delle immagini

Le immagini dei gel analitici colorati con i coloranti in fluorescenza SYPRO Ruby, Pro-Q Diamond e quelle dei gel preparativi colorati con il colorante Coomassie EZBlue sono state acquisite ad alta risoluzione mediante il sistema ChemiDoc MP (Bio-Rad Laboratories, Hercules, CA, USA) dotato di una CCD (*Charge-Coupled Device*) camera ad alta risoluzione e di una tecnologia di rilevamento ad alta sensibilità in grado di supportare più metodi di rilevazione consentendo di ottenere quantificazioni per sistemi di rilevazione sia in fluorescenza che in colorimetria. Il sistema precalibrato prevede la messa a fuoco precisa per qualsiasi impostazione dello zoom; e il sistema di funzionamento automatizzato assicura prestazioni

costanti, riproducibili, e ad alto rendimento. Il sistema è controllato dal software Image Lab per ottimizzare le prestazioni per una veloce, integrata e automatizzata cattura, analisi ed esportazione delle immagini acquisite.

Per i gel colorati in fluorescenza con il SYPRO Ruby è stato applicato un tempo di esposizione automatico di 0,697 secondi (s) (che viene consigliato dalla ditta produttrice per le bande intense) utilizzando il trans illuminatore UV come fonte di eccitazione e il filtro di emissione 605/50.

Per i gel colorati in fluorescenza con il Pro-Q Diamond è stato applicato un tempo di esposizione automatico di 18.4 s (che viene consigliato dalla ditta produttrice per le bande intense) utilizzando il trans illuminatore UV come fonte di eccitazione e il filtro di emissione 605/50.

Per i gel colorati in Coomassie è stato applicato un tempo di esposizione automatico di 0,1 s (che viene consigliato dalla ditta produttrice per le bande intense) impostando il protocollo *White epi illumination*.

Tutte le immagini dei gel bidimensionali sono state analizzate per la quantificazione degli spot con il software SameSpots (TotalLab).

Per le analisi analitiche, è stato selezionato un gel di riferimento rispetto al quale sono stati allineati tutti gli altri. Dopo che il software ha effettuato automaticamente il controllo di qualità delle immagini (saturazione, dimensioni, sfondo), queste sono state processate per il rilevamento degli spot e l'operazione di normalizzazione con gli strumenti automatici seguendo il protocollo del programma. L'operazione di normalizzazione è necessaria negli esperimenti di proteomica per consentire il confronto tra gel contenenti quantitativi proteici diversi. Tutti gli spot sono stati ricontrollati manualmente per eliminare dall'analisi gli artefatti tecnici rilevati erroneamente dal software, ad esempio particelle di polvere di forma irregolare e bolle d'aria di sfondo.

L'analisi statistica è stata effettuata attraverso il modulo di analisi statistica multivariata del software SameSpots sui volumi log-normalizzati per tutti gli spot, calcolando il parametro *Fold Variation*, cioè la differenza dei volumi log-normalizzati tra i due gruppi, ottenuta confrontando i valori medi dei volumi log-normalizzati di ogni singolo spot in tutti i campioni all'interno di un gruppo e il parametro p di ANOVA, ottenuti dall'analisi della varianza unidirezionale di tutti gli spot.

Solamente gli spot che hanno presentato differenze di *Fold Variation* $\geq 2,0$ e un valore p di ANOVA $\leq 0,05$ sono stati considerati statisticamente significativi e sono stati presi in considerazione per le successive fasi dell'analisi.

1.2.8. Digestione triptica

Tutti gli spot che hanno mostrato variazioni statisticamente significative sono stati selezionati per essere asportati dal gel e sottoposti a digestione triptica al fine di identificare la proteina presente in ciascuno spot. La digestione triptica porta alla scissione selettiva dei legami peptidici sul lato carbossilico dei residui di arginina e lisina, dando origine a frammenti altamente caratterizzanti per ciascuna proteina (peptidi triptici). Dalle sequenze di tali peptidi triptici, ottenute mediante analisi MS/MS, è possibile identificare in modo univoco la proteina di partenza, effettuando una ricerca sulle banche dati di sequenze proteiche se la sua sequenza è stata precedentemente depositata in banca dati.

Gli spot sono stati asportati esclusivamente dai gel colorati in Coomassie, dato che il maggior quantitativo proteico caricato in questi gel (150 µg di proteine totali) offre una più alta possibilità di successo nell'identificazione anche degli spot meno rappresentati.

Gli spot con variazioni statisticamente significative sono stati asportati dal gel mediante l'uso di un bisturi sterile e trasferiti singolarmente all'interno di microprovette. Tutti gli spot sono stati sottoposti a una fase di lavaggio con una soluzione 0,1M ammonio bicarbonato (AMBIC) pH 8.0 e ACN in rapporto 1:1 (v/v) con un'incubazione di 30 min a temperatura ambiente, ripetendo due volte questo passaggio per incrementare la rimozione del colorante.

La soluzione è stata quindi rimossa e ad ogni spot è stato aggiunto un volume di ACN puro opportuno per coprire completamente il frammento di gel (circa 100 µL). Dopo un'incubazione di 10 min per disidratare il gel, la soluzione di ACN è stata eliminata ed è stato aggiunto ad ogni spot un egual volume di una soluzione 10 mM DTT in AMBIC incubando a 65 °C per 30 min.

La soluzione di DTT è stata quindi rimossa ed ad ogni spot è stato nuovamente addizionato un egual volume di ACN puro incubando per 10 min. Una volta eliminato l'ACN, si è addizionato un egual volume di una soluzione 55 mM IAM in AMBIC incubando per 30 min a 30 °C al buio (la IAM è fotosensibile). Infine la soluzione è stata rimossa ed è stato aggiunto nuovamente un egual volume di ACN puro per ulteriori 10 min. Il gel disidratato è così pronto per la digestione triptica.

Per la digestione triptica è stato utilizzato il kit Trypsin Singles Proteomics Grade (Sigma-Aldrich). La tripsina liofilizzata è stata risospesa prima dell'uso in 5 µL del tampone fornito con il kit (Trypsin Solubilization Reagent) addizionati con 45 µL del tampone di digestione (AMBIC e ACN ultrapuro in rapporto 10:1) in modo da

ottenere una concentrazione finale di tripsina pari a 20 µg/mL. Circa 10 µL di tale soluzione di tripsina (si è utilizzato un volume variabile in funzione del presunto quantitativo proteico di ciascuno spot) sono stati addizionati a ciascuno spot e dopo aggiunta di ulteriori 70 µL di tampone di digestione le provette sono state centrifugate per 5 min a 5000 g per spingere tutti i frammenti di gel sul fondo della provetta e quindi incubate a 37 °C per tutta la notte.

Il giorno seguente, il supernatante contenente i peptidi triptici è stato prelevato da ciascuna provetta, e trasferito in una nuova provetta pulita. La reazione di digestione enzimatica è stata bloccata acidificando la soluzione con acido formico (FA concentrazione finale 0,1%), per inibire l'azione della tripsina e favorire la ionizzazione dei frammenti proteici. La soluzione è stata congelata, liofilizzata, e opportunamente risospesa per l'analisi in MS.

Gli esperimenti HPLC-ESI-MS/MS ad alta risoluzione sono stati effettuati utilizzando un sistema Ultimate 3000 RSLC nano collegato ad una colonna cromatografica Zorbax 300 SB-C18 (150 mm x 1 mm, diametro delle particelle 3.5 mm) (Agilent Technologies, Santa Clara, CA) accoppiato ad un LTQ Orbitrap ELITE (Thermo Fisher Scientific, Waltham, MA, USA). Sono stati utilizzati i seguenti eluenti: A) soluzione acquosa di FA 0,1% (v/v); B) 0,1% (v/v) FA in ACN/H₂O 80/20 v/v.

Utilizzando un flusso di 50 µL/min, nella fase iniziale dell'analisi (0-2 min) la fase mobile utilizzata è costituita dal 5% della fase B, nella fase successiva (2-40 min) viene applicato un gradiente lineare dal 5 al 70% di B, da 40 a 45 min un gradiente lineare dal 70 al 99% di B.

Gli spettri MS sono stati raccolti con risoluzione di 120000 in un intervallo di valori m/z da 350 a 2000. Nella modalità di acquisizione dei dati applicata (*data-dependent mode*) sono stati selezionati i cinque più intensi ioni multicarica che sono stati frammentati nella trappola ionica utilizzando la dissociazione indotta dalla collisione (CID, 35% di energia di collisione normalizzata).

E' stata applicata al capillare una temperatura di 300 ° C, e una tensione di 4,0 kV.

Per il riconoscimento delle proteine presenti negli spot selezionati, le sequenze dei peptidi triptici ottenute con gli esperimenti di MS e MS/MS ad alta risoluzione sono state confrontate con le sequenze proteiche riportate nel database UniProtKB/Swiss-Prot Homo Sapiens.

I dati MS/MS sono stati elaborati dal software Proteome Discoverer (versione 1.4.1.14 Thermo Fisher Scientific), che utilizza SEQUEST HT come motore di ricerca (University of Washington, USA, licensed to Thermo Electron Corp. San

Jose, CA, USA). Sono stati impostati come parametri di ricerca 10 ppm di tolleranza per gli ioni precursori e 0,5 Da per gli ioni prodotto, 2 mancati tagli, carbamidometilazione delle cisteine (57,02 Da) come modificazione fissa, ossidazione della metionina (15,99 Da) e fosforilazione della serina, treonina e tirosina (+ 79.96 da) come modificazione variabile. Inoltre è stato impostato un minimo di due peptidi identificati ed un *false discovery rate* (FDR) calcolato sotto il 5% e per l'analisi dei dati è stato impostato il filtro ad alta confidenza.

1.3 Risultati e Discussione

Al fine di ottenere informazioni in grado di integrare i risultati ottenuti utilizzando l'approccio HPLC-ESI-MS, completando così il quadro proteomico della saliva dei neonati prematuri, in questo lavoro di tesi sono state analizzate mediante elettroforesi bidimensionale le frazioni insolubili in acido della saliva intera non stimolata, raccolte da 7 neonati prematuri con PCA alla nascita di 168-196 giorni (24-28 settimane). Immediatamente dopo il prelievo, la saliva è stata addizionata con un ugual volume di TFA 0,2%. Questo trattamento è fondamentale per inibire l'azione enzimatica delle proteasi, ridurre la viscosità della frazione solubile in acido, e favorire la ionizzazione delle proteine, rendendo la frazione solubile in acido idonea all'analisi in MS. Questo trattamento induce però la precipitazione di un certo numero di proteine, che non possono, di conseguenza, essere analizzate in LC-MS, ma possono essere analizzate in 2DE dopo risolubilizzazione.

Nel primo set di esperimenti, il precipitato ottenuto dopo trattamento con TFA è stato trattato immediatamente con la soluzione denaturante per risolubilizzare le proteine prima di procedere con l'analisi in 2-DE. In Figura 6 è mostrata un'immagine rappresentativa di un gel 2-DE della frazione insolubile in acido della saliva intera di un neonato prematuro. Al fine di identificare tutte le proteine maggiormente rappresentate, 45 spot sono stati selezionati, escissi dal gel e sottoposti a digestione triptica *in-gel*. Dopo aver effettuato gli esperimenti di HPLC-MS/MS ad alta risoluzione sulle miscele dei peptidi triptici ottenuti da ciascuno spot, i dati ottenuti sono stati analizzati utilizzando il software Proteome Discoverer e sono state identificate 32 differenti proteine, riportate in Tabella 1. Alcune delle proteine identificate (proteina S100A9, S100A8, cistatine, peroxiredoxina-6) sono state precedentemente identificate anche nella frazione solubile in acido dei campioni di saliva di neonati prematuri (Castagnola, 2011b). Probabilmente queste proteine si ripartiscono tra la frazione solubile e la frazione insolubile in acido in funzione sia della loro concentrazione, ma anche della loro interazione con dei componenti, presenti nella saliva intera, che precipitano dopo l'aggiunta del TFA allo 0,2%.

Osservando l'immagine del gel riportata in Figura 6, emerge chiaramente che alcune proteine sono state identificate in due o più spot. Questo suggerisce che queste proteine sono presenti in diverse proteoforme. L'accoppiamento della separazione 2-DE e dell'analisi MS/MS dei peptidi triptici è particolarmente efficace nella caratterizzazione di queste proteoforme, a differenza di quanto accade utilizzando altri approcci, come per esempio l'approccio *shot-gun* mediante il quale

è molto difficile riconoscere per esempio le forme troncate delle proteine (Messana, 2013).

Alcune proteine sono state identificate in due o tre spot, localizzati in corrispondenza dello stesso PM ma a pI differenti, che appaiono nel gel come treni di spot (Fig. 6). Molto spesso questo pattern è dovuto a proteoforme che differiscono nel loro grado di fosforilazione. Purtroppo la presenza dei gruppi fosfato non è facilmente rilevabile mediante analisi MS/MS a causa di fenomeni di “*in source decay*” che causano idrolisi in sorgente del legame estere tra il gruppo fosfato e i residui di Ser o Thr della catena polipeptidica e, nelle condizioni sperimentali utilizzate in questo lavoro di tesi non siamo riusciti ad evidenziare i peptidi fosforilati per confermare che la differente migrazioni degli spot sia dovuta alla presenza di uno o più gruppi fosforici.

Molte delle proteine identificate nei gel 2-DE della frazione insolubile in acido della saliva dei neonati prematuri presentano un valore di PM medio-alto (45-66 kDa) e sono localizzate nella parte superiore del gel (Fig. 6). Tra queste proteine riportiamo diverse citocheratine, appartenente al tipo I (spot 79 e 94) e al tipo II (spot 74, 75 e 78), annexina A1 (spot 31 e 32) e annexina A2 (spot 36 e 83), actina citoplasmatica 1 (spot 88), serpina B3 (spot 33) e serpina B5 (spot 89). Sono state inoltre identificate la gliceraldeide-3-fosfato deidrogenasi (G3PD) nello spot 37, l' α -enolasi nello spot 35, l'albumina serica nello spot 92 e la cornulina negli spot 72 e 73.

Numerose proteine identificate presentano valori di PM medio-bassi (Fig. 6): tra queste, citiamo la proteina antagonista del recettore per l'interleuchina-1 (spot 28), la proteina battericida che aumenta la permeabilità delle cellule batteriche (spot 29), l'*heat shock* protein β -1 (HSP, spot 25 e 26), la proteina 14-3-3 sigma (spot 38) e diversi enzimi tra cui la glutatione S-transferasi P (spot 30), la triosofosfato isomerasi (spot 80) e la peroxiredoxina-6 (spot 82).

Tra le proteine a basso PM, sono state identificate due proteoforme della proteina S100A8 (spot 8 e 9) e diverse proteoforme della proteina S100A9 (spot 1, 2 e 11) la cistatina B (spot 6), la tioredoxina (spot 12), la proteina legante gli acidi grassi (FABP, spot 3), la galectina-7 (spot 81) e la proteina *calmodulina-like 3* (spot 14).

Tutte le proteine identificate in questo lavoro di tesi nella frazione insolubile in acido della saliva dei neonati prematuri sono state precedentemente evidenziate nella saliva intera degli adulti (Jessie, 2010; Hu, 2005), e quindi, a differenza della frazione solubile in acido, che presenta numerose proteine non rilevate negli adulti, nei pellet non sono state evidenziate proteine esclusive dei neonati prematuri.

Al fine di evidenziare differenze quantitative nelle proteine salivari tra i neonati prematuri, i neonati a termine e gli adulti, i gel 2-DE ottenuti dall'analisi delle frazioni insolubili in acido della saliva di 7 neonati prematuri (PCA alla nascita 169-197 giorni) sono stati confrontati con i gel 2-DE ottenuti da 6 neonati a termine (PCA alla nascita 273-294 giorni) e 6 soggetti adulti, analizzati nelle medesime condizioni sperimentali. Per questo set di esperimenti, i precipitati sono stati sottoposti a lavaggi con H₂O prima di procedere alla ri-solubilizzazione per l'analisi 2-DE, in modo da diminuire la forza ionica dei campioni, poiché la saliva dei neonati e dei bambini nel primo anno d'età contiene una maggiore concentrazione di ioni Na⁺, K⁺, Ca⁺⁺, Mg⁺⁺, Cl⁻ rispetto a quella degli adulti (Ben-Aryeh, 1984). In questo modo è stato possibile evitare l'interferenza delle alte forze ioniche sulla migrazione delle proteine nella prima dimensione della 2-DE.

Nella Figura 7 è riportata un'immagine rappresentativa dei gel 2-DE di ciascun gruppo. Dopo l'allineamento e l'analisi di tutte le immagini ottenute, effettuata utilizzando il software SameSpots, è stato possibile rilevare la presenza di 50 spot differenzialmente espressi che sono stati asportati e sottoposti a digestione triptica *in gel*. Dopo aver analizzato la miscela di peptidi triptici ottenuti mediante esperimenti MS/MS ad alta risoluzione, è stata effettuata l'analisi dei dati mediante il software Proteome Discoverer, che ha permesso di identificare 19 proteine (Tabella 2). Alcune di queste proteine sono state riconosciute in più di uno spot, rivelando la presenza di diverse proteoforme. Per alcuni spot, l'identificazione non è stata possibile a causa della bassa concentrazione proteica, che ha portato, dopo la digestione triptica *in gel*, ad una concentrazione dei peptidi triptici al di sotto del limite di rilevabilità dello spettrometro di massa utilizzato. In altri casi, l'analisi dei peptidi triptici non ha permesso l'identificazione di una sola proteina, ma di più proteine, ciascuna con un elevato numero di peptidi univoci. In questi casi di identificazione non univoca, abbiamo riportato in Tabella 2 la proteina come non identificata (NI).

Tre proteine hanno mostrato un aumento statisticamente significativo nei neonati a termine rispetto ai neonati prematuri e gli adulti: la proteina battericida che aumenta la permeabilità delle cellule batteriche, identificata nello spot 29, due proteoforme dell'annexina A1, identificate negli spot 31 e 32, e la citocheratina 13 (di tipo I), riconosciuta nello spot 79.

Numerose proteine sono risultate aumentate in modo statisticamente significativo nei gel 2-DE degli adulti rispetto ai gel dei neonati pretermine e a termine: FABP (spot 3), S100A6 (spot 57), S100A7 (spot 44), due proteoforme delle proteine

S100A9 (spot 11 e 1B), diverse proteoforme della proteina indotta dalla prolattina (PIP, identificata negli spot 15, 16 e 17), la catena kappa delle immunoglobuline (22 e 45), due proteoforme della cistatina SN (spot 27 e 34), la cistatina S/SA (spot 19) e diverse proteoforme dell' α -amilasi 1 (spot 51, 53, 54, 56 e 39).

Due proteoforme dell'albumina (spot 64 e 65) hanno mostrato un aumento statisticamente significativo (5,1 e 4,4 volte, rispettivamente) nel gel 2-DE della saliva del neonato pretermine rispetto ai neonati a termine e agli adulti. Dato che abbiamo stabilito di considerare statisticamente significative solo le variazioni di concentrazione che risultino $\geq 2,0$ e che presentino un valore p di Anova $\leq 0,05$, la proteoforma della cistatina S, identificata nello spot 20, e la proteoforma della proteina S100A8, identificata nello spot 9, che risultano aumentate nella saliva del neonato pretermine di 1,9 e 1,5 volte rispetto ai neonati a termine e agli adulti, non sono state considerate significative.

La famiglia delle proteine S100 è costituita da 24 piccole (9-13 kDa) proteine acide, che legano il Ca^{2+} e che sono specifiche dei vertebrati. Prendono il nome di S100 a causa della loro solubilità in soluzioni sature di solfato di ammonio. Sono state identificate da Moore nel 1965 come una frazione proteica del cervello abbondante a livello delle cellule gliali, ma successivamente sono state identificate in una gamma diversificata di tessuti e in alcuni fluidi corporei, tra cui siero, urina, plasma, liquido seminale, saliva, espettorato e liquido cerebrospinale.

Le proteine S100 sono implicate anche nella trasduzione del segnale del Ca^{2+} che legano grazie alla presenza di un specifico sito di tipo EF-hand all'estremità C-terminale e di un inusuale sito EF-hand all'estremità N-terminale (Isobe, 1978) separati da una regione a cerniera. Il sito all'estremità C-terminale lega il Ca^{2+} con una affinità 100 volte superiore rispetto al sito all'estremità N-terminale (Zimmer 2003). Il dominio C-terminale è variabile tra le S100 ed è responsabile dell'interazione selettiva con i target specifici di ciascuna proteina della famiglia. Le S100 esistono nelle cellule come omo-dimeri anti paralleli e etero-dimeri; i monomeri sono uniti da legami non covalenti (Brodersen 1998).

In seguito al legame con il Ca^{2+} , vanno incontro a modificazioni conformazionali; le eliche si organizzano a formare una "fenditura" in ciascun monomero e il residuo che presenta questa fenditura crea un sito di riconoscimento per le proteine target (Rety, 1999).

Diverse proteine S100 agiscono come regolatori intracellulari, altre come proteine di segnalazione extracellulare e possono essere sia secreti dalle cellule che rilasciati a regolare l'attività delle cellule bersaglio. All'interno delle cellule, sono

coinvolte nell'omeostasi del Ca^{2+} , nel metabolismo energetico, nell'infiammazione e nella migrazione e invasione cellulare attraverso interazioni che stabiliscono con una varietà di proteine bersaglio tra cui sono compresi enzimi, subunità del citoscheletro, recettori, fattori di trascrizione e acidi nucleici. Le S100 extracellulari agiscono sia in modo autocrino che paracrino mediante l'attivazione di recettori di superficie, recettori accoppiati a proteine G, recettori scavenger, o proteoglicani eparan solfato e N-glicani. È stato dimostrato che sono implicate nella regolazione della proliferazione, del differenziamento, della sopravvivenza e della migrazione cellulare sia in condizioni normali che in condizioni patologiche, nell'infiammazione e nella riparazione dei tessuti. Inoltre esercitano attività antimicrobica.

Le proteine S100 mancano della sequenza leader e non sono secrete attraverso la classica via del Golgi, tuttavia il loro meccanismo di secrezione rimane ancora poco chiaro. Le S100A8/S100A9 possono venir rilasciate passivamente dalle cellule mieloidi necrotiche, o possono venir secrete in modo attivo a seguito della traslocazione dalla membrana, in un processo che richiede una rete di microtubuli e l'attivazione della proteina chinasi C.

L'S100A9 è una proteina associata agli stati infiammatori, altamente espressa nei neutrofili. In un precedente lavoro, quattro isoforme di questa proteina sono state descritte mediante analisi HPLC-ESI-MS nella frazione solubile in acido della saliva dei neonati prematuri: due isoforme (*long-type*) risultano acetilate all'estremità N-terminale, dopo rimozione della metionina e differiscono per la presenza di un gruppo fosfato legato al penultimo residuo (Thr_{112}), mentre le altre due proteoforme (*short-type*), anch'esse acetilate all'estremità N-terminale, dopo la rimozione dei primi cinque residui amminoacidici (MTCKM), differiscono per la presenza di un gruppo fosfato legato al penultimo residuo (Thr_{108}) (Castagnola, 2011b). Nelle analisi in 2-DE della saliva di soggetti adulti riportate in diversi articoli presenti in letteratura, la proteina S100A9 è stata più volte identificata in diversi spot, ma non sono state evidenziate le differenze tra le diverse proteoforme (Vitorino, 2004; Jessie, 2010). Oltre ad essere stata identificata nella frazione solubile in acido dei campioni di saliva di neonati prematuri, l'S100A9 è risultata presente anche nella frazione insolubile in acido. In questa frazione, in particolare, l'S100A9 è stata identificata in diversi spot (1A, 1B, 2, 11, Tabella 2) dei gel 2-DE della saliva di neonati prematuri (Fig. 7A), neonati a termine (Fig. 7B), ed adulti (Fig. 7C).

La colorazione in fluorescenza specifica per le fosfoproteine, effettuata con il colorante Pro-Q Diamond (Figura 8A), sui gel 2-DE della frazione insolubile in acido della saliva di un soggetto adulto, seguita dalla colorazione per le proteine totali con il SYPRO-Ruby, effettuata sullo stesso gel (Figura 8B), ci ha permesso di rilevare che gli spot 11 e 1A risultano fosforilati. Questo risultato è in accordo con quanto riportato recentemente da Pavon e coll. (2012) che hanno analizzato mediante 2-DE le diverse proteoforme della proteina S100A9 delle cellule mononucleari del sangue periferico di soggetti adulti. In questo lavoro gli autori hanno riportato una diminuzione del pI della proteoforma *short-type* non fosforilata di 0.016 unità di pH rispetto alla *long-type* non fosforilata, a causa della perdita della Lys₄. Un'identica diminuzione di pI è stata osservata dagli stessi autori anche tra la proteoforma *long-type* fosforilata e non-fosforilata. Utilizzando un'IPG strip con un intervallo di pH 3-10, gli autori hanno dimostrato che la proteoforma *short-type* non fosforilata è localizzata sotto la *long-type* fosforilata e che la proteoforma fosforilata *short-type* è localizzata in una posizione leggermente più acida. Nel gel 2-DE della frazione insolubile in acido della saliva intera di un soggetto adulto, utilizzando il colorante per le proteine totali SYPRO Ruby (Fig. 8B), abbiamo ottenuto un pattern delle S100A9 in accordo con quello riportato da Pavon, e, utilizzando il colorante specifico per le fosfoproteine Pro-Q Diamond (Fig. 8A), abbiamo potuto evidenziare che gli spot 1A e 11 sono fosforilati e corrispondono rispettivamente alle proteoforme *long-* e *short-type* fosforilate, mentre gli spot 2 e 1B, che non si colorano con il Pro- Diamond, non sono fosforilati e corrispondono rispettivamente alla proteoforma *long-* e *short-type* non fosforilata della S100A9.

Le due proteoforme della proteina S100A9 (spot 11 e 1B) che sono risultate sovraespresse nella frazione insolubile in acido della saliva degli adulti rispetto ai neonati prematuri e i neonati a termine (2,0 e 2,1 volte, rispettivamente) corrispondono alla proteoforma *short-type* fosforilata e non-fosforilata (spot 11 e 1B).

Per rimuovere molecole interferenti che potrebbero impedire di rilevare in maniera ottimale le proteoforme di proteine diverse con pI e PM simili, dopo colorazione sequenziale con il Pro-Q Diamond e il SYPRO-Ruby, il precipitato ottenuto dopo trattamento della saliva intera con TFA, è stato trattato con TCA e acetone, prima della sua risolubilizzazione. In questo modo è stato possibile precipitare la quasi totalità delle proteine, rimuovendo numerose molecole non proteiche interferenti nella successiva migrazione elettroforetica.

La proteina S100A8, altamente espressa nei neutrofili insieme alla proteina S100A9, è stata identificata nel gel 2-DE in 2 spot (8 e 9, Figura 7), caratterizzati

dalla stessa mobilità elettroforetica nella seconda dimensione ma localizzati in corrispondenza di un pI differente, in quanto lo spot 8 presenta un pI leggermente più basico. L'unica PTM riportata sul database Uniprot per le S100A8 è la S-nitrosilazione della Cys₄₂ (Lim, 2008). Diversi autori hanno riportato l'ossidazione *in vitro* della Cys₄₂ ad acido sulfenico/sulfinico/sulfonico (Gomes, 2013; Lim, 2009) e l'ossidazione della Cys₄₂ ad acido sulfonico e sulfinico è stata osservata nella frazione solubile in acido della saliva (Cabras, 2015, in preparazione). Al fine di confermare o escludere l'ossidazione o la S-nitrosilazione della Cys₄₂, gli spettri MS/MS dei digeriti triptici degli spot 8 e 9 sono stati analizzati manualmente ma il residuo di Cys₄₂ è stato rilevato esclusivamente nella forma carbamidometilata, così come ci si aspettava a seguito della riduzione con il DTT e dell'alchilazione con la IAM effettuata durante il passaggio di equilibratura delle IPG-strip attuato al termine della prima dimensione della 2-DE e ripetuto prima della digestione triptica *in-gel*. Di conseguenza la presenza delle PTM ipotizzate è stata esclusa. La colorazione specifica per le fosfoproteine realizzata con il Pro-Q Diamond seguita dalla colorazione per le proteine totali effettuata con il SYPRO-Ruby (Figure 8A e 8B) ha confermato che gli spot 8 e 9 non risultano fosforilati.

Considerando che la copertura dell'analisi ESI-MS/MS dei digeriti triptici è risultata pari all'87% (Tabella 2) per entrambi gli spot e che i peptidi T1, T3 e T9 (C-terminale) sono gli unici peptidi che non è stato possibile sequenziare, possiamo supporre che la PTM responsabile della variazione di pI sia localizzata proprio in uno di questi peptidi.

Gli spot 9 e 8 sono risultati sovra-espressi nella saliva dei neonati prematuri (1,5 e 1,4 volte) rispetto ai neonati a termine e agli adulti, ma non in modo statisticamente significativo.

L'S100A7 è una piccola proteina di 11.4 kDa che forma omo-dimeri che legano ioni Ca²⁺ e ioni Zn²⁺ (Brodersen, 1999). E' riscontrabile intracellularmente (a livello citoplasmatico e nucleare) o extracellularmente, in forma secreta.

E' stato riportato che nell'ambiente intracellulare la proteina S100A7 interagisce con la proteina che lega gli acidi grassi, isoforma dell'epidermide (E-FABP) (Hagens, 1999a; Ruse, 2003). Sebbene l'E-FABP sia coinvolta nel metabolismo e nel trasporto lipidico, il ruolo dell'interazione intracellulare tra queste due proteine non è noto. Per quanto riguarda l'ambiente extracellulare, la proteina S100A7 è stata rilevata nell'essudato della psoriasi a placche, così come la E-FABP, ed è stato suggerito che possa svolgere un'azione antibatterica (Glaser, 2005; Hagens, 1999b).

Sia l'S100A7 che l'E-FABP sono state identificate nei gel 2-DE (rispettivamente spot 44 e 3, Fig. 7) e sono risultate sovra-espresse negli adulti rispettivamente di 4,0 e 2,4 volte (Tabella 2).

Nella frazione solubile in acido della saliva dei neonati pretermine è stato determinato che la proteoforma della S100A7 più rappresentata è la D₂₇, mentre sono state rilevate solo piccole quantità di E₂₇ (Castagnola, 2011b). Nella frazione insolubile in acido, nonostante la copertura dei digeriti triptici ottenuta a seguito dell'analisi ESI-MS/MS sia risultata pari al 72,3% (Tabella 2), non sono state ottenute informazioni sulla sequenza del peptide T5, in cui è localizzata questa sostituzione, e, di conseguenza, non è stato possibile determinare la proteoforma presente.

La proteina S100A6 è una piccola proteina di circa 11 kDa, espressa in modo cellulo-specifica e tessuto-specifico, preferenzialmente nella fase G1 del ciclo cellulare, dopo la nascita e durante lo sviluppo (Tiu, 2000). È implicata nella regolazione di diversi fenomeni cellulari, quali la proliferazione, la differenziazione e l'apoptosi (Filipek, 2008). Tuttavia, la funzione biologica di questa proteina non è chiara e i ruoli proposti risultano contraddittori. Nei gel 2-DE mostrati in Fig. 7, la S100A6 è risultata sovra-espresa negli adulti di 3,0 volte (spot 57).

L' α -amilasi 1, enzima che scinde i legami α 1 \rightarrow 4 glicosidici nei carboidrati complessi, è una delle proteine salivari più abbondanti. Nella saliva intera quest'enzima è stato identificato in numerose proteoforme (Hirtz, 2005) e in un elevato numero di frammenti proteolitici. Dato che la sua concentrazione nella saliva aumenta dopo i pasti, per evitare che il riconoscimento delle proteine meno rappresentate che presentano pI e PM simili a quelli dell'amilasi sia notevolmente difficoltoso, i campioni di saliva sono stati raccolti a digiuno. Oltre alla sua importante funzione nella digestione, l'amilasi svolge anche un'azione antinfiammatoria (Hirtz, 2005) e antibatterica (Rudney, 1995). L'amilasi è risultata presente negli spot 39, 51, 53, 54 e 56 (Fig. 7), tra questi, è ragionevole supporre che gli spot 56 e 39, in base alla posizione nel gel 2-DE, siano relativi alla proteina nella sua forma intera, e che gli spot 51, 53 e 54 corrispondano a frammenti proteolitici di α -amilasi, già evidenziati in precedenti lavori (Hirtz, 2005).

La proteina indotta dalla prolattina (PIP) è una proteina secretoria espressa nelle ghiandole lacrimali, salivari, e sudoripare. La prolattina e gli androgeni inducono la sua espressione, mentre gli estrogeni la inibiscono (Hassan, 2009). Anche se la sua esatta funzione è ancora da definire, è stato evidenziato un suo ruolo nell'inibizione

della crescita batterica, anche grazie alla sua localizzazione strategica nei siti che rappresentano le prime aree d'ingresso degli organismi patogeni.

Dopo rimozione del peptide segnale, la PIP è presente nella saliva con un PM di 11 kDa, e in diverse proteoforme N-glicosilate a PM maggiore (Yao, 2003). Nei gel 2-DE riportati in Fig. 7C, la PIP è stata identificata negli spot 15, 16 e 17, tutti risultati sovra-espressi negli adulti rispetto ai neonati prematuri e a termine rispettivamente di 4,3, 3,5 e 2,1 volte.

La porzione costante della catena leggera κ delle immunoglobuline è stata identificata negli spot 22 e 45 del gel 2-DE mostrato in Fig. 7C. Entrambi gli spot sono risultati sovra-espressi negli adulti rispettivamente di 2,2 e 4,4 volte. Poiché le catene leggere kappa sono comuni alle cinque classi di immunoglobuline, l'aumento della loro concentrazione, rilevato nella saliva degli adulti, deve essere considerato semplicemente un indice di un aumento della concentrazione di anticorpi.

L'annexina A1 è una proteina citosolica che lega il calcio. Anche se la sua funzione non è ancora stata definita in modo inequivocabile, sembra essere coinvolta nel traffico e nella fusione vescicolare (Patel, 2005). È stato suggerito che l'annexina A1 esibisca proprietà anti-infiammatorie *in vitro* e che svolga un ruolo chiave nella difesa immunitaria innata (Vong, 2007).

Nei nostri esperimenti l'annexina A1 è stata identificata negli spot 31 e 32 (Fig. 7B) localizzati allo stesso PM ma differenti per pI. Entrambi gli spot risultano sovra-espressi nei neonati a termine rispetto ai neonati prematuri e agli adulti (rispettivamente di 3,7 e 3,6 volte).

La proteina battericida che aumenta la permeabilità delle cellule batteriche (*bactericidal permeability-increasing protein*, BPI) è una proteina secretoria facente parte della famiglia delle PLUNC (*palate, lung and nasal epithelium clone*, clone del palato, polmone e dell'epitelio nasale). Viene espressa sia nell'epitelio delle vie aeree superiori, dove ricopre la superficie dell'epitelio e delle ciglia, sia nelle cellule della mucosa sottostante e nei dotti delle ghiandole associate con le vie aeree superiori. La sua presenza è stata dimostrata, oltre che nella saliva e nel fluido nasale, anche nelle ghiandole minori del naso, nella porzione posteriore della lingua e nelle tonsille (Bingle, 2011; Vargas, 2008). La BPI ha un ruolo nella risposta immunitaria innata nelle regioni della bocca, del naso e dei polmoni, siti esposti ad una significativa esposizione batterica, e la sua attività battericida è spiegata grazie all'elevata affinità della sua porzione basica N-terminale, che a pH fisiologico risulta carica positivamente, per i lipopolisaccaridi specifici della parete dei batteri Gram-negativi, provvisti di cariche negative.

La BPI, identificata nello spot 29 del gel 2-DE riportato in Fig. 7B, è risultata sovra-espressa (2,2 volte) nei neonati a termine.

Le citocheratine sono filamenti intermedi espressi per lo più dalle cellule epiteliali. Comprendono una vasta gamma di proteine, suddivise nel tipo I (acide) e nel tipo II (basiche o neutre). La loro espressione varia tra i diversi tipi di epiteli e nelle diverse fasi dello sviluppo. La citocheratina 13 (di tipo I) è stata identificata nello spot 79 del gel 2-DE riportato in Fig. 7C e risulta altamente espressa nell'epitelio interno squamoso non cheratinizzato del cavo orale, delle tonsille, della laringe e dell'esofago, mentre la sua espressione risulta bassa in tutti i tipi di cellule delle ghiandole salivari (Azevedo, 2008). Questa proteoforma è risultata sovra-espressa 4,3 volte nei neonati a termine, rispetto ai neonati prematuri e agli adulti, forse a causa di una maggiore desquamazione della mucosa orale nei neonati a termine.

Le cistatine S, SA e SN appartengono alla famiglia 2 delle cistatine, che comprende diversi inibitori delle proteasi a cisteina che possono esercitare un ruolo importante nella protezione del cavo orale da agenti patogeni e nel controllo della catepsina lisosomiale (Bobek, 1992). Nell'analisi 2-DE della frazione insolubile in acido della saliva, la cistatina salivare SN è stata identificata in due spot (27 e 34, Fig. 7C), caratterizzati da una simile mobilità nella seconda dimensione ma da un pI diverso. Nella banca dati UniProt le uniche PTMs riportate per questa proteina sono due ponti disolfuro (94-104 e 118-138) (Ryan, 2010) e, di conseguenza, gli atomi di S, impegnati nel ponte disolfuro, non possono subire ulteriori ossidazioni. Nessuna fosforilazione è stata evidenziata finora sulla cistatina SN, anche se è stata rilevata una sequenza consenso per la fosforilazione da parte di una Golgi chinasi sul residuo di Ser₉₈ (SFE, Dickinson, 2002). Utilizzando la colorazione specifica per le fosfoproteine sui gel 2-DE (Figura 8A e 8B), abbiamo confermato l'assenza di fosforilazioni sulle due proteoforme della cistatina SN.

Tre polimorfismi che coinvolgono una variazione di pI sono stati riportati per questa proteina: N₁₂₉-> D, R₁₃₁-> M, K₁₃₅ -> N (Peterson, 2010). L'analisi degli spettri MS/MS dei peptidi triptici ottenuti dagli spot 27 e 34 ha permesso di escludere la presenza della sostituzione N₁₂₉-> D, mentre non è stato possibile escludere la presenza delle altre due sostituzioni, in quanto la porzione C-terminale della proteina non è stata sequenziata negli esperimenti di MS/MS. In diverse analisi 2-DE della saliva di adulti, la cistatina SN è stata identificata in tre spot, ma le differenze tra le diverse proteoforme non sono state descritte (Hirtz, 2005; Hu, 2005; Vitorino 2004). I dati ottenuti in questo lavoro (Tabella 2) riportano la sovra-espressione (rispettivamente 2,6 e 5,6 volte) delle due proteoforme della cistatina

SN (spot 27 e 34) nella frazione insolubile in acido della saliva degli adulti rispetto ai neonati a termine e pretermine.

La cistatina S e SA sono state identificate nello spot 19 del gel 2-DE degli adulti con alta copertura e con un elevato numero di peptidi unici (Fig. 7C e Tabella 2). La difficoltà nell'identificazione univoca di una delle due proteine è probabilmente dipendente dalla co-migrazione delle due proteine che presentano pI e PM molto simili. Lo spot 19 è risultato sovra-espresso (3,8 volte) negli adulti e questo risultato è in accordo con i dati ottenuti sulla frazione solubile in acido dei neonati prematuri e a termine in cui le cistatine S non sono state riscontrate alla nascita, ma solo a partire da un anno di età (Messana, 2015), e risultano aumentate nei bambini di 6-9 anni (Cabras, 2009). Considerando che la cistatina S può essere fosforilata sulla Ser-3 (cistatina S1) o sulla Ser-1 e sulla Ser-3 (cistatina S2), la proteoforma presente nello spot 19 potrebbe essere la S1, che nella frazione solubile in acido è la forma prevalente (Messana, 2015).

L'albumina è una proteina presente nella saliva intera, mentre la sua concentrazione nella saliva parotidea o sottomandibolare/sublinguale è molto bassa. Questa differenza può essere spiegata considerando il passaggio di componenti del siero che si riversano nella saliva attraverso il fluido gengivale crevicolare. L'albumina è stata identificata negli spot 64, 65 e 69 dei gel 2-DE (Fig. 7B). Gli spot 64 e 65 sono risultati sovra-espressi nei neonati pretermine rispettivamente 5,1 e 4,4 volte. Studi precedenti, che hanno mostrato un aumento nella concentrazione dell'albumina nella saliva intera di bambini di 3 mesi di età rispetto ai 6 mesi, hanno attribuito questa variazione ad una aumentata permeabilità della mucosa nei bambini di 3 mesi di età (Morzel, 2011).

Inoltre, è stato condotto uno studio longitudinale analizzando le frazioni insolubili in acido della saliva intera di 7 neonati prematuri, ottenute immediatamente dopo la nascita e nelle settimane successive (PCA compresa tra 169 e 259 giorni), al fine di evidenziare variazioni quantitative in specifiche proteine della frazione salivare insolubile in acido in funzione della PCA.

In particolare, abbiamo voluto verificare se la diminuzione di concentrazione misurata per 25 proteine della frazione solubile in acido della saliva di neonati prematuri, dalla nascita nelle settimane successive (Castagnola, 2011b), fosse presente anche per sette di queste 25 proteine, identificate anche nella frazione insolubile in acido della saliva. Le sette proteine analizzate nella frazione insolubile in acido corrispondono a 4 diverse proteoforme della S100A9, a 2 proteoforme della S100A8 e all'S100A7. A tale scopo, abbiamo analizzato, mediante 2-DE, 41

frazioni insolubili in acido della saliva di 7 neonati prematuri a diverse PCA, e, dopo allineamento delle immagini, l'intensità dei 7 spot è stata quantificata in tutti i gel mediante il software SameSpots. I risultati sono riportati in Fig. 9 per ciascuna proteina in funzione della PCA: i grafici mostrano che per nessuna delle 7 proteine considerate è riscontrabile una diminuzione significativa della concentrazione all'aumentare della PCA. Questo andamento inatteso potrebbe essere dovuto ad un'interazione di queste proteine, secondo un determinato rapporto, con una componente specifica presente nella saliva intera che precipita dopo l'aggiunta del TFA allo 0,2%. La concentrazione costante di tale componente alle diverse PCA avrebbe come risultato l'assenza di variazioni significative nella concentrazione delle proteine in esame in funzione dalla PCA.

1.4 Conclusioni

In questo lavoro di tesi, l'accoppiamento della 2-DE alla spettrometria di massa ad alta risoluzione ha permesso di ottenere la caratterizzazione comparativa della frazione insolubile in acido della saliva di neonati prematuri, neonati a termine e adulti, evidenziando diverse differenze tra questi campioni. La caratterizzazione proteomica della saliva dei neonati prematuri, precedentemente effettuata mediante analisi HPLC-ESI-MS e MS/MS, esclusivamente sulla frazione solubile in acido, aveva evidenziato la presenza di numerose proteine non riscontrabili nella saliva degli adulti (Castagnola, 2011b). Questo lavoro di tesi ha permesso di integrare i risultati ottenuti nello studio su citato, aggiungendo nuove informazioni relative a proteine che non era stato possibile analizzare proprio per la loro insolubilità nelle condizioni sperimentali dell'analisi ESI-MS. In questo lavoro è stata ottenuta per la prima volta una mappa bidimensionale della frazione insolubile in acido della saliva di neonati pretermine, che ha consentito l'identificazione e la caratterizzazione di interessanti proteoforme. Sono state, infatti, evidenziate un certo numero di proteine identificate in diversi spot, tra queste la proteina S100A9, la proteina S100A8, la proteina indotta dalla prolattina e alcune cistatine. Quattro spot del gel 2-DE della frazione insolubile in acido della saliva di adulti sono stati assegnati a specifiche proteoforme della proteina S100A9, dopo l'identificazione con analisi MS/MS, utilizzando la colorazione sequenziale del gel 2-DE prima con il colorante in fluorescenza Pro-Q Diamond, per evidenziare le proteine fosforilate, e poi con il colorante in fluorescenza SYPRO Ruby, per evidenziare le proteine totali.

Nella frazione insolubile in acido sia degli adulti che dei neonati sono presenti numerose proteine che originano dalla desquamazione dell'epitelio della mucosa orale e dalle cellule del sangue, a differenza della frazione solubile in acido, che contiene numerose proteine secrete dalle ghiandole salivari maggiori. Numerose proteine che risultano aumentate nella frazione insolubile in acido degli adulti sono coinvolte nella risposta immunitaria (catene kappa delle immunoglobuline, PIP, cistatine, S100A9), e, la loro maggiore concentrazione negli adulti è in accordo con la debole risposta immunitaria, sia innata che adattativa, dei neonati. Un'eccezione in tal senso sembra essere rappresentata dalla proteina BPI, componente umorale dell'immunità innata, che risulta aumentata nei neonati a termine. Il suo aumento potrebbe essere collegato alla significativa esposizione batterica dei neonati

attraverso le vie aeree, in un momento in cui la risposta immunologica acquisita è solo parziale.

Considerata la non invasività del prelievo di saliva e alla luce della sua rilevanza come fluido diagnostico, i risultati ottenuti in questo lavoro di tesi forniscono un importante contributo per una più ampia e completa caratterizzazione del proteoma salivare, dato che finora la frazione proteica insolubile in acido non era stata oggetto di studio proprio a causa dell'insolubilità delle proteine presenti.

Tabella 1: Elenco delle proteine identificate nella frazione insolubile in acido della saliva di un neonato pretermine mediante analisi HPLC-ESI-MS/MS effettuata con uno spettrometro di massa LTQ-Orbitrap ad alta risoluzione dopo digestione triptica in-gel degli spot della 2-DE

Spot #	Codice Uniprot	Proteina	Score ^(b)	Copertura ^(b)	# Peptidi unici ^(b)	# Peptidi ^(b)	^(a) PM [kDa]	^(a) pI
1	P06702	S100-A9	15.84	43.86%	4	4	13.1	5.7
2	P06702	S100-A9	16.32	43.86%	4	4	13.1	5.7
3	Q01469	Proteina che lega gli acidi grassi, isoforma epidermica	40.78	65.19%	10	11	15	6.8
5	P69892	Gamma globina	106.1	88.44%	1	13	16	6.7
6	P04080	Cistatina-B	26.61	64.29%	4	4	11.1	7
7	P69892	Gamma globina	198.7	88.44%	1	17	16	6.7
8	P05109	S100-A8	54.7	69.89%	7	7	10.8	6.5
9	P05109	S100-A8	15.52	53.76%	4	4	10.8	6.5
11	P06702	S100-A9	12.18	43.86%	4	4	13.1	5.7
12	P10599	Tioredoxina	3.5	20.95%	2	2	11.6	4.8
14	P27482	Calmodulina-like 3	10.67	26.85%	3	3	16.9	4.3
25	P04792	Heat shock protein beta-1	75.25	68.78%	10	10	22.8	6
26	P04793	Heat shock protein beta-1	17.79	48.78%	6	6	22.8	6
28	P18510	Proteina antagonista del recettore per l'interleuchina-1	53.63	77.62%	7	7	17.1	5.4
29	Q9NP55	BPI-proteina battericida che aumenta la permeabilità delle cellule batteriche	15.07	18.75%	3	3	24.7	5.4
30	P09211	Glutazione S-transferasi P	80.41	46.67%	7	7	23.2	5.4
31	P04083	Annexina A1	143.2	60.40%	20	20	38.6	6.6
32	P04083	Annexina A1	29.53	29.77%	9	9	38.6	6.6
33	P29508	Serpina B3	27.82	26.15%	8	8	44.5	6.4
35	P06733	Alfa-enolasi	62.82	44.70%	16	16	47	7
36	P07355	Annexina A2	44.47	51.03%	13	13	38.4	7.6
37	P04406	Gliceraldeide-3-fosfato deidrogenasi	7.7	13.43%	3	3	35.9	8.6
38	P31947	Proteina 14-3-3 sigma	93.67	66.94%	11	14	27.8	4.7
50	P05109	S100-A8	12.88	53.76%	4	4	10.8	6.5
66	P19013	Citocheratina 4 (di tipo II)	57.94	37.27%	17	17	57.2	6.3
67	P13646	Citocheratina 13 (di tipo I)	81.46	50.66%	17	20	49.6	4.9
68	P13645	Citocheratina 10 (di tipo I)	8.15	6.85%	3	3	58.8	5.1
72	Q9UBG3	Cornulina	57.4	57.17%	11	11	53.5	5.7
73	Q9UBG3	Cornulina	10.97	14.75%	3	3	53.5	5.7
74	P19013	Citocheratina 4 (di tipo II)	189.5	75.66%	37	42	57.3	6.3
75	P19013	Citocheratina 4 (di tipo II)	150.7	64.23%	30	32	57.3	6.3
78	P19013	Citocheratina 4 (di tipo II)	117.7	57.30%	25	27	57.3	6.3
79	P13646	Citocheratina 13 (di tipo I)	162.2	57.64%	23	28	49.6	4.9
80	P60174	Triosofosfato isomerasi	14.1	31.73%	5	5	30.8	5.7
81	P47929	Galectina 7	18.21	46.32%	4	4	14.9	7
82	P30041	Peroxiredoxina 6	5.2	9.38%	2	2	24.9	6
83	A6NMY6	Annexina A2	6.7	6.19%	2	2	38.5	6.5
84	P13646	Citocheratina 13 (di tipo I)	11.6	10.24%	4	4	49.6	4.9
86	P27797	Calreticulina	7.87	15.11%	3	3	46.4	4.3
87	P07237	Disulfuro-isomerasi	1.87	3.94%	2	2	55.3	4.7
88	P60709	Actina Icitoplasmica	30.56	43.47%	10	10	41.7	5.3
89	P36952	Serpina B5	8.28	9.60%	3	3	42.1	5.7
90	P30740	Inibitore dell'elastasi leucocitaria	12.17	14.25%	4	4	42.7	5.9
92	P02768	Albumina sierica	20.74	14.61%	7	7	66.4	5.7
94	P13646	Citocheratina 13 (di tipo I)	117.9	60.70%	20	23	49.6	4.9

a = dati ottenuti da Uniprot

b = dati ottenuti dal software Proteome Discoverer 1.4.1.14

Tabella 2: Elenco delle proteine identificate ^(a) nella frazione insolubile in acido della saliva di neonati pretermine, a termine e adulti attraverso analisi HPLC-ESI-MS/MS dopo digestione triptica in-gel degli spot della 2-DE

Spot #	Codice Uniprot	Proteina	Punteggio ^(b)	Copertura ^(b)	# Peptidi unici ^(b)	# Peptidi ^(b)	PM [kDa] ^(c)	pI ^(c)	Identificata in	ANOVA (p) ^(d)	Variazione * ^(d)	
1A	P06702	S100-A9	15.8	43.86%	4	4	13.1	5.7		NS	1.7	↑adulti
1B	P06702	S100-A9	72.1	67.54%	6	10	13.1	5.7	adulti	0.012	2.1	↑adulti
2	P06702	S100-A9	72.1	45.61%	6	10	13.1	5.7		NS	1.5	↑adulti
3	Q01469	Proteina che lega gli acidi grassi, isoforma epidermica	40.8	65.19%	10	11	15.0	6.8	adulti/pretermine	1.88E-04	2.4	↑adulti
4		NI								1.63E-04	3.1	↑adulti
5	P69892	Gamma globina 2	246.2	84.35%	1	11	16.0	6.7	a termine	NS	1.9	↑pretermine
	P69891	Gamma globina 1	245.5	84.35%	1	11	16.0	6.7	a termine			
8	P05109	S100-A8	149.5	87.10%	10	12	10.8	6.5	a termine	NS	1.4	↑pretermine
9	P05109	S100-A8	168.2	87.10%	8	15	10.8	6.5	a termine	0.05	1.5	↑pretermine
10		NI								5.31E-04	4.4	↑adulti
11	P06702	S100-A9	160.3	78.95%	9	9	13.1	5.7	adulti	0.014	2	↑adulti
	P12273	Proteina indotta dalla prolattina	91.7	67.81%	8	8	13.5	5.4				
15	P12273	Proteina indotta dalla prolattina	228.5	69.86%	10	10	13.5	5.4	adulti	2.65E-07	4.3	↑adulti
16	P12273	Proteina indotta dalla prolattina	109.3	69.86%	9	9	13.5	5.4	adulti	1.47E-04	3.5	↑adulti
17	P12273	Proteina indotta dalla prolattina	113.9	69.86%	9	9	13.5	5.4	adulti	9.96E-04	2.1	↑adulti
18		NI								0.013	3.4	↑adulti
19	P01036	Cistatina-S	207.0	63.12%	7	8	14.2	4.8	adulti	3.81E-04	3.8	↑adulti
	P09228	Cistatina-SA	50.6	56.03%	6	7	14.3	4.8	adulti			
20	P01036	Cistatina-S	80.9	63.12%	7	8	14.2	4.8	adulti	0.037	1.9	↑pretermine
	P05387	Proteina acida ribosomiale P2 (60S)	39.6	85.22%	7	7	11.7	4.4				
21		NI								0.039	2.7	↑adulti
22	P01834	Catena kappa delle Ig (porz. costante)	105.2	80.19%	6	6	11.6	5.6	adulti	2.314E-05	2.2	↑adulti
23		NI								6.45E-04	2	↑adulti
24		NI								0.001	2.9	↑adulti
27	P01037	Cistatina-SN	213.8	61.70%	7	10	14.3	6.9	adulti	0.006	2.6	↑adulti
29	Q9NP55	BPI proteina battericida che aumenta la permeabilità delle cellule batteriche	157.6	54.30%	7	7	24.7	5.4	a termine	0.003	2.2	↑a termine
31	P04083	Annexina A1	162.4	61.85%	20	20	38.6	6.6	a termine	0.007	3.7	↑a termine
32	P04083	Annexina A1	206.6	59.25%	20	20	38.6	6.6	a termine	0.008	3.6	↑a termine
34	P01037	Cistatina-SN	71.6	60.99%	6	8	14.3	6.9	adulti	7.07E-07	5.6	↑adulti

NS= Proteine senza differenze statisticamente significative nella saliva delle tre classi considerate (neonati pretermaturi, neonati a termine e adulti)

NI= proteine non identificate

*Solo gli spot che differiscono di almeno 2,0 volte nei volumi (log-normalizzati) e che presentano un valore p di ANOVA ≤ 0.05 sono stati considerati statisticamente significativi.

(a) = i numeri corrispondono a quelli riportati nelle Figure 6, 7 e 8

(b) = dati ottenuti dal software Proteome Discoverer 1.4.1.14

(c) = dati ottenuti da Uniprot

Tabella 2: Elenco delle proteine identificate ^(a) nella frazione insolubile in acido della saliva di neonati pretermine, a termine e adulti attraverso analisi HPLC-ESI-MS/MS dopo digestione triptica in-gel degli spot della 2-DE

Spot #	Codice Uniprot	Proteina	Punteggio ^(b)	Copertura ^(b)	# Peptidi unici ^(b)	# Peptidi ^(b)	PM [kDa] ^(c)	pI ^(c)	Identificata in	ANOVA (p) ^(d)	Variazione * ^(d)	
38	P31947	Proteina 14-3-3 sigma	141.1	85.48%	13	21	27.8	4.7	a termine	0.008	1.6	↑ a termine
39	P04745	Alfa-amilasi 1	262.0	57.73%	21	21	55.9	6.3	adulti	8.71E-04	5.4	↑adulti
40		NI								0.005	2.3	↑adulti
44	P31151	S100-A7	90.9	72.28%	11	11	11.3	6.3	adulti	0.024	4	↑adulti
45	P01834	Catena kappa delle Ig (porz. costante)	200.2	85.85%	8	8	11.6	5.6	adulti	3.938E-06	4.4	↑adulti
	P0CG05	Catena lambda-2 delle Ig (porz. costante)	81.7	79.25%	2	6	11.3	6.9				
	P0CG04	Catena lambda-1 delle Ig (porz. costante)	76.9	79.25%	2	6	11.3	7.9				
46		NI								0.009	5.2	↑adulti
47		NI								4.63E-04	3.4	↑adulti
48		NI								1.58E-06	3.3	↑adulti
49		NI								0.002	3.3	↑adulti
51	P04745	Alfa-amilasi 1	145.8	57.14%	16	21	55.9	6.3	adulti	0.047	3.3	↑adulti
52		NI								2.795E-06	4.7	↑adulti
53	P04745	Alfa-amilasi 1	42.2	22.90%	7	7	55.9	6.3	adulti	0.003	2.8	↑adulti
54	P04745	Alfa-amilasi 1	84.4	54.01%	16	16	55.9	6.3	adulti	0.013	2.5	↑adulti
55		NI								1.53E-04	4.2	↑adulti
56	P04745	Alfa-amilasi 1	213.1	763.21%	5	23	55.9	6.3	adulti	0.006	3.9	↑adulti
57	P06703	S100-A6	69.7	40.00%	4	5	10.2	5.3	a termine/ adulti	0.011	3	↑adulti
63		NI								NS	1.2	↑adulti
64	P02768	Albumina sierica	668.2	73.56%	49	49	66.4	5.7	a termine	0.014	5.1	↑pretermine
65	P02768	Albumina sierica	834.9	77.34%	54	54	66.4	5.7	a termine	0.02	4.4	↑pretermine
69	P02768	Albumina sierica	866.0	77.01%	52	52	66.4	5.7	a termine	NS	3.1	↑pretermine
70	P01833	Recettore polimerico delle Ig	130.9	41.88%	21	21	81.3	5.6	a termine	NS	2.3	↑adulti
	P02768	Albumina sierica	61.5	33.99%	17	17	66.4	5.7				
71	P01833	Recettore polimerico delle Ig	184.3	42.41%	26	26	81.3	5.6	a termine	NS	2.5	↑adulti
	P04264	Citocheratina I (di tipo II)	100.2	34.01%	15	18	65.9	8.2				
79	P13646	Citocheratina 13 (di tipo I)	605.9	66.59%	28	28	49.6	5.0	adulti	1.62E-04	4.3	↑ a termine

NS= Proteine senza differenze statisticamente significative nella saliva delle tre classi considerate (neonati pretermaturi, neonati a termine e adulti)

NI= proteine non identificate

*Solo gli spot che differiscono di almeno 2,0 volte nei volumi (log-normalizzati) e che presentano un valore p di ANOVA ≤ 0.05 sono stati considerati statisticamente significativi.

(a) = i numeri corrispondono a quelli riportati nelle Figure 6, 7 e 8

(b) = dati ottenuti dal software Proteome Discoverer 1.4.1.14

(c) = dati ottenuti da Uniprot

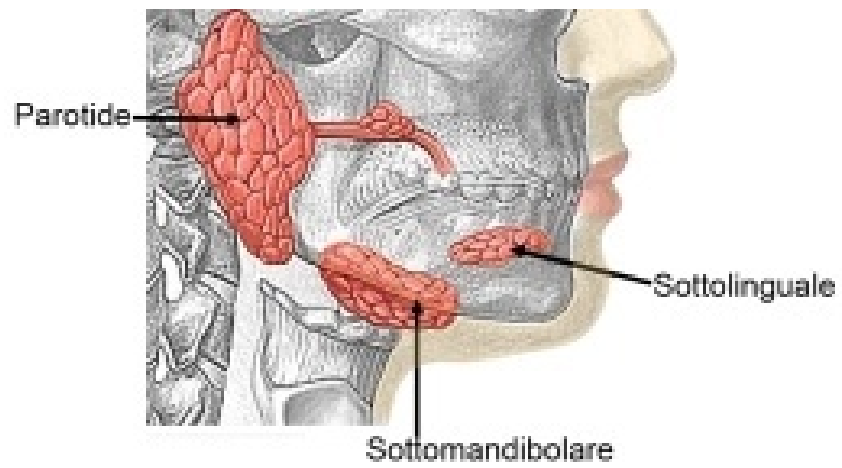


Figura 4: Localizzazione delle ghiandole salivari maggiori
(Tratto da <http://superagatoide.altervista.org/bocca.html>)

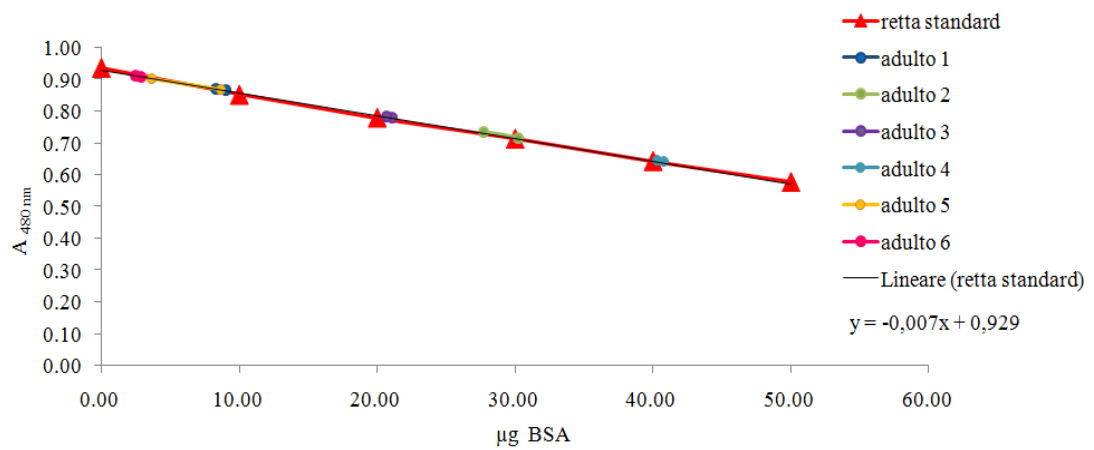


Figura 5: Determinazione del quantitativo proteico delle frazioni insolubili in acido di 6 campioni di saliva di soggetti adulti utilizzando il 2D-Quant Protein Kit e l'albumina serica bovina come proteina di riferimento

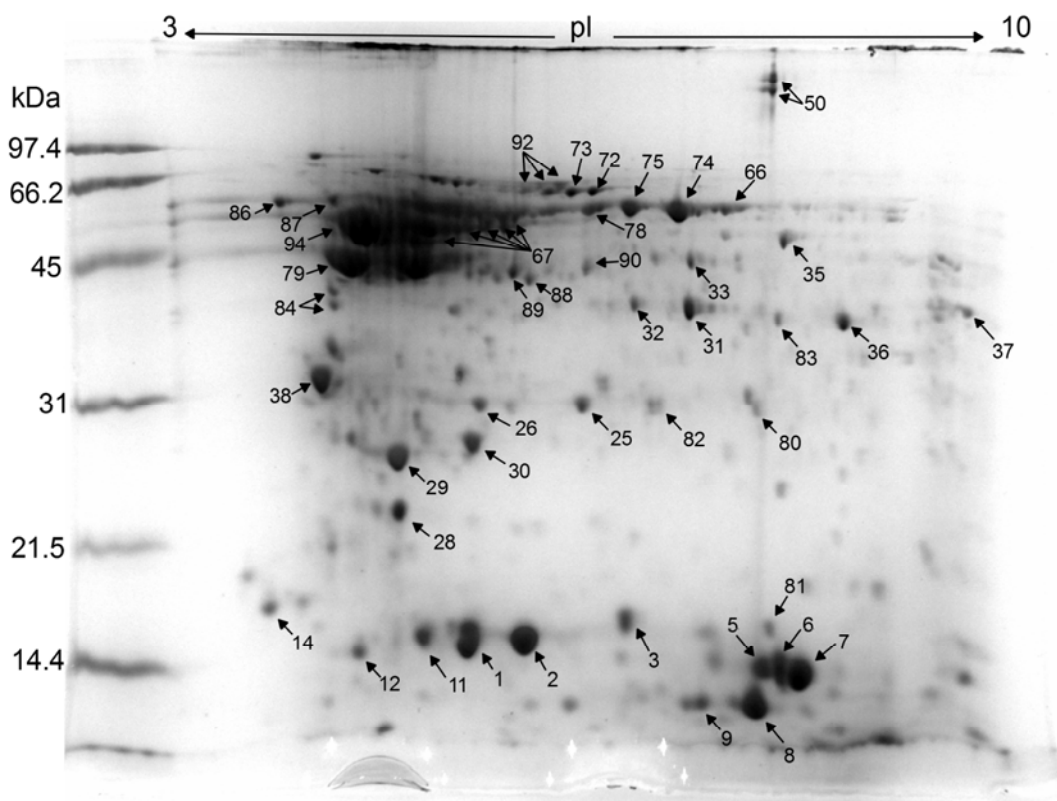


Figura 6: Immagine del gel 2-DE ottenuto analizzando la frazione insolubile in acido di un campione di saliva di un neonato prematuro. La prima dimensione è stata eseguita su una IPG strip da 7 cm (pH 3-10 NL) dopo reidratazione passiva. La seconda dimensione è stata eseguita su un gel "home-made" Tris-HCl, T=12,5% utilizzando il colorante Coomassie colloidale EZBlue (Sigma-Aldrich). Le proteine sono state identificate dopo digestione triptica in-gel mediante analisi MS/MS utilizzando l'LTQ-Orbitrap ad alta risoluzione. I risultati delle identificazioni sono riportati in Tabella 1.

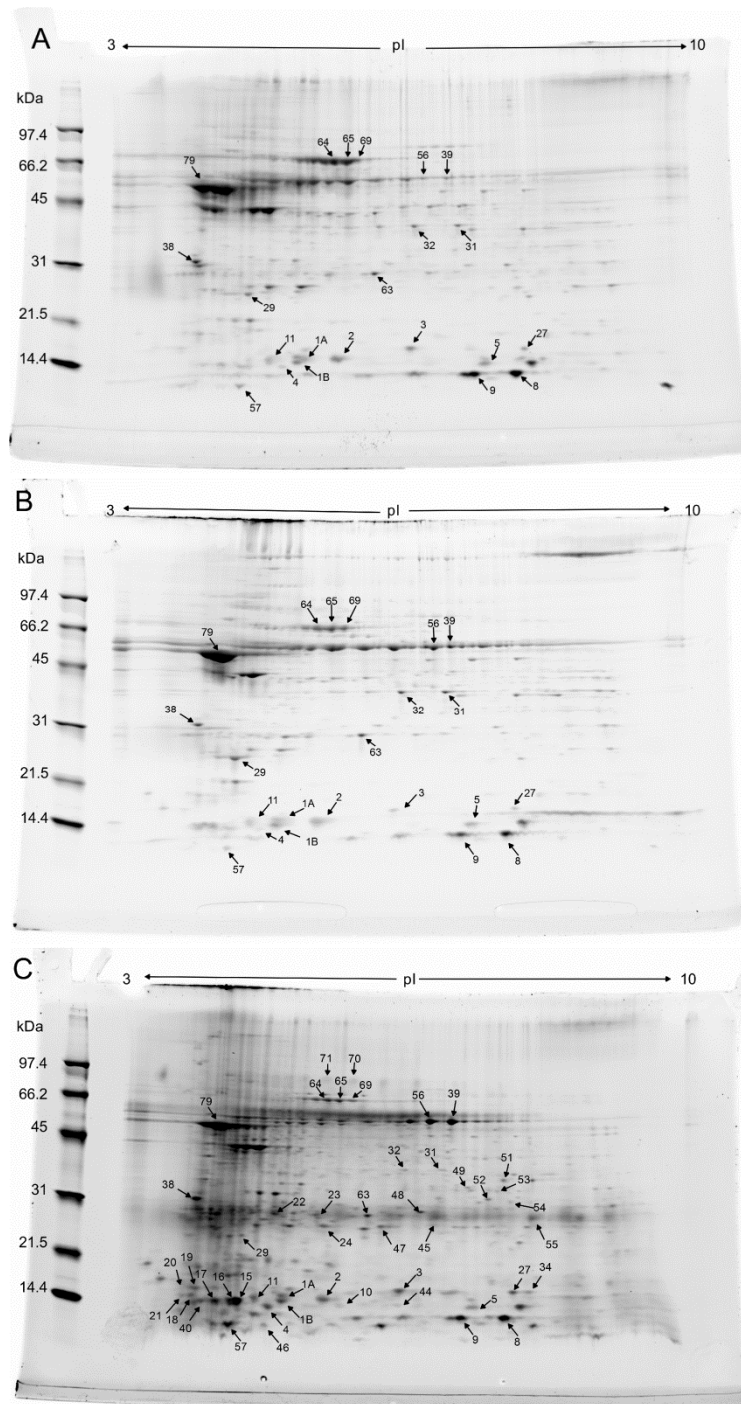


Figura 7: Immagini rappresentative dei gel 2-DE della frazione insolubile in acido di campioni di saliva di neonati prematuri (pannello A), neonati a termine (pannello B) e adulti (pannello C). Sono mostrati i gel di riferimento scelti dal software SameSpots per l'analisi delle immagini. La prima dimensione è stata eseguita su una IPG strip da 11 cm (pH 3-10 NL) dopo reidratazione passiva con 25 µg di proteine totali. La seconda dimensione è stata eseguita su un gel Criterion Tris-HCl T=10-20%. La colorazione è stata effettuata con il colorante in fluorescenza SYPRO Ruby. Le proteine sono state identificate dopo digestione triptica in-gel mediante analisi MS/MS utilizzando l'LTQ Orbitrap ad alta risoluzione. I risultati delle identificazioni sono riportati in Tabella 2.

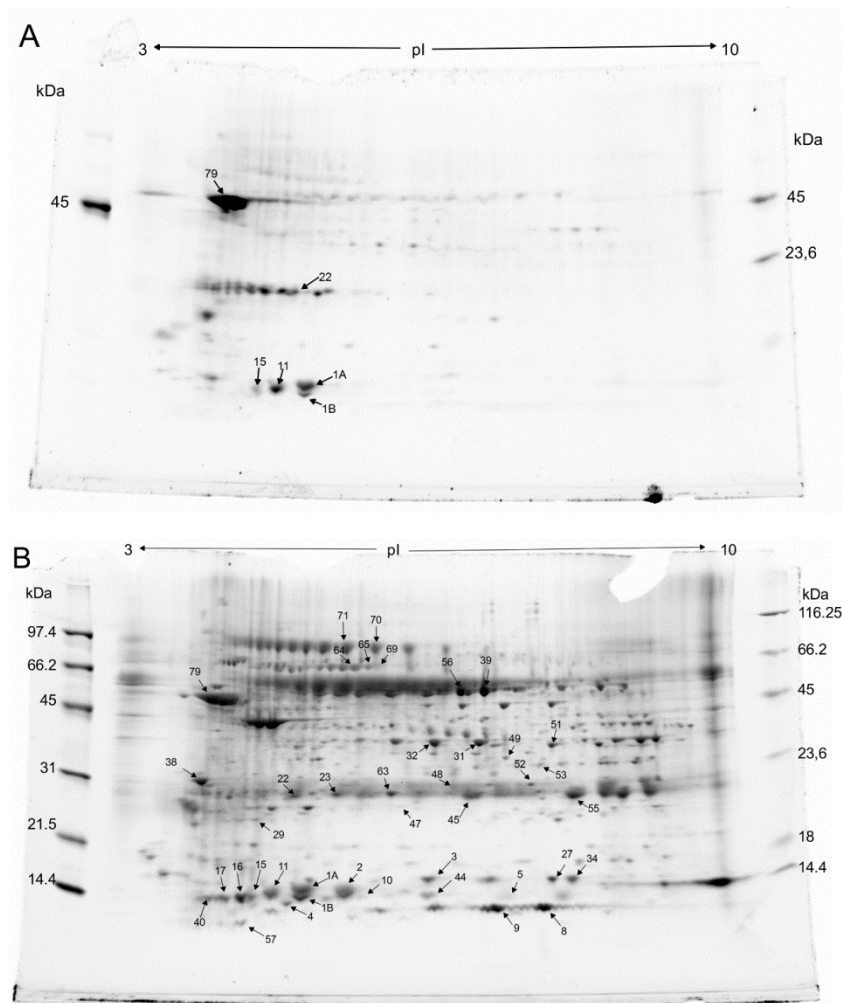


Figura 8: Immagini del gel 2-DE della frazione insolubile in acido di un campione di saliva di un soggetto adulto, dopo trattamento con 10% TCA, 20 mM DTT in acetone. Il gel ottenuto è stato colorato con il Pro-Q Diamond (pannello A) e successivamente con il SYPRO Ruby (pannello B). La prima dimensione è stata eseguita su una IPG strip da 11 cm (pH 3-10 NL) dopo reidratazione passiva con 25 μ g di proteine totali. La seconda dimensione è stata eseguita su un gel Criterion Tris-HCl T=10-20%. L'allineamento di questi due gel con quelli precedentemente analizzati, ha permesso il riconoscimento delle proteine (i numeri corrispondono a quelli riportati in Tabella 2)

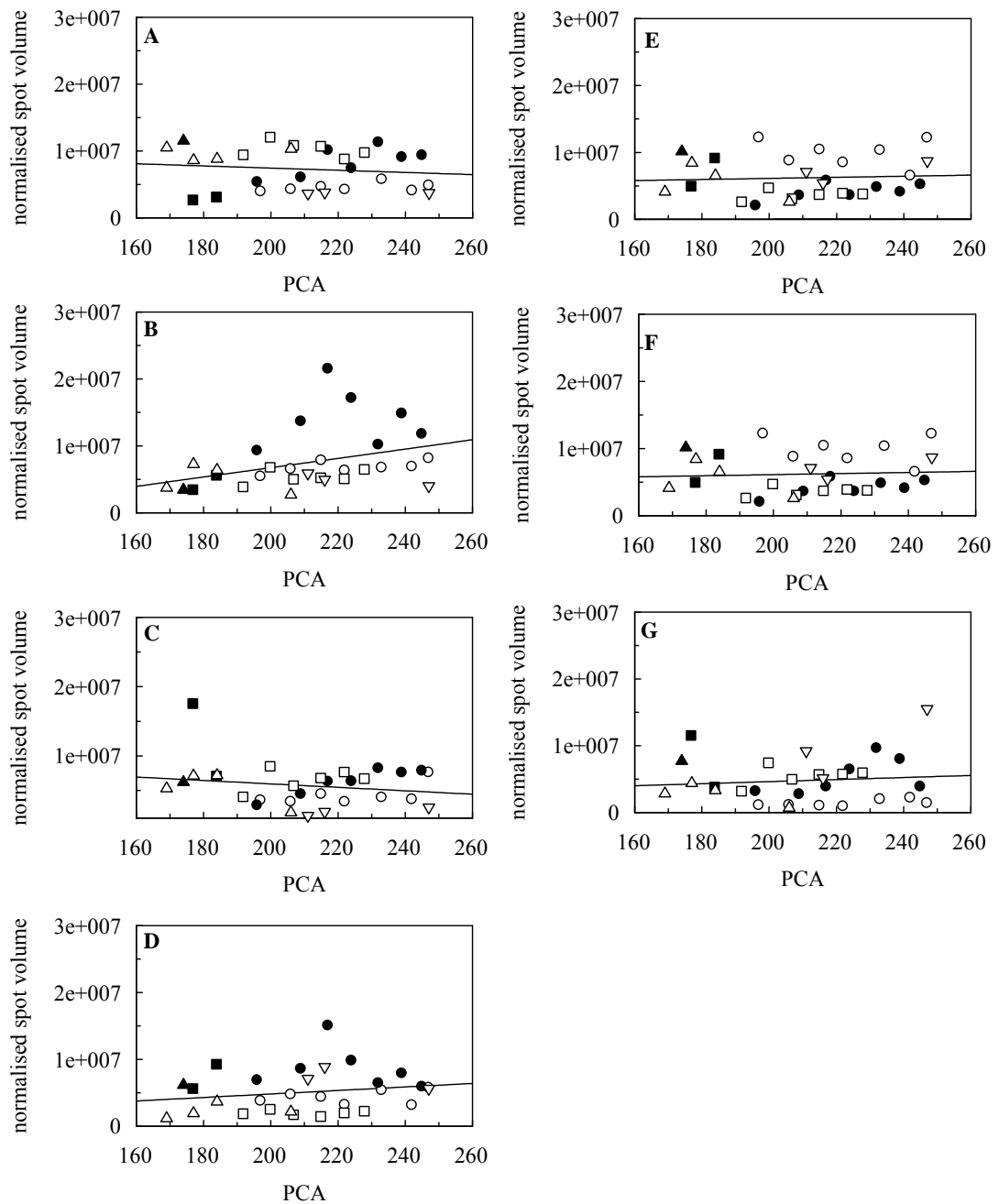


Figura 9: Volumi normalizzati degli spot di 7 proteine ottenuti dopo analisi 2-DE sulla frazione insolubile in acido della saliva totale di 7 neonati pretermine, in funzione dell'età post-concezionale (PCA). La quantificazione dei volumi normalizzati è stata effettuata con il software SameSpots dopo allineamento di 41 gel 2-DE. Pannello A: spot 1B (S100A9); B: spot 2 (S100A9); C: spot 1A (S100A9); D: spot 11 (S100A9); E: spot 8 (S100A8); F: spot 9 (S100A9); G: spot 44 (S100A7).

Parte II

Caratterizzazione proteomica della frazione insolubile in acido di tessuti tumorali cerebrali infantili di Medulloblastoma e Astrocitoma Pilocitico mediante analisi in elettroforesi bidimensionale accoppiata alla spettrometria di massa HPLC-ESI-MS/MS ad alta risoluzione

2.1 Introduzione

La proteomica è stata ampiamente applicata in numerosissimi campi biomedici tra cui la ricerca sui tumori (oncoproteomica), l'oncologia del sistema nervoso centrale e in particolare la ricerca sui tumori pediatrici, consentendo una descrizione più dettagliata del proteoma del tessuto tumorale grazie all'alto numero di informazioni ottenute (Anagnostopoulos, 2014). Infatti, le informazioni che possono essere ricavate dall'analisi dettagliata delle alterazioni molecolari descritte mediante l'approccio proteomico offrono una preziosa possibilità nella pianificazione razionale delle strategie terapeutiche (Anagnostopoulos, 2011).

I tumori del Sistema Nervoso Centrale (SNC), essendo le neoplasie più letali, hanno da sempre suscitato una notevole attenzione, soprattutto per quanto riguarda lo sviluppo dei tumori nella popolazione giovanile, in cui le conseguenze dei disturbi possono essere devastanti (Huse, 2010). L'analisi dei tessuti biotici può essere un importante strumento nell'investigazione dei meccanismi molecolari coinvolti nell'insorgenza e nella progressione dei tumori cerebrali pediatrici e nella scoperta di potenziali biomarcatori.

I tumori sono la seconda più comune causa di morte nei bambini nei paesi sviluppati. In particolare, i tumori cerebrali sono le neoplasie solide più comuni nell'infanzia, colpiscono circa il 20% della popolazione oncologica al di sotto dei 15 anni (Kaatsch, 2010) e sono la maggiore causa di mortalità correlata ai tumori nell'infanzia.

Anche se le caratteristiche istopatologiche delle neoplasie del SNC del bambino risultano simili a quelle riscontrate nell'adulto, l'incidenza delle varie forme neoplastiche riscontrate si differenzia ampiamente in funzione dell'età. I tumori del SNC più frequenti nell'adulto sono l'astrocitoma anaplastico, il glioblastoma, i meningiomi, le metastasi e i tumori ipofisari. Nei bambini gli astrocitomi tendono ad essere istologicamente e biologicamente benigni, i meningiomi e gli adenomi sono rari e le metastasi sono sporadiche. Contemporaneamente vi è un aumento considerevole di lesioni etero formative, comuni in età pediatrica e pressoché sconosciute negli adulti. Un'altra differenza sostanziale riguarda la distribuzione anatomica delle neoplasie poiché gli emisferi cerebrali sono la sede preferenziale in età adulta mentre tra i tumori cerebrali in età pediatrica, quelli della fossa cranica posteriore (PCF) sono i più frequenti.

La fossa cranica (Figura 10) è localizzata alla base della faccia interna del cranio, è delimitata anteriormente dal limite posteriore della fossa cranica media, e

posteriormente è separata dalla volta per mezzo della linea nucale superiore fino alla protuberanza occipitale esterna.

I tumori della fossa cranica posteriore sono classificati in tre gruppi principali in base alla loro origine, localizzazione e caratteristiche istopatologiche (Choux, 1999): medulloblastoma (MB), astrocitoma pilocitico (PA) ed ependimoma, e sono descritti come disturbi multifattoriali dovuti sia ad una predisposizione genetica sia a fattori ambientali (Schroeder, 2014; Ichimura, 2012). Sebbene la loro eterogeneità morfologica e molecolare sia stata studiata a fondo, la loro eziologia rimane ancora oggi sconosciuta.

Gli studi indirizzati alla caratterizzazione proteomica dei tessuti biopsici dei tumori cerebrali pediatrici a tutt'oggi risultano ancora in numero limitato. Zanini e coll. hanno analizzato tessuti tumorali prelevati da quattro differenti istotipi di MB utilizzando la 2-DE accoppiata al MALDI-TOF-MS descrivendo una differente espressione proteica associata ai diversi istotipi di MB, che presentano solo un limitato numero di proteine in comune (Zanini, 2011). Lo stesso gruppo di ricerca ha riportato un interessante caso di studio in cui è stato analizzato il profilo di espressione proteica di un campione di MB anaplastico al momento della diagnosi e in recidiva dopo radioterapia (Zanini, 2010). La radioterapia comporta una drastica riduzione del contenuto proteico totale ed induce l'espressione di poche nuove proteine insieme al mantenimento di alcune specie conservate. Tra quest'ultime è stato confermato il ruolo della statmina come marcatore di aggressività, osservato precedentemente anche in uno studio di Kuo e coll. (Kuo, 2009), che riporta una forte correlazione tra la sovra-espressione di questa proteina e la diffusione del tumore, ascrivendo alla statmina un importante ruolo come biomarcatore prognostico.

Anagnostopoulos e coll. hanno applicato un approccio complementare bioinformatico e proteomico per classificare e caratterizzare il PA infantile, rispetto al tessuto cerebrale sano, sottolineando il coinvolgimento di gruppi di proteine piuttosto che singole specie nello sviluppo della malattia. Nel PA è stata riportata una sottoregolazione della proibitina e una sovra-espressione della vimentina (Anagnostopoulos, 2011). Recentemente, attraverso l'utilizzo di tecniche come la 2-DE, il MALDI-TOF e sistemi RT² miRNA PCR Array, sono stati analizzati i pattern di espressione globale delle proteine e dei microRNA degli astrocitomi pediatrici di differenti gradi (I, II e IV) (Ruiz Esparza-Garrido, 2013). Tra i 200 spot analizzati, 49 sono risultati differenzialmente espressi tra i tumori astrocitici

confrontati con tessuti sani mentre 39 sono risultati differentemente espressi tra astrocitomi di alto e basso grado.

2.1.1 Medulloblastoma

Il medulloblastoma, un tumore neuro ectodermico primitivo del cervelletto, è tra i più aggressivi del SNC. È il più comune tumore maligno cerebrale nell'infanzia e rappresenta circa il 20% di tutti i tumori cerebrali infantili (Packer, 1990) con un picco di incidenza tra i 4 e i 7 anni (Barlett, 2013), mentre solo raramente viene diagnosticato negli adulti.

È classificato come lesione maligna di grado IV secondo l'Organizzazione Mondiale della Sanità (*World Health Organization-WHO*), e nonostante aggressive terapie multimodali che includono interventi chirurgici, radioterapia e chemioterapia, il tasso di sopravvivenza nei 5 anni non supera il 60% (Taylor, 2003).

Sulla base delle linee guida di classificazione dettate dall' WHO vengono distinti cinque diversi istotipi di MB (Louis, 2007):

- classico
- desmoplastico/nodulare
- caratterizzato da estensiva nodularità
- anaplastico
- a grandi cellule.

Il MB si sviluppa in una regione del cervello coinvolta nell'integrazione della percezione sensoriale e del controllo del movimento (Anagnostopoulos, 2015). A livello molecolare è evidente che diverse vie sono coinvolte simultaneamente nello sviluppo cerebrale e nella formazione del MB (Louis, 2007), tuttavia, non è del tutto chiaro se sono le cellule staminali multipotenti, le cellule progenitrici a sviluppo limitato o altre cellule a dar luogo a queste neoplasie pediatriche.

Istologicamente, i MB sono caratterizzati da fitti raggruppamenti di piccole e rotonde cellule ipercromatiche che tipicamente mostrano un notevole attività mitotica (Anagnostopoulos, 2015). Nonostante tutti e cinque i diversi istotipi di MB rientrino nella più aggressiva designazione WHO (grado IV), le loro distinte caratteristiche morfologiche sono state associate a sostanziali differenze anche nel comportamento biologico. I sottotipi nodulare/desmoplastico sono caratterizzati da estensiva nodularità e sono associati ad una prognosi relativamente favorevole (Rutkowski, 2009; Sure, 1995) mentre gli istotipi a grandi cellule e anaplastico

tendono a esibire un corso clinico più aggressivo con una più alta incidenza di disturbi metastatici nei neurasse tramite lo spazio sub-aracnoideo (Brown, 2000; Giangaspero, 1992; Giangaspero, 2006; McManamy, 2003).

Studi citogenetici e di ibridazione genomica comparativa hanno portato all'identificazione di molte anomalie cromosomiche nei MB, tra le più frequenti ci sono l'isocromosomia 17q, spesso in combinazione con perdita del 17p, che è stata riscontrata nel 30-50% dei casi (Kool, 2008).

Recenti studi di genomica e dei profili di espressione genetica hanno permesso di chiarire il background molecolare del MB infantile (Thompson, 2006) ma nonostante ciò gli studi proteomici sono ad oggi ancora limitati.

2.1.2 Astrocitoma pilocitico

Gli astrocitomi sono i più comuni tumori cerebrali infantili, rappresentano circa il 20% dei tumori cerebrali della popolazione pediatrica con una incidenza dello 0.5-0.6 ogni 100000 persone l'anno (Fleming, 2012). Si manifestano in pazienti giovani (da pochi mesi d'età fino a 20 anni), con un picco di incidenza tra i 6 e i 9 anni. L'astrocitoma pilocitico è considerato quello con la migliore prognosi tra tutti gli astrocitomi in quanto generalmente non è una lesione infiltrante e presenta una crescita lenta, che raramente mostra una progressione a lesione maligna anche se può permeare i tessuti circostanti (Sadighi, 2013). Di solito viene trattato con successo, con una prognosi favorevole a seguito di terapie appropriate che includono la rimozione chirurgica e la chemioterapia.

Gli astrocitomi sono suddivisi in quattro categorie secondo le loro caratteristiche istologiche e sulla base delle linee guida di classificazione WHO:

- astrocitoma pilocitico (Grado I),
- astrocitoma diffuso (Grado II),
- astrocitoma anaplastico (Grado III),
- glioblastoma multiforme (Grado IV) .

In contrasto con la crescente conoscenza delle anomalie genetiche nei gliomi di grado elevato, relativamente poco si conosce riguardo all'espressione delle proteine nel PA infantile. Il lavoro pubblicato da Anagnostopoulos e coll. fornisce un'analisi sul PA infantile da un punto di vista proteomico, genomico e bioinformatico (Anagnostopoulos, 2011), che risulta essere il primo studio proteomico in questo settore. La patogenesi del PA non è ancora ben chiara: tradizionalmente si ritiene che gli astrocitomi derivino dagli astrociti (cellule neurali costituenti della nevroglia il cui nome è dovuto al fatto che presentano delle ramificazioni originatesi dal

nucleo e che conferiscono loro una forma stellata) e da loro cellule precursori. L'origine degli astrocitomi è determinata da alterazioni geniche caratteristiche, compresa l'attivazione di proto-oncogeni e l'inattivazione di geni oncosoppressori che svolgono un ruolo importante nella crescita cellulare, nell'apoptosi, nella motilità cellulare e nelle vie di invasione. In particolare sembra giocare un ruolo importante la perdita della funzione nel gene oncosoppressore p53, dovuta a mutazioni, che avviene nelle fasi iniziali della tumorigenesi nel 30-40% dei casi di astrocitoma (Nozaki, 1999). Considerando però che il PA progredisce raramente a gradi di malignità più elevati (Kleihues, 1993), è stato ipotizzato che la mutazione nel gene oncosoppressore p53 non sia rilevante nella tumorigenesi e quindi il PA risulterebbe un'entità geneticamente distinta dai gradi più elevati (II-IV) (Nozaki, 1999).

2.2 Materiali e Metodi

I campioni di tessuto cerebrale utilizzati in questo studio sono stati prelevati durante 8 interventi chirurgici su pazienti ricoverati nel reparto di Neurochirurgia Pediatrica dell'Università Cattolica di Roma, di età compresa tra i 2 e i 12 anni. Parte dei tessuti biotipici sono stati sottoposti ad esame istopatologico, ed è stata accertata la diagnosi di MB in 4 casi e di PA in altri 4 (vedi Tabella 3). La porzione di campione non utilizzato per l'analisi istologica è stata congelata a -80°C , e in seguito riunita con gli altri campioni appartenenti alla stessa tipologia di tumore per formare due *pool*, siglati MB e PA, che sono stati successivamente trattati per poter essere utilizzati in un studio proteomico realizzato grazie ad una collaborazione tra l'Università Cattolica del Sacro Cuore di Roma e l'Università di Cagliari, di cui in questa tesi è descritta la parte relativa all'analisi della frazione insolubile in acido mediante elettroforesi bidimensionale.

Il protocollo e la scheda per la richiesta del consenso informato per la partecipazione alla ricerca descritta in questo studio sono stati approvati dal Comitato Etico del Dipartimento di Pediatria e dal Comitato Etico Medico della Facoltà di Medicina dell'Università Cattolica di Roma, in accordo con quanto previsto dalla Dichiarazione di Helsinki. Il consenso alla partecipazione è stato firmato dai genitori dei bambini che hanno partecipato allo studio.

A causa del ridotto quantitativo di materiale disponibile, non è stato possibile procedere all'analisi dei singoli tessuti prelevati mediante biopsia, ma è stato necessario analizzare i due *pool* di tessuti tumorali.

Ciascun *pool* di tessuto tumorale, dopo scongelamento, è stato pesato ed addizionato di un appropriato volume di una soluzione acquosa di ACN ($\text{H}_2\text{O}:\text{ACN}$ in rapporto 70:30) contenente 0,4% (v/v) di TFA al fine di ottenere una concentrazione di 0,2 mg di tessuto/ μL . I campioni sono stati omogenati per 60 sec in ghiaccio utilizzando l'ULTRA-TURRAX T10 (IKA-Werke GmbH & Co. KG, Staufen, Germany). Gli omogenati sono stati quindi sonicati (Sonifer 450, Branson Ultrasonics, Danbury, USA) per 10 sec e in seguito centrifugati a 24000 g a 4°C per 15 min. Il supernatante, contenente la frazione proteica solubile in acido, è stato separato dal pellet, contenente la frazione proteica insolubile in acido, ed entrambi i campioni sono stati conservati a -80°C prima di essere sottoposti alle diverse analisi proteomiche.

I supernatanti sono stati analizzati in HPLC-ESI-MS/MS presso l'Istituto di Biochimica e Biochimica Clinica dell'Università Cattolica del S. Cuore di Roma,

mentre i precipitati sono stati analizzati mediante 2-DE nei laboratori della sezione Biomedica del Dipartimento di Scienze della Vita e dell'Ambiente dell'Università di Cagliari. In Figura 12 è riportato lo schema di lavoro seguito durante le analisi proteomiche.

A ciascun pool di pellet sono stati addizionati 5 mL di una soluzione di precipitazione contenente 10% (p/v) TCA, 20 mM DTT in acetone. Le sospensioni così ottenute sono state omogenate con l'ULTRA-TURRAX T25 basic (IKA-Werke GmbH & Co. KG, Staufen, Germany) eseguendo 3 cicli consecutivi da 30 sec ciascuno. I campioni così trattati sono stati incubati a $-20\text{ }^{\circ}\text{C}$ per una notte, al fine di ottenere la precipitazione delle proteine. Il giorno seguente i campioni sono stati centrifugati a 24000 g per 30 min a $4\text{ }^{\circ}\text{C}$. Al termine della centrifugazione è stato eliminato il supernatante ed il pellet è stato sottoposto a 3 lavaggi consecutivi con una soluzione 20 mM DTT in acetone freddo ($-20\text{ }^{\circ}\text{C}$), al fine di rimuovere il TCA dal campione, centrifugando ogni volta a 24000 g per 30 min a $4\text{ }^{\circ}\text{C}$ ed eliminando il supernatante al termine di ogni lavaggio. Il precipitato ottenuto è stato posizionato sotto cappa aspirante a temperatura ambiente per circa 3 ore per permettere l'evaporazione dei residui di acetone.

I pellet ottenuti nella sezione precedente sono stati neutralizzati con l'aggiunta di 5 μL di NaOH 0,1M vorticando per 2 min. Sono stati quindi addizionati ad ogni campione 250 μL di soluzione di risolubilizzazione (DeStreak Rehydration Solution) e i campioni sono stati incubati 3 ore a temperatura ambiente, vorticando occasionalmente.

Al termine dell'incubazione, i campioni sono stati centrifugati a 24000 g per 10 min a $4\text{ }^{\circ}\text{C}$ al fine di rimuovere il materiale insolubile. Il supernatante ottenuto è stato prelevato, trasferito in una nuova provetta, misurato e portato ad un volume finale di 230 μL con l'aggiunta di un opportuno volume di soluzione di risolubilizzazione.

La determinazione della concentrazione proteica totale dei campioni è stata effettuata utilizzando il 2D-Quant Protein Kit, come descritto nella sezione 1.2.3.

In 11 è riportata la retta standard costruita mediante il 2-D Quant kit con le soluzioni di albumina bovina serica a concentrazione nota, che è stata utilizzata per ricavare il quantitativo proteico dei campioni a concentrazione incognita esaminati.

I due pool MB e PA hanno mostrato un contenuto proteico di 43,6 e 45 μg rispettivamente, nei 15 μL di campione utilizzati per il saggio, perciò il quantitativo proteico totale nel volume complessivo di campione di 200 μL è risultato essere 580 μg per il campione MB e 600 μg per il campione PA.

Il campione costituito dalla frazione insolubile in acido del pool di MB e di PA, ottenuto dopo opportuno trattamento con TCA, e contenente circa 600 µg di proteine è stato risolubilizzato in 200 µL di DeStreak, a cui è stato aggiunto lo 0,5% di anfoliti che coprono l'intervallo di pH 3-10.

Tale campione è stato utilizzato per analisi bidimensionali sia a finalità analitica che preparativa. Sono state effettuate tre repliche tecniche delle analisi bidimensionali analitiche sui campioni PA ed MB per aumentare l'affidabilità e la riproducibilità dei risultati ottenuti.

Per tutte le elettroforesi bidimensionali sono state utilizzate IPG-strip (ReadyStrips IPG, Bio-Rad) da 11 cm contenenti un gradiente di pH immobilizzato 3-10 non lineare. Per la realizzazione dei gel analitici, le strip sono state reidratate prelevando circa 8 µL di ciascun campione corrispondenti a 25 µg di proteine totali, mentre per la realizzazione dei gel preparativi sono stati utilizzati 48 µL di ciascun campione corrispondenti a 150 µg di proteine.

Il protocollo utilizzato per l'IEF, in entrambi i casi, è analogo a quello descritto in Sezione 1.2.4.

Per l'SDS-PAGE sono stati utilizzati gel precast Criterion (Bio-Rad) con parametro T in gradiente dal 10 al 20% in modo che le dimensioni delle maglie del gel di poliacrilammide siano sempre più strette man mano che si passa dalla sommità del gel fino alla sua estremità inferiore, per poter separare efficacemente sia le proteine ad alto che quelle a basso PM, consentendo di ottenere un aumento nella risoluzione della separazione. Il protocollo seguito per l'analisi SDS-PAGE è analogo a quello riportato nella sezione 1.2.5.

Al termine della corsa elettroforetica, si è effettuata la fase di colorazione. I gel analitici sono stati colorati in fluorescenza con il SYPRO Ruby, mentre i gel preparativi sono stati colorati con il Coomassie colloidale EZBlue seguendo i protocolli riportati nella sezione 1.2.6. Le immagini dei gel sia analitici che preparativi sono state acquisite ad alta risoluzione mediante il sistema ChemiDoc MP, come descritto nella sezione 1.2.7. Le immagini delle tre repliche dei gel analitici di PA e MB sono state analizzate con il software SameSpots per la quantificazione degli spot. Tutti gli spot che hanno mostrato variazioni statisticamente significative nei gel analitici sono stati selezionati per essere asportati dal gel preparativo ed essere sottoposti a digestione triptica al fine di identificare la proteina presente in ciascuno spot secondo il protocollo riportato nella sezione 1.2.8. Per il riconoscimento delle proteine presenti nei campioni, le sequenze dei peptidi triptici ottenuti con gli esperimenti di MS e MS/MS ad alta

risoluzione sono state confrontate con le sequenze proteiche riportate nella banca dati SwissProt Homo Sapiens proteome (Uniprot_homo+sapiens_reviewed_noisiform, rilasciato il 19 Agosto 2013), con l'ausilio del programma Proteome Discoverer 1.4.0, impostando il filtro ad alta confidenza.

La classificazione Gene Ontology (GO) delle proteine identificate nella frazione insolubile in acido dei due pool tumorali in base alle classi di proteine, ai processi biologici e alle funzioni molecolari è stata eseguita utilizzando il software PANTHER Classification System (**P**rotein **A**nalysis **T**hrough **E**volutionary **R**elationships, versione 9.0, <http://www.pantherdb.org>).

L'analisi statistica è stata effettuata attraverso il modulo di analisi statistica multivariata del software SameSpots sui volumi log-normalizzati per tutti gli spot, calcolando il parametro *Fold Variation*, cioè la differenza dei volumi log-normalizzati tra i due gruppi, ottenuta confrontando i valori medi dei volumi log-normalizzati di ogni singolo spot in tutti i campioni all'interno di un gruppo e il parametro p di ANOVA, ottenuto dall'analisi della varianza unidirezionale di tutti gli spot.

Solamente gli spot che hanno presentato un valore di *Fold Variation* $\geq 2,0$ e un valore p di ANOVA $\leq 0,05$ sono stati considerati statisticamente significativi e sono stati presi in considerazione per le successive fasi dell'analisi.

2.3 *Risultati e Discussione*

Lo scopo di questo studio è la caratterizzazione comparativa del proteoma del medulloblastoma, il più frequente tumore maligno dell'età pediatrica, e dell'astrocitoma pilocitico, meno aggressivo. Questa ricerca, frutto di una collaborazione tra l'Università Cattolica del Sacro Cuore di Roma e l'Università di Cagliari, ha utilizzato una piattaforma proteomica integrata con l'utilizzo delle tecniche HPLC e dell'elettroforesi bidimensionale accoppiate alla spettrometria di massa ESI ad alta risoluzione. Grazie all'utilizzo di questa piattaforma, i campioni sono stati analizzati sia mediante l'approccio top-down che mediante l'approccio bottom-up, e l'integrazione delle diverse tecniche di indagine proteomica ha permesso una più ampia caratterizzazione dei tessuti tumorali, sfruttando la complementarità dei due approcci. La ricerca è stata effettuata su due pool di tessuti biotici, ottenuti raggruppando insieme i tessuti che all'esame istopatologico sono risultati della stessa tipologia, dato che, a causa della difficoltà di reperimento di questa tipologia di campioni, non è stato possibile disporre di un quantitativo di materiale sufficiente per procedere all'analisi di ciascun campione singolarmente. I due pool sono stati processati secondo lo schema riportato in Figura 12 che descrive i diversi passaggi applicati. Dopo aggiunta di TFA, le proteine solubili in acido sono state separate per centrifugazione da quelle insolubili in acido. Sulla frazione proteica solubile in acido è stata applicata la piattaforma proteomica integrata top-down/bottom-up, mentre sulla frazione proteica insolubile in acido, contenente un alto numero di proteine ad alto PM, è stata applicata l'analisi bottom-up, analizzando i peptidi triptici ottenuti per digestione di singoli spot del gel bidimensionale, e tale analisi viene descritta dettagliatamente in questa tesi.

In Figura 13A e 13B sono riportate le immagini dei gel bidimensionali ottenuti rispettivamente dai campioni di PA ed MB. Dopo aver analizzato i gel mediante il software SameSpots è stato possibile evidenziare un certo numero di spot sovra-espresi o sotto-espresi in un tessuto rispetto all'altro con differenze statisticamente significative. Questi spot sono stati escissi dai gel e sottoposti a digestione triptica per procedere all'identificazione delle proteine, ottenuta dopo analisi dei peptidi triptici mediante esperimenti di HPLC-ESI-MS/MS utilizzando uno spettrometro di massa ad alta risoluzione.

In particolare, le sequenze dei peptidi triptici ottenuti sono state confrontate con le sequenze proteiche riportate nella banca dati SwissProt Homo Sapiens proteome con l'ausilio del programma Proteome Discoverer 1.4.0. Nella Tabella 4a e 4b sono

riportate tutte le proteine identificate con questa strategia, unitamente ai diversi parametri ottenuti con l'analisi effettuata per mezzo del software SameSpots.

Utilizzando questo software, i gel sono stati allineati e il volume di ciascuno spot è stato quantificato e confrontato con quello dello spot localizzato nella stessa posizione nell'altro gruppo di gel al fine di valutare l'espressione differenziale delle diverse proteine nelle due tipologie di gel.

Quasi tutti gli spot identificati nei campioni di PA ed MB contengono singole proteine, ma in alcuni casi l'analisi in spettrometria di massa ha riconosciuto diverse proteine all'interno dello stesso spot, evidenziando la possibile co-localizzazione di più proteine. Alcune proteine sono state invece riconosciute in più spot che risultano localizzati nel gel come treni di spot caratterizzati da un PM molto simile ma differente pI: sono state identificate diverse proteoforme della proteina gliale fibrillare acida (GFAP), della malato deidrogenasi mitocondriale, della peroxiredoxina-1 e della ribonucleoproteina nucleare eterogenea A2/B1. Questo suggerisce la presenza di PTM e nel caso specifico, molto probabilmente, di fosforilazioni.

Tra le proteine identificate, 7 sono risultate sovra-espresse nel campione di MB rispetto al campione di PA: peroxiredoxina-1, ribonucleoproteina nucleare eterogenea A2/B1, malato deidrogenasi, nucleoside difosfato chinasi A, peptidil-prolil cis-trans isomerasi A, fruttosio bifosfato aldolasi A, glutatione S-transferasi P (Tabella 4a). Nel campione di PA sono risultate sovra-espresse 8 proteine: GFAP, catena B dell' α -cristallina, catena leggera della ferritina, serotransferrina, annexina A5, proteoforma cerebrale della proteina che lega gli acidi grassi, sorcina e apolipoproteina AI (Tabella 4b).

Alcune delle proteine identificate nelle frazione proteica insolubile in acido sono state identificate anche nel corrispondente supernatante (ubiquitina, GFAP, subunità β dell'emoglobina, actina 1 citoplasmatica, gliceraldeide-3 fosfato deidrogenasi, tubulina, peptidil-prolil cis-trans isomerasi A, α enolasi e serotransferrina), ed è ragionevole ritenere che queste proteine si ripartiscano tra le due fasi in funzione della loro concentrazione o di peculiari proprietà fisico-chimiche.

La peroxiredoxina-1 è stata identificata in due spot (#64, #65) del gel 2-DE di MB localizzati allo stesso PM ma a diverso pI (Figura 13B). Lo spot con un maggiore pI (# 64) ha mostrato una *fold variation* aumentata di 3,8 volte in MB rispetto al corrispondente punto di PA, e quello con minor pI (#65), che corrisponde probabilmente alla proteoforma mono-fosforilata, ha mostrato un aumento di 2,8

volte (Figura 14). Questi dati sono in accordo con i risultati ottenuti applicando la strategia bottom-up all'analisi dei supernatanti in cui la peroxiredoxina-1 è stata riconosciuta come proteina esclusiva del MB.

Le peroxiredoxine rappresentano una famiglia di proteine coinvolte nella riduzione del perossido di idrogeno, e quindi nella regolazione dello stato redox cellulare; partecipano ai meccanismi di difesa antiossidanti della cellula, ai meccanismi di regolazione della proliferazione cellulare e all'aumento della resistenza ai farmaci nelle cellule in coltura (Lehtonem, 2004).

In precedenti lavori la peroxiredoxina-1 è stata descritta essere sovra-espressa nel glioblastoma multiforme (Odreman, 2005), in cui è stato osservato un rapporto tra questa proteina e la progressione del tumore, suggerendo un suo possibile utilizzo come nuovo marcatore per il glioblastoma multiforme. In uno studio sull'espressione della peroxiredoxina-1 nei tumori tiroidei umani Yanagawa e coll. hanno suggerito che questa proteina giochi un possibile ruolo come marcatore tumorale per discriminare tra le diverse tipologie di tessuti tumorali in base all'osservazione che la peroxiredoxina-1 è risultata sovra-espressa nelle neoplasie follicolari, mentre il suo livello di espressione rimane normale nei carcinomi papillari (Yanagawa, 1999). La peroxiredoxina-1 è risultata sovra-espressa anche nel cancro del polmone (Lehtonem, 2004), nel carcinoma mammario (Karihtala, 2003), nel mesoentelioma maligno (Kinnula, 2002) nelle linee cellulari D238 e DAOY di MB (Peyrl, 2003).

La ribonucleoproteina nucleare eterogenea A2/B1 (hnRNP A2/B1) è stata identificata in due spot (#70, #72) del gel 2-DE di MB. Entrambi gli spot sono localizzati in corrispondenza dello stesso PM ma presentano un differente pI (Figura 13B). Lo spot con un maggiore pI (#70) ha mostrato una *fold variation* aumentata di 6,0 volte in MB rispetto alla corrispondente area nel gel di PA mentre quello con pI inferiore (#72), che corrisponde probabilmente alla proteoforma mono-fosforilata, ha mostrato un aumento di 2,8 volte (Figura 14). Questa proteina è una componente importante del complesso del core hnRNP A2/B1 che si trova nel nucleo delle cellule dei mammiferi. È una proteina che lega l'RNA ed è coinvolta nello splicing dell'RNA nel nucleo e nel trasporto dell'mRNA dal nucleo al citoplasma (Mayeda, 1994). È stata descritta precedentemente in lavori effettuati sulle linee cellulari D238 e DAOY20 di MB (Peyrl, 2003) ed è stata riportata essere sovra-espressa anche in altre tipologie tumorali: tumore al pancreas (Yan-Sanders, 2002), tumore del polmone non a piccole cellule (Sueoka, 2001), carcinoma orale a cellule squamose (Goto, 1999). Nel tumore alla mammella è stata descritta

l'espressione di hnRNP-A2/B1 nella maggioranza dei tumori primari invasivi rispetto al 10% dei tessuti mammari normali, ma ancora non risulta chiaro il ruolo preciso di questa proteina nella cancerogenesi al seno (Zhou, 2001).

La proteoforma mitocondriale della malato-deidrogenasi è stata identificata in due spot (#68 e #69) del gel 2-DE di MB. Entrambi gli spot sono localizzati in corrispondenza dello stesso PM ma presentano un differente pI (Figura 13 B). Lo spot caratterizzato da un pI maggiore (# 68) ha mostrato un aumento di 3,2 in MB rispetto alla corrispondente area di PA, e quello con pI inferiore (# 69), che corrisponde probabilmente alla proteoforma mono-fosforilata, ha mostrato un aumento di 2,9 volte. Dalla visione tridimensionale degli spot riportata in Figura 14 è possibile notare come gli spot siano presenti in quantità molto bassa nel gel 2-DE di PA. Anche questa proteina era stata precedentemente riconosciuta in esperimenti condotti sulle linee cellulari D238 e DAOY di MB (Peyrl, 2003).

Le proteine nucleoside difosfato chinasi A (NDK A) e B (NDK B) sono state identificate negli spot #38 e #34, rispettivamente del gel 2-DE di MB (Figura 13B). La proteina NDK A risulta sovra-espressa (3,0 volte) in MB rispetto alla corrispondente area del gel di PA (Figura 14), mentre la NDK B non ha mostrato un aumento significativo. In un precedente lavoro, le catene polipeptidiche A e B della nucleoside difosfato chinasi sono state identificate nelle linee cellulari D238 di MB (Peyrl, 2003). È stato suggerito che NDK potrebbe svolgere un duplice ruolo nella progressione del tumore infatti, in tumori primari è risultata sovra-espressa nelle fasi iniziali, mentre la sua riduzione nelle fasi più tardive è stata collegata all'aggressività del tumore e al suo potenziale metastatico (Prabhu, 2012).

La proteina peptidil-prolil cis-trans isomerasi A (PPIase) è stata identificata nello spot #35 del gel 2-DE di MB (Figura 13B), dove ha mostrato un aumento di 2,3 volte rispetto alla corrispondente area del gel di PA. Questa proteina è stata identificata anche nella frazione solubile in acido analizzata con l'approccio bottom-up, in cui è risultata essere una proteina esclusiva di MB. Nella corrispondente area del gel 2-DE di PA è stata rilevata in quantità estremamente piccole (Figura 14). La PPIase è una foldasi, accelera il ripiegamento *in vivo* delle proteine e catalizza l'isomerizzazione cis-trans dei legami peptidici imidici che la prolina forma negli oligopeptidi.

La PPIase è stata precedentemente identificata in esperimenti condotti su campioni di MB (istotipo classico e istotipi ad estesa modularità) (Zanini, 2011), sulle linee cellulari DAOY (Peyrl, 2003) e in studi condotti sul PA infantile (Anagnostopoulos, 2011).

La proteina Glutathione S-transferasi P (GST-pi) è stata identificata nello spot #44 dei gel di PA e MB (Figura 13A e 13B), con un aumento di 2,4 volte nel gel di MB rispetto alla corrispondente area del gel di PA (Figura 14). Questa proteina è stata identificata anche in esperimenti condotti sulle linee cellulari di medulloblastoma DAOY e D238 (Peyrl, 2003). GST-pi partecipa all'attacco nucleofilo dell'atomo di zolfo del glutathione sui centri elettrofili di vari composti endogeni e xenobiotici. Può essere considerato una proteina marker per la rilevazione di tossicità chimica e carcinogenesi, poiché durante la carcinogenesi chimica è espressa abbondantemente nelle cellule tumorali.

Tra le proteine che sono risultate sovra-espresse nei gel di PA, la GFAP è stata identificata come l'unica proteina presente in una serie di spot (5a, 5b, 5c e 5d, Figura 13A) localizzati nella parte acida ad alto PM del gel 2-DE di PA. A causa della loro vicinanza e della loro abbondanza, non è stato possibile asportare separatamente i singoli spot dal gel, ma sono stati escissi e sottoposti a digestione triptica in un campione unico. Gli spot 5a, 5b, 5c e 5d sono risultati sovra-espresi 7,1, 9,4, 8,4 e 3,0 volte rispettivamente, nel gel 2-DE di PA rispetto alla corrispondente area del gel di MB, dove gli spot appaiono presenti in quantità molto bassa (Figura 14). I risultati dell'analisi dei digeriti triptici hanno mostrato che la GFAP è l'unica proteina presente in questo treno di spot e questo risultato suggerisce che si tratti di proteoforme fosforilate.

La GFAP è la principale proteina dei filamenti intermedi gliali, è specifica degli astrociti maturi ed è un importante componente del citoscheletro. Recenti scoperte sulla biologia degli astrociti e su nuove funzioni dei filamenti intermedi hanno aumentato l'interesse per le funzioni della GFAP. Il suo ruolo strutturale negli astrociti è ampiamente accettato, ma nel corso degli anni, questa proteina ha dimostrato di essere coinvolta in diverse funzioni svolte dagli astrociti nei processi di rigenerazione, plasticità sinaptica e gliosi reattiva. Inoltre, sono state identificate differenti sottopopolazioni di astrociti, che possono svolgere diversi compiti nella fisiologia e nella patologia cerebrale e che vengono classificati non solo per il loro aspetto spaziale e temporale, ma anche per l'espressione di specifiche proteine dei filamenti intermedi, tra cui proprio diverse isoforme di GFAP. La presenza di queste proteoforme aumenta la complessità del citoscheletro degli astrociti ed è probabile che tale presenza sottintenda a specifiche funzioni (Middeldorp, 2011).

La GFAP è stata inoltre identificata anche in altri due spot localizzati ad un PM inferiore e a pI più acido (#54 e #51, Figura 13A), che corrispondono a frammenti proteolitici della proteina. Anche questi spot sono risultati sovra-espresi (4,3 e 5,7

volte, rispettivamente) nel gel 2-DE di PA. Frammenti proteolitici di GFAP sono stati precedentemente identificati in diversi lavori (Chumbalkar, 2005; Khalil, 2007; Narayan, 1986), in cui è stato ipotizzato che questa proteina subisca proteolisi cellulare specifica per un'aumentata destabilizzazione a seguito del rapido turnover a cui è soggetta nelle cellule (Chumbalkar, 2005). La presenza di frammenti di GFAP con PM inferiore è stata rilevata anche negli esperimenti condotti sulla frazione solubile in acido con gli approcci top-down/bottom-up in cui è stata riscontrata la sovra-espressione di alcuni dei frammenti identificati in PA.

GFAP è un noto marcatore per il decadimento neuronale e i danni cerebrali. Distingue gli astrociti dalle altre cellule gliali durante la fase di sviluppo del sistema nervoso centrale e i suoi livelli sembrano rispondere dinamicamente all'invecchiamento e alle lesioni neurodegenerative: molti studi hanno dimostrato che la quantità di GFAP generalmente aumenta nei tumori cerebrali in quanto si può accumulare mano a mano che la proporzione di astrociti aumenta con la crescita del tumore, e nel caso di malattie neurodegenerative, come il morbo di Alzheimer (Korolainen, 2005). In uno studio precedente è stata riportata l'assenza di variazione di concentrazione di GFAP tra i tessuti di PA infantile e i tessuti cerebrali normali (Anagnostopoulos, 2011), mentre è stata riportata una differente espressione tra i gliomi di diverso grado e in particolare una sotto-espressione dei livelli di GFAP nei gliomi di grado elevato che potrebbe indicare un potenziale ruolo di questa proteina come marcatore specifico per i tumori di basso grado (Odreman, 2005).

La catena B dell' α cristallina è stata identificata nello spot #41 del gel 2-DE di PA (Figura 13A) e presenta un aumento di 3,2 volte nel gel di PA rispetto alla stessa area del gel 2-DE di MB (Figura 14), dove lo spot risulta quasi non rilevabile. Questa osservazione è in accordo con i dati precedentemente riportati da Peyrl e coll. che non riportano questa proteina nelle linee cellulari DAOY e D238 di MB (Peyrl, 2003). La catena B dell' α cristallina è stata descritta come sovra-espressa in varie malattie neurologiche e diversi tumori tra cui gliomi (Odreman, 2005), tumore della mammella, della prostata e carcinoma delle cellule renali (Pinder, 1994; Takashi, 1998). Tuttavia, dati precedenti ottenuti confrontando il livello di espressione di questa proteina tra astrocitoma di grado basso ed elevato non risultano in accordo: in uno studio è stata riportata la sovra-espressione della catena B dell' α cristallina negli astrocitomi di grado elevato (Odreman, 2005), mentre un altro studio la descrive come sovra-espressa negli astrocitomi di grado basso (Iwadate, 2004).

Infine, la proteoforma cerebrale della proteina che lega gli acidi grassi (B-FABP) è stata identificata nello spot #29 del gel 2-DE di PA (Figura 13A) che ha mostrato un aumento di 2,0 rispetto alla corrispondente area del gel 2-DE di MB (Figura 14), dove è riscontrabile in quantità molto bassa. B-FABP è normalmente espressa nelle cellule gliali radiali, dove svolge un ruolo nella creazione della rete di fibre gliali radiali richiesta per la migrazione dei neuroni (Mita, 2007). B-FABP è stata descritta in lavori precedenti sugli astrocitomi e in alcune linee cellulari di glioma maligno che co-esprimono gli mRNA di GFAP e di B-FABP (Godbout, 1998).

La classificazione Gene Ontology (GO) delle proteine identificate nella frazione insolubile in acido dei due tessuti tumorali per classi di proteine, processi biologici e funzioni molecolari (Figura 15), ha confermato la presenza di proteine con attività molecolare antiossidante esclusivamente nel pool di MB, oltre all'incremento dei valori percentuali delle proteine coinvolte nei processi metabolici e nei processi del sistema immunitario. Importanti differenze sono state determinate nella classificazione per classi di proteine: il campione MB presenta una maggiore percentuale di proteine leganti gli acidi nucleici, di isomerasi e di liasi. Nel campione PA è stata invece rilevata la presenza di proteasi, idrolasi, proteine recettoriali, proteine di riserva e proteine strutturali trans membrana con funzione di regolazione dei segnali, ad esempio proteine leganti il calcio.

2.4 Conclusioni

L'utilizzo di una piattaforma proteomica integrata top-down/bottom-up, basata sull'accoppiamento della cromatografia liquida e della 2-DE con la spettrometria di massa LTQ-Orbitrap ad alta risoluzione, ha permesso di effettuare un'indagine comparativa sui profili di espressione proteica di due tumori cerebrali (medulloblastoma e astrocitoma pilocitico), frequenti in età pediatrica, e caratterizzati da diversa origine istologica e con differente grado di malignità. Le due piattaforme proteomiche si sono ben integrate nella caratterizzazione del proteoma dei due tumori cerebrali, permettendo l'analisi di una vasta gamma di masse molecolari e l'identificazione di interessanti proteoforme.

In particolare, l'analisi della frazione proteica insolubile in acido, effettuata mediante elettroforesi bidimensionale accoppiata alla spettrometria di massa ad alta risoluzione, ha permesso l'identificazione di diverse proteine esclusive di ciascun tessuto tumorale e di alcune proteine in comune tra i due diversi istotipi tumorali. Inoltre, va sottolineato che l'elettroforesi bidimensionale ha consentito l'identificazione di specie proteiche insolubili nelle condizioni sperimentali utilizzate per l'analisi in HPLC-ESI-MS.

I dati così ottenuti possono contribuire alla caratterizzazione del proteoma dei tumori infantili della fossa cranica posteriore: le proteine esclusive di un particolare istotipo tumorale potrebbero essere rilevanti per l'identificazione di specifici biomarcatori del tumore, mentre le proteine in comuni ai due istotipi, potrebbero fornire una visione molecolare complessiva dei tessuti tumorali cerebrali.

Esperimenti futuri su un numero statisticamente significativo di campioni individuali di entrambi gli istotipi tumorali potranno essere eseguiti al fine di verificare la significatività dei risultati finora ottenuti, in modo da valutare le variazioni inter-individuali, le differenze tra i diversi sottotipi molecolari tumorali e confermare il ruolo di potenziali biomarcatori della patologia o di malignità che potrebbero avere alcune proteine.

Il contributo apportato da questo studio è significativo in relazione alla crescente necessità di disporre di terapie sempre più avanzate, poiché l'analisi dei profili di espressione proteica dei tessuti tumorali può costituire un utile strumento per lo studio delle malattie a livello molecolare, individuando possibili biomarcatori e bersagli terapeutici, utili per lo sviluppo di terapie innovative, in particolare nel campo dei tumori cerebrali pediatrici, dove ad oggi ancora pochi risultano gli studi effettuati.

Tabella 3: Classificazione degli otto campioni biotipici esaminati in questo studio

Paziente	Età (anni)	Diagnosi
1	12	Astrocitoma Pilocitico, I WHO
2	2	Astrocitoma Pilocitico, I WHO
3	2	Astrocitoma Pilocitico, I WHO
4	10	Astrocitoma Pilocitico, I WHO
5	8	Medulloblastoma (classico), IV WHO
6	6	Medulloblastoma (classico/anaplastico), IV WHO
7	11	Medulloblastoma (classico), IV WHO
8	4	Medulloblastoma (classico/anaplastico), IV WHO

WHO = World Health Organization

Tabella 4a: Elenco della proteine della frazione insolubile in acido che sono risultate statisticamente sovra-esprese nel campione di Medulloblastoma (MB) rispetto al campione di Astrocitoma pilocitico (PA), dopo digestione triptica in-gel degli spot della 2-DE e analisi HPLC-ESI-MS/MS

Spot #	Codice Uniprot	Proteina	PM [kDa] ^(c)	pI ^(c)	Punteggio ^(b)	Copertura ^(b)	# Peptidi unici ^(b)	# Peptidi ^(b)	Identificata in	ANOVA (p) ^(d)	Variazione* ^(d)	
4	P62158	Calmodulina	16.7	4.1	60.8	61.74%	4	4	MB, PA	2.89E-02	1.7	↑ in MB
8	P04264	Citocheratina 1 (di tipo II)	65.9	8.2	18.6	6.99%	3	4	PA	NS	1.9	↑ in MB
10	Q9BVA1	Beta Tubulina	49.9	4.8	324.6	50.34%	4	19	MB	NS	1.9	↑ in MB
16	P11142	Heat shock cognate 71 kDa	70.7	5.4	12.9	12.69%	9	9	MB	NS	1.4	↑ in MB
20	P06733	Alfa-enolasi	47.0	7.0	9.9	19.12%	5	5	PA	8.35E-03	1.4	↑ in MB
21	P16152	Carbonil reductasi [NADPH] 1	30.2	8.6	6.5	9.75%	2	2	PA	8.64E-04	1.6	↑ in MB
23	P04406	Gliceraldeide-3-fosfato deidrogenasi	35.9	8.6	7.4	11.64%	2	2	PA	NS	1.5	↑ in MB
34	P22392	Nucleoside difosfato chinasi B	17.2	8.6	43.4	44.08%	5	5	MB	2.00E-02	1.8	↑ in MB
35	P62937	Peptidil-prolil cis-trans isomerasi A	18.0	7.7	60.6	47.27%	6	6	MB	2.59E-05	2.3	↑ in MB
38	P15531	Nucleoside difosfato chinase A	17.0	5.8	98.4	81.58%	10	10	MB	3.38E-05	3.0	↑ in MB
44	P09211	Glutatione S-transferasi P	23.2	5.4	39.4	37.14%	4	4	MB, PA	8.09E-04	2.4	↑ in MB
52	P04075	Fruttosio-bisfosfato aldolasi A	39.3	8.4	16.2	25.00%	5	5	MB	5.77E-04	2.2	↑ in MB
62	P0CG48	Ubiquitina	8.6	6.6	36.4	64.38%	5	5	MB	NS	2.1	↑ in MB
64	Q06830	Peroxiredoxina-1	21.9	8.3	122.5	57.79%	11	11	MB	2.00E-03	3.8	↑ in MB
65	Q06830	Peroxiredoxina-1	2.0	8.3	166.1	54.77%	14	14	MB	3.59E-04	2.8	↑ in MB
68	P40926	Malato deidrogenasi mitocondriale	32.9	8.5	125.3	46.45%	12	12	MB	2.00E-03	3.2	↑ in MB
69	P40926	Malato deidrogenasi mitocondriale	32.9	8.5	89.7	36.69%	8	8	MB	2.00E-03	2.9	↑ in MB
70	P22626	Ribonucleoproteina eterogenea A2/B1	37.4	9.0	11.4	7.08%	2	2	MB	9.86E-05	6.0	↑ in MB
72	P22626	Ribonucleoproteina eterogenea A2/B1	37.4	9.0	69.0	37.11%	8	8	MB	1.40E-02	2.8	↑ in MB

NS= Proteine senza differenze statisticamente significative in MB e PA

*Solo gli spot che differiscono di almeno 2,0 volte nei volumi (log-normalizzati) e con un valore p di ANOVA ≤ 0.05 sono stati considerati statisticamente significativi.

(a) = i numeri corrispondono a quelli riportati nelle Figure 13 e 14

(b) = dati ottenuti dal software Proteome Discoverer 1.4.0

(c) = dati ottenuti da Uniprot

Tabella 4b: Elenco delle proteine della frazione insolubile in acido che sono risultate statisticamente sovra-esprese nel campione di Astrocitoma pilocitico (PA) rispetto al campione di Medulloblastoma (MB), dopo digestione triptica in-gel degli spot della 2-DE e analisi HPLC-ESI-MS/MS

Spot #	Codice Uniprot	Proteina	PM [kDa] ^(c)	pI ^(c)	Punteggio ^(b)	Copertura ^(b)	# Peptidi unici ^(b)	# Peptidi ^(b)	Identificata in	ANOVA (p) ^(d)	Variazione ^{*(d)}
5a	P14136	Proteina Gliale fibrillare acida	49.9	5.4	146.6	28.47%	10	10	PA	4.57E-04	7.1 ↑ in PA
5b	P14136	Proteina Gliale fibrillare acida	49.9	5.4	146.6	28.47%	10	10	PA	1.79E-04	9.4 ↑ in PA
5c	P14136	Proteina Gliale fibrillare acida	49.9	5.4	146.6	28.47%	10	10	PA	9.49E-04	8.4 ↑ in PA
5d	P14136	Proteina Gliale fibrillare acida	49.9	5.4	146.6	28.47%	10	10	PA	7.00E-03	3.0 ↑ in PA
6	P60709	Actina 1 citoplasmatica	41.7	5.3	185.9	44.00%	6	11	MB	NS	1.4 ↑ in PA
29	O15540	Proteina che lega gli acidi grassi, isoforma cerebrale	14.8	5.4	87.3	72.73%	7	7	PA	2.80E-02	2.0 ↑ in PA
30	P02787	Serotransferrina	75.1	6.7	72.1	24.93%	13	13	PA	8.02E-04	5.1 ↑ in PA
41	P02511	Catena B dell'Alfa-cristallina	20.2	6.8	12.1	17.71%	3	3	PA	4.43E-04	3.2 ↑ in PA
42	P02792	Catena leggera della Ferritina	19.9	5.5	29.9	28.57%	3	3	PA	1.19E-04	3.4 ↑ in PA
43	P30626	Sorcina	21.7	5.3	30.2	43.43%	6	6	PA	4.60E-04	2.3 ↑ in PA
46	P02647	Apolipoproteina A-I	28.1	5.3	73.2	31.84%	7	7	PA	5.00E-03	2.8 ↑ in PA
50	P08758	Annexina A5	35.8	4.9	103.1	45.63%	10	10	PA	4.30E-04	2.4 ↑ in PA
51	P14136	Proteina Gliale fibrillare acida	49.9	5.4	103.4	22.45%	8	8	PA	5.00E-03	4.3 ↑ in PA
54	P14136	Proteina Gliale fibrillare acida	49.9	5.4	246.6	41.67%	17	17	PA	2.00E-03	5.7 ↑ in PA
55	P10809	Heat shock mitocondriale (60 kDa)	57.9	5.2	350.0	49.91%	25	25	MB	NS	1.5 ↑ in PA
57	P11021	Proteina regolata dal glucosio (78 kDa)	70.4	5.0	450.2	42.05%	23	25	MB, PA	NS	1.7 ↑ in PA

Tabella 4c: Elenco delle proteine della frazione insolubile in acido che NON sono differenzialmente espresse nei due pool tumorali

Spot #	Codice Uniprot	Proteina	PM [kDa] ^(c)	pI ^(c)	Punteggio ^(b)	Copertura ^(b)	# Peptidi unici ^(b)	# Peptidi ^(b)	Identificata in	ANOVA (p) ^(d)	Variazione ^{*(d)}
1	P60709	Actina 1 citoplasmatica	41.7	5.29	491.7	55.47%	7	20	MB, PA	NS	NS
2	P68871	Beta globina	15.9	6.8	127.5	54.42%	7	7	MB, PA	NS	NS
3	P68871	Beta globina	15.9	6.8	293.9	62.59%	8	9	MB, PA	NS	NS
26	P68871	Beta globina	15.9	6.8	50.9	34.69%	4	4	MB, PA	NS	NS
80	P60709	Actina 1 citoplasmatica	41.7	5.3	250.1	54.13%	16	16	MB	NS	NS

NS= Proteine senza differenze statisticamente significative in MB e PA

*Solo gli spot che differiscono di almeno 2,0 volte nei volumi (log-normalizzati) e con un valore p di ANOVA ≤ 0.05 sono stati considerati statisticamente significativi.

(a) = i numeri corrispondono a quelli riportati nelle Figure 13 e 14

(b) = dati ottenuti dal software Proteome Discoverer 1.4.0

(c) = dati ottenuti da Uniprot

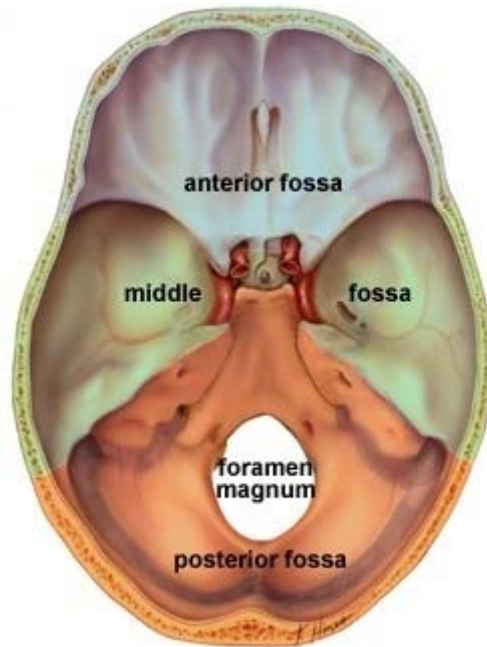


Figura 10: Particolare anatomico della base cranica umana
(tratto da <http://www.mayfieldclinic.com>)

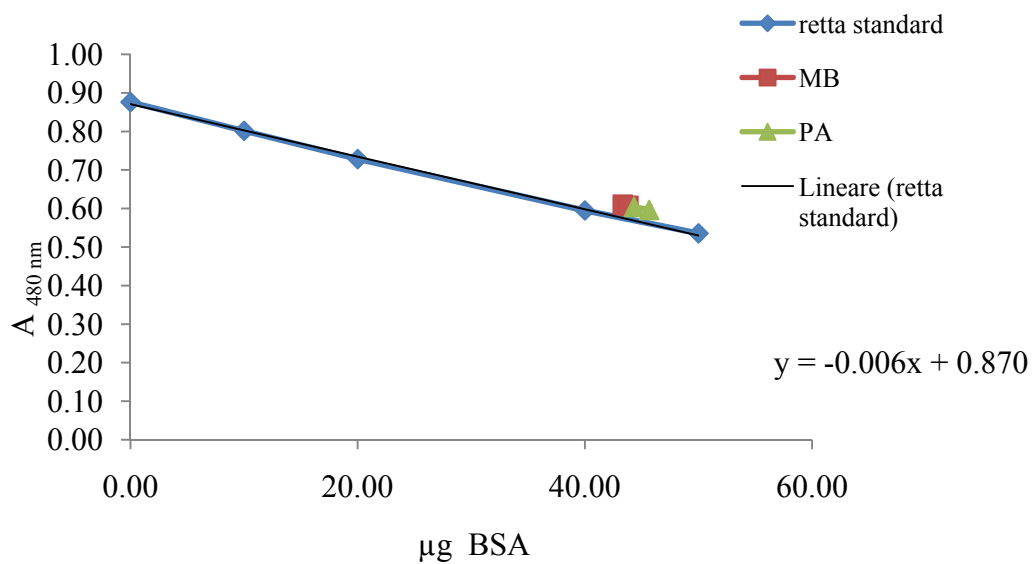


Figura 11: Determinazione fotometrica del quantitativo proteico dei campioni di MB e PA mediante il 2D-Quant Protein Kit

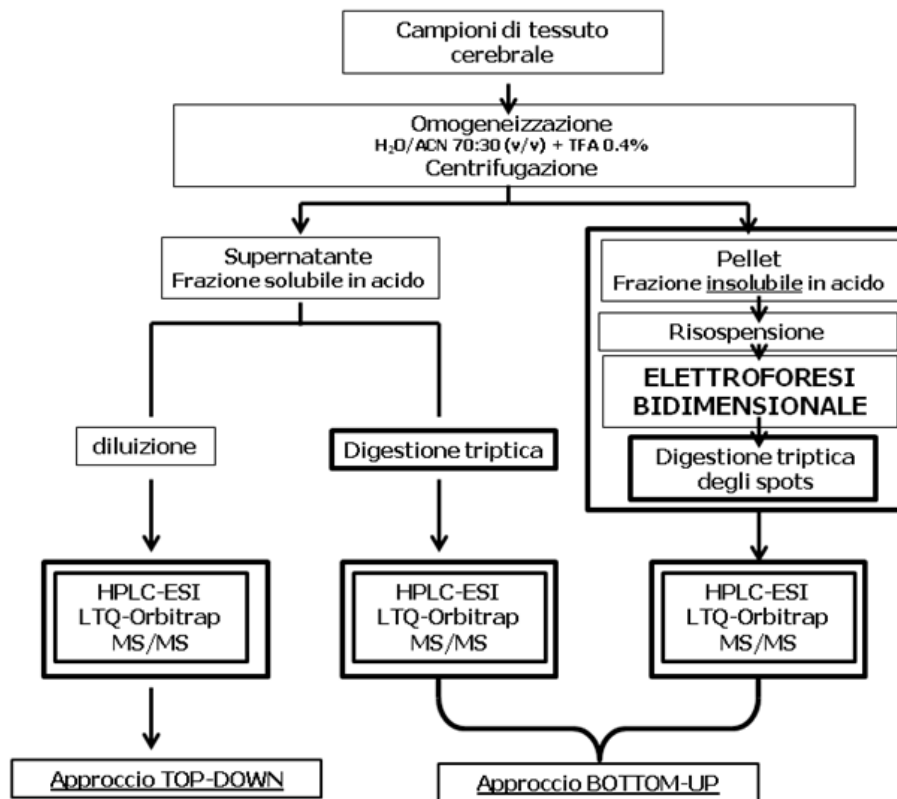


Figura 12: Schema seguito per l'analisi proteomica dei campioni di tessuto tumorale cerebrale

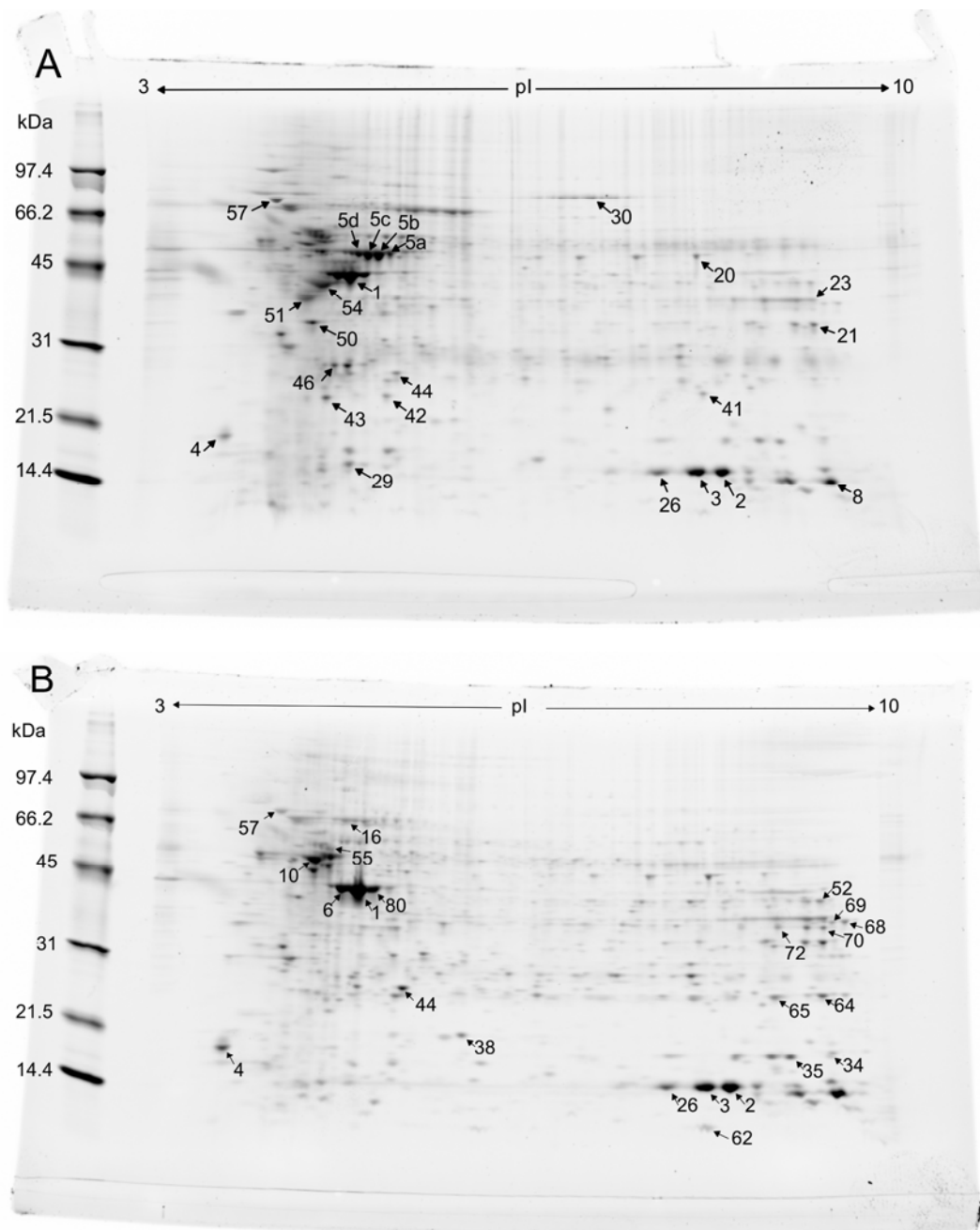


Figura 13: Immagini rappresentative dei gel 2-DE della frazione insolubile in acido dei campioni di astrocitoma pilocitico (PA, pannello A) e medulloblastoma (MB, pannello B). La prima dimensione è stata eseguita su una IPG strip da 11 cm (pH 3-10 NL) dopo reidratazione passiva con 25 μ g di proteine totali. La seconda dimensione è stata eseguita su un gel Criterion Tris-HCl T=10-20%. La colorazione è stata effettuata con il colorante in fluorescenza SYPRO Ruby. Le proteine sono state identificate dopo digestione triptica in-gel mediante analisi HPLC-ESI-MS/MS utilizzando l'LTQ Orbitrap ad alta risoluzione. I risultati delle identificazioni sono riportati in Tabella 4.

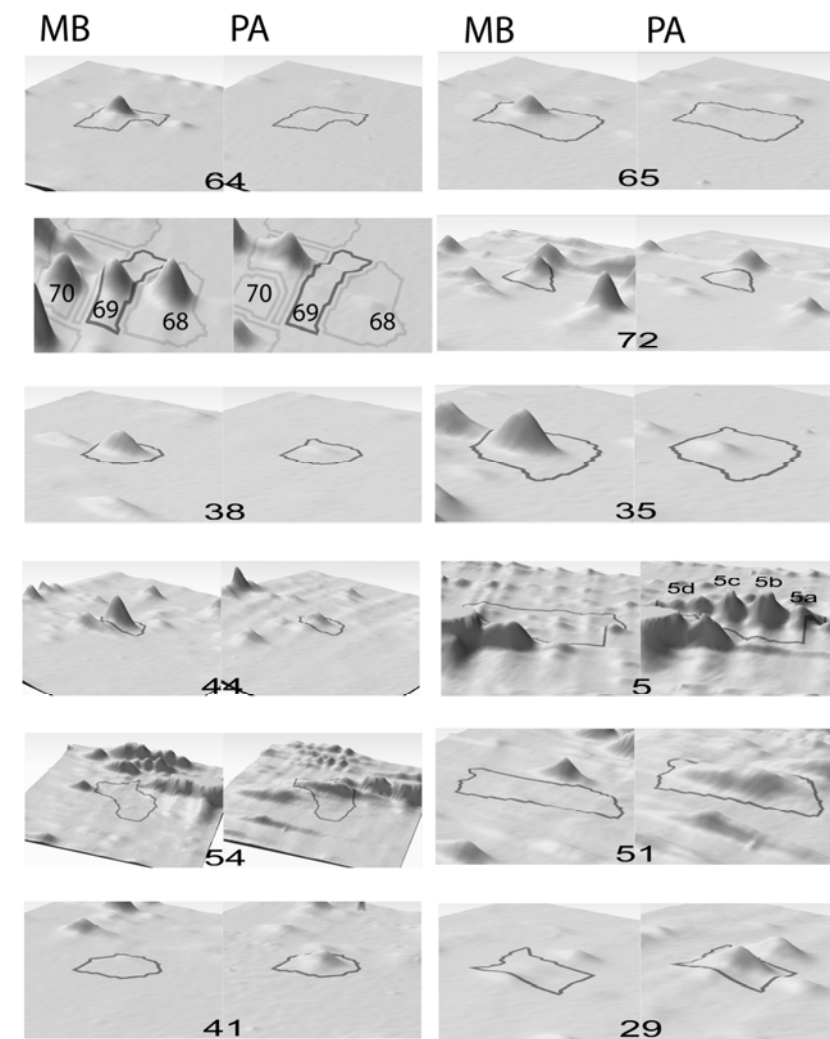


Figura 14: Visione 3D di spot selezionati relativi a proteine sovra-esprese nel campione PA o MB. Per ciascuno spot sono stati confrontate le aree corrispondenti dei gel 2-DE di PA e MB mostrati in Figura 13: il picco MB è riportato a sinistra, il picco PA a destra.

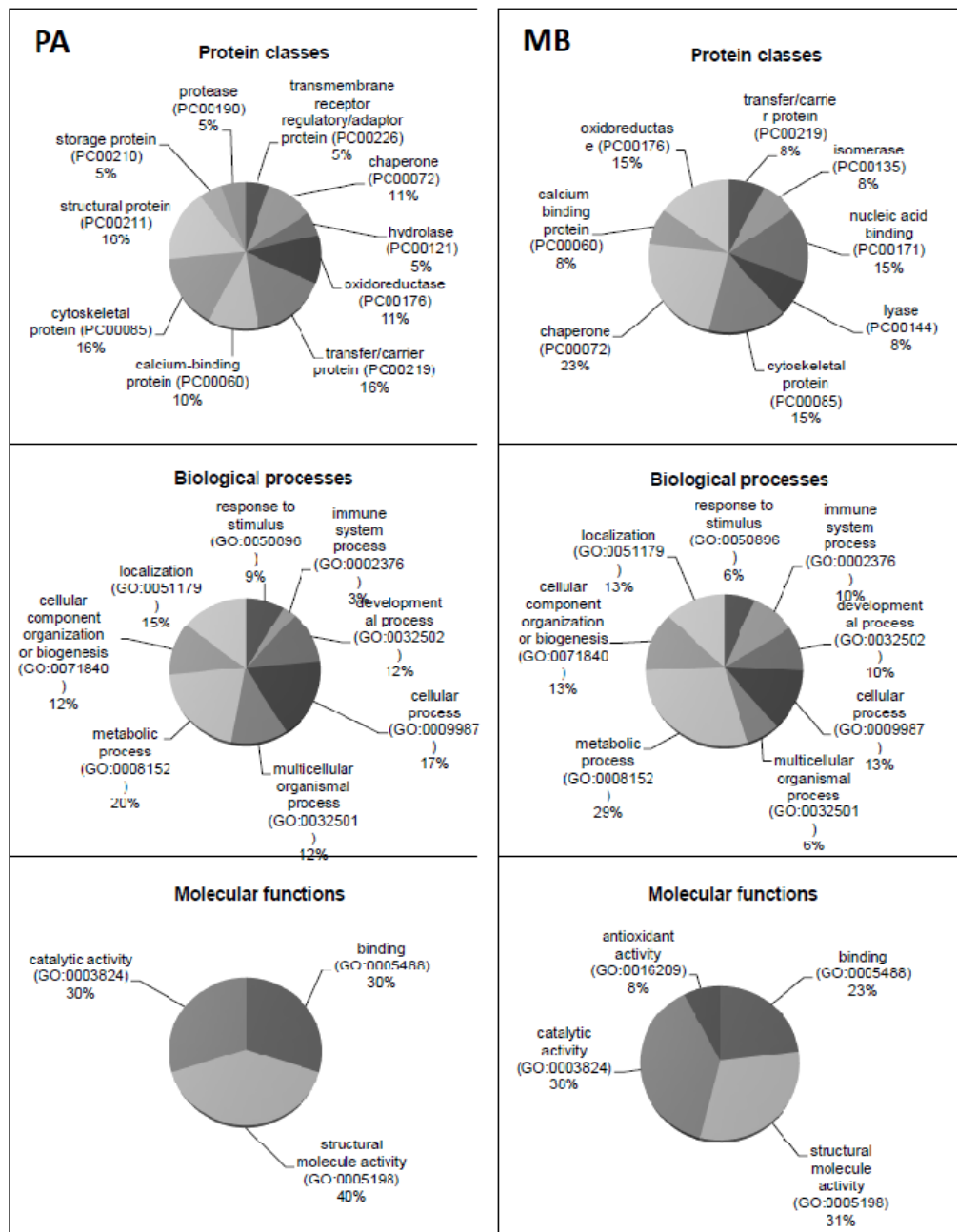


Figura 15: Classificazione Gene Ontology delle proteine identificate nella frazione insolubile in acido di PA ed MB mediante analisi HPLC-ESI-MS/MS ad alta risoluzione su LTQ Orbitrap, dopo analisi 2-DE. La classificazione è stata effettuata secondo le classi di proteine, i processi biologici e le funzioni molecolari, utilizzando il sistema di classificazione PANTHER (Protein ANalysis THrough Evolutionary Relationships, versione 9.0, <http://www.pantherdb.org>).

Bibliografia

Amado F, Lobo MJ, Domingues P, Duarte JA, Vitorino R. 2010. Salivary peptidomics. *Expert Rev Proteomics* 7(5): 709–21.

Anagnostopoulos AK, Papathanassiou C, Karamolegou K, *et al.* 2015. Proteomic studies of pediatric medulloblastoma tumors with 17p deletion. *J Proteome Res.* 14(2):1076-88

Anagnostopoulos AK, Tsangaris GT. 2014. The proteomics of pediatric brain tumors. *Expert Rev proteomics* 11(5): 641–48.

Anagnostopoulos AK, Dimas KS, Papathanassiou C, *et al.* 2011. Proteomics studies of childhood pilocytic astrocytoma. *J Proteome Res* 10(5): 2555–65.

Aps JK, Martens LC. 2005. Review: the physiology of saliva and transfer of drugs into saliva. *Forensic Sci Int.* 150(2-3): 119–31.

Azevedo RS, de Almeida OP, Kowalski LP, Pires FR. 2008. Comparative cytokeratin expression in the different cell types of salivary gland mucoepidermoid carcinoma. *Head Neck Pathol* 2(4): 257–64.

Bandhakavi S, Stone MD, Onsongo G, Van Riper SK, Griffin TJ. 2009. A dynamic range compression and three-dimensional peptide fractionation analysis platform expands proteome coverage and the diagnostic potential of whole saliva. *J Proteome Res* 8(12): 612–24.

Bartlett F, Kortmann R, Saran F. 2013. Medulloblastoma statement of search strategies used and sources of information. *Clin Oncol* 25: 36–45.

Ben-Aryeh H, Lapid S, Szargel R, Benderly A, Gutman D. 1984. Composition of whole unstimulated saliva of human infants. *Arch Oral Biol.* 29(5): 357–62.

Bingle L, Bingle CD. 2011. Distribution of Human PLUNC/BPI Fold-Containing (BPIF) Proteins. *Biochem Soc Trans.* 39(4): 1023–27.

Bobek LA, Levine MJ. 1992. Cystatins-inhibitors of cysteine proteinases. *Crit Rev Oral Biol Med* 3(4): 307–32.

Brodersen DE, Etzerodt M, Madsen P, *et al.* 1998. EF-Hands at atomic resolution: the structure of human psoriasin (S100A7) solved by mad-phasing. *Structure* 6(4): 477–89.

- Brown HG, Kepner JL, Perlman EJ, *et al.* 2000. Large cell/anaplastic medulloblastomas: a pediatric oncology group study. *J Neuropathol Exp Neurol.* 59(10): 857–65.
- Burgess RR. 2009. Protein precipitation techniques. In *Methods Enzymol.* 463:331-42
- Cabras T, Pisano E, Boi R, *et al.* 2009. Age-dependent modifications of the human salivary secretory protein complex. *J. Proteome Res* 8(8): 4126–34.
- Cabras T, Iavarone F, Manconi B, *et al.* 2014. Top-down analytical platforms for the characterization of the human salivary proteome. *Bioanalysis* 6(4): 563–81.
- Castagnola M, Picciotti PM, Messana I, *et al.* 2011a. Potential applications of human saliva as diagnostic fluid. *Acta Otorhinolaryngol Ital* 31 (6): 347–57.
- Castagnola M, Inzitari R, Fanali C, *et al.* 2011b. The surprising composition of the salivary proteome of preterm human newborn. *Mol Cell Proteomics* 10(1): M110.003467.
- Castagnola M, Cabras T, Iavarone F, *et al.* 2012. The human salivary proteome: a critical overview of the results obtained by different proteomic platforms. *Expert Rev of Proteomics* 9(1): 33–46.
- Castle D, Castle A. 1998. Intracellular transport and secretion of salivary proteins. *Crit Rev Oral Biol Med*:9(1): 4–22.
- Chevalier F. 2010. Highlights on the capacities of ‘gel-based’ proteomics. *Proteome Sci* 8: 1-10.
- Choux M, Di Rocco C, Hockley AD, Walker ML. 1999. Pediatric neurosurgery. Churchill Livingstone ed.
- Chumbalkar VC, Subhashini C, Dhople VM, *et al.* 2005. Differential protein expression in human gliomas and molecular insights. *Proteomics* 5(4): 1167–77.
- Cui W, Rohrs HW, Gross ML. 2011. Top-down mass spectrometry: recent developments, applications and perspectives. *Analyst.* 136(19): 3854–64.
- Dawes C. 1975. Circadian rhythms in the flow rate and composition of unstimulated and stimulated human submandibular saliva. *J. Physiol* 244(2): 535–48.
- Dickinson DP. 2002. Salivary (SD-Type) Cystatins: over one billion years in the making-but to what purpose? *Crit Rev Oral Biol Med* 13(6): 485–508.

- Ferguson DB, Botchway CA. 1980. A comparison of circadian variation in the flow rate and composition of stimulated human parotid, submandibular and whole salivas from the same individuals. *Arch Oral Biol* 25(8-9): 559–68.
- Filipek A, Michowski W, Kuznicki J. 2008. Involvement of S100A6 (calcyclin) and its binding partners in intracellular signaling pathways. *Adv Enzyme Regul* 48: 225–39.
- Fleming AJ, Chi SN. 2012. Brain tumors in children. *Curr Probl Pediatr Adolesc Health Care* 42(4): 80–103.
- Giangaspero F, Rigobello L, Badiali M, *et al.* 1992. Large-cell medulloblastomas. A distinct variant with highly aggressive behavior. *Am J Sur Pathol* 16(7): 687–93.
- Giangaspero F, Wellek S, Masuoka J, *et al.* 2006. Stratification of medulloblastoma on the basis of histopathological grading. *Acta Neuropathol* 112(1): 5–12.
- Gläser R, Harder J, Lange H, *et al.* 2005. Antimicrobial psoriasin (S100A7) protects human skin from *Escherichia coli* infection. *Nat Immunol* 6(1): 57–64.
- Godbout RD, Bisgrove DA, Shkolny D, Day RS 3rd. 1998. Correlation of B-FABP and GFAP expression in malignant glioma. *Oncogene* 16(15): 1955–62.
- Goebel C, Mackay LG, Vickers ER, Mather LE. 2000. Determination of Defensin HNP-1, HNP-2, and HNP-3 in human saliva by using LC/MS. *Peptides* 21(6): 757–65.
- Gomes LH, Raftery MJ, Yan WX, *et al.* 2013. S100A8 and S100A9-oxidant scavengers in inflammation. *Free Radic Biol Med* 58: 170–86.
- Görg A, Obermaier C, Boguth G, *et al.* 2000. The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* 21(6): 1037–53.
- Görg A, Weiss W, Dunn MJ. 2004. Current two-dimensional electrophoresis technology for proteomics. *Proteomics* 4(12): 3665–85.
- Goto Y, Sueoka E, Chiba H, Fujiki H. 1999. Significance of heterogeneous nuclear ribonucleoprotein b1 as a new early detection marker for oral squamous cell carcinoma. *Jpn J Cancer Res.* 90(12): 1358–63.
- Greenough C, Jenkins RE, Kitteringham NR, *et al.* 2004. A method for the rapid depletion of albumin and immunoglobulin from human plasma. *Proteomics* 4(10): 3107–11.

- Hagens G, Masouyé I, Augsburger E, *et al.* 1999a. Calcium-binding protein S100a7 and epidermal-type fatty acid-binding protein are associated in the cytosol of human keratinocytes. *Biochem J* 339 (2): 419–27.
- Hagens G, Roulin K, Hotz R, *et al.* 1999b. Probable interaction between S100A7 and E-FABP in the cytosol of human keratinocytes from psoriatic scales. *Mol Cell Biochem* 192(1-2): 123–28.
- Hassan MI, Waheed A, Yadav S, Singh TP, Ahmad F. 2009. Prolactin inducible protein in cancer, fertility and immunoregulation: structure, function and its clinical implications. *Cell Mol Life Sciences*: 66(3): 447–59.
- Heizmann, CW, Cox JA. 1998. New perspectives on S100 proteins: a multi-functional Ca(2+)-, Zn(2+)- and Cu(2+)-binding protein family. *Biometals* 11(4): 383–97.
- Hirtz C, Chevalier F, Centeno D, *et al.* 2005. MS characterization of multiple forms of alpha-amylase in human saliva. *Proteomics* 5(17): 4597–607.
- Holbrook IB, Molan PC, 1975. The identification of a peptide in human parotid saliva particularly active in enhancing the glycolytic activity of the salivary micro-organisms. *Biochem J* 149(2): 489–92.
- Hu S, Xie Y, Ramachandran P, *et al.* 2005. Large-scale identification of proteins in human salivary proteome by liquid chromatography/mass spectrometry and two-dimensional gel electrophoresis-mass spectrometry. *Proteomics* 5(6): 1714–28.
- Hu S, Loo JA, Wong DT. 2006. Human body fluid proteome analysis. *Proteomics* 6 (23): 6326–53.
- Huff T, Müller CS, Otto AM, Netzker R, Hannappel E. 2001. Beta-thymosins, small acidic peptides with multiple functions. *Int J Biochem Cell Biol.* 33(3): 205–20.
- Humphrey SP, Williamson RT. 2001. A review of saliva: normal composition, flow, and function. *J Prosthet Dent.* 85(2): 162–69.
- Huse JT, Holland EC. 2010. Targeting brain cancer: advances in the molecular pathology of malignant glioma and medulloblastoma. *Nat Rev Cancer* 10(5): 319–31.
- Ichimura K, Nishikawa R, Matsutani M. 2012. Molecular markers in pediatric. *Neuro Oncol* 14(S4): 90–99.

- Inzitari R, Vento G, Capoluongo E, *et al.* 2007. Proteomic analysis of salivary acidic proline-rich proteins in human preterm and at-term newborns. *J Proteome Res* 6(4): 1371–77.
- Isemura S. 2000. Nucleotide sequence of gene PBII encoding salivary proline-rich protein P-B. *J Biochem.* 127(3): 393–98.
- Isobe T, Okuyama T, 1978. The amino-acid sequence of S-100 protein (PAP I-b protein) and its relation to the calcium-binding proteins. *Eur J Biochem.* 89(2): 379–88.
- Iwadate Y, Sakaida T, Hiwasa T, *et al.* 2004. Molecular classification and survival prediction in human gliomas based on proteome analysis molecular classification and survival prediction in human gliomas based on proteome analysis. *Cancer Res* 64 (7): 2496–2501.
- Jessie K, Pang WW, Haji Z, Rahim A, Hashim OH. 2010. proteomic analysis of whole human saliva detects enhanced expression of interleukin-1 receptor antagonist, thioredoxin and lipocalin-1 in cigarette smokers compared to non-smokers. *Int J Mol Sci* 11(11): 4488–505.
- Kaatsch P. 2010. Epidemiology of childhood cancer. *Cancer Treat Rev.* 36(4): 277–85.
- Karihtala P, Mäntyniemi A, Kang SW, Kinnula VL, Soini Y. 2003. Peroxiredoxins in breast carcinoma. *Clin Cancer Res.* 9(9): 3418–24.
- Khalil A. 2007. Biomarker Discovery: A proteomic approach for brain cancer profiling. *Cancer Sci.* 98(2): 201–13.
- Kinnula VL, Lehtonen S, Sormunen R, *et al.* 2002. Overexpression of peroxiredoxins I, II, III, V, and VI in malignant mesothelioma. *J Pathol.* 196(3): 316–23.
- Kleihues P, Burger PC, Scheithauer BW. 1993. The new WHO classification of brain tumours. *Brain Pathol* 3(3): 255–68.
- Kool M, Koster J, Bunt J, *et al.* 2008. integrated genomics identifies five medulloblastoma subtypes with distinct genetic profiles, pathway signatures and clinicopathological features. *PLoS One* 3(8): e3088.
- Korolainen MA, Auriola S, Nyman TA, Alafuzoff I, Pirttilä T. 2005. Proteomic analysis of glial fibrillary acidic protein in alzheimer's disease and aging brain. *Neurobiol Dis.* 20(3): 858–70.

- Kuo MF, Wang HS, Kuo QT, *et al.* 2009. High expression of stathmin protein predicts a fulminant course in medulloblastoma. *J Neurosurg. Pediatr.* 4(1): 74–80.
- Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227(5259): 680-85.
- Lehtonen ST, Svensk AM, Soini Y, *et al.* 2004. Peroxiredoxins, a novel protein family in lung cancer. *Int J Cancer* 111(4): 514–21.
- Lim SY, Raftery M, Cai H, *et al.* 2008. S-Nitrosylated S100A8: novel anti-inflammatory properties. *J Immunol.* 181(8): 5627–36.
- Louis DN, Ohgaki H, Wiestler OD, *et al.* 2007. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol* 114(2): 97–109.
- Manconi B, Cabras T, Pisano E, *et al.* 2013. Modifications of the acidic soluble salivary proteome in human children from birth to the age of 48 months investigated by a top-down HPLC-ESI-MS platform. *J Proteomics* 91: 536–43.
- Mandel ID. 1987. The functions of saliva. *J Dent Res.* 66: 623–27.
- Mayeda A, Munroe SH, Cáceres JF, Krainer AR. 1994. Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. *EMBO J* 13(22): 5483–95.
- McLafferty FW, Breuker K, Jin M, Han X, *et al.* 2007. Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics. *FEBS J.* 274(24): 6256–68.
- McManamy CS, Lamont JM, Taylor RE, *et al.* 2003. Morphophenotypic variation predicts clinical behavior in childhood non-desmoplastic medulloblastomas. *J Neuropathol Exp Neurol* 62(6): 627–32.
- Messana I, Inzitari R, Fanali C, Cabras T, Castagnola M. 2008. Facts and artifacts in proteomics of body fluids. What proteomics of saliva is telling us? *J Sep Sci* 31(11): 1948–63.
- Messana I, Cabras T, Iavarone F, *et al.* 2013. Unraveling the different proteomic platforms. *J Sep Sci* 36(1): 128–39.
- Messana I, Cabras T, Iavarone F, *et al.* 2015. Chrono-proteomics of human saliva: variations of the salivary proteome during human development. *J Proteome Res.* 14(4):1666-77

- Middeldorp J, Hol EM. 2011. GFAP in health and disease. *Progr Neurobiol* 93(3): 421–43.
- Mita R, Coles JE, Glubrecht DD, *et al.* 2007. B-FABP-expressing radial glial cells: the malignant glioma cell of origin? *Neoplasia* 9(9): 734–44.
- Moore BW. 1965. A soluble protein characteristic of the nervous system. *Biochem Biophys Res Commun.*19(6): 739–44.
- Morzell M, Palicki O, Chabanet C, *et al.* 2011. Saliva electrophoretic protein profiles in infants: changes with age and impact of teeth eruption and diet transition. *Arch Oral Biol* 56(7): 634–42.
- Narayan RK, Heydorn WE, Creed GJ, Jacobowitz DM. 1986. Protein patterns in various malignant human brain tumors by two- dimensional gel electrophoresis. *Cancer Res.* 46(9): 4685–94.
- Nesvizhskii AI, Aebersold R. 2005. Interpretation of shotgun proteomic data: the protein inference problem. *Mol Cell Proteomics* : 4(10): 1419–40.
- Nozaki M, Tada M, Kobayashi H, *et al.* 1999. Roles of the functional loss of p53 and other genes in astrocytoma tumorigenesis and progression. *Neuro Oncol.*1(2): 124–37.
- O’Farrell PH. 1975. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 250(10): 4007–21.
- Odreman F, Vindigni M, Gonzales ML, *et al.* 2005. Proteomic studies on low- and high-grade human brain astrocytomas. *J Proteome Res* 4(3): 698–708.
- Packer RJ. 1990. Chemotherapy for medulloblastoma/primitive neuroectodermal tumors of the posterior fossa. *Ann Neurol* 28(6): 823–28.
- Pandey A, Mann M. 2000. Proteomics to study genes and genomes. *Nature* 405(6788): 837–46.
- Patel DR, Isas JM, Ladokhin AS, *et al.* 2005. The conserved core domains of annexins a1, a2, a5, and b12 can be divided into two groups with different Ca²⁺-dependent membrane-binding properties. *Biochemistry* 44(8): 2833–44.

Pavón EJ, García-Rodríguez S, Zumaquero E, *et al.* 2012. Increased expression and phosphorylation of the two S100a9 isoforms in mononuclear cells from patients with systemic lupus erythematosus: a proteomic signature for circulating low-density granulocytes. *J Proteomics* 75(6): 1778–91.

Peterson TA, Adadey A, Santana-Cruz I, *et al.* 2010. DMDM: Domain Mapping of Disease Mutations. *Bioinformatics* 26(19): 2458–59.

Peyrl A, Krapfenbauer K, Slavic I, *et al.* 2003. Protein Profiles of medulloblastoma cell lines DAOY and D283: identification of tumor-related proteins and principles. *Proteomics* 3(9): 1781–1800.

Pinder SE, Balsitis M, Ellis IO, *et al.* 1994. The expression of alpha b-crystallin in epithelial tumours: a useful tumour marker? *J Pathol* 174(3): 209–15.

Prabhu VV, Siddikuzzaman, Grace VM, Guruvayoorappan C. 2012. Targeting tumor metastasis by regulating Nm23 gene expression. *Asian Pac J Cancer Prev.* 13(8): 3539–48.

Réty S, Sopkova J, Renouard M, *et al.* 1999. The crystal structure of a complex of p11 with the annexin II N-terminal peptide. *Nat Struct Biol.* 6(1): 89–95.

Rudney JD, Ji Z, Larson CJ, Liljemark WF, Hickey KL. 1995. Saliva protein binding to layers of oral streptococci in vitro and in vivo. *J Dent Res.* 74(6): 1280–88.

Ruiz Esparza-Garrido R, Velázquez-Flores MÁ, Diegopérez-Ramírez J, *et al.* 2013. A proteomic approach of pediatric astrocytomas: miRNAs and network insight. *J Proteomics* 94: 162–75.

Ruse M, Broome AM, Eckert RL. 2003. S100A7 (psoriasin) interacts with epidermal fatty acid binding protein and localizes in focal adhesion-like structures in cultured keratinocytes. *J Invest Dermatol.* 121(1): 132–41.

Rutkowski S, Gerber NU, von Hoff K, *et al.* 2009. Treatment of early childhood medulloblastoma by postoperative chemotherapy and deferred radiotherapy. *Neuro Oncol* 11(2): 201–10.

Ryan CM, Souda P, Halgand F, *et al.* 2010. Confident assignment of intact mass tags to human salivary cystatins using top-down fourier-transform ion cyclotron resonance mass spectrometry. *J Am Soc Mass Spectrom,* 21(6): 908–17.

Sadighi Z, Slopis J. 2013. Pilocytic Astrocytoma: A disease with evolving molecular heterogeneity. *J Child Neurol* 28(5): 625–32.

Schlesinger DH, Hay DI. 1977. Complete covalent structure of statherin, a tyrosine-rich acidic peptide which inhibits calcium phosphate precipitation from human parotid saliva. *J Biol Chem.* 252(5): 1689–95.

Schroeder K, Gururangan S. 2014. Molecular variants and mutations in medulloblastoma. *Pharmacogenomics Pers Med.* 7: 43–51.

Smith LM, Kelleher NL, Consortium for Top Down Proteomics. 2013. Proteoform: a single term describing protein complexity. *Nat Methods* 10(3):186-7

Sreebny LM. 2000. Saliva in health and disease: an appraisal and update. *Int Dent J.* 50(3): 140–61.

Sueoka E, Sueoka N, Goto Y, *et al.* 2001. Heterogeneous nuclear ribonucleoprotein b1 as early cancer biomarker for occult cancer of human lungs and bronchial dysplasia. *Cancer Res* 61(5)1896–1902.

Sure U, Berghorn WJ, Bertalanffy H, *et al.* 1995. Staging, scoring and grading of medulloblastoma - a postoperative prognosis predicting system based on the cases of a single institute. *Acta Neurochir.* 132(1-3): 59–65.

Takashi M, Katsuno S, Sakata T, Ohshima S, Kato K. 1998. Different concentrations of two small stress proteins, *ab* crystallin and hsp27 in human urological tumor tissues. *Urol Res* 26(6): 395–99.

Taylor MD, Northcott PA, Korshunov A, *et al.* 2012. Molecular subgroups of medulloblastoma: the current consensus. *Acta Neuropathol* 123(4): 465–72.

Taylor RE, Bailey CC, Robinson K., *et al.* 2003. Results of a randomized study of preradiation chemotherapy versus radiotherapy alone for non metastatic medulloblastoma: the international society of pediatric oncology/United Kingdom children's cancer study group PNET-3 study. *J Clin Oncol* 21(8): 1581–91.

Thompson MC1, Fuller C, Hogg TL, *et al.* 2006. Genomics identifies medulloblastoma subgroups that are enriched for specific genetic alterations. *J Clin Oncol* 24(12): 1924–31.

Tipton JD, Tran JC, Catherman AD, *et al.* 2011. Analysis of intact protein isoforms by mass spectrometry. *J Biol Chem* 286(29): 25451–58.

- Tiu SC, Chan WY, Heizmann CW, *et al.* 2000. Differential expression of S100B and S100A6(1) in the human fetal and aged cerebral cortex. *Brain Res Dev Brain Res.* 119(2): 159–68.
- Troxler RF, Offner GD, Xu T, Vanderspek JC, Oppenheim FG. 1990. Structural relationship between human salivary histatins. *J Dent Res* 69(1): 2–6.
- Unlü M, Morgan ME Minden, JS. 1997. Difference Gel Electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* 18(11): 2071–77.
- Vargas PA, Speight PM, Bingle CD, Barrett AW, Bingle L. 2008. Expression of PLUNC family members in benign and malignant salivary gland tumours. *Oral Dis* 14(7): 613–19.
- Vitorino R, Lobo MJ, Ferrer-Correira AJ, *et al.* 2004. Identification of human whole saliva protein components using proteomics. *Proteomics* 4(4): 1109–15.
- Vong L, D'Acquisto F, Pederzoli-Ribeil M, *et al.* 2007. Annexin 1 cleavage in activated neutrophils: a pivotal role for proteinase 3. *J Biol Chem.* 282(41): 29998–4.
- Weist S, Eravci M, Broedel O, *et al.* 2008. Results and reliability of protein quantification for two-dimensional gel electrophoresis strongly depend on the type of protein sample and the method employed. *Proteomics* 8(16): 3389–96.
- Wilkins, MR, Pasquali C, Appel, RD, *et al.* 1996. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Bio/technology* 14(1): 61–65.
- Wilkins MR, Sanchez JC, Gooley AA, *et al.* 1996. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev* 13: 19–50.
- Yamada M, Murakami K, Wallingford JC, Yuki Y. 2002. Identification of low-abundance proteins of bovine colostrum and mature milk using two-dimensional electrophoresis followed by microsequencing and mass spectrometry. *Electrophoresis* 23(7-8): 1153–60.
- Yanagawa T, Ishikawa T, Ishii T, *et al.* 1999. Peroxiredoxin I expression in human thyroid tumors. *Cancer Lett* 145(1-2): 127–32.

- Yan-Sanders Y, Hammons GJ, Lyn-Cook BD. 2002. Increased expression of heterogeneous nuclear ribonucleoprotein a2/b1 (hnRNP) in pancreatic tissue from smokers and pancreatic tumor cells. *Cancer Lett* 183(2): 215–20.
- Yao Y, Berg EA, Costello CE, Troxler RF, Oppenheim FG. 2003. Identification of protein components in human acquired enamel pellicle and whole saliva using novel proteomics approaches. *J Biol Chem* 278(7): 5300–5308.
- Zanini C, Mandili G, Pulerà F, *et al.* 2009. Immunohistochemical and proteomic profile of melanotic medulloblastoma. *Pediatr Blood Cancer* 52(7): 875–77.
- Zanini C, Mandili G, Baci D, *et al.* 2010. Proteomic profile modification of anaplastic medulloblastoma after in-vivo radiotherapy: a case study. *J Cancer Ther* 01(02): 97–103.
- Zanini C, Mandili G, Bertin D, *et al.* 2011. Analysis of different medulloblastoma histotypes by two-dimensional gel and MALDI-TOF. *Childs Nerv Syst* 27(12): 2077–85.
- Zhou J, Allred DC, Avis I, *et al.* 2001. Differential expression of the early lung cancer detection marker, heterogeneous nuclear ribonucleoprotein-A2/B1 (hnRNP-A2/B1) in normal breast and neoplastic breast cancer. *Breast Cancer Res Treat* 66(3): 217–24.
- Zimmer DB, Wright Sadosky P, Weber DJ. 2003. Molecular mechanisms of S100-target protein interactions. *Microsc Res Tech* 60(6): 552–59.

Articoli scientifici pubblicati

- Chrono-proteomics of human saliva: variations of the salivary proteome during human development.

Messana I, Cabras T, Iavarone F, Manconi B, Huang L, Martelli C, Olinas A, Sanna MT, Pisano E, Sanna M, **Arba M**, D'Alessandro A, Desiderio C, Vitali A, Pirolli D, Tirone C, Lio A, Vento G, Romagnoli C, Cordaro M, Manni A, Gallenzi P, Fiorita A, Scarano E, Calò L, Passali GC, Picciotti PM, Paludetti G, Fanos V, Faa G, Castagnola M.

J Proteome Res. (2015) 14(4):1666-77.

- Integrated proteomic platforms for the comparative characterization of medulloblastoma and pilocytic astrocytoma pediatric brain tumors: a preliminary study.

Claudia Martelli, Federica Iavarone, Luca D'Angelo, **Morena Arba**, Federica Vincenzoni, Ilaria Inserra, Daniela Delfino, Diana Valeria Rossetti, Marta Caretto, Luca Massimi, Gianpiero Tamburrini, Concezio Di Rocco, Massimo Caldarelli, Irene Messana, Massimo Castagnola, Maria Teresa Sanna, Claudia Desiderio.

Mol. Biosyst. (2015) Apr 24 [Epub ahead of print]

Articolo in preparazione

- Proteomic characterization of acid-insoluble fraction of whole saliva from preterm human newborns

M. Arba, F. Vincenzoni, F. Iavarone, G. Vento, T. Cabras, M. Castagnola, I. Messana, M.T. Sanna

Comunicazioni a congressi

Presentazione poster al 9^a *Italian Proteomics Association (ItPA) Annual Congress*
Napoli 24-27 Giugno 2014:

- Salivary proteome changes in subjects affected by celiac disease and type 1 diabetes

Morena Arba, Elisabetta Pisano, Carlo Ripoli, Federica Iavarone, Federica Vincenzoni, Massimo Castagnola, Tiziana Cabras, Irene Messana, Maria T. Sanna

- Top-down proteomics for the investigation of saliva during human development

Tiziana Cabras, Federica Iavarone, Barbara Manconi, Elisabetta Pisano, Liling Huang, Alessandra Olianias, Maria Teresa Sanna, Monica Sanna, **Morena Arba**, Alfredo D'Alessandro, Claudia Desiderio, Claudia Martelli, Alberto Vitali, Chiara Tirone, Alessandra Lio, Giovanni Vento, Costantino Romagnoli, Massimo Cordaro, Armando Manni, Antonella Fiorita, Emanuele Scarano, Lea Calò, Giulio Cesare Passali, Pasqualina Picciotti, Gaetano Paludetti, Vassilios Fanos, Gavino Faa, Irene Messana, Massimo Castagnola.

- Proteomic integrated platforms for the comparative characterization of medulloblastoma and pilocytic astrocytoma pediatric brain tumors

Claudia Martelli, Maria Teresa Sanna, **Morena Arba**, Irene Messana, Federica Iavarone, Massimo Castagnola, Luca D'Angelo, Marta Caretto, Gianpiero Tamburrini, Concezio Di Rocco, Massimo Caldarelli, Claudia Desiderio.

Presentazione poster al *10th European Symposium on Saliva* - Egmond aan Zee (the Netherlands) 14-17 Maggio 2014:

- Different activation of kinases responsible for salivary proteins phosphorylation during human development

I. Messana, F. Iavarone, T. Cabras, B. Manconi, E. Pisano, L. Huang, A. Olianias, M. T. Sanna, M. Sanna, **M. Arba**, A. D'Alessandro, C. Desiderio, C. Martelli, A. Vitali, C. Tirone, A. Lio, G. Vento, C. Romagnoli, M. Cordaro, A. Manni, A. Fiorita, E. Scarano, L. Calò, G. C. Passali, P. Picciotti, G. Paludetti, V. Fanos, G. Faa, M. Castagnola.

Presentazione poster al *8th Italian Proteomics Association (ItPA) Annual Congress* Padova 18- 21 Giugno 2013:

- Proteomic analysis of the insoluble fraction of preterm human newborn saliva

Morena Arba , Federica Iavarone, Federica Vincenzoni, Liling Huang, Tiziana Cabras, Barbara Manconi, Alessandra Olianias, Massimo Castagnola, Irene Messana and Maria Teresa Sanna

- Proteomic analysis of human salivary pellet

Liling Huang, Federica Iavarone, Federica Vincenzoni, Maria Teresa Sanna, Tiziana Cabras, Barbara Manconi, **Morena Arba**, Irene Messana, Massimo Castagnola.