

UNIVERSITA' DEGLI STUDI DI NAPOLI

FEDERICO II

Dottorato di ricerca in Scienze Chimiche

XXV Ciclo



**Studio delle interazioni proteine-acidi nucleici mediante approcci di proteomica
funzionale**

Tutore

Prof. Pietro Pucci

Cotutore

Dott.ssa Maria Monti

Coordinatore

Prof. Lucio Previtiera

Dottoranda

Carla Iannone

*Non sono i frutti della ricerca scientifica
che elevano un uomo ed arricchiscono la
sua natura, ma la necessità di capire e il
lavoro intellettuale.*

Albert Einstein.

Indice

A. Riassunto	6
1 Introduzione	10
1.1 L'approccio genomico	10
1.2 Il passaggio dall'era genomica all'era proteomica	11
1.3 Gen-“omica”, gen-“omica” funzionale e prote-“omica”	13
1.4 Metodiche e tecniche in proteomica funzionale	15
1.4.1 Proteomica funzionale	16
1.4.2 La proteomica da un punto di vista critico	17
1.4.3 Metodologie per lo studio delle interazioni acidi nucleici-proteine	18
1.4.3.1 Studi in vitro	18
1.4.3.2 Studi in vivo	22
1.5 Le interazioni proteina-proteina nelle malattie umane	24
1.6 Spettrometria di massa e proteomica	25
1.6.1 Le metodologie in spettrometria di massa	26
1.6.2 La strumentazione	27
1.7 Database pubblici per le interazioni proteina-proteina	29
1.8 Scopo del progetto di tesi	30
Bibliografia	32
2 Messa a punto della metodologia EMSA/MS-MS per lo studio degli interattori del promotore del gene cftr	35
2.1 Fibrosi Cistica	35
2.1.1 La regolazione dell'espressione del gene cftr	39
2.2 Materiali e metodi	43
2.2.1 Materiali	43
2.2.2 Metodi	43
2.2.2.1 Colture cellulari	43
2.2.2.2 Lisi cellulare frazionata	44
2.2.2.3 Western Blotting	45
2.2.2.4 Preparazione delle sonde oligonucleotidiche	45
2.2.2.4.1 Quantificazione delle sonde oligonucleotidiche	46
2.2.2.4.2 Annealing delle sonde oligonucleotidiche	46
2.2.2.5 Gel Nativo	46
2.2.2.5.1 Tecniche di visualizzazione delle bande dal gel	47

2.2.2.6	Idrolisi in situ	48
2.2.2.6.1	Rimozione delle bande dal gel	48
2.2.2.6.2	Riduzione ed alchilazione delle proteine in situ	48
2.2.2.6.3	Digestione in situ delle proteine presenti nel gel	48
2.2.2.6.4	Estrazione dei peptidi	48
2.2.2.7	Identificazione	49
2.2.2.7.1	Analisi mediante Matrix Assisted Laser Desorption Ionization (MALDI/MS)	49
2.2.2.7.2	Analisi mediante LC/MS-MS	49
2.2.2.7.3	Identificazione mediante Mascot	50
2.3	Risultati	52
2.3.1	Messa a punto del protocollo EMSA/MS-MS	52
2.3.1.1	Ottimizzazione del rapporto sonda oligonucleotidica/proteine nucleari per l'esperimento EMSA/MS-MS	53
2.3.1.2	Sonda e sua marcatura	55
2.3.1.3	Emsa con spiazzamento	56
2.3.1.4	Emsa preparativi	59
2.4	Discussione	68
	Bibliografia	72
3.	Studio degli interattori del mRNA di ENPP1 mediante approcci di proteomica funzionale	75
3.1	Resistenza insulinica	75
3.2	Materiali e Metodi	80
3.2.1	Materiali	80
3.2.2	Metodi	80
3.2.2.1	Colture cellulari	80
3.2.2.2	Estrazione delle proteine	80
3.2.2.3	Isolamento di complessi proteici mediante tecniche di Affinity	81
3.2.2.4	SDS-PAGE	82
3.2.2.5	SDS-PAGE preparativa	83
3.2.2.6	Colorazione	83
3.3	Risultati	84
3.3.1	Influenza di GALNT2 sull'espressione di ENPP1	92
3.4	Discussione	95

Bibliografia	99
4Messa a punto della metodologia ChIP/MS-MS per lo studio della proteina	
CBX7	101
4.1 CBX7 e la carcinogenesi tiroidea	101
4.1.2 CBX7 e i geni Polycomb	102
4.2 Materiali e Metodi	107
4.2.1 Materiali	107
4.2.2 Metodi	107
4.2.2.1 Colture cellulari	107
4.2.2.2 Crosslink proteine nucleari/cromatina	107
4.2.2.3 Lisi cellulare frazionata	108
4.2.2.4 Isolamento di complessi proteici mediante tecniche di Chromatin Immunoprecipitation	108
4.3 Risultati	112
4.3.1 Messa a punto del protocollo di ChIP/MS-MS	112
4.3.2 Ottimizzazione della strategia e prove di eluizione	113
4.3.3 ChIP/MS-MS Preparativa	115
4.4 Discussione	124
Bibliografia	126

A. Riassunto

Una completa descrizione della complessa rete di meccanismi cellulari e l'uso di network per predire la maggior parte dei comportamenti cellulari sono i principali obiettivi della system biology. Un ruolo chiave nella biologia contemporanea è svolto dalla proteomica funzionale che si propone di fornire delucidazioni sulle funzioni delle proteine e sulla definizione dei meccanismi cellulari a livello molecolare. Il conseguimento di questi obiettivi dipende principalmente dall'identificazione di proteine all'interno di complessi funzionali in vivo. L'isolamento delle proteine che interagiscono tra loro è effettuato mediante procedure di cromatografia di affinità o di immuno-precipitazione nelle quali l'esca proteica ed i suoi specifici *partners* possono essere isolati utilizzando uno specifico ligando immobilizzato su supporti solidi insolubili. Questi approcci conducono all'identificazione di molte proteine che appartengono a complessi distinti dotati di funzioni biologiche diverse. Attribuire l'appartenenza di ogni proteina ad un specifico complesso costituisce un ingente problema che può essere solo parzialmente risolto usando database di interazione proteina-proteina ed informazioni di letteratura.

È possibile che lo sviluppo delle metodologie di pre-frazionamento per separare complessi proteici preservando le loro interazioni native rappresenti una soluzione essenziale per il futuro della proteomica funzionale. La pre-purificazione dei complessi può essere intrapresa solamente in condizioni native sulla base delle loro caratteristiche chimico-fisiche, come la grandezza e dimensione (cromatografia di gel-filtrazione), la densità (ultracentrifugazione), ecc. In seguito al pre-frazionamento, il complesso associato ad una specifica funzione biologica può essere isolato usando tecniche di purificazione per affinità.

Approcci di proteomica funzionale capaci di descrivere singole proteine appartenenti a complessi coinvolti in specifiche funzioni cellulari avrà un forte impatto su futuri studi della system biology.

In particolare il progetto di tesi di dottorato si è articolato nei seguenti punti:

- ✓ Messa a punto di una strategia EMSA/MS_MS.
- ✓ Messa a punto di una strategia ChIP/MS_MS.
- ✓ Messa a punto di una strategia di RNA Chromatography Affinity.

✓ **Messa a punto di una strategia EMSA/MS-MS**

L'EMSA è una tecnica che consente di verificare in vitro il legame di una proteina specifica con un tratto di DNA, generalmente marcato radioattivamente o con fluorofori, basandosi sulla differente mobilità (shift) che il complesso DNA-proteina mostra durante l'elettroforesi su gel nativo rispetto alla sequenza di DNA non legato. Nell'approccio classico, la proteina generalmente viene aggiunta pura in forma ricombinante, ma è possibile anche effettuare questo saggio partendo da un estratto cellulare. In quest'ultimo caso la conferma che l'eventuale ritardo della mobilità elettroforetica sia dovuto alla presenza della proteina di interesse viene realizzata aggiungendo un anticorpo capace di riconoscerla e legarla in maniera specifica. Questa aggiunta provoca un ulteriore ritardo nella migrazione elettroforetica definito super shift. L'approccio classico EMSA è stato modificato al fine di identificare le proteine legate alla sonda ritardata non più attraverso l'impiego di anticorpi mirati, quanto piuttosto attraverso tecniche di spettrometria di massa che portassero all'identificazione dell'intero complesso legato al DNA e non solo della specifica proteina. La regione oligonucleotidica di interesse, marcata con un fluoroforo, è stata incubata con un estratto proteico per favorire il legame con gli specifici interattori e sottoposta ad elettroforesi in condizioni native. La banda che ha mostrato uno shift rispetto alla sonda isolata è stata escissa da gel e sottoposta alla classica procedura di identificazione. Questo nuovo approccio è stato utilizzato per investigare fattori proteici interagenti con sequenze oligonucleotidiche riproducenti alcune regioni del promotore di *cftr*, gene coinvolto nell'insorgenza della Fibrosi Cistica. L'attenzione si è focalizzata su questa particolare regione genica in quanto, in seguito alle analisi del gene *cftr* in vari pazienti affetti da questa patologia, è emerso che alcune regioni del promotore presentavano mutazioni correlate alla patologia. Sono pertanto state analizzate sia le regioni del promotore nella forma wild-type che in quella mutata.

✓ **Messa a punto di una strategia ChIP/MS-MS**

La ChIP è una metodologia largamente utilizzata in studi di biologia molecolare relativi alla struttura e alla regolazione della cromatina attraverso l'identificazione delle regioni del DNA legate in vivo da particolari fattori proteici, fattori trascrizionali, istoni, ecc. Fino ad oggi la principale finalità di un esperimento ChIP è stata l'identificazione delle regioni di DNA legate in vivo da una specifica proteina. Il protocollo classico di ChIP prevede la stabilizzazione dei legami nativi proteina-DNA mediante l'impiego di reattivi cross-linking seguita dall'immunoprecipitazione dell'intero complesso proteina-DNA utilizzando un anticorpo specifico per la proteina di interesse. Le regioni specifiche di DNA così purificate

vengono identificate solitamente mediante amplificazione con PCR e sequenziamento genico. Nell'ambito di questo progetto di ricerca, questa metodologia è stata opportunamente modificata al fine di avere non solo informazioni circa le specifiche sequenze oligonucleotidiche a cui si lega la proteina di interesse, ma di identificare i componenti proteici presenti nel complesso proteina-DNA immunoprecipitato, mediante l'impiego di metodologie di spettrometria di massa tandem accoppiate a cromatografia capillare a fase inversa (nano-LCMSMS). In questo modo è possibile identificare tutte le proteine che in vivo, legandosi funzionalmente ad un tratto di DNA, agiscono in maniera concertata alla proteina di interesse. Questo nuovo approccio sperimentale è stato impiegato nello studio del fattore trascrizionale Cbx 7 (Chromobox protein homolog 7). E' stato dimostrato che il gene *cbx7* è drasticamente down-regolato in carcinoma tiroideo ed altri tumori gravi. Inoltre, il livello d'espressione di Cbx7 in tumori tiroidei risulta inversamente relazionato con il grado di malignità della neoplasia.

✓ **Messa a punto di una strategia di RNA Affinity Chromatography Affinity**

L'idea alla base di questo approccio consiste nella possibilità di utilizzare sequenze di RNA sintetiche, che riproducono regioni oligonucleotidiche coinvolte in vivo in importanti processi di regolazione, come esche per isolare da estratti cellulari i complessi multiproteici che presentano alta affinità di legame per le sequenze stesse. I componenti proteici di questi complessi sono successivamente frazionati mediante SDS-PAGE ed identificati utilizzando metodologie di spettrometria di massa. La strategia appena descritta è stata impiegata per l'identificazione delle proteine capaci di legare la regione 3'UTR (UnTranslated Region) dell'mRNA di ENPP1. ENPP1 è un ectoenzima di membrana, che se iper-espresso in tessuti bersaglio causa resistenza insulinica, in molti modelli cellulari ed animali. Nonostante il suo potenziale ruolo patogenetico nella insulino-resistenza poco si sa dei meccanismi molecolari che modulano i livelli di espressione di ENPP1. I processi che vanno dallo splicing del RNA messaggero alla sua traduzione sul ribosoma rappresentano un fondamentale stadio di controllo dell'espressione genica. Questi meccanismi di regolazione post-trascrizionale, di cui solo recentemente si è cominciato ad apprezzarne l'importanza, sono modulati dall'interazione di fattori proteici (elementi in trans) con sequenze specifiche (elementi in cis) non tradotte dell'mRNA. Queste regioni sono soprattutto caratterizzate da sequenze ricche in adenina e uracile che legano proteine specifiche con la conseguente stabilizzazione o degradazione dell'mRNA, e la differente regolazione dell'espressione genica. Dati recenti individuano il 3'UTR mRNA di ENPP1 come un potenziale sito regolatorio importante nella

fase di traduzione della proteina, suggerendo che fattori in trans interagiscono con l'mRNA di ENPP1 modulandone ed influenzandone la sua espressione.

1 Introduzione

1.1 L'approccio genomico

Alcuni dei grandi progetti di ricerca che hanno caratterizzato le ultime decadi del XX secolo hanno avuto come obiettivo la determinazione della sequenza dell'intero patrimonio genetico degli organismi. Il più grande di questi progetti, ambiziosamente teorizzato per la prima volta da Charles DeLisi nel 1985 e successivamente denominato "*The human-genome project*", aveva l'ambizioso obiettivo di produrre una sequenza accurata del DNA euromatico umano. Nel giugno del 2000, l'IHGSC (*International Human Genome Sequencing Consortium*) un consorzio internazionale di soggetti pubblici e la Celera Genomics, un'azienda privata americana, hanno contemporaneamente annunciato una prossima pubblicazione della prima bozza del genoma umano, resa effettivamente disponibile nel febbraio del 2001, ponendo una pietra miliare nello studio della genetica e, più in generale, della biologia [1, 2]. Ad oggi le nostre conoscenze del genoma umano coprono circa il 99% del DNA euromatico per un totale di 2.9 Gb, con un tasso di errore stimato in 1 base su 10^5 . Stime attendibili concludono che solo 1-1.5% del genoma sia codificante per proteine [243 IHGSC 2004]. In teoria la conoscenza dei dati genomici consente, sulla base dell'universalità del codice genetico, di determinare la sequenza amminoacidica codificata da tutti i geni e quindi, in linea di principio, la struttura primaria di tutte le proteine di un organismo. Purtroppo la definizione stessa dei *loci* genici in un genoma complesso come quello umano è estremamente difficile. In organismi più semplici, infatti, la presenza dei *loci* genici è facilmente rilevata grazie alla presenza di lunghe *Open Reading Frame* (ORF). I *loci* genici umani, al contrario, presentano piccoli esoni, codificanti in media per soli 50 amminoacidi, intervallati da lunghi introni, che possono eccedere le 10 kB. La larga prevalenza degli introni sugli esoni impedisce di fatto il riconoscimento diretto delle ORF, obbligando ad integrare nell'analisi computerizzata il complesso ed inaccurato riconoscimento degli introni. Un'accurata identificazione dei geni umani si deve quindi basare sullo studio delle librerie di cDNA e sulla analisi per omologia con i genomi di altri organismi. Parallelamente al progetto genoma umano, infatti, molti altri progetti genoma sono stati avviati: in questo modo si spera non solo di aumentare la conoscenza degli organismi utili all'uomo (alcuni esempi sono *Bos taurus*, *Gallus gallus*, *Oryza sativa*), ma anche di migliorare la conoscenza del genoma umano mediante studi comparativi (*Pan troglodytes*, *Mus musculus*, *Rattus norvegicus*). Attualmente sono sequenziati i genomi di almeno dieci organismi eucarioti, e almeno altri 200 sono in via di completamento. La

comparazione per omologia dei genomi è molto efficiente nel riconoscimento di geni fortemente conservati, come ad esempio quelli del metabolismo basale, ma è molto meno sensibile per la rilevazione di geni in evoluzione rapida, che, sfortunatamente, includono i geni per la differenziazione delle specie, per la determinazione sessuale, per il controllo dell'attività cellulare. La nostra conoscenza degli organismi viventi non dipende unicamente dal numero di geni identificati, ma piuttosto dalla accuratezza con la quale si definisce la funzione di ognuno dei *loci* ossia la posizione cromosomica in cui risiede il gene per un particolare carattere. Il complesso lavoro del riconoscimento dei *loci* genici deve, quindi, essere seguito dalla accurata definizione della funzione del gene, sia esso codificante per proteine o semplicemente per RNA. La difficoltà nella definizione del ruolo dei geni cresce esponenzialmente con la complessità dell'organismo studiato. L'ambiguità nella definizione della funzione dei singoli geni inizia con il fallimento del paradigma 1 gene = 1 proteina = 1 funzione cellulare. L'evoluzione verso gli organismi superiori è infatti dovuta, più che all'utilizzo di un maggior numero di funzioni geniche, allo sviluppo della loro complessità, in maniera da aumentarne la flessibilità di funzionamento e la possibilità di modulazione. In funzione della complessità dell'organismo, ad un gene possono corrispondere uno o molti prodotti genici, ognuno dei quali può svolgere molteplici funzioni cellulari. Questo implica che la sequenza del genoma non è in grado da sola di definire in dettaglio gli "attori" che esistono in una cellula e che ne realizzano le funzioni biologiche. Questo quadro è ulteriormente complicato dalla sostanziale mancanza di informazioni che definiscano il livello e la sequenza temporale dell'espressione del gene, la localizzazione dei prodotti di trascrizione e traduzione, l'interazione fra questi per formare i complessi network che garantiscono il corretto svolgimento delle funzioni biologiche [2].

1.2 Il passaggio dall'era genomica all'era proteomica

La determinazione della sequenza del genoma di vari organismi, incluso quello dell'uomo, ha paradossalmente contribuito ad un rinnovato interesse per lo studio delle proteine. Infatti, la conoscenza dell'intero DNA umano costituisce un enorme patrimonio di informazioni, ma deve essere considerato un punto di partenza nella comprensione del funzionamento della macchina cellulare. La sfida della ricerca biologica si è quindi di nuovo rivolta alle proteine, originando la cosiddetta "era proteomica" con l'obiettivo di studiare l'intero complesso delle proteine (il proteoma) espresse da un genoma [3]. La proteomica, o meglio la Scienza del Proteoma, dunque costituisce una "filosofia" del tutto nuova nello studio delle proteine. L'attenzione dei ricercatori non è più focalizzata all'analisi in grande dettaglio della singola

molecola proteica purificata ma si indirizza allo studio simultaneo di un gran numero di proteine all'interno del loro contesto biologico. In tal modo si cerca di ottenere informazioni sulle variazioni quali-quantitative, sul comportamento e sulle interazioni delle proteine nel loro habitat naturale. Il genoma umano permette di predire le proteine che possono essere potenzialmente generate, ma non dove e quando o a che livello; non può informarci sul tipo di cellule in cui le proteine saranno espresse o in quali fasi di sviluppo o di differenziazione questo accadrà, né può prendere in considerazione l'enorme diversificazione di struttura che risulta. A livello genico, questa complessità è realizzata mediante meccanismi di regolazione e controllo della trascrizione più complessi, in grado di rispondere a molteplici stimoli a breve e a lungo termine. A livello del trascritto questa complessità si manifesta nel *processing* degli mRNA, che consente una fine e rapida regolazione della attività traduzionale e una modulazione delle attività proteiche presenti mediante lo *splicing* combinatoriale che porta alla rimozione degli introni ed alla congiunzione degli esoni nell'RNA. Infine, a livello proteico, ciò si riflette in proteine multidominio, appartenenti a famiglie codificate a volte da un solo gene, con un gran numero di funzioni biologiche e altrettanti siti di modulazione delle loro attività. La regolazione dei processi cellulari richiede una miriade di comandi, positivi e negativi, per mantenere sotto controllo le reazioni che prendono luogo ed essere sicuri che nessun evento cruciale avvenga a un tempo o in un luogo inappropriato. Noi sappiamo oggi che il segnale principale che orchestra queste reazioni si basa sulla modifica delle proteine [4]. Negli ultimi anni è stato ampiamente riconosciuto che il genoma rappresenta solo il primo livello di complessità. Le funzioni biologiche non sono solo regolate da un genoma statico, ma anche da una popolazione dinamica di proteine mediante un gioco di regolazione di geni e di proteine con influenze extracellulari [5]. Funzioni cellulari vitali come la replicazione del DNA, la trascrizione e la traduzione dell'mRNA richiedono l'azione coordinata di un certo numero di proteine che vengono assemblate in una matrice di complessi multiproteici di composizione e struttura diversa. Allo stesso modo, i processi biologici sono orchestrati e regolamentati mediante reti di *signalling* dinamici che connettono stimoli chimici o fisici a specifiche molecole effettrici. L'analisi dei complessi proteici e delle reti d'interazione proteina-proteina ed acidi nucleici-proteine sono quindi di centrale importanza nella ricerca biologica dell'era postgenomica [6]. Anche se le proteine sono i principali effettori del messaggio genomico, esse provvedono solo ad una parte del messaggio. Tuttavia, le proteine raramente agiscono da sole. Complesse funzioni cellulari sono quasi sempre il risultato dell'azione coordinata di molte proteine che agiscono in assemblaggi molecolari o pathway per realizzare una

determinata attività. Si pensa che gli schemi forniti dalle reti d'interazione proteica possano contribuire alla nostra comprensione sui rapporti tra genomi e fenotipi. Da una prospettiva evuzionistica, l'apparente complessità osservata a livello di un organismo non necessariamente deve riguardare le dimensioni del suo genoma. Basti pensare, infatti, che l'uomo possiede 31.780 geni, più o meno quanti ne ha il vermicciattolo *Caenorhabditis elegans*, il cui corpo è costituito di solo mille cellule, e decisamente meno di quanti ne abbia una piantina alta una diecina di centimetri, la *Arabidopsis thaliana*, organismi dalla struttura e dal comportamento decisamente meno complessi dei nostri. Il motivo è un'evidenza palese della non validità del dogma centrale secondo cui a un gene corrisponde una proteina. Se, come ormai è noto, nelle cellule eucariotiche ogni gene produce fra tre o più prodotti proteici, il numero di proteine in un organismo superiore può variare da 50000 a mezzo milione.

1.3 Gen-“omica”, gen-“omica” funzionale e prote-“omica”

Negli ultimi anni, una sconcertante e talvolta sciocca terminologia è sorta in seguito gli esperimenti biologici su larga scala. La Fig. 1.1 da una visione semplificata delle principali categorie [3].

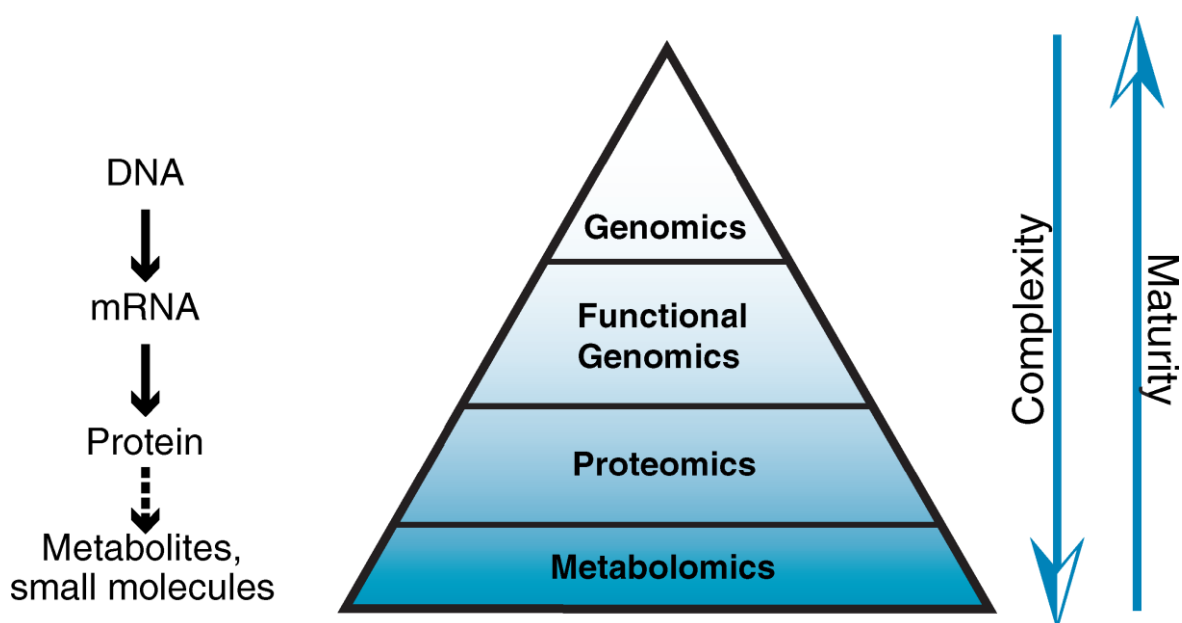


Figura 1.1 Relazione schematica delle diverse discipline “omiche” in relazione al flusso di informazione dal genoma attraverso i trascritti alle proteine e piccole molecole. Muovendo dalla genomica alla proteomica, la complessità aumenta drasticamente mentre la maturità della tecnologia diminuisce.

La *genomica* mira a determinare la sequenza cromosomica lineare di un organismo modello, così come le differenze nelle sequenze tra gli individui. Annotazione di genomi, compresa la definizione delle sequenze codificanti e sequenze di regolazione, fanno comunque parte di quella che è definita genomica. Il livello successivo verso il basso procede con i trascritti [RNA messaggero (mRNA)]. Non vi è necessariamente una relazione uno-a-uno tra la sequenza del genoma e il trascritto; le cellule usano lo splicing alternativo per generare diversi trascritti da un gene e tutta una serie di meccanismi per controllare la produzione, lo stato della trascrizione, e la degradazione dell'mRNA. Inoltre per la misura globale dei livelli di trascrizione in miriadi di condizioni, la genomica funzionale comprende diverse strategie in cui è eliminato il messaggio e sono osservati gli effetti cellulari conseguenti. Questa parte della *genomica funzionale*, quando eseguita su organismi interi, è anche chiamata *fenomica* perché è osservato il fenotipo di un organismo. Le proteine, i vettori principali dell'attività biologica, rappresentano il successivo livello di complessità. La funzione delle proteine dipende dalla precisa sequenza di aminoacidi, dalle modifiche (soprattutto quelle di regolazione, come la fosforilazione), dalla struttura tridimensionale (3D), dalla concentrazione proteica, dall'associazione con altre proteine, e dall'ambiente extracellulare. Di conseguenza, la *proteomica* è volta a stabilire la struttura di una proteina, le modifiche, la localizzazione e le interazioni proteina-proteina in aggiunta ai livelli di espressione proteica. Attualmente la spettrometria di massa (MS) è la tecnologia più versatile per poter misurare direttamente le proteine endogene [5]. Tecnologie non basate sulla spettrometria di massa in proteomica coinvolgono sistemi basati su array (ad esempio, il saggio del doppio ibrido per le interazioni proteina-proteina) o su strumenti strutturali e di imaging. Molte proteine sono enzimi o regolano la funzione degli enzimi che in ultima analisi determinano il livello di piccole molecole (metaboliti). Una visione globale della cellula richiede che siano misurate e modellate anche i metaboliti. La corrispondente disciplina su larga scala, che è la più avanzata nel campo dell'ingegneria e delle analisi di flusso del metabolita, è talvolta chiamato *metabolomica*. Approcci sperimentali in metabolomica spesso coinvolgono la spettrometria di massa, come in proteomica.

Queste definizioni non si devono però considerare assolute. Se da un lato infatti vi è generale un accordo sulla nomenclatura e la delimitazione di queste discipline, dall'altro diversi approcci sperimentali in queste categorie spesso si sovrappongono. Pertanto nessuno degli approcci sopra citati può da sola delucidare la funzione biologica. Piuttosto, ogni disciplina con la sua peculiarità può fornire un contributo a chiarire un quadro alquanto complesso.

1.4 Metodiche e tecniche in proteomica funzionale

Il termine *proteoma* fu inizialmente coniato per definire il complemento proteico di un genoma. Nel tempo l'accezione ha assunto nuove sfumature, ed oggi il termine proteoma connota lo studio della dinamica globale della funzione proteica, che include non solo l'espressione, la localizzazione e lo stato di modifica post-traduzionale, ma anche le interazioni, i meccanismi di controllo e le funzionalità biologiche in genere. Attualmente gli studi di proteomica sono focalizzati su due aree principali, la proteomica di espressione che tende alla definizione qualitativa e quantitativa dell'aumento e/o diminuzione dei livelli di proteine in dipendenza di opportuni stimoli e/o di condizioni fisio-patologiche, e la proteomica funzionale che tenta di identificare componenti di compartimenti cellulari, complessi multiproteici e vie di trasduzione del segnale. Tuttavia i moderni studi di proteomica funzionale sono volti non solo allo studio delle interazioni proteina-proteina ma anche allo studio delle interazioni tra complessi multiproteici ed acidi nucleici al fine di definire non solo la funzione biologica di specifiche proteine ma anche la loro influenza sulla regolazione dell'espressione genica. L'interazione DNA/RNA-proteina è un fenomeno estremamente delicato ed importante, in quanto tutte le funzioni cellulari dipendono dal binding di specifiche proteine con gli acidi nucleici. Tali proteine per legarsi al DNA e/o RNA devono essere in grado di riconoscere una sequenza nucleotidica specifica (sequenza consensus) e quindi di legare un numero limitato di nucleotidi. La complessità di questo tipo di interazione è evidente, basta pensare che eventuali "anomalie" o di queste famiglie di proteine o delle sequenza consensus sono alla base dell'insorgenza di patologie particolarmente gravi.

L'idea alla base di questo tipo di studio consiste nella possibilità di utilizzare approcci di diversa natura, analitici o anche preparativi, sia in vivo che in vitro, che consentano di isolare i componenti dei complessi proteici interagenti in maniera specifica con determinate sequenze oligonucleotidiche ed identificarli

Conseguentemente le metodiche in proteomica si sono arricchite di nuove tecniche, derivate dalla chimica delle proteine, dalla biologia cellulare ma, soprattutto, dalla spettrometria di massa. Una trattazione completa delle metodiche attualmente disponibili in proteomica si estenderebbe, quindi, ben oltre i confini di questa introduzione. Per questo motivo qui si presenta solo una breve rassegna delle principali metodologie e tecniche d'uso comune in proteomica funzionale.

1.4.1. Proteomica funzionale

La capacità di ogni organismo di sopravvivere dipende da diversi complessi meccanismi cellulari che soddisfano tutte le funzioni necessarie per la vita delle cellule. Inoltre, le cellule devono essere in grado di modificare i modelli di espressione genica in risposta ai segnali extra e intracellulari al fine di garantire che l'importo corretto del sottoinsieme appropriato di geni è espresso nel momento giusto [6, 7]. In passato, i meccanismi cellulari e di trasduzione del segnale sono stati pensati come un insieme di processi isolati o percorsi lineari dove le funzioni biologiche si susseguivano una dopo l'altra in un modello tempo-dipendente. Oggi, tuttavia, è ben noto che i meccanismi biologici fondamentali sono esercitati da un gran numero di proteine raccolte insieme per formare temporaneamente o transientemente grandi complessi poliproteici funzionali [8], o pathways cellulari [9]. Una volta svolte le loro funzioni, questi complessi dissociano formando singoli componenti liberi di associare ad altri complessi per svolgere attività distinte. Per complicare ulteriormente il quadro, meccanismi diversi e molti pathways di signalling sono in grado di interagire l'uno con l'altro in modo da formare una rete complessa, che integra contemporaneamente segnali extracellulari e intracellulari. Questi processi dinamici costituiscono i network molecolari attraverso cui le informazioni di signalling fluiscono all'interno della cellula [10, 11]. Inoltre oggi più che mai risulta evidente il fatto che per comprendere in che modo specifiche proteine influenzino la regolazione dell'espressione genica è necessario individuare ed isolare questi complessi multiproteici durante la loro interazione con acidi nucleici. Una descrizione completa di questo complesso intrico di vie di trasduzione del segnale e l'uso di network per predire l'intera gamma di comportamenti cellulari sono gli obiettivi principali della system biology [12, 13]. Un ruolo chiave nel campo delle scienze biologiche contemporanee e future può essere svolto dalla proteomica funzionale, un settore emergente nel campo della ricerca proteomica indirizzata verso il chiarimento delle funzioni proteiche e la definizione dei meccanismi cellulari a livello molecolare [14, 15]. Una descrizione completa dei processi cellulari è infatti strettamente dipendente da una chiara definizione dei complessi multi proteici che vi prendono parte e delle singole componenti proteiche coinvolte in queste entità funzionali. L'associazione di una proteina sconosciuta con partners appartenenti ad uno specifico complesso di proteine coinvolto in un meccanismo particolare potrebbe fortemente suggerire la sua funzione biologica [16]. Inoltre, una descrizione dettagliata dei pathways di signalling cellulare potrebbe trarre grande beneficio dalla delucidazione delle interazioni proteina-proteina in vivo.

1.4.2. La proteomica da un punto di vista critico

L'assemblaggio di proteine per formare complessi funzionali è un processo dinamico che avviene al momento opportuno. Il complesso transiente dissocia subito dopo che il meccanismo biologico è stato completato e i singoli componenti possono partecipare alla formazione di altri complessi guidati da segnali specifici. Inoltre, i processi cellulari si verificano in luoghi ben definiti all'interno della cellula e quindi i complessi funzionali assemblano al momento giusto e nel posto giusto per svolgere correttamente una specifica funzione. Esperimenti di proteomica funzionale volti a descrivere il network di processi cellulari dovrebbero quindi prendere in considerazione con attenzione la distribuzione nel tempo e la localizzazione intracellulare di ogni specifico meccanismo che si vuole esaminare.

La Biologia moderna sta vivendo una situazione paradossale in cui molto spesso la sequenza della proteina, il gene corrispondente codificante e anche i meccanismi di regolazione relativi all'espressione genica sono noti, ma il ruolo biologico della proteina non è affatto chiaro. Questo apparente paradosso suggerisce che vi è la necessità di cambiare il nostro approccio alla comprensione del funzionamento della cellula. Dobbiamo comprendere che l'attività di una proteina e il suo ruolo biologico potrebbero essere due aspetti molto diversi. L'attività di una proteina descrive la proteina stessa, sia essa un enzima, un trasportatore, un recettore, uno chaperone, uno scaffold, ecc, ma questa attività può essere esercitata in processi biologici completamente diversi a seconda della necessità della cellula. La funzione biologica di una proteina è quindi legata al processo specifico a cui la proteina sta in realtà partecipando. Di conseguenza, poiché la stessa proteina è in grado di assemblarsi con partner diversi per formare diversi complessi funzionali, ne consegue che una proteina può avere una sola attività, ma diverse funzioni biologiche.

È relativamente facile prevedere che la proteomica funzionale avrà un ruolo fondamentale negli studi futuri della system biology. Una descrizione accurata di componenti proteici all'interno di complessi funzionali è un pre-requisito fondamentale per comprendere il funzionamento della cellula e per svelare la complessa rete di pathways di signalling a livello molecolare. L'interpretazione dei dati di proteomica costituisce attualmente un grave problema, poiché non è sempre possibile assegnare esattamente una proteina identificata ad uno specifico complesso riducendo così notevolmente l'impatto di questi studi. Lo sviluppo di metodi di pre-frazionamento dotati di risoluzione adeguata per separare i singoli complessi di proteine riduce la complessità del campione ed il background della proteina rendendo l'interpretazione dei dati relativamente semplice. Esperimenti specifici possono

essere progettati per dare un significato funzionale ai risultati della proteomica. Vi è molto spazio per la creatività nel collegare strategie biologiche, cellulari e molecolari, con il potente approccio della proteomica per risolvere questioni che non potevano essere affrontate in precedenza.

Un ulteriore obiettivo da perseguire in futuro, sarà il tentativo di effettuare esperimenti di proteomica funzionale in tempo reale nei sistemi in vivo mediante la produzione di modelli animali recanti una versione taggata dell'esca proteica. Animali omozigote, quando vitali, forniranno i tessuti e / o le cellule progenitrici per l'isolamento di complessi funzionali che si formeranno in vivo, in modo da rivelare l'identità dei loro singoli componenti ed affrontare la questione del loro funzionamento reale in vivo.

1.4.3 Metodologie per lo studio delle interazioni acidi nucleici-proteine

1.4.3.1 Studi in vitro

Lo studio delle interazioni DNA/RNA-proteina è stato fino ad oggi affrontato mediante approcci di diversa natura, sia in vitro che in vivo. In particolare, tra le tecniche più comunemente utilizzate in vitro vi è senza dubbio l'EMSA (Electrophoretic Mobility Shift Assay). Si tratta di un saggio di mobilità elettroforetica mediante il quale è possibile verificare il legame di una proteina specifica con un tratto di DNA (sonda), semplicemente osservando un ritardo (shift) del probe oligonucleotidico in presenza della proteina[17, 18]. Questo saggio è utilizzato soprattutto per lo studio di fattori trascrizionali che possono essere aggiunti o puri in forma ricombinante, se già si ipotizza il loro legame alla sequenza di DNA di interesse, oppure possono essere analizzati nel contesto dell'estratto cellulare di appartenenza al fine di identificare DNA binding proteins ancora sconosciute. In quest'ultimo caso, in seguito alla formulazione di ipotesi funzionali, è possibile poi confermare la presenza delle proteine ipotizzate mediante super shift. Sperimentalmente si ricorre all'utilizzo di un anticorpo che riconosca in maniera specifica la putativa proteina. L'anticorpo viene incubato con il complesso "DNA-estratto proteico", se vi è riconoscimento antigene anticorpo si osserverà un ulteriore ritardo della mobilità elettroforetica della sonda.

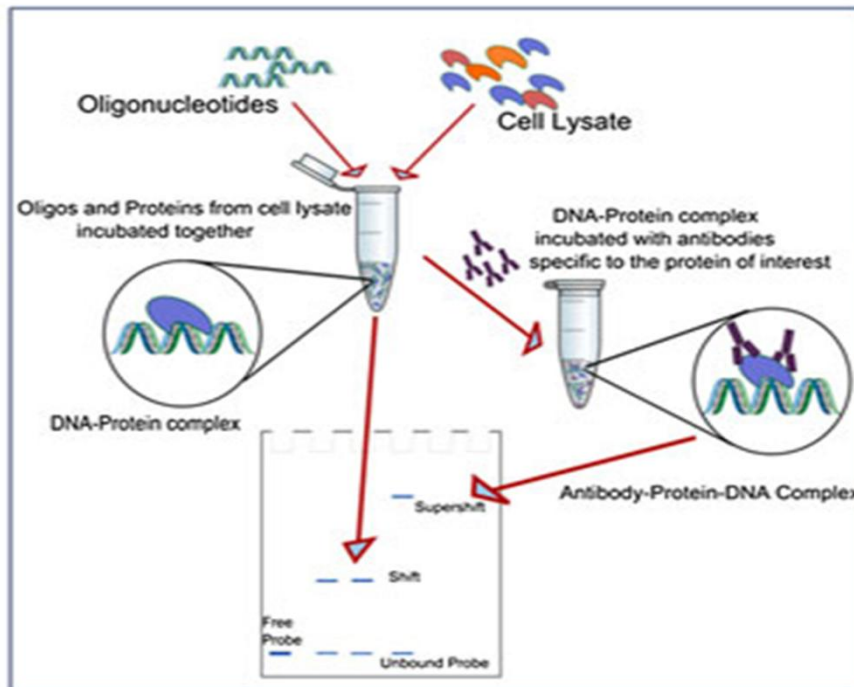


Figura. 1.2 Strategia di un esperimento EMSA.

Per poter seguire la corsa elettroforetica ed apprezzare lo shift generalmente il frammento oligonucleotidico di interesse viene marcato o radioattivamente (dATP con P32) oppure con un fluoroforo (Cy3, Cy5) [19].

In teoria il saggio EMSA sembra essere rapido e non particolarmente complesso, in realtà invece c'è bisogno di ottimizzare diversi parametri affinché l'esperimento abbia un esito positivo:

In primo luogo bisogna valutare la "lunghezza" della sonda che si vuole investigare, in principio si riteneva che il doppio strand di DNA dovesse essere costituito da un minimo di 20 ad un massimo di 200 coppie di basi; in realtà studi relativamente recenti hanno dimostrato che in condizioni ottimali il frammento oligonucleotidico non debba superare le 50 coppie di basi. Infatti se una sequenza troppo lunga viene incubata con un estratto proteico nucleare vi è la possibilità che anche proteine che non mostrano affinità con il probe vi si leghino in maniera aspecifica. Inoltre probe più piccoli consentono una maggiore risoluzione elettroforetica e quindi una maggiore capacità di apprezzare lo shift.

Un'altra importante considerazione va fatta in merito al tampone di binding da utilizzare per favorire in vitro il legame tra la sonda oligonucleotidica e le proteine.

I parametri critici che devono essere considerati sono: il pH del tampone, la concentrazione di ioni mono e bi valenti, la possibilità di utilizzare o meno dei competitori specifici per il

DNA (poly dI, dC, etc.), al fine di inibire il legame di proteine aspecifiche, ed infine la quantità di glicerolo da utilizzare nel tampone di binding.

E' di fondamentale importanza, inoltre, definire le più opportune condizioni elettroforetiche. La risoluzione dei complessi DNA- proteine, infatti, dipende anche dalla capacità di questi ultimi di non dissociare durante la corsa sul gel.

In ultima analisi poi non bisogna dimenticare che mentre sono state "standardizzate" le quantità di proteina ricombinante da incubare con la sonda, poco ancora si sa sulle quantità di estratti proteici da utilizzare in saggi di questo tipo.

Oltre alle finalità precedentemente descritte, va ricordato che la metodologia EMSA consente inoltre di determinare la relativa affinità di un fattore trascrizionale per una o più sonde oligonucleotidiche, oppure per comparare l'affinità di differenti fattori trascrizionali per il medesimo probe. L'affinità della proteina di interesse per la sequenza oligonucleotidica viene determinata valutando il rapporto, utilizzando diverse quantità di probe, tra la sonda legata rispetto alla quantità di sonda non ritenuta.

In generale possiamo asserire che i vantaggi principali di un saggio di questo tipo derivano dal fatto che tale tecnica è in grado di risolvere i complessi DNA-proteine in diversi rapporti stechiometrici o conformazionali; inoltre l'estratto nucleare di interesse può essere utilizzato tal quale senza necessitare di purificazioni a monte.

Quando invece non si hanno indicazioni su quali possano essere i possibili ligandi proteici di una determinata sequenza oligonucleotidica, si ricorre a strategie alternative, come la purificazione delle proteine mediante cromatografia di affinità.

Tale metodologia sfrutta l'affinità intrinseca di un ligando, in questo caso costituito da una sequenza di DNA e/o RNA. Queste regioni oligonucleotidiche, amplificate per PCR possono essere immobilizzate covalentemente su supporti inerti disponibili in commercio e utilizzati per l'isolamento delle proteine coinvolte nel legame specifico alla sequenza nucleotidica [20-22]. Anche un tag di biotina può essere aggiunto a oligonucleotidi sintetici; l'immobilizzazione dell'esca sul supporto solido sfrutterà quindi il vantaggio dalla sua alta affinità con le molecole di streptavidina [23] La molecola ligando è covalentemente legata ad un supporto insolubile (di solito sfere di agarosio/Sefarosio) e usata come esca per legare selettivamente la proteina di interesse ed i suoi partners ("pull down"). Teoricamente non ci sono limitazioni nella scelta delle molecole di legante che può essere impiegato in esperimenti di pull-down. Numerosi esempi sono stati riportati utilizzando substrati modificati e/o inibitori di proteine selezionate con specifiche attività enzimatiche [24-26]. In altri casi sono stati utilizzati anche farmaci nativi e/o modificati indirizzati verso specifici

target proteici [27]. A questo proposito, una questione particolarmente interessante che deriva da questa strategia consiste nella possibilità di individuare nuove molecole bersaglio precedentemente sconosciute per un farmaco specifico. La strategia di fishing ha trovato grande applicazione nell'isolamento di proteine leganti acidi nucleici, in studi su complessi di trascrizione e/o di traduzione o sulla macchina dello splicing. (Fig. 1.3).

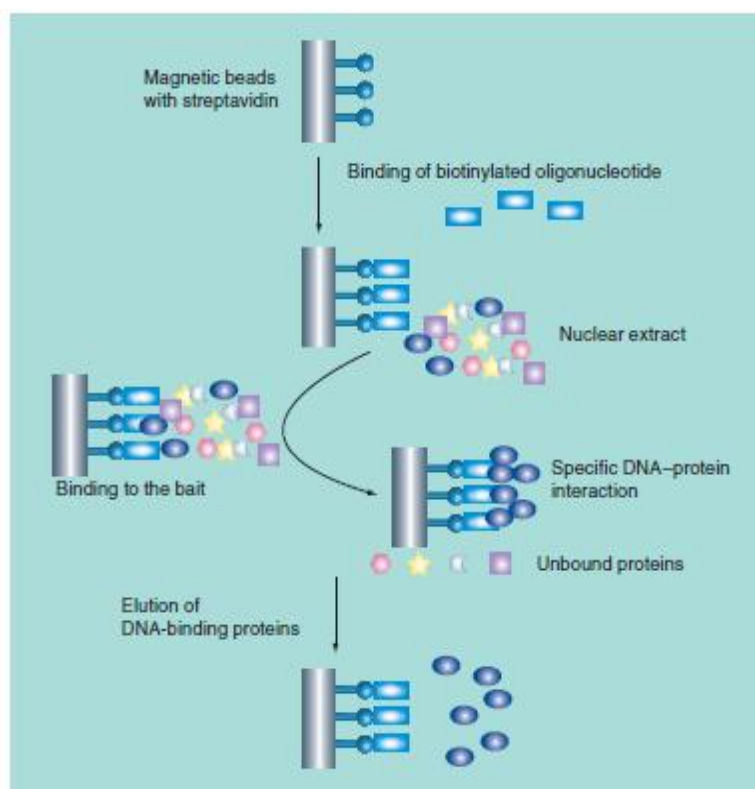


Figura 1.3 Strategia di isolamento di proteine che legano il DNA.

Le componenti proteiche espressamente riconosciute dall'esca possono quindi essere eluite e frazionate mediante SDS-PAGE. Le bande proteiche sono escisse dal gel, digerite con enzimi proteolitici e le miscele peptidiche derivanti analizzate mediante tecniche LC-MS/MS, che portano alla identificazione delle proteine attraverso la ricerca in database [28]. Questa sequenza sperimentale, SDS-PAGE/ digestione in situ (su gel)/LC-MS/MS, di solito è preferibile ad altri metodi esistenti, perché l' SDS-PAGE, anche se dotata di una risoluzione molto bassa, è abbastanza sensibile da permettere l'analisi di piccole quantità di complessi proteici isolati attraverso procedure di affinity e/o immunoprecipitazione. Quindi, poiché una singola banda da gel monodimensionale spesso può contenere numerose proteine diverse, le metodologie LC-MS/MS, essendo in grado di analizzare miscele di peptidi molto complesse, sono abitualmente utilizzate per l'identificazione delle proteine. La DNA

Affinity Chromatography nonostante la sua vasta applicazione presenta dei limiti sperimentali intrinseci di cui non si può non tenere conto.

In primo luogo dobbiamo ricordare che le proteine che interagiscono in maniera prevalente con il DNA sono i Fattori Trascrizionali, i quali generalmente non solo sono under-espressi ma si legano in maniera transiente agli acidi nucleici. Per ovviare a questa difficoltà generalmente si procede ad una multerimizzazione dell'acido nucleico di interesse al fine di aumentare i siti di binding per questa classe di proteine, contestualmente però aumenta la probabilità da parte di proteine aspecifiche di legare la sequenza oligonucleotidica di interesse. Quest'ultimo rappresenta un notevole limite sperimentale. Per allontanare questi contaminanti si ricorre spesso all'utilizzo di lavaggi stringenti, per esempio utilizzando tamponi ad alta salinità, i quali mettono a repentaglio il debole e transiente legame dei fattori trascrizionali alla sonda di interesse.

Tuttavia questa metodologia non necessita a priori di ipotesi funzionali, ovvero non richiede la conoscenza delle sequenze oligonucleotidiche necessarie per il legame di determinate proteine, anzi ci permette in tempi rapidi l'isolamento e l'identificazione dei complessi proteici interagenti con le sequenze di interesse e di formulare ipotesi funzionali che potranno poi essere validate mediante metodologie proprie della proteomica funzionale.

1.4.3.2 Studi in vivo

Nonostante una notevole quantità di dati biochimici ottenuti con sistemi in vitro, informazioni dettagliate sulle interazioni fra i fattori trascrizionali e i loro targets in vivo sono solo recentemente emerse, grazie alla tecnica dell'immunoprecipitazione della cromatina (ChIP). D'altra parte, la sequenza del genoma umano apre nuovi orizzonti per comprendere la strategia dei fattori trascrizionali, grazie all'identificazione dei loro target genomici e del complesso dei geni da essi regolati. La prova formale che un gene è sotto il controllo di un fattore trascrizionale è data dalla dimostrazione che detto fattore sia effettivamente legato in vivo a un elemento funzionalmente importante del promotore o dell'enhancer del gene in questione. L'immunoprecipitazione della cromatina è un saggio ineludibile per studiare queste interazioni. Il protocollo di ChIP prevede la stabilizzazione dei legami nativi proteina-DNA mediante l'impiego di reattivi cross-linking seguita dall'immunoprecipitazione dell'intero complesso "proteina-DNA" utilizzando un anticorpo specifico per la proteina di interesse. Le regioni specifiche di DNA purificate vengono identificate solitamente mediante amplificazione con PCR e sequenziamento genico [29, 18].

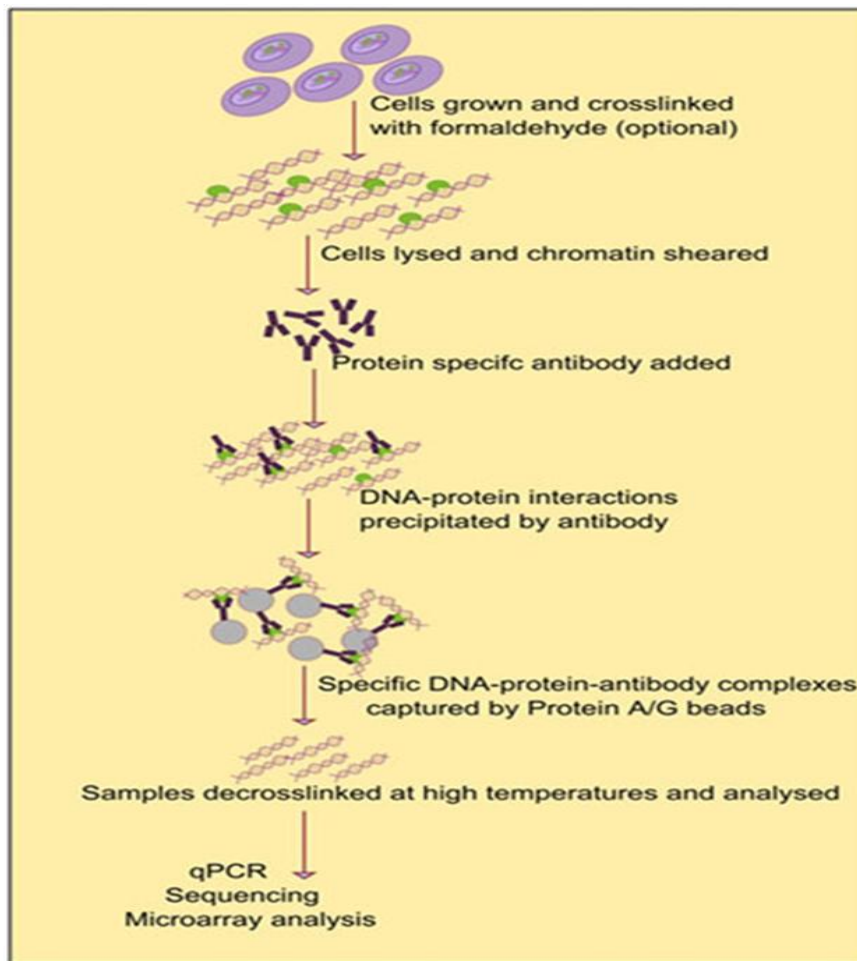


Figura. 1.4 Strategia di un esperimento ChIP.

Un importante aggiornamento della tecnica è la possibilità di effettuare due immunoprecipitazioni sequenziali con due anticorpi diretti contro fattori diversi (Re-ChIP): si otterrà così un arricchimento in frammenti che sono legati contemporaneamente dai due fattori.

La ChIP può essere anche utilizzata non solo per verificare che un fattore sia legato a una sequenza nota, ma anche per individuare nuovi target: la tecnica viene modificata in modo tale da utilizzare il DNA immunoprecipitato come sonda per ibridizzare un vetrino contenente frammenti di DNA genomico (chips). Questa variante prende il nome di "ChIP on chip". Inizialmente messa a punto in lievito, è stata recentemente applicata a sistemi umani.

Un'altra tecnica è quella dei Microarray, la quale a differenza della ChIP che permette di studiare una sola sequenza che lega una proteina, ci consente di analizzare un gran numero di sequenze.

La tecnica dei microarray nasce negli anni 90 al fine di indagare i livelli di espressione dei geni in una determinata linea cellulare. Essa è in grado di stabilire se un determinato gene è attivo in quella cellula, cioè se l'RNA corrispondente viene effettivamente prodotto e in quale misura. Oggi è possibile stabilire l'espressione per migliaia di geni contemporaneamente, confrontando allo stesso tempo due diversi tessuti, ad esempio campioni provenienti da pazienti diversi (uno sano e uno ammalato) o dallo stesso paziente (prima e dopo la cura). Un microarray, o DNA-chip, è un sistema miniaturizzato in cui su un apposito supporto vengono immobilizzati acidi nucleici a sequenza nota corrispondenti ai geni che si vogliono studiare. Da un campione biologico si preleva un'opportuna quantità di RNA, che è il prodotto intermedio del processo che dai geni conduce alle proteine e che ha la caratteristica di riconoscere le sonde complementari alle quali si lega chimicamente. L'RNA viene sottoposto ad un processo di etichettatura con marker radioattivi o fluorescenti in modo da poterlo rilevare in fase successiva. Infine si procede alla ibridazione, dove la miscela di RNA viene messa in contatto con il microarray in modo che le sue molecole si leghino alle rispettive sonde. Il microarray viene poi letto da uno scanner, rendendo un'immagine in cui la luminosità o il colore di ogni punto è proporzionale alla quantità di RNA legatosi alla sonda in quel punto. L'immagine è acquisita ed elaborata al computer per produrre dati "raw image", una misura dell'espressione genica per ogni gene presente sul chip.

Sebbene una forte spinta all'applicazione delle ChIP, in tutte le sue varianti, è derivata dalla disponibilità commerciale di una vasta gamma di anticorpi impiegabili con buoni risultati nell'immunoprecipitazione della cromatina, fino ad oggi essa è stata applicata quasi esclusivamente all'analisi del DNA. Soltanto in questi ultimi anni, con l'avvento della proteomica funzionale, si è cominciato ad impiegare la ChIP come strumento per l'identificazione delle proteine che in vivo sono legate funzionalmente ad un tratto di DNA.

1.5 Le interazioni proteina-proteina nelle malattie umane

“La vita è una relazione tra molecole, non la proprietà di ogni molecola. Così è per la malattia, che mette in pericolo la vita”, scrissero Zuckerkandl and Pauling (1962) nel loro capitolo su *'Molecular disease, evolution and genic heterogeneity'*. Sono trascorsi circa 50 anni e ancora siamo lontani dal capire i meccanismi molecolari della maggior parte delle malattie, nonché le interazioni molecolari che si stabiliscono in organismi sani e malati [30]. Infatti, le proteine non funzionano in modo isolato, ma piuttosto nella cellula, interagiscono prevalentemente con altre proteine, ma anche con altre molecole come il DNA, l'RNA ed i

metaboliti. Così, gli studi delle proteine e delle loro interazioni sono essenziali per capire il loro ruolo all'interno della cellula. In genere, il termine '*protein interaction*' comprende una vasta gamma di eventi, come complessi transitori e stabili, oltre che interazioni fisiche e funzionali. Le interazioni proteiche sono coinvolte nel *signaling*, nei network metabolici etc. Una migliore comprensione delle interazioni proteiche, sia con altre proteine o con DNA, RNA, molecole piccole o di membrana, potrebbe rivelare il meccanismo molecolare dei processi che portano alle malattie, processi estremamente complessi, così come lo sono le proteine e le interazioni coinvolte in esse.

Mutazioni all'interfaccia di interazione delle proteine (o siti correlati, ad esempio siti attivi, i siti di legame allosterico) evidentemente potrebbero disturbare l'interazione della proteina. In realtà, le proteine sono continuamente sintetizzate e degradate in amminoacidi. La cinetica dei processi e le dinamiche dei network devono essere considerate per comprendere come la perturbazione delle proteine e delle loro interazioni conducano alla malattia. Infine, è importante considerare anche il contesto specifico (tessuto, stadio di malattia e di risposta) in cui avviene l'interazione proteica. E' chiaro poi, che esiste ancora un gap tra l'identificazione del network di proteine associate alla malattia e la comprensione completa del meccanismo della malattia. Questo gap, sfortunatamente, è riempito con più domande che risposte. Siamo ancora lontani dal comprendere l'eziologia della maggior parte delle malattie, ma ulteriori rilevanti progressi in ambito tecnologico e sperimentale, come quelli riguardanti le interazioni proteiche, la struttura delle proteine, l'espressione genica, insieme agli strumenti di calcolo per organizzare, visualizzare e integrare questi dati forniranno un passo avanti in quella direzione.

Idealmente, dal momento che gli approcci strutturali e del network sono complementari, la combinazione di questi due ha il potenziale di essere un eccellente quadro per lo studio dei meccanismi delle malattie.

1.6 Spettrometria di massa e proteomica

L'introduzione della spettrometria di massa nello studio delle proteine ha permesso di raggiungere con relativa facilità risultati precedentemente fuori portata (ad esempio, il proteoma del parassita malarico, il proteoma di batteri, la determinazione completa dei costituenti di organelli). La spettrometria di massa si presta, inoltre, allo studio delle interazioni fisiche fra le proteine nelle complesse reti determinate dalla formazione di legami dinamici fra le proteine stesse (ad esempio, l'analisi globale dei complessi multiproteici in lievito e dei macrocomplessi multiproteici quali l'apparato che effettua lo

splicing dell'mRNA). Infine, come già ricordato, la spettrometria di massa consente lo studio delle modificazioni post-traduzionali che regolano e modificano le funzioni delle proteine. Anche la ricerca clinica ha cominciato ad avvalersi della spettrometria di massa, con risultati estremamente promettenti. Con questa tecnologia è infatti possibile ottenere “profili” proteomici da campioni biologici, cioè profili che rappresentano un gran numero di proteine presenti in un campione o in una sua frazione. L'applicazione di queste metodiche alle proteine sieriche ha permesso di identificare profili caratteristici (fingerprints) di gravi patologie quali il carcinoma dell'ovaio o della prostata. Si tratta di patologie in cui la diagnosi precoce ha un'importanza vitale e i profili caratteristici ottenuti dalla sieroproteomica sembrano costituire predittori diagnostici di sensibilità e accuratezza grandemente superiori a quelli ottenuti con altre tecniche attualmente in uso.

1.6.1 Le metodologie in spettrometria di massa

L'identificazione di una proteina precedentemente purificata, o di proteine da miscele complesse, separate solitamente mediante elettroforesi su gel SDS-PAGE mono e bidimensionali, può essere effettuata mediante la cosiddetta procedura del mass fingerprint. Il principio fondamentale su cui esse si basano sfrutta l'idea che il pattern proteolitico di ciascuna proteina sia univocamente dettato dalla specifica sequenza amminoacidica, e che pertanto risulti identificativo della proteina in esame [31].

Questo metodo prevede l'idrolisi chimica o enzimatica direttamente sulle bande proteiche escisse da gel e la successiva analisi di massa dei peptidi rilasciati, mediante MALDI MS. I pesi molecolari ottenuti sono introdotti in opportuni programmi [32], sviluppati per l'identificazione di proteine, attraverso ricerche in banche dati telematiche. L'insieme dei dati raccolti, il mass fingerprint per l'appunto, viene confrontato con una digestione “virtuale” effettuata tramite software di tutte le proteine contenute nelle banche dati genomiche. La determinazione accurata della massa di un numero sufficiente di peptidi identificherà una e una sola proteina fra quelle presenti nelle banche dati. In altre parole, solo una proteina potrà dare origine all'insieme dei peptidi aventi le masse osservate.

La risposta può essere ambigua e non univoca se una serie di parametri non sono soddisfatti: numero di peptidi non sufficiente, presenza di contaminanti peptidici dovuti alla manipolazione del campione, scarsa risoluzione elettroforetica. In tal caso sono utilizzate tecniche di spettrometria di massa tandem accoppiate alla cromatografia liquida capillare. Le miscele peptidiche prodotte dall'idrolisi chimica o enzimatica sono separate mediante HPLC capillare e le frazioni eluite dalla colonna sono direttamente introdotte nella sorgente ESI

dello spettrometro. Gli ioni isolati da un primo analizzatore passano in una camera di collisione. Gli ioni frammento prodotti sono selezionati da un secondo analizzatore. I frammenti carichi prodotti danno informazioni in merito alla struttura primaria del peptide selezionato. La strumentazione scelta per questo tipo di indagine prevede l'utilizzo di un accoppiamento tra due analizzatori separati da una camera di collisione oppure di una trappola ionica. In tal caso le banche dati sono interrogate con dati relativi non solo alle masse peptiche, ma anche alla loro sequenza. E' stato dimostrato che due soli peptidi sono sufficienti per identificare una proteina in modo non ambiguo [33].

Con questo metodo, i peptidi vengono frammentati mediante collisione con gas inerti (Collision-Induced Dissociation, CID); in appropriate condizioni i peptidi si frammentano in corrispondenza dei legami fra gli amminoacidi (legami peptidici); l'insieme dei frammenti risultanti dalla CID di un singolo peptide, poniamo ad esempio, di dieci amminoacidi, sarà costituito dallo stesso peptide che avrà subito la perdita di uno, due, tre amminoacidi e così via. La differenza di massa fra il peptide originario, il frammento di nove amminoacidi, quello di otto, di sette, ecc., corrisponde al peso molecolare di ciascun amminoacido e quindi fornisce la sequenza del peptide, sia pure parziale. L'informazione di sequenza in alcuni casi può confermare l'identificazione mediante mass fingerprint, le conferisce grande affidabilità e può essere utilizzata per ulteriori ricerche, in altri casi come detto può fornire informazioni che diversamente non si sarebbero potute ottenere. Analogamente, lo studio delle modificazioni posttraduzionali delle proteine viene effettuato misurando le variazioni di massa di peptidi dovute all'aggiunta di uno o più gruppi chimici quali, ad esempio, il gruppo fosfato nel caso della fosforilazione.

I profili proteomici di miscele di proteine sono ottenuti misurando il peso molecolare di un campione complesso di proteine. In questo caso non si ha la formazione di un solo "picco" di massa, cioè la misurazione di un singolo peso molecolare, ma di numerosissimi picchi, ciascuno corrispondente a una singola proteina. Sebbene in questo caso non sia possibile determinare direttamente l'identità delle proteine coinvolte, il profilo complessivo o alcune sue parti possono essere dotate di una specificità tale da risultare diagnostici, ad esempio, di una specie batterica, dello stato funzionale di una cellula o di una patologia in atto.

1.6.2 La strumentazione

La spettrometria di massa si impone come tecnica elettiva per le identificazioni proteiche nel panorama scientifico proteomico. Ciò è dovuto allo sviluppo di strumentazioni innovative e

all'accumulo di una notevole quantità di informazioni in merito a sequenze geniche e proteiche, raccolte in banche dati. La spettrometria di massa si avvale di numerose e differenti strumentazioni, caratterizzate da vantaggi e limiti. Gli spettrometri di massa si compongono di sorgenti, necessarie alla ionizzazione e volatilizzazione del campione in esame, analizzatori che separano gli ioni in base al rapporto massa/carica dell'analita, ed un rivelatore che determina il numero di ioni per ogni rapporto massa/carica trovato. Se in passato il suo impiego era limitato all'analisi di molecole di piccole e medie dimensioni, oggi è una tecnica fondamentale per lo studio delle macromolecole biologiche. Il limite, dovuto ai sistemi di analisi, risiedeva nella difficoltà di trasferire biomolecole polari cariche, di alto peso molecolare, allo stato gassoso. Nuove tecniche di ionizzazione hanno aperto la strada allo studio delle macromolecole biologiche mediante spettrometria di massa: electrospray ionization (ESI) e matrix-assisted laser ionization (MALDI) [34-36].

Ambedue le tecniche di ionizzazione sono accoppiate ai seguenti analizzatori: time of flight (TOF), quadrupoli (Q), ion trap (IT). Le configurazioni comunemente selezionate in base alle esigenze sperimentali sono: MALDI TOF, ESI IT, ESI Q, ESI Q-TOF.

Nella tecnologia MALDI il campione è analizzato in fase solida. La tecnica di ionizzazione, inducendo la formazione preferenziale di ioni monocarica, consente l'analisi diretta di miscele complesse. Il campione in esame co-cristallizza in presenza di una matrice, un composto aromatico acido, che media il trasferimento di energia da un raggio laser (UV o raramente IR) al campione stesso. Il sistema aromatico della matrice è eccitato dal laser e, tornando allo stato fondamentale, trasferisce energia al campione con essa cocristallizzato.

Detta energia è sufficiente a promuovere la formazione di uno ione monocarica MH^+ e ad indurre il desorbimento dal cristallo originario. A definire una tecnica di spettrometria di massa non è solo il metodo di ionizzazione scelto, ma anche il tipo di analizzatore, in cui gli ioni prodotti sono separati in base al rapporto massa/carica. Ad una sorgente MALDI è accoppiato un analizzatore a tempo di volo (TOF), fisicamente costituito da un tubo d'acciaio, di lunghezza compresa tra 1 e 2 metri. Il principio su cui si realizza la separazione sfrutta il fatto che gli ioni generati in sorgente, avendo energia cinetica $1/2mv^2$, si muovono con velocità inversamente proporzionali alla radice quadrata della massa e pertanto raggiungeranno il rivelatore in tempi successivi, in ordine di massa crescente.

Nella tecnologia ESI, il campione è analizzato in fase liquida. Un flusso di gas inerte in una camera di ionizzazione, ad alta temperatura e a pressione inferiore a quella atmosferica, favorisce l'evaporazione del solvente, che contiene l'analita carico. L'evaporazione del solvente porta ad una drastica diminuzione del volume: le forze repulsive tra molecole

cariche prevalgono rispetto alle forze di tensione superficiale, determinando una repentina nebulizzazione. Un'opportuna differenza di potenziale indirizza gli ioni verso l'analizzatore. Durante il processo di ionizzazione, le molecole proteiche e/o i peptidi vengono protonati da un numero variabile di protoni, in corrispondenza di residui basici o all'estremità Nterminale. Si forma una popolazione di residui a carica multipla $(M+nH)^{n+}$ che, separati da un analizzatore, danno luogo a distribuzioni di tipo gaussiano dei segnali m/z , ciascuna delle quali può essere trasformata su scala di massa reale da un algoritmo matematico.

Poiché l'analita è in fase liquida, la tecnica di ionizzazione descritta è accoppiata con una di frazionamento a monte: la cromatografia liquida. Questo tipo di strumentazione è generalmente configurato con un analizzatore quadrupolare o con una trappola ionica. Il primo consta di quattro barre metalliche con geometria cilindrica, alle quali è applicato un campo elettrico costante. Il movimento dello ione dipende da un campo elettrico oscillante a radio-frequenze. Solo lo ione con un definito rapporto m/z percorrerà, ad un dato valore del campo elettrico, una traiettoria stabile che gli consente l'uscita dal quadrupolo. Nella trappola ionica l'idea di base è opposta. E' un quadrupolo chiuso, al cui interno gli ioni si muovono con una traiettoria stabile, per effetto di un campo elettrico. Gli ioni con un dato rapporto m/z sono selezionati dal campo elettrico che, rendendo instabile la loro traiettoria all'interno dell'analizzatore li spinge ad uscire verso il detector.

1.7 Database pubblici per le interazioni proteina-proteina

Le proteine e le loro interazioni sono sempre state studiate una per una. Decadi di approcci "una per volta" hanno fornito informazioni funzionali per solo il 5-10% di tutti i geni predetti [37]. Più recentemente, la delucidazione del pieno repertorio di geni in diversi organismi e i rilevanti progressi nel campo della genomica funzionale e della proteomica hanno spianato la strada per le analisi delle interazioni proteina-proteina. Con lo sviluppo di queste metodologie su larga scala, è stata generata una lista di potenziali target.

Pertanto è nata la necessità di cercare in letteratura per scoprire se alcune delle proteine identificate come partners di una determinata proteina esca fossero già state identificate da altri ricercatori. Così sono stati sviluppati una serie di database che integrano i dati delle interazioni proteina-proteina provenienti da diverse fonti, quali interazioni fisiche o genetiche, con dati provenienti dalla letteratura. Di solito i ricercatori iniziano questo tipo di approfondimento di ricerca con PubMed. Tuttavia esistono database pubblici molto utili, quali ad es. HRPD, IntAct e BIND.

La “Human Protein Reference Database” (HRPD) rappresenta una piattaforma centralizzata per visualizzare ed integrare le informazioni inerenti l'architettura di dominio, le modificazioni post-traduzionali, le reti di interazione e di associazione per ogni proteina del proteoma umano. Tutte le informazioni in HRPD sono estratte manualmente dalla letteratura da biologi esperti che leggono, interpretano e analizzano i dati pubblicati. IntAct è un database d'interazioni proteina-proteina disponibile liberamente in rete. Tutte le interazioni sono estrapolate dalla letteratura e da comunicazioni provenienti direttamente dagli utenti. Interazioni che derivano da studi in larga scala di solito non sono recuperate mediante ricerche con l'utilizzo di parole chiave in PubMed, mentre è facile trovarle in Intact. BIND invece è un database progettato per memorizzare le descrizioni complete di interazioni, complessi molecolari e pathways.

1.8 Scopo del progetto di tesi

Le interazioni acidi nucleici-proteina sono basilari non solo per definire la funzione biologica di specifiche proteine ma anche per comprendere la loro influenza sulla regolazione dell'espressione genica.

Il lavoro svolto durante il progetto di tesi è stato, dunque, incentrato sullo sviluppo di metodologie innovative in proteomica funzionale per lo studio delle interazioni proteine-acidi nucleici. A tale scopo le metodologie classiche per lo studio di tali interazioni (EMSA, ChIP, ecc...) sono state opportunamente modificate per avere informazioni sulla natura delle proteine presenti nei complessi multiproteici in grado di legare specificamente una sequenza di DNA o RNA. Il lavoro si è articolato come segue:

✓ **Messa a punto di una strategia EMSA/MS-MS**

L'attenzione si è focalizzata sul promotore del gene *cftr* in quanto, in seguito all'analisi di quest'ultimo in vari pazienti affetti da Fibrosi Cistica, è emerso che alcune regioni del promotore presentavano mutazioni correlate alla patologia. Sono pertanto state analizzate sia le regioni del promotore nella forma wild-type che in quella mutata. Per identificare le proteine che interagivano con le regioni geniche di interesse è stato messo a punto un saggio EMSA che consentisse di identificare le proteine legate alla sonda ritardata non più attraverso l'impiego di anticorpi mirati, quanto piuttosto attraverso tecniche di proteomica che portassero all'identificazione dell'intero complesso legato al DNA e non solo della specifica proteina.

✓ **Messa a punto di una strategia ChIP/MS-MS**

Il classico protocollo ChIP è stato modificato al fine di avere non solo informazioni circa le specifiche sequenze oligonucleotidiche a cui si lega la proteina di interesse, ma di identificare i componenti proteici presenti nel complesso proteine-DNA immunoprecipitato, mediante l'impiego di metodologie di spettrometria di massa tandem accoppiate a cromatografia capillare a fase inversa (nano-LCMSMS).

Questo nuovo approccio sperimentale è stato impiegato nello studio del fattore trascrizionale Cbx 7 (Chromobox protein homolog 7). E' stato dimostrato che il gene *cbx7* è drasticamente down-regolato in carcinoma tiroideo ed altri tumori gravi. Inoltre, il livello d'espressione di Cbx7 in tumori tiroidei risulta inversamente relazionato con il grado di malignità della neoplasia.

✓ Messa a punto di una strategia di RNA Chromatography Affinity

Lo scopo di questo lavoro è stato quello di identificare le eventuali proteine che legano il 3'UTR di PC-1 (meglio conosciuta come ENPP1), proteina che causa resistenza all'insulina, contribuendo alla modulazione del suo mRNA, e che potrebbero quindi essere la causa della sua over-espressione e del suo ruolo nell'insulino-resistenza e nelle complicanze correlate. Come approccio sperimentale si è ricorso all'utilizzo della Cromatografia di affinità utilizzando non una sonda di DNA, come prevede la metodologia classica, ma utilizzando una sequenza di RNA.

Il compito più impegnativo in questo tipo di esperimenti è stato quello di mettere a punto delle procedure opportune, sia in *vivo* che in *vitro*, che ci consentissero di isolare ed identificare i complessi proteici interagenti con geni implicati in diverse patologie.

Bibliografia

1. International Human Genome Sequencing Consortium: "Initial sequencing and analysis of the human genome." *Nature* 409, 860-922, 2001
2. Venter C. J. et al: "The Sequence of the Human Genome." *Science* 291, 1304-1351, 2001
3. Carmen L. de Hoog and Matthias Mann. PROTEOMICS. *Annu. Rev. Genomics Hum. Genet.* 5, 267-93, 2004.
4. Hochstrasser DF, Sanchez JC, Appel RD. Proteomics and its trends facing nature's complexity. *Proteomics*, 2, 807-812, 2002.
5. Dahlberg AE. Ribosome structure. The ribosome in action. *Science* 292, 868-69, 2001.
6. Kobayashi Y, Weigel D. Move on up, it's time for change-mobile signals controlling photoperiod-dependent flowering. *Genes Dev.*, 21, 2371-2384, 2007.
7. Qiu Z, Ghosh A. A calcium-dependent switch in a CREST-BRG1 complex regulates activity-dependent gene expression. *Neuron*, 60, 775-787, 2008.
8. Michaud GA, Snyder M. Proteomic approaches for the global analysis of proteins. *Biotechniques*, 33, 1308-16, 2002.
9. Souchelnytskyi S. Proteomics in studies of signal transduction in epithelial cells. *J Mammary Gland Biol.*, 7, 359 - 371, 2002.
10. de la Fuente van Bentem S, Mentzen WI, de la Fuente A, Hirt H. Towards functional phosphoproteomics by mapping differential phosphorylation events in signalling networks. *Proteomics*, 8, 4453-4465, 2008.
11. Lecaudey V, Cakan-Akdogan G, Norton WH, Gilmour D. Dynamic Fgf signaling couples morphogenesis and migration in the zebrafish lateral line primordium. *Development*, 135, 2695-705, 2008.
12. Pfeifer AC, Timmer J, Klingmüller U. Systems biology of JAK/STAT signalling. *Assays Biochem*, 45, 109-20, 2008.
13. Zak DE, Aderem A. Systems biology of innate immunity. *Immunol Rev.*, 227, 264-282, 2009.
14. Schlieker CD, Van der Veen AG, Damon JR, Spooner E, Ploegh HL. A functional proteomics approach links the ubiquitin-related modifier Urm1 to a tRNA modification pathway. *Proc Natl Acad Sci U S A.*, 105, 18255-18260, 2008.
15. Souza TA, Chen X, Guo Y, Sava P, Zhang J, Hill JJ, Yaworsky PJ, Qiu Y. Mol. Proteomic identification and functional validation of activins and bone morphogenetic protein 11 as candidate novel muscle mass regulators. *Endocrinol*, 22, 2689-2702, 2008.

16. Brown KA, Ham AJ, Clark CN, Meller N, Law BK, Chytil A, Cheng N, Pietenpol JA, Moses HL. Identification of novel Smad2 and Smad3 associated proteins in response to TGF-beta1. *J Cell Biochem*,105, 596-611, 2008.
17. Lance M. Hellman and Michael G. Fried. Electrophoretic Mobility Shift Assay (EMSA) for Detecting Protein-Nucleic Acid Interactions. *Nat Protoc.*; 2(8), 1849–1861, 2007
18. Bipasha D, Sameer T, Shruti K, Mainak C, Sahil G, Chanchal M, Vibha R. DNA–protein interactions: methods for detection and analysis. *Mol Cell Biochem*. 365, 279–299, 2012.
19. Takuo O, Sumiyo E, Masahiko H, Shozo K, Hideo A. Design of a Fluorescent Electrophoretic Mobility Shift Assay Improved for a Quantitative and Multiple Analysis of Protein-DNA Complexes. *Biosci. Biotechnol Biochem*, 66, 2732-2734, 2002.
20. Russo A, Cirulli C, Amoresano A, Pucci P, Pietropaolo C, Russo G. Cis-acting sequences and trans-acting factors in the localization of mRNA for mitochondrial ribosomal proteins. *Biochim Biophys Acta*, 1779, 820-829, 2008.
21. Lin JY, Li ML, Huang PN, Chien KY, Horng JT, Shih SR. Heterogeneous nuclear ribonuclear protein K interacts with the enterovirus 71 5' untranslated region and participates in virus replication. *J Gen Virol*, 89, 2540-2549, 2008.
22. Medugno L, Costanzo P, Lupo A, Monti M, Florio F, Pucci P, Izzo P. A novel zinc finger transcriptional repressor, ZNF224, interacts with the negative regulatory element (AldA-NRE) and inhibits gene expression. *FEBS Lett*, 534, 93-100, 2003.
23. Hall-Pogar T, Liang S, Hague LK, Lutz CS. Specific trans-acting proteins interact with auxiliary RNA polyadenylation elements in the COX-2 3'-UTR. *RNA*, 13, 1103-15, 2007.
24. Klein T, P Geurink P, S Overkleeft H, K Kauffman H, Bischoff R. Functional Proteomics on Zinc-Dependent Metalloproteinases using Inhibitor Probes. *Chem. Med Chem*. 2008 Dec 12.
25. Rix U, Hantschel O, Dürnberger G et al. Chemical proteomic profiles of the BCR-ABL inhibitors imatinib, nilotinib, and dasatinib reveal novel kinase and nonkinase targets. *Blood*, 110, 4055-4063, 2007.
26. Rensing Rix LL, Rix U, Colinge J, Hantschel O, Bennett KL, Stranzl T, Müller A, Baumgartner C, Valent P, Augustin M, Till JH, Superti-Furga G. Global target profile of the kinase inhibitor bosutinib in primary chronic myeloid leukemia cells. *Leukemia*. 2008 Nov 27.
27. Aye TT, Mohammed S, Toorn HW, Veen TA, Heyden MA, Scholten A, Heck AJ. Selectivity in enrichment of PKA RI and RII isoforms and their interactors using modified cAMP affinity resins. *Mol Cell Proteomics*. 2008 Dec 31.

28. Lambert JP, Mitchell L, Rudner A, Baetz K and Figeys D., 2008.
29. Schuch et al. Site-specific chromatin immunoprecipitation a selective method to individually analyze neighboring transcription factor binding sites in vivo. BMC Research Notes, 2012.
30. Kann MG. Protein interactions and disease: computational approaches to uncover the etiology of diseases. Brief Bioinform, Sep;8(5), 333-46, 2007.
31. Cottrell JS. Protein identification by peptide mass fingerprinting. Pept Res, 7(3), 115-124, 1994.
32. Yates JR. Database searching using mass spectrometry data. Electrophoresis, 19(6), 893-900, 1998.
33. Godovac-Zimmermann J, Brown LR. Perspective for mass spectrometry and functional proteomics. Mass Spectrom Rev, 20, 1-57, 2001.
34. Mann M, Hendrickson RC, Pandey A. Analysis of proteins and proteomes by mass spectrometry. Annu. Rev. Biochem, 70, 437-73, 2001.
35. Aebersold R. A mass spectrometric journey into protein and proteome research. J Am Soc Mass Spectrom, 14(7), 685-95, 2003.
36. Lin D, Tabb DL, Yates JR. Large-scale protein identification using mass spectrometry. Biochim Biophys Acta, 1646(1-2),1-10, 2003.
37. Vidal M. Interactome modeling. FEBS Lett, 579:1834–1838, 2005.

2. Messa a punto della metodologia EMSA/MS-MS per lo studio degli interattori del promotore del gene *cftr*.

2.1 Fibrosi Cistica

La Fibrosi Cistica (FC) è la malattia genetica ad esito letale più frequente nella popolazione caucasica, è trasmessa con meccanismo autosomico recessivo e ne è affetto un neonato ogni 2500-2700 nati vivi; la frequenza dei portatori sani è di 1:25 [1], e rende questi individui immuni dal contrarre il colera. Questo fenomeno spiega perché la frequenza dei portatori della mutazione sia così elevata. Trattandosi di una malattia autosomica recessiva, gli individui affetti da FC devono presentare mutazioni su entrambi gli alleli. Gli eterozigoti recanti un allele mutato ed uno sano non risulteranno essere affetti dalla malattia. Se invece l'individuo eredita da ciascun genitore un allele mutato per la FC, anche con diverse mutazioni, in quel caso si parlerà di eterozigote composto e manifesterà la malattia.

La malattia dipende dalle alterazioni di una glicoproteina di membrana di 1480 amminoacidi, espressa e funzionante nella porzione apicale della membrana delle cellule epiteliali, denominata Cystic Fibrosis Transmembrane Conductance Regulator (CFTR), che svolge funzioni di canale ionico del cloro (Fig. 2.1).

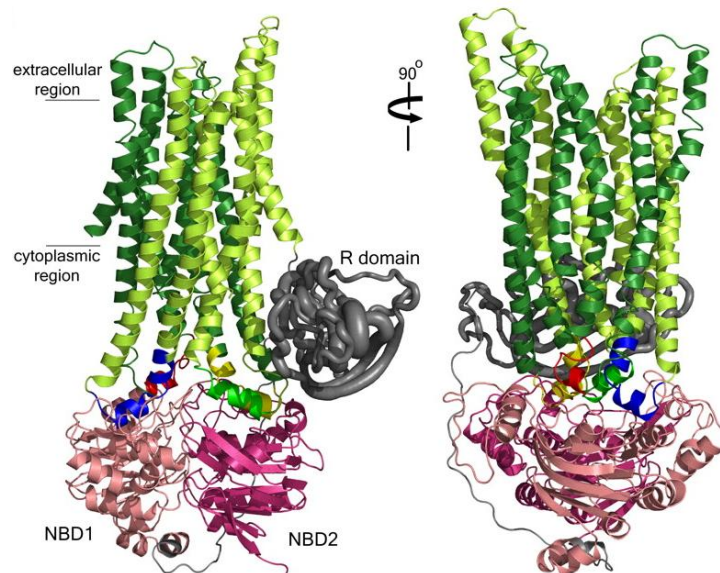


Figura 2.1 Struttura tridimensionale della glicoproteina CFTR

Nei soggetti in cui la proteina CFTR non presenta alterazioni, i secreti salini vengono adeguatamente idratati e fluidificati poiché gli ioni cloro e sodio si associano al transito di acqua. Nei soggetti in cui la proteina è alterata, le secrezioni diventano più dense, causando i sintomi sistemici della malattia [2]. Le secrezioni dense ostruiscono i bronchi favorendo lo

sviluppo di infezioni, ostruiscono i dotti pancreatici, impedendo agli enzimi digestivi di giungere all'intestino, provocando malassorbimento e deficit di accrescimento; a livello epatico rendono la bile densa, danneggiando il fegato. Inoltre il riassorbimento di ioni cloro e sodio è compromesso anche a livello delle ghiandole sudoripare. I pazienti CF presentano uno squilibrio elettrolitico, a cui è imputabile l'eccessiva quantità di NaCl presente nel sudore. Questa manifestazione della patologia viene utilizzata a scopo diagnostico: è infatti possibile effettuare un test del sudore che prevede la stimolazione delle ghiandole con pilocarpina.

Il gene che codifica per la proteina CFTR, mappato nel 1989 [3], e sequenziato nel 1991 [4], è localizzato sul cromosoma 7, in posizione q31.2, ed è composto da 27 esoni (Fig. 2.2).

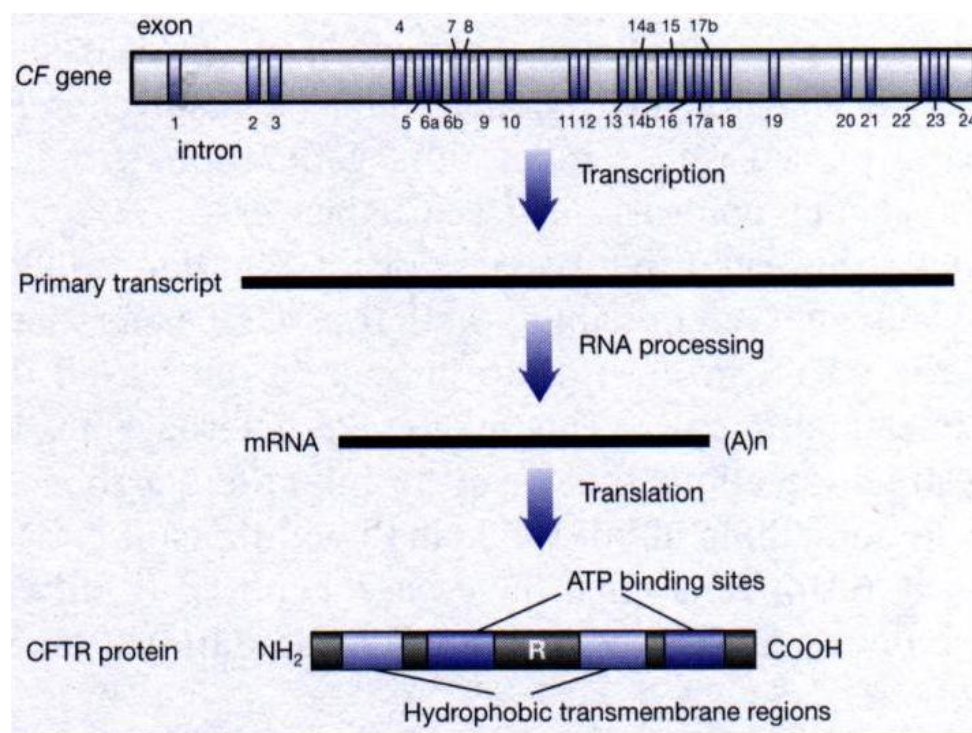


Figura 2.2 Schema riassuntivo della formazione della proteina CFTR a partire dal suo gene.

Ad oggi, sono state identificate oltre 1500 diverse mutazioni causative di CF: alcune sono diffuse ubiquitariamente (es. $\Delta F508del$, N1303K), altre sono specifiche di ristretti ambiti etnico-geografici e la maggior parte sono "private", cioè confinate a gruppi geografici o familiari [5].

In base al tipo di mutazione si hanno diversi effetti sulla proteina CFTR: alcune mutazioni fanno sì che essa non venga prodotta affatto, altre permettono che venga prodotta una proteina poco funzionante o ridotta in quantità. Ad oggi, non si conosce l'effetto ultimo di tutte le mutazioni sulla proteina e quindi sulle loro conseguenze cliniche.

Pur essendo molto eterogenee, le mutazioni CF sono state raggruppate in 5 classi in base all'alterazione funzionale della proteina CFTR di cui sono responsabili [6].

Le mutazioni appartenenti alle classi I, II e III alterano maggiormente il destino della proteina, non consentendone affatto la produzione (classe I) o producendo una proteina molto difettosa (classe II e III); quelle di classe IV consentono la sintesi di una proteina difettosa ma capace di svolgere seppure in piccolissima misura la sua funzione; quelle di classe V permettono la produzione di una certa quota, anche se piccola, di proteina normale. Non si conoscono ancora sufficientemente le relazioni fra le mutazioni e l'interessamento di altri organi oltre al pancreas, quali polmoni, fegato, intestino, apparato riproduttivo. In particolare, non sono ancora noti gli effetti delle mutazioni sul polmone, sebbene i problemi più seri siano in genere a carico di quest'organo, in quanto la viscosità del muco rende difficoltosa la rimozione delle particelle inalate da parte delle cilia bronchiali, favorendo così il manifestarsi di infezioni.

Si è visto che gli effetti genetici sul singolo malato dipendono, oltre che dall'interazione nella coppia di mutazioni CFTR che costituiscono il genotipo, nel caso che una mutazione sia di una classe e l'altra di un'altra, anche dall'influsso dei geni modificatori delle mutazioni del gene CFTR, di recente scoperti. Questi geni possono aggravare o alleggerire l'effetto delle mutazioni [6].

A causa del differente impatto sulla produzione della proteina, le numerose mutazioni causative fino ad ora riscontrate in CFTR sono state oggetto di studio da parte di molti gruppi, al fine di generare una possibile correlazione tra il difetto molecolare del gene CFTR e l'eterogeneità clinica della malattia.

E' noto, infatti, che la Fibrosi Cistica è una malattia cronica, evolutiva, con prognosi severa, che si manifesta, però, in maniera variabile da paziente a paziente, con un decorso clinico non sempre prevedibile sulla base degli esami clinici e molecolari. L'analisi molecolare di CFTR è di fondamentale importanza per identificare i portatori sani della malattia, del tutto asintomatici. Per questo motivo essa è oggetto di continui studi tesi a migliorare la sensibilità e l'accuratezza dei test genetici. Vengono sottoposti ad analisi molecolare del gene tutti quei soggetti a maggior rischio di essere portatori (consanguinei di un affetto) e i partner di soggetti già risultati portatori sani di mutazioni del gene CFTR. Da studi condotti

è emerso che nel 92% dei casi il manifestarsi della patologia è da attribuirsi alla presenza di una mutazione presente in una delle regioni codificanti di CFTR.

Nel restante 8%, invece, non è stata riscontrata alcuna mutazione correlabile con l'insorgerenza della patologia.

Per spiegare dove possano essere localizzate le mutazioni ancora non identificate sono state formulate varie ipotesi, tra cui:

a) La presenza di grossi riarrangiamenti del gene CFTR [7], che, se presenti in eterozigosi, non sono riscontrabili mediante le tecniche di scanning genomico.

b) La presenza di mutazioni localizzate in regioni non ancora analizzate come il promotore, gli introni o altre regioni non codificanti, le cui mutazioni potrebbero alterare siti coinvolti nella regolazione dell'espressione del gene CFTR.

c) La presenza di fattori genetici che, ereditati in maniera indipendente da CFTR, potrebbero modulare l'espressione fenotipica della malattia [8].

Per quanto concerne l'eventuale presenza di riarrangiamenti genici, negli ultimi anni, sono state messe a punto complesse tecniche basate sulla PCR quantitativa e sulla elettroforesi capillare, che permettono l'identificazione di grosse delezioni e/o di inserzioni; si tratta di indagini molto specialistiche ancora poco utilizzate, che consentono, in tempi relativamente rapidi, l'identificazione di mutazioni che in passato sfuggivano nel corso dell'analisi molecolare del gene [9]. Lo studio di mutazioni in regioni non tradotte o dei fattori genici esterni, costituiscono, ad oggi i due filoni principali di ricerca per la Fibrosi Cistica, in quanto una delucidazione di tutti i meccanismi con cui diversi fattori, solo in parte noti, contribuiscono alla corretta espressione e regolazione del gene CFTR rappresentano la base conoscitiva indispensabile per una più efficace e corretta diagnosi di CF. Infatti l'insorgerenza della malattia, potrebbe essere dovuta al malfunzionamento o all'assenza di uno di questi fattori. La conoscenza dettagliata dei meccanismi molecolari dell'espressione di CFTR potrebbe quindi portare al miglioramento della diagnosi molecolare, alla predizione del fenotipo clinico e alla creazione di terapie innovative per la Fibrosi Cistica. Dai dati presenti in letteratura risultano ancora poco chiari i meccanismi con cui CFTR viene espresso e risulta evidente che la maggior parte degli elementi che conferiscono una regolazione temporale o tessuto-specifica all'espressione risiedono in parte nella regione del promotore.

2.1.1 La regolazione dell'espressione del gene cfr

I meccanismi che regolano l'espressione del gene *cftr* sono stati oggetto di numerosi studi negli ultimi anni proprio perché è possibile che in una parte dei pazienti CF la malattia sia causata da alterazioni di questi stessi meccanismi. Questi meccanismi si possono suddividere in quelli di controllo tissutale e temporale dell'espressione genica.

E' stato osservato che le regioni maggiormente coinvolte nella regolazione dell'espressione di CFTR sono il promotore, i siti di inizio della trascrizione, l'elemento di regolazione CCAAT [9]. I primi meccanismi ad essere analizzati sono stati quelli coinvolti nella regolazione mediata da AMPciclico [10], mentre negli ultimi anni si sta cercando di far luce sul ruolo della regolazione ormonale [11] ed alcune regioni del gene, come le DHS, siti ipersensibili alla DNasi I potenzialmente coinvolti nella regolazione dell'espressione genica essendo regioni protette dal nucleo soma [12].

L'analisi della sequenza genomica a monte dell'esone 1 del gene CFTR ha rivelato che non è presente l'elemento TATA box, ma che ci sono siti multipli di inizio della trascrizione e molti potenziali siti di legame di proteine Sp1 e AP1. Inoltre, in questa regione sono stati identificati un elemento responsivo al cAMP (CRE), un sito di legame a C/EBP e molti elementi responsivi a glucocorticoidi.

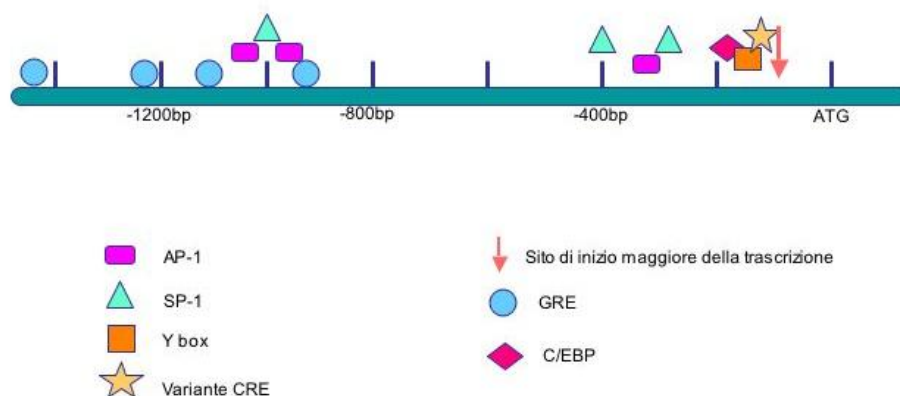


Figura 2.3 La regione del promotore di CFTR

La regione promotrice di CFTR è stata identificata mediante esperimenti di delezione del promotore effettuati utilizzando costrutti di fusione con un gene reporter: sono stati identificati due elementi positivi che agiscono in cis, uno dei quali corrisponde a una piccola sequenza del promotore da -226 a +98, e un elemento negativo (AP-1). Al 5' ci sono due

sequenze ricche di GC da -273 a -190 che assomigliano a siti di legame per la proteina Sp1: l'eliminazione di questi siti rende il promotore inattivo. L'elemento negativo è stato localizzato subito a monte della sequenza minima del promotore da -345 a -277 ed elimina completamente l'espressione quando è incluso nel costrutto del gene reporter [13, 14].

In linee cellulari, in cui CFTR presenta bassi livelli di espressione, è stata evidenziata una regione CpG sensibile alla metilazione e al 5' sono stati descritti elementi regolatori che agiscono in maniera cooperativa [15]. Sempre al 5', è stato identificato un elemento regolatorio che diminuisce l'attività del promotore CFTR dopo il legame con NF-kB[16].

Un'altra regione ben caratterizzata è quella dell'elemento invertito CCAAT (Y box) localizzato tra -132 e -119 a monte del sito di inizio della traduzione (Fig. 2.4), poichè la sua rimozione riduce i livelli del trascritto di CFTR [17].

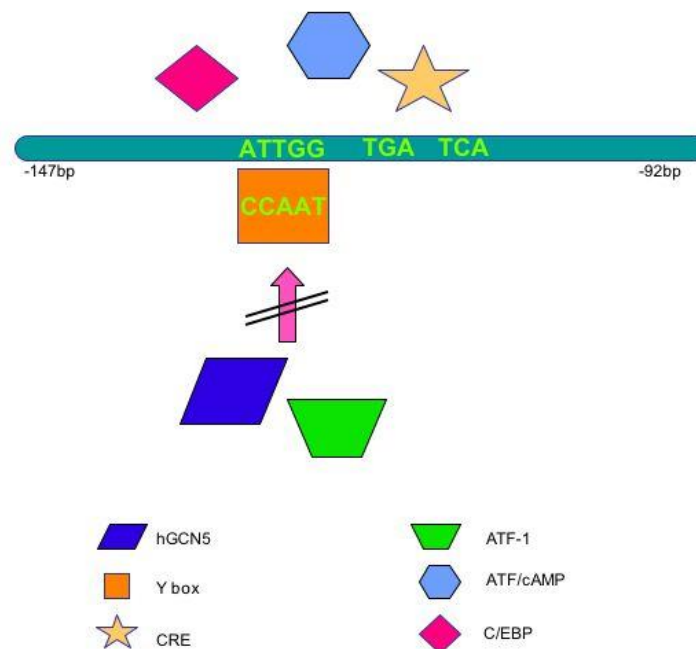


Figura 2.4 Esperimenti condotti con Electrophoretic mobility shift assays (EMSA) hanno permesso di identificare che C/EBP e ATF/cAMP formano complessi con CCAAT necessari per l'attivazione della trascrizione mediata da cAMP. La proteina CCAAT reprime la trascrizione di CFTR poichè impedisce il legame degli attivatori della trascrizione hGCN5 e ATF-1. La proteina CDP recluta l'istone deacetilasi, modificando la struttura della cromatina in uno stato inattivo per la trascrizione e, inoltre, interagisce con la proteina legante CREB e con il fattore associato alla proteina legante CREB, che hanno attività istone acetiltransferasica.

L'importanza funzionale degli elementi CRE e CCAAT è supportata dalla conservazione di una regione di 34 bp che comprende questi elementi nel gene CFTR dell'uomo, del topo, del ratto [18], della pecora [19], del maiale e della mucca [20].

Da quanto descritto finora si può stabilire che la maggior parte degli elementi che conferiscono una regolazione quantitativa e tessuto-specifica all'espressione di CFTR risiedono nella regione del promotore.

Poco ancora si conosce però su come l'espressione di *cftr* possa essere regolata dalle 6000 coppie di basi che sono a monte del sito di inizio della trascrizione. L'interesse per questa regione nasce dal fatto che in pazienti CF in cui non si riscontrava nessuna mutazione all'interno delle canoniche sequenze codificanti è stata riscontrata, invece, una mutazione all'interno di questa regione. In particolare all'interno di questa sequenza oligonucleotidica sono state identificate ben 23 mutazioni, 9 delle quali identificate solo recentemente [21] (Fig.2.5).

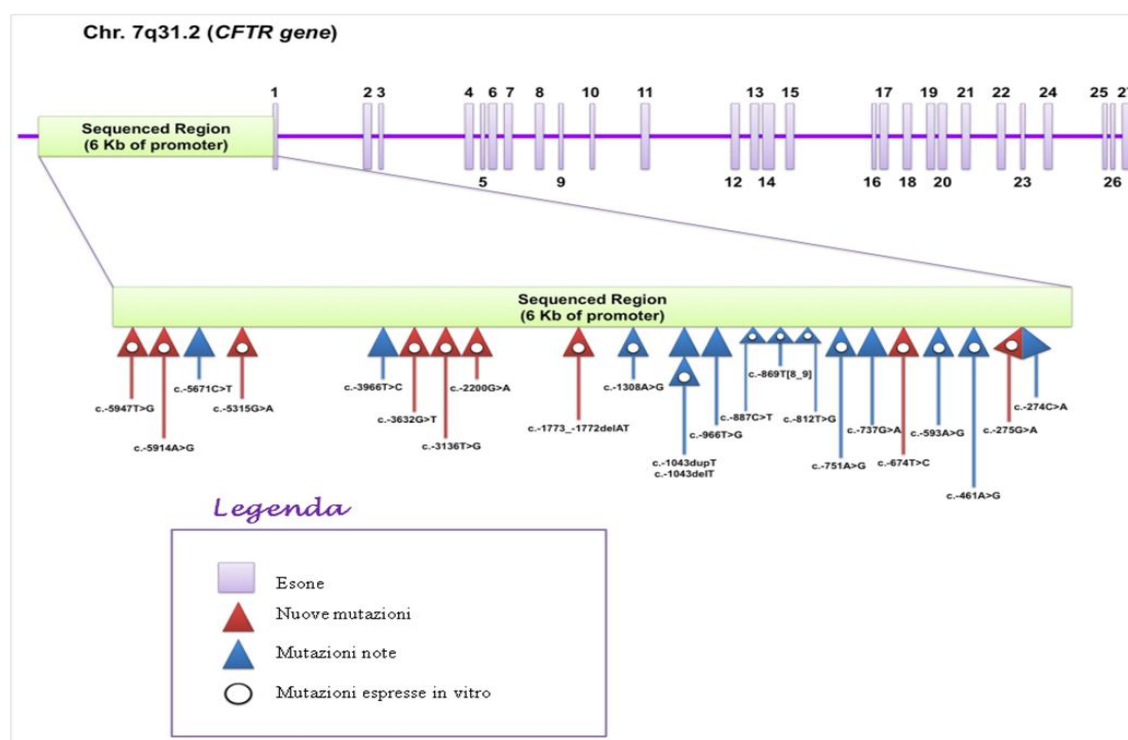


Figura 2.5 Mutazioni identificate nella regione a monte del sito di inizio di trascrizione di *cftr*.

Queste mutazioni potrebbero alterare il legame di specifici complessi proteici, alterando così l'espressione del gene *cftr*. La conoscenza degli interattori di questa particolare regione genica risulta quindi uno strumento estremamente importante per spiegare l'insorgenza della patologia laddove non siano presenti mutazioni all'interno delle regioni codificanti del gene. In particolare, questo progetto di ricerca, ha previsto l'analisi di due regioni, una a -1640 e l'altra a -3500 coppie di basi dal sito di inizio della trascrizione, contenenti in un caso una

delezione di due basi azotate e nell'altra una mutazione puntiforme, sia nella loro forma wild-type (wt) che in quella mutata (mut) (Fig.2.6).

-1640 forma WT: GTGAGTTCAATCACATGTCTGGGAAAAGTC

-1640 forma MUT: GTGAGTTCAATCACGTCTGGGAAAAGTC

-3500 forma WT: TGGCCCACTGTTGAAAGAGCTGTGGCTGTTC

-3500 forma MUT: TGGCCCACTGTTGAAATAGCTGTGGCTGTTC

Figura 2.6 Sequenze wild-type e mutate utilizzate nel presente progetto.

2.2 Materiali e Metodi

2.2.1 Materiali

Il detergente usato per l'estrazione proteica è il Triton X-100 della J.T. Baker.

I prodotti della SIGMA utilizzati sono:

il ditionitrotolo (DTT);

la tripsina ;

Tutti i materiali utilizzati per le tecniche elettroforetiche, 30% Acrilammide/ Bis, N,N,N',N'-tetrametilendiammina (TEMED), , ammoniopersolfato (APS).

La miscela di sequenze oligonucleotidiche a peso molecolare noto “5X Green GoTaq” usata come standard nella migrazione elettroforetica è della PROMEGA.

Il tampone usato per la corsa elettroforetica Tris Borate EDTA (TBE) è della MP Biomedicals.

La iodoacetammide (IAM) e l'ammonio bicarbonato sono della FLUKA.

Gli inibitori di proteasi sono della ROCHE.

2.2.2 Metodi

2.2.2.1 Colture cellulare

Le cellule utilizzate negli esperimenti eseguiti durante questo lavoro di tesi sono state fornite dalla facility di colture cellulari del CEINGE.

Le due linee cellulari utilizzate sono: Human hepatocellular carcinoma cells (HepG2) ed epithelioid carcinoma pancreatic ducts cells (Panc1). Gli studi funzionali sono stati condotti su due diverse linee cellulari in quanto cellule diverse esprimono differenti livelli di *cftr*, questo quindi influenza il pattern di interattori. L'espressione di *cftr*, nelle due linee cellulari di interesse, è stata valutata mediante analisi PCR e i risultati sono stati normalizzati facendo riferimento all' mRNA di GAPDH (Fig. 2.7)

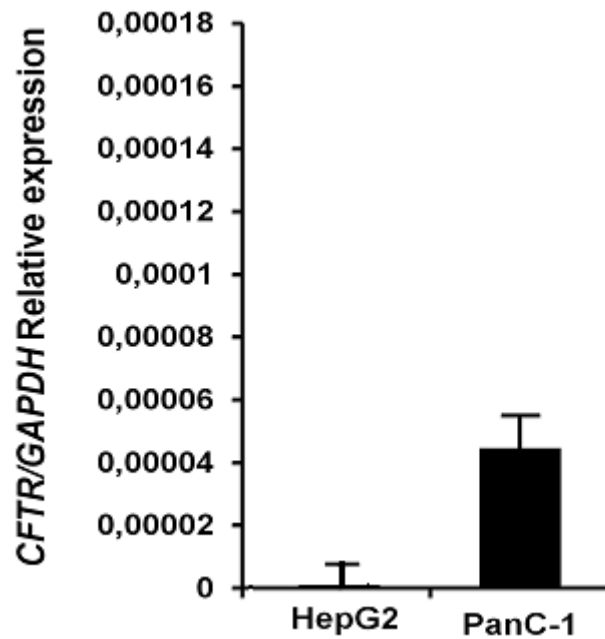


Figura 2.7 Livelli di espressione di *cftr* nelle due linee cellulari utilizzate nel presente progetto.

2.2.2.2 Lisi cellulare frazionata

Al fine di ottenere una separazione dei complessi proteici nucleari, da quelli citosolici, è stata effettuata una lisi cellulare frazionata. A tal proposito sono stati preparati due tamponi di lisi, uno a basso sale (10 mM HEPES / 10mM KCl / 0,1mM EDTA / 0,1 mM EGTA / 0,5 mM PMSF / NP-40 0,5% ed inibitori di proteasi) ed uno a concentrazione salina più elevata (20 mM. HEPES / 150 mM NaCl / 1mM EDTA / 1 mM EGTA / 0,5 mM PMSF ed inibitori di proteasi). Il primo punto del protocollo di lisi ha previsto l'aggiunta del tampone a basso sale, in rapporto di 1:40 rispetto al pellet cellulare, che è stato, poi, immediatamente messo in ghiaccio per 15 minuti. Al termine è stato aggiunto, direttamente sul pellet, dell'NP-40 la cui concentrazione finale è stata pari allo 0,5%. Il pellet, così trattato, è stato dapprima messo su ruota, a velocità moderata, per 5 minuti a 4 °C e poi centrifugato a 2000 rpm per altri 5 minuti. Dal surnatante è stato recuperato l'estratto. Al fine di eliminare eventuali tracce di proteine citosoliche dal pellet nucleare, quest'ultimo è stato lavato almeno due volte nel tampone di lisi a basso sale. A questo punto si è proceduto alla lisi dei nuclei, aggiungendo al pellet il tampone ad alta concentrazione salina in rapporto 1:5 in volume. Il tutto è stato messo in ghiaccio per 5 minuti alla fine dei quali si è proceduto con 15 minuti di vigorosa agitazione e successive centrifugazione per 30 minuti a 13000 rpm. Alla fine si recupera l'estratto nucleare. La concentrazione degli estratti proteici, sia quello nucleare che

quello citosolico, è stata determinata attraverso il saggio di Bradford, utilizzando l'albumina di siero bovina (BSA) come standard. Le assorbanze sono state misurate a 595 nm. Sono state determinate 10 diverse concentrazioni di BSA. Tali punti sono stati utilizzati per costruire una retta di taratura.

2.2.2.3 Western Blotting

L'esito della lisi cellulare frazionata è stato verificato preliminarmente con un esperimento di western blotting. Questa tecnica di immunorivelazione ha permesso di verificare la separazione delle proteine citosoliche da quelle nucleari.

Tale tecnica ha previsto un frazionamento elettroforetico del campione, seguito da un trasferimento su membrana di nitrocellulosa delle proteine intrappolate nelle maglie del gel e il riconoscimento delle proteine d'interesse mediante l'impiego di anticorpi specifici. Il gel (di dimensione 10x10 cm e di spessore 1.5 mm) viene posto a contatto con una membrana di nitrocellulosa all'interno di un sandwich composto da 2 fogli di carta di 3 mm (10x8 cm) e 2 spugnette del sistema di trasferimento (Bio-Rad Protean) ed immerso nel tampone di trasferimento (Tris-glicina). Il trasferimento è stato effettuato a voltaggio costante a 50 V per 2 ore a 4°C. Dopo il trasferimento la membrana è stata immersa in una soluzione di latte no fat al 5% in PBS per tutta la notte a 4°C. La porzione della membrana ad alti pesi molecolari, da 250 a 50 KDa, e quella da 37 a 10 KDa sono state incubate con gli specifici anticorpi primari, in modo tale da far riconoscere sia le proteine nucleari che quelle citosoliche, in una soluzione di latte 5% in PBS 1X più Tween 0.1%; l'incubazione è seguita da tre lavaggi con una soluzione di PBS più Tween 0.1%, al termine dei quali fa seguito l'incubazione con gli anticorpi secondari, che riconoscono gli anticorpi primari. La rivelazione è stata effettuata mediante una reazione di chemiluminescenza (ECL) utilizzando reagenti specifici; i filtri sono poi impiegati per impressionare una lastra fotografica.

Come anticorpi primari sono stati utilizzati Lamin B, per le proteine nucleari, diluito 1:500 in una soluzione di latte 1% in PBS più Tween 0.1%, e β -actina, per le proteine citosoliche, diluito 1:1000 in una soluzione di latte 1% in PBS più Tween 0.1%.

2.2.2.4 Preparazione delle sonde oligonucleotidiche

Le sonde oligonucleotidiche di interesse sono state fornite dalla Facility di Sintesi Oligonucleotidica del CEINGE. Tutti i probe sono stati marcati all'estremità 5' con la Cy5.

2.2.2.4.1 Quantificazione delle sonde oligonucleotidiche

La quantificazione delle sonde di DNA avviene tramite corsa elettroforetica oppure attraverso quantificazione al nanodrop. Per la corsa elettroforetica si prepara, in un eppendorf, il campione oligonucleotidico annilato addizionato con xilene cianolo o blu di bromofenolo. Il campione viene poi caricato su gel di agarosio al 1%. In un pozzetto adiacente viene caricata una miscela contenente frammenti di DNA a concentrazione nota (STANDARD). Per confronto si estrapola il valore di concentrazione del campione in esame. Per una quantificazione più precisa si effettua un'ulteriore quantificazione allo spettrofotometro nanodrop.

2.2.2.4.2 Annealing delle sonde oligonucleotidiche

Le sonde oligonucleotidiche sono state risospese in 200 µl di acqua sterile. In eppendorf sono state introdotte le stesse quantità di sonda forward e reverse, la miscela è stata riscaldata per 10 minuti a 95°C. I campioni sono poi stati lasciati raffreddare, fino ad arrivare a temperatura ambiente.

2.2.2.5 Gel Nativo

I gel nativi, detti anche gel non denaturanti, consentono di separare non solo i complessi DNA- proteine, ma anche complessi RNA- proteine, proteine-peptidi, etc...

La miscela di molecole di DNA e di complessi Proteine-DNA, nel campo elettrico, migra verso il polo positivo, e la mobilità elettroforetica dei singoli componenti varia in base al loro diverso peso molecolare: le molecole a maggior peso molecolare migrano più lentamente rispetto a quelle a minor peso molecolare. Il risultato della corsa elettroforetica consiste in una serie di bande. Sul gel viene caricata anche una miscela di frammenti di DNA a peso molecolare noto (MARKER) che servirà per il calcolo della lunghezza in basi delle bande incognite

I gel nativi possono essere sia in poliacrilammide-TBE (Tris Borate EDTA) che in agarosio-TAE (Tris Acetato). Per i gel di agarosio bisogna ricordare che prima della polimerizzazione è necessario aggiungere un colorante, il Bromuro di Etidio (EtBr), che consentirà al termine della corsa la visualizzazione delle bande.

I gel in poliacrilammide sono usati per separare frammenti di DNA < 500 bp e hanno un'elevata risoluzione, ma sono più complicati da preparare e più difficili da maneggiare

rispetto ai gel di agarosio. Quest'ultimi, invece, sono semplici da preparare e sono tipicamente usati per separare frammenti di dimensioni variabili da poche centinaia di basi a 20 kb. Essi sono largamente utilizzati per la maggior parte delle analisi routinarie su DNA.

In questo tipo di elettroforesi come tamponi da corsa vengono utilizzati TBE e TAE, in quanto entrambi hanno pH basico, ciò rende ancora più negativo il gruppo fosfato del DNA aiutandolo, quindi, nella sua migrazione verso l'anodo. Il TBE, generalmente è da preferirsi al TAE, in quanto contiene sodio borato ed EDTA che serve a chelare i cationi bivalenti, ed impedire che possano attivare le numerose nucleasi che potrebbero digerire il DNA.

La miscela da separare viene caricata sul gel con l'aggiunta di coloranti (blu di bromofenolo, xilene cianolo), ciò permette di seguire visivamente il caricamento del campione nel pozzetto e la migrazione del campione. Inoltre viene aggiunto del glicerolo per appesantire il campione e far sì che si compatti all'interno del pozzetto.

2.2.2.5.1 Tecniche di visualizzazione delle bande da gel

Una volta terminata la corsa elettroforetica le bande del DNA devono essere visualizzate all'interno della matrice del gel. Esistono differenti coloranti che si legano agli acidi nucleici con meccanismi diversi e che sono in grado di essere rivelati mediante sorgenti luminose di diverso tipo (raggi ultravioletti, luce visibile, etc.). Per il gel di agarosio si utilizza abitualmente il bromuro di etidio (EtBr), un agente intercalante che si lega al DNA e può essere rivelato con un transilluminatore UV. La molecola del bromuro di etidio è piatta e molto simile a quella delle basi azotate per cui si intercala tra una base azotata e quella successiva nella doppia elica e inoltre emette fluorescenza arancione se eccitata da luce ultravioletta con lunghezza d'onda compresa tra 254 e 306 nm. Anche per il gel di poliacrilammide è possibile effettuare una colorazione con bromuro di etidio oppure si può ricorrere alla colorazione argentea, metodo circa 100 volte più sensibile, anche se necessita di tempi operativi più lunghi. In maniera alternativa, laddove è possibile, si può optare per una marcatura del probe oligonucleotidico con molecole fluorescenti (Cy5, Cy3). La visualizzazione delle bande avviene irradiando il gel alla adeguata lunghezza d'onda.

2.2.2.6 Idrolisi in situ

2.2.2.6.1 Rimozione delle bande dal gel

Dopo la separazione elettroforetica, le bande di interesse sono state tagliate dal gel con un bisturi, sminuzzate e poste in eppendorf. Sono state quindi trattate con 50 µl di acetonitrile; dopo 15 minuti il solvente è stato rimosso e sono stati aggiunti 50 µl di una soluzione di NH_4HCO_3 50 mM, pH 8.0. Dopo 5 minuti le particelle di gel sono state centrifugate, il solvente rimosso e quindi eliminato.

2.2.2.6.2 Riduzione ed alchilazione delle proteine in situ

I campioni proteici presenti nelle bandine di gel tagliate sono stati sottoposti ad un trattamento con agenti riducenti ed, infine, le cisteine presenti sono state alchilate. A tale scopo le bande di gel sono state trattate con 50 µl di una soluzione di ditione 10 mM in NH_4HCO_3 50 mM ed incubate per 45 minuti a 56° C. Al termine della riduzione, l'eccesso di soluzione riducente è stato rimosso e i pezzetti di gel sono stati disidratati con acetonitrile; successivamente essi sono reidratati con una soluzione NH_4HCO_3 50 mM, contenente iodoacetammide 55 mM; la reazione è stata condotta al buio e a temperatura ambiente, al fine di alchilare i gruppi tiolici presenti. Passati 30 minuti la iodoacetammide in eccesso è stata rimossa e le particelle sono state nuovamente trattate con acetonitrile per disidratarle. Esse sono state infine sottoposte a lavaggi NH_4HCO_3 50 mM/ acetonitrile come precedentemente descritto.

2.2.2.6.3 Digestione in situ delle proteine presenti nel gel

Le bande di gel disidratate sono state trattate con 10 µl di una soluzione di tripsina (10 ng/µl) in AMBIC 50 mM pH 8 e poste a 4°C per 1 ora. A questo punto, è stato aggiunto NH_4HCO_3 50 mM, in un volume minimo ma tale da coprire i pezzettini di gel; le eppendorf sono quindi state poste a 37° C per 12 ore.

2.2.2.6.4 Estrazione dei peptidi

La miscela di idrolisi è stata centrifugata alla massima velocità per pochi secondi ed il surnatante contenente la miscela peptidica, raccolto, posto in eppendorf ed acidificato con 2-5 µl di TFA 20% (1 µl di TFA 20% per ogni µl di surnatante).

Ai rimanenti pezzettini di gel è stata aggiunta una piccola quantità di acetonitrile, lasciando il tutto a temperatura ambiente per circa 15 minuti al fine di estrarre completamente

eventuali peptidi ancora trattenuti nelle maglie del gel. Alla fine dell'incubazione si è centrifugato e recuperato l'estratto, che riunito al primo surnatante è stato quindi liofilizzato.

2.2.2.7 Identificazione

2.2.2.7.1 Analisi mediante Matrix Assisted Laser Desorption Ionization (MALDI/MS)

Le miscele peptidiche liofilizzate sono state risospese in 10 μ l di TFA 2% e caricate negli opportuni pozzetti della piastra portacampioni dello spettrometro, mescolate ad 1 μ l di matrice (acido α -ciano-4-idrossicinnamico 10 mg/mL sciolto in una miscela di 70% acetonitrile e 30% Acido citrico 50 mM).

La miscela è stata fatta essiccare all'aria, per consentire la formazione dei cristalli di matrice in cui il campione proteico resta intrappolato. Infine, la piastra portacampioni è stata introdotta nella camera di ionizzazione dove in condizioni di alto vuoto il campione è stato bombardato mediante un raggio laser pulsato tra 1 e 10 ns. L'intervallo di m/z analizzato è stato calibrato usando standard interni aventi un opportuno valore di massa.

Gli spettri sono stati acquisiti utilizzando uno spettrometro di massa MALDI Voyager DE-PRO (Applied Biosystems) a sorgente di ionizzazione laser equipaggiato con analizzatore a tempo di volo in grado di separare gli ioni in modalità reflectron.

2.2.2.7.2 Analisi mediante LC-MS/MS

Le miscele peptidiche sono state analizzate attraverso LCMS/MS, utilizzando uno spettrometro di massa ibrido, Q-TOF (Waters), dotato di sorgente Z-spray e accoppiato on-line con un sistema di cromatografia capillare. Dopo il caricamento, la miscela peptidica (6 μ l) è stata prima concentrata e poi desalificata a 10 μ l/min su una precolonna C18 a fase inversa (Waters) usando 0.2% di acido formico come eluente (Tampone C).

Il campione è stato trasferito on-line su una colonna capillare C18 a fase inversa (75 μ m, 20 cm) e un flusso di 4 μ l/min dove è stato frazionato usando un gradiente lineare di eluente B (0.2% acido formico in 95% acetonitrile) in A (0.2% acido formico, 5% acetonitrile in acqua) dal 7 al 50% in 64 minuti.

Alternativamente l'analisi LCMS/MS è stata condotta mediante un sistema LC/MSD Trappola XCT Ultra (Agilent Technologies), equipaggiato con un sistema HPLC 1100 ed un chip cube (Agilent Technologies). Dopo il caricamento, la miscela peptidica (8 μ l in 0.2% acido formico) è stata prima concentrata e poi desalificata a 4 μ l/min su una colonna di arricchimento inversa (chip, Agilent Technologies), con 0.1% acido formico (CHOOH)

come eluente. Il campione è stato poi frazionato su una colonna capillare C18 a fase inversa (75 µm, 43 mm) ed un flusso di 200nl/min, con un gradiente lineare di B in A dal 5 al 60% in 50 min.

In entrambi i casi l'eluato è stato introdotto nello spettrometro di massa programmato nella modalità MS/MS data-dependent, dove ogni spettro di massa (intervallo 400-2000 Da/e) è stato seguito da uno o più spettri di massa tandem ottenuti a partire dagli ioni parenti più intensi eluiti in ogni frazione cromatografica.

Una rampa di collisioni energetiche tra 30 e 60 eV è applicata in base alla massa e alla carica del precursore ionico. Gli spettri di frammentazione così ottenuti sono stati analizzati con il programma Protein Lynx del Mass Lynx 4.0 per gli esperimenti effettuati sul Q-TOF e con il programma Analysis List qualora fosse stata utilizzata la Trappola. Tali programmi generano delle peak list che vengono inserite in banche dati disponibili in rete conducendo all'identificazione della proteina.

2.2.2.7.3 Identificazione mediante Mascot

I valori di massa dei peptidi ottenuti in seguito all'analisi mediante MALDI/MS-MS sono stati inseriti nella banca dati di Mascot utilizzando l'opzione Peptide Mass Fingerprint.

Il programma di ricerca mima l'esperimento calcolando i possibili valori di m/z teorici ottenuti dalla frammentazione dei peptidi, originati dalla digestione proteolitica, per ogni proteina la cui sequenza è presente nella banca dati. I valori di massa dei frammenti ritrovati nello spettro sono quindi paragonati con quelli teorici.

La ricerca viene effettuata impostando dei parametri che sono:

- Banca dati : NCBI
- Tassonomia: Homo Sapiens
- Enzima: Tripsina
- Modifiche fisse : carbossiammidometilazione
- Modifiche variabili: ossidazione delle metionine e ciclizzazione della glutammina all'N-terminale

- Errore sul valore di massa del peptide: 600 ppm
- Errore sul valore di massa dello ione frammento: 0,6 Da
- Carica degli ioni : +2, +3

Il report restituito in output dal software contiene informazioni sulle proteine identificate. Il numero di accesso della proteina alla banca dati permette di ottenere la sua sequenza, la quale viene utilizzata poi con un'altro software, Blast, che allinea sequenze proteiche

presenti in diverse banche dati. Blast consente di ricavare per ogni proteina analizzata il codice identificativo della stessa proteina relativa alla banca dati Swiss Prot. Questa operazione è stata effettuata allo scopo di ottenere quante più informazioni, funzionali e non, sulle proteine identificate, infatti la Swiss Prot è una banca dati molto annotata, ma non viene aggiornata con la stessa frequenza di NCBI, e quindi facendo riferimento solo ad essa, si correrebbe il rischio di perdere informazioni. Gli allineamenti delle sequenze proteiche sono state effettuate consultando banche dati di proteina in Homo Sapiens. Le proteine identificate nella corsia di controllo sono state sottratte a quelle identificate nella corsia del campione.

2.3 Risultati

2.3.1 Messa a punto del protocollo di EMSA/MS-MS

Lo studio dei meccanismi cellulari che coinvolgono la regione a monte del sito di inizio trascrizionale del gene *cftr* è stato affrontato utilizzando un particolare e innovativo approccio di Proteomica Funzionale. A questo proposito molti sono gli studi e gli sforzi che abbiamo condotto al fine di modificare il classico approccio EMSA. L'obiettivo principale è stato quello di variare questa metodologia al fine di isolare ed identificare, attraverso la spettrometria di massa, le proteine che interagiscono in maniera specifica con una determinata sequenza oligonucleotidica.

La metodologia EMSA è, tra le tecniche analitiche volte allo studio delle interazioni proteine-acidi nucleici, una delle più utilizzate. È stata originariamente sviluppata con lo scopo di quantificare le interazioni tra DNA e proteine (Fried & Crothers, 1981; Garner & Revzin, 1981) e da allora si è evoluta per essere adattata a scopi diversi, compresa l'individuazione e la quantificazione di interazioni RNA-proteina. In condizioni adeguate, tuttavia, l'EMSA può anche essere utilizzata per la determinazione dell'affinità di legame di una particolare proteina rispetto ad una sequenza oligonucleotidica, in studi cinetici e stechiometrici.

Ad oggi l'EMSA trova largo impiego soprattutto nella caratterizzazione di fattori trascrizionali, che legano il DNA in maniera sequenza specifica. Il grande successo dell'EMSA risiede in primo luogo nella vasta gamma di anticorpi che sono oggi disponibili sul mercato e dai costi relativamente contenuti. Inoltre da un punto di vista sperimentale questa tecnica ha notevoli vantaggi: è veloce, semplice, sensibile, richiede basse quantità di proteina ricombinante o estratto proteico ed infine i probe oligonucleotidici utilizzati sono piccoli (20 / 30 bp).

Il limite principale di questa metodologia consiste nel fatto che l'EMSA viene utilizzata solo se si hanno delle ipotesi funzionali tali per cui si decide di investigare un determinato fattore proteico, che si suppone avere affinità per la sequenza di DNA e/o RNA di interesse. L'EMSA nella sua accezione classica non permette di fare indagini ad ampio spettro identificando tutte le proteine che interagiscono con la sonda che si vuole analizzare. L'identificazione di questi complessi proteici rappresenta un nodo cruciale per lo sviluppo di studi funzionali che consentano di delucidare i meccanismi molecolari che sono alla base della regolazione dell'espressione genica. Molte malattie genetiche sono generate da mutazioni puntiformi o delezioni di poche basi azotate, risulta quindi fondamentale capire

come tutto ciò va ad influenzare il binding di alcune proteine piuttosto che altre, e capire quindi quali sono i fattori proteici coinvolti nell'insorgenza di determinate patologie. L'obiettivo principale di questo progetto di ricerca è stato la messa a punto di una strategia che andasse a modificare il classico saggio EMSA, in modo da poterlo accoppiare alla Spettrometria di Massa. Il principale vantaggio che deriva dall'accoppiamento EMSA/MS risiede nella possibilità di identificare tutte le proteine interagenti con il probe di interesse, grazie all'elevata sensibilità della Spettrometria di Massa, e formulare a posteriori ipotesi funzionali che possono poi essere ulteriormente validate con altri esperimenti di proteomica funzionale. Diverse sono stati gli studi preliminari condotti per la messa a punto di questa strategia innovativa.

La prima fase, senza dubbio la più delicata e complicata, è consistita nella definizione del miglior rapporto di estratto proteico nucleare per ng di sonda, infatti un eccesso di estratto proteico potrebbe comportare la formazione di legami aspecifici tra proteine ed il probe di interesse, oltre ad una difficoltà di migrazione su gel nativo di agarosio. Oltretutto va ricordato che uno dei principali problemi derivanti dall'accoppiamento EMSA/MS è costituito dalle esigue quantità di materiale proteico che è possibile caricare su gel nativo.

Infine particolare cura è stata richiesta nella preparazione dei probe, infatti nel 90% dei casi le sonde oligonucleotiche sono marcate radioattivamente. Questo approccio risulta essere incompatibile con l'utilizzo degli spettrometri di massa. Pertanto sono stati condotti diversi esperimenti allo scopo di comprendere quale fosse la migliore marcatura per il probe di DNA da analizzare.

2.3.1.1 Ottimizzazione del rapporto sonda oligonucleotidica/proteine nucleari per l'esperimento di EMSA/MS-MS

La prima fase è consistita nella definizione del miglior rapporto tra la sonda oligonucleotidica e i µg di estratto proteico nucleare. A questo scopo si è deciso di utilizzare una linea cellulare a noi nota, in modo da condurre l'esperimento contro una specifica proteina. In particolare si è proceduto con la lisi nucleare frazionata di cellule A459 di adenocarcinoma polmonare. Tra le proteine espresse da questa linea cellulare vi è la *Activator protein 1* (AP1), pertanto in questa fase di ottimizzazione della procedura come sonda oligonucleotidica è stata utilizzata la *sequenza consensus* riconosciuta da AP-1.

API forward: TGTGATGACTCAGGTTTG

API reverse: CAAACCTGAGTCATCACA

Preliminarmente si è proceduto con un esperimento di western blot (Fig 2.8) , utilizzando uno specifico anticorpo per AP-1, per confermare la presenza di questa proteina nell'estratto nucleare di interesse.

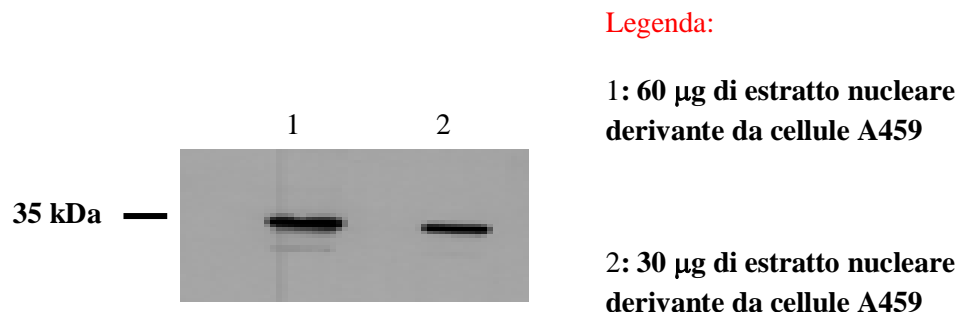
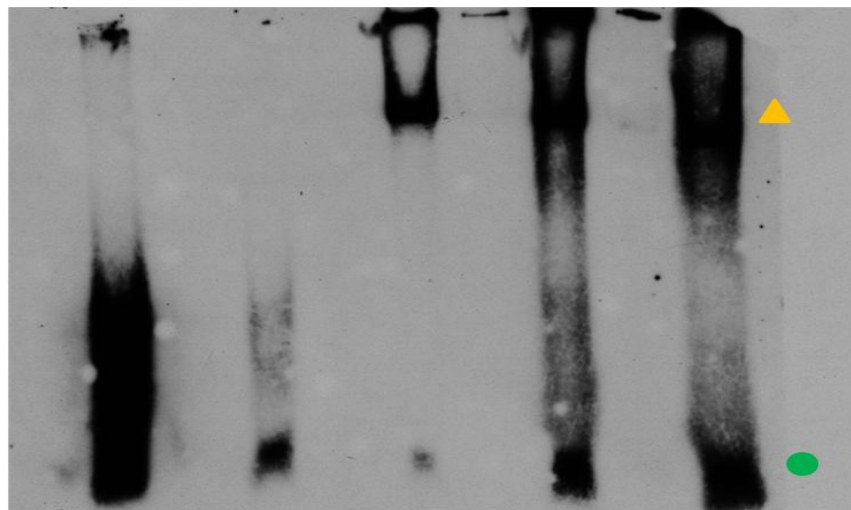


Figura 2.8 Western blot per confermare la presenza di AP-1 nell'estratto nucleare di interesse.

Sono stati sperimentati tre diversi rapporti sonda oligonucleotidica/estratto: 5 e 10 µg di proteine nucleari per 60 e 200 ng di sonda marcata con biotina. La sequenza oligonucleotidica è stata incubata con l'estratto proteico nucleare; per favorire il legame con gli specifici interattori è stato utilizzato un tampone di binding (25 mM Hepes pH 7.6, 50 mM KCl, 12.5 mM MgCl₂, 1mM DTT, 20% glicerolo, 0.1% triton). L'incubazione è stata condotta a temperatura ambiente, per 30 minuti.

Al termine dell'incubazione i campioni sono stati caricati su gel nativo di poliacrilammide al 6% (10x10 cm di dimensione e 1.5 mm di spessore). La corsa elettroforetica è stata condotta a 100 V cost per 1h . Le proteine presenti sul gel sono state quindi trasferite su membrana di nitrocellulosa. Il segnale relativo alla proteina esca è stato rivelato mediante metodi chemiluminescenti. (Fig. 2.9)

Sonda	+200ng	+60ng	+60ng	+200ng	+200ng
E.N.	-	-	+10µg	+10µg	+5µg



Legenda:

- ▲ Binding DNA-proteine
- DNA non legato

Figura 2.9 Wester Blotting al fine di ottimizzare il rapporto sonda oligonucleotidica-estratto proteico.

Sul gel, oltre ai tre diversi rapporti sonda oligonucleotidica/estratto proteico nucleare, sono state caricate anche due diverse quantità di sonda libera, al fine di visualizzare non solo la migrazione della sonda libera, ma anche di capire se eccessive quantità di sonda potessero causare problemi durante la corsa elettroforetica. Come si può osservare, anche in seguito all'incubazione con l'estratto, vi è sempre una certa quantità di probe che non viene legato. Il migliore risultato è stato ottenuto incubando 60 ng di sonda oligonucleotidica con 10 µg di estratto proteico nucleare come evidenziato dal gel in Figura 2.9.

2.3.1.2 Sonda e sua marcatura

Alla luce dei risultati ottenuti con il precedente esperimento, si è passati a saggi EMSA preparativi volti a stabilire se vi fosse o meno interazione tra alcuni componenti degli estratti proteici nucleari, derivanti da cellule HepG2 e Panc1, con le due sequenze oligonucleotidiche di interesse, sia nella loro forma wild type che mutata. Stavolta però si è preferito marcare le sonde oligonucleotidiche piuttosto che con la biotina con un fluoroforo. Questa diversa strategia sperimentale è stata scelta per due motivi fondamentali:

- Diminuire i tempi e i costi dell'esperimento. La marcatura con il fluoroforo consente infatti, immediatamente dopo aver terminato la corsa elettroforetica di visualizzare se è

avvenuto o meno il binding DNA-proteine, evitando l'utilizzo di anticorpi e reattivi per la chemiluminescenza.

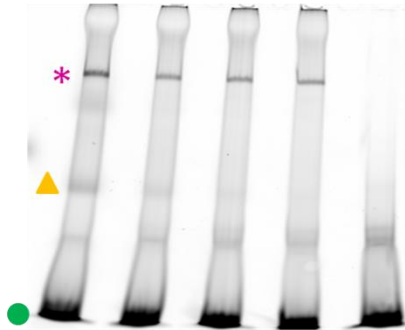
- Una marcatura di questo tipo consente una più facile e veloce estrazione delle bande di interesse da gel. A tale scopo infatti va ricordato che i campioni sottoposti a questo particolare tipo di saggio EMSA non sono addizionati con coloranti, che generalmente consentono di seguire la corsa elettroforetica. Utilizzando una sonda fluorescente contestualmente con la scansione è possibile estrarre dal gel le bande di interesse e procedere con l'estrazione delle proteine.

2.3.1.3 Emsa con spiazzamento

Prima di procedere con gli esperimenti preparativi, la fase di messa a punto della metodologia ha previsto esperimenti di competizione. Al fine di valutare la specificità di interazione DNA-proteine sono stati effettuati dei saggi EMSA in cui sono state utilizzate, in vari rapporti, sia le sonde marcate con Cy5 (sonde calde) che le medesime sonde non marcate (sonde fredde). La presenza della sonda fredda, in quantità crescente, ha fatto sì che vi fosse competizione tra le sonde marcate e le medesime sonde non marcate, per il binding specifico di alcune proteine. Nel caso in cui il binding DNA-proteine fosse stato specifico infatti, i complessi proteici di interesse avrebbero legato statisticamente la sonda non marcata perché in eccesso, di conseguenza in seguito alla scansione del gel la fluorescenza relativa allo shift sarebbe diminuita con l'aumento della sonda fredda. Dai profili elettroforetici (Fig. 2.10 e 2.11) si è potuto stabilire se vi era o meno l'interazione DNA-proteine, e in caso positivo quali fossero le bande di interesse.

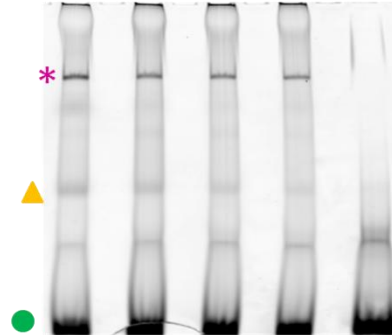
HepG2 con sonda -1640 wt

Sonda calda	+	+	+	+	+
Sonda fredda	-	+300ng	+600ng	+1.2µg	-
E.N.	+	+	+	+	-



HepG2 con sonda -1640 mut

Sonda calda	+	+	+	+	+
Sonda fredda	-	+300ng	+600ng	+1.2µg	-
E.N.	+	+	+	+	-

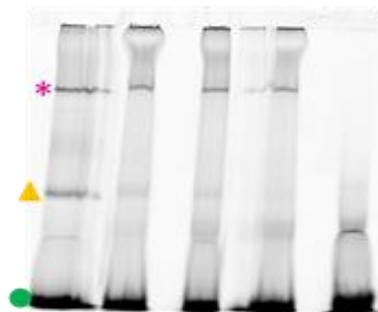


Legenda:

- * Binding DNA-proteine aspecifico
- ▲ Binding DNA-proteine specifico
- DNA non legato

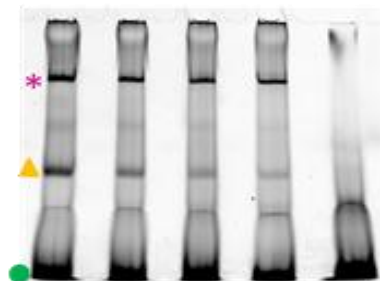
Panc1 con sonda -1640 wt

Sonda calda	+	+	+	+	+
Sonda fredda	-	+300ng	+600ng	+1.2µg	-
E.N.	+	+	+	+	-



Panc1 con sonda -1640 mut

Sonda calda	+	+	+	+	+
Sonda fredda	-	+300ng	+600ng	+1.2µg	-
E.N.	+	+	+	+	-



Legenda:

- * Binding DNA-proteine aspecifico
- ▲ Binding DNA-proteine specifico
- DNA non legato

Figura.2.10 Emsa con spiazzamento, per le sonde -1640 wt e mut.

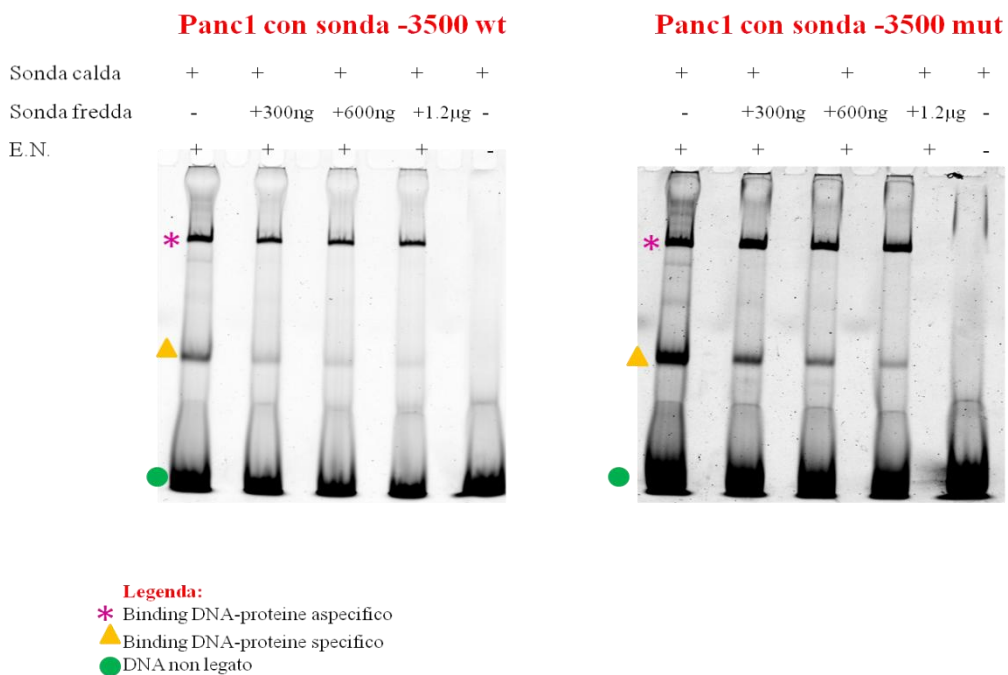
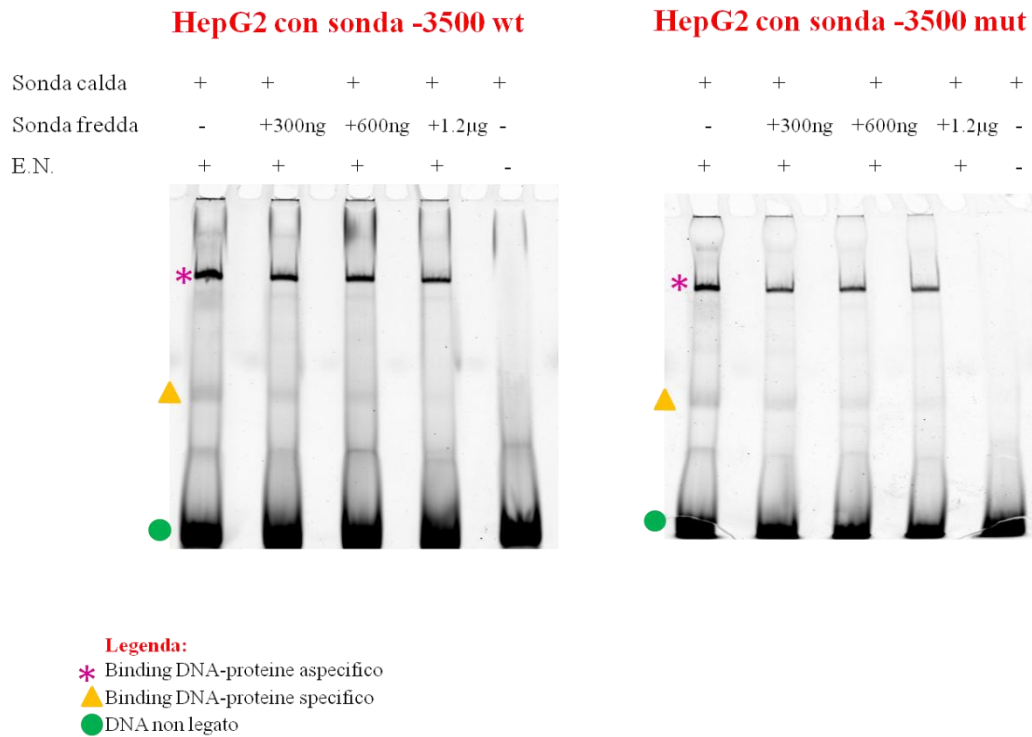


Figura.2.11 Emsa con spiazzamento, per le sonde -3500 wt e mut.

Per questi esperimenti analitici si è deciso di utilizzare, per la corsa elettroforetica, dei gel nativi di agarosio al 6% di dimensione maggiori (18 x 20 cm di dimensione, 1,5 mm di spessore) rispetto a quello mostrato in Fig. . Questo ha consentito una migliore risoluzione .

Dal confronto dei vari EMSA, appare evidente come i profili elettroforetici siano simili tra di loro. In particolare, in tutti e quattro gli esperimenti, si osserva ad alto peso molecolare la formazione di complessi DNA-proteine (*) che non vengono spiazzati per competizione con la sonda fredda, al contrario di quanto accade invece per i complessi a più basso peso molecolare (▲). Questo dato induce a pensare che la formazione dei complessi ad alto peso molecolare sia da imputare all'interazione di proteine che in maniera aspecifica si legano alle sonde oligonucleotidiche. La nostra attenzione si è pertanto focalizzata sui complessi a più basso peso molecolare. In particolare risulta evidente come la sequenza oligonucleotica -1640, sia quando incubata con l'estratto nucleare derivante da cellule HepG2 che con quello derivante da cellule Panc1, nella sua forma wild type viene spiazzata più facilmente rispetto alla forma mutata. Questo risultato lascia ipotizzare che il probe mutato sia legato da proteine che presentano, per esso, una maggiore affinità rispetto alle proteine interagenti con l'analoga sonda nella forma wild type. Inoltre a parità di condizioni, con gli estratti derivanti da Panc1 la banda ritardata risulta essere più intensa rispetto a quanto accade con gli estratti derivanti da HepG2. E' possibile, quindi, che in Panc1, a parità di proteine totali, siano basalmente più espresse quelle che legano le sonde di interesse rispetto a HepG2. La presenza di una maggiore quantità di proteine potenzialmente leganti la regione del promotore potrebbe giustificare la maggiore espressione residua di cftr in Panc1 rispetto a HepG2.

2.3.1.4 Emsa preparativi

Alla luce di quanto ottenuto con gli esperimenti precedenti, si è deciso di effettuare dei saggi EMSA preparativi, che consentissero l'identificazione delle proteine interagenti con le sonde oligonucleotidiche di interesse. Per ciascun esperimento sono stati utilizzati circa 5 milioni di cellule, sia per la linea cellulare HepG2 che per Panc1, in modo da poter utilizzare successivamente gli estratti nucleari per l'incubazione con le sequenze oligonucleotidiche. A tale fine su entrambe le linee cellulari è stato condotta una lisi cellulare frazionata. La concentrazione dell'estratto proteico nucleare è stata determinata attraverso il saggio di Bradford. Per tutti i saggi EMSA si è stabilito di incubare 180 µg di proteine nucleari con 1,1 µg di sonda oligonucleotidica marcata. Si è scelto di aumentare la quantità di estratto proteico da utilizzare, rispetto a quella utilizzata negli esperimenti analitici, in quanto sebbene la Spettrometria di massa sia una tecnica molto sensibile, i fattori trascrizionali normalmente risultano essere poco espressi. Quantità inferiori ai 180 µg non avrebbero consentito una buona identificazione degli interattori. Al fine di individuare le proteine che

in maniera aspecifica si legassero alla sonda, si è scelto di utilizzare come controllo la medesima quantità di estratto proteico nucleare non incubato con la sequenza oligonucleotidica di interesse. Terminati i 30 minuti di incubazione estratto proteico nucleare-sonda, in presenza del binding buffer, i campioni sono stati separati su gel nativo in poliacrilammide al 6%. Si è deciso di suddividere il campione da analizzare in più pozzetti, in quanto durante la migrazione elettroforetica potevano esserci fenomeni di diffusione dovuti all'eccessiva quantità di materiale proteico. Per seguire la corsa elettroforetica, non avendo addizionato ai campioni nessun colorante, si è caricato nel primo pozzetto di ciascun gel EMSA, una miscela di due frammenti di DNA a peso molecolare noto. Il gel è stato fatto correre per un 1h, applicando inizialmente un voltaggio di 200V ed aumentandolo gradualmente fino ad arrivare a 250V. La corsa è stata eseguita al buio, in quanto le sonde utilizzate risultavano essere fotosensibili per la presenza della Cy5.

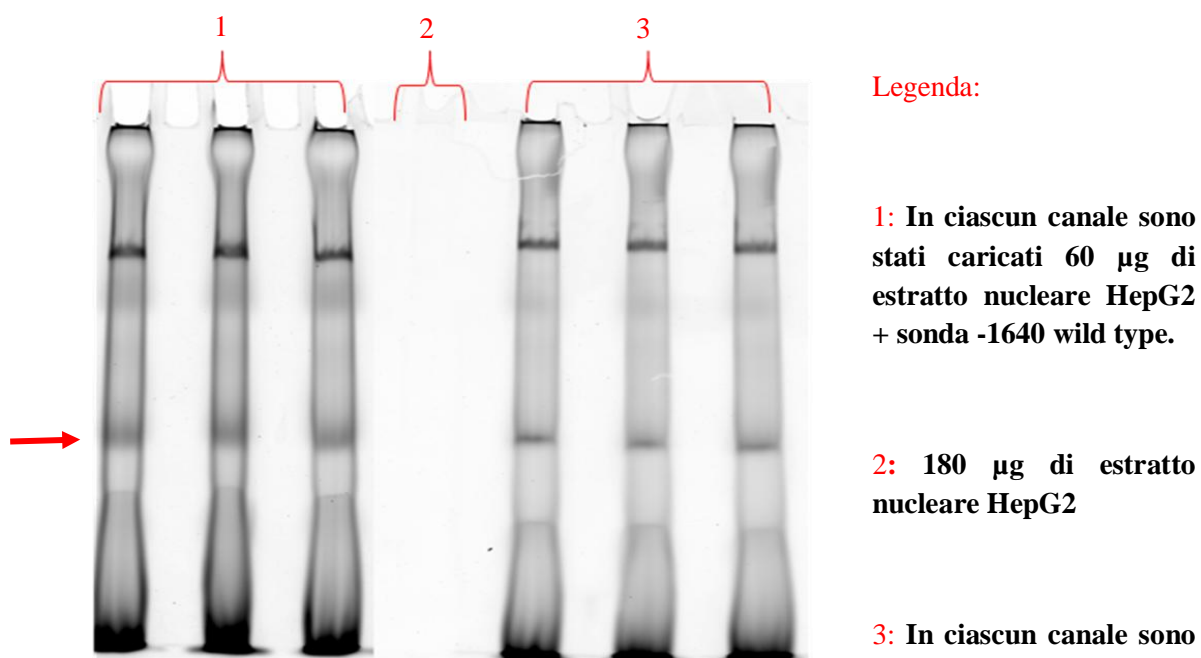


Figura 2.12 Esperimento EMSA preparativo. Proteine nucleari derivanti da cellule HepG2 con sonda -1640 wild type e mutata. Da ciascun canale sono state escisse le bande all'altezza indicata della freccia.

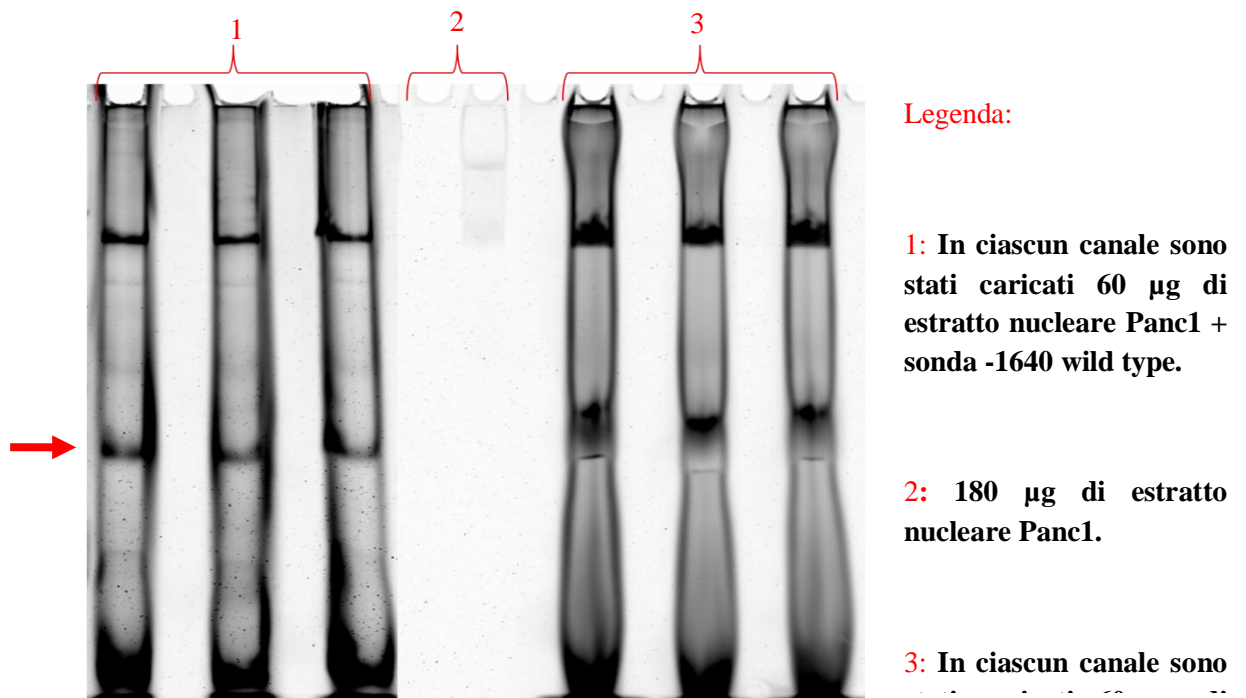


Figura 2.13 Esperimento EMSA preparativo. Proteine nucleari derivanti da cellule Panc1 con sonda -1640 wild type e mutata. Da ciascun canale sono state escisse le bande all'altezza indicata della freccia.

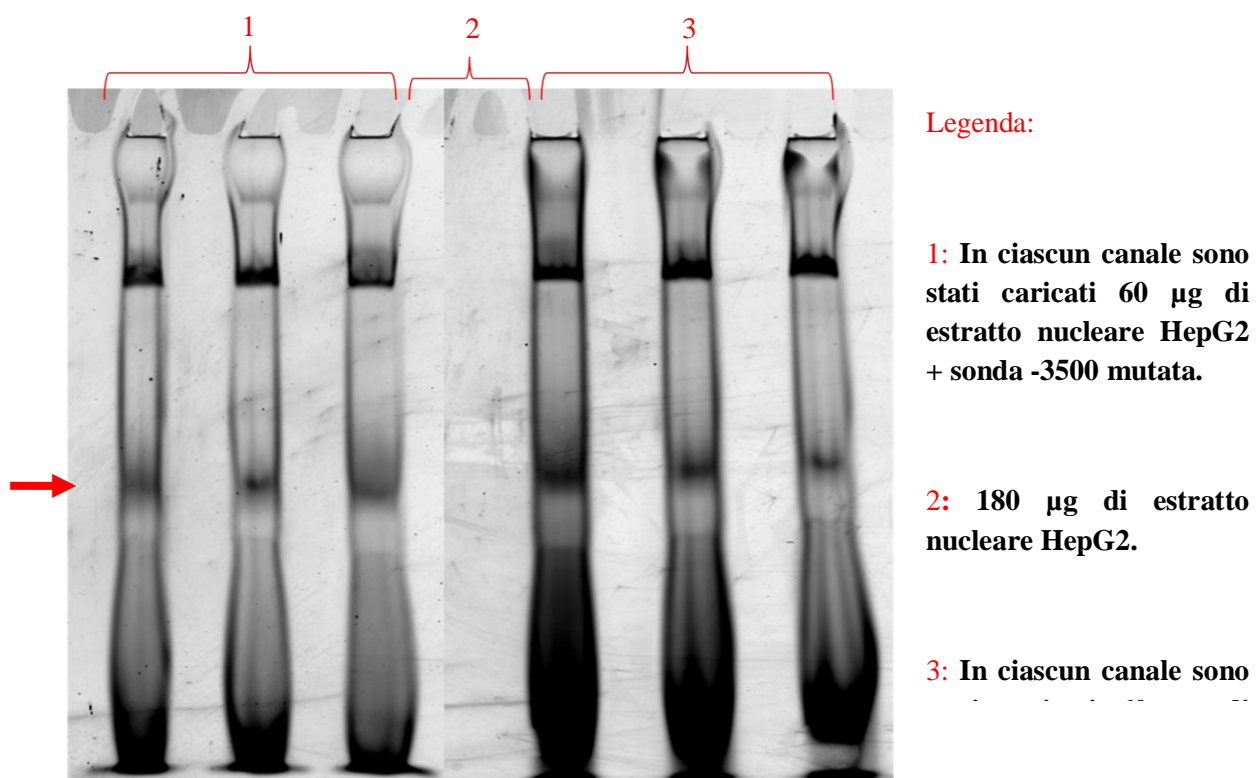
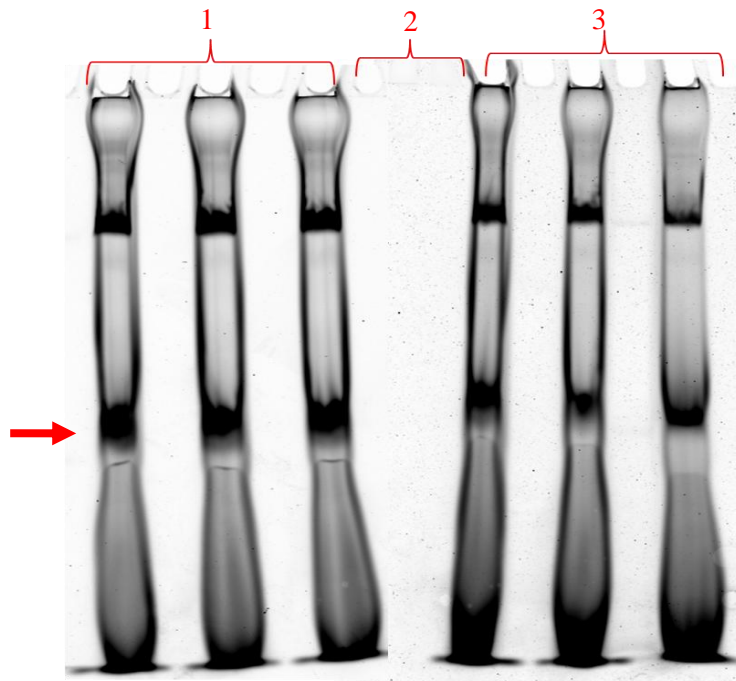


Figura 2.14 Esperimento EMSA preparativo. Proteine nucleari derivanti da cellule HepG2 con sonda -3500 wild type e mutata. Da ciascun canale sono state escisse le bande all'altezza indicata della freccia.



Legenda:

1: In ciascun canale sono stati caricati 60 µg di estratto nucleare Panc1 + sonda -3500 mutata.

2: 180 µg di estratto nucleare Panc1.

3: In ciascun canale sono

Figura 2.15 Esperimento EMSA preparativo. Proteine nucleari derivanti da cellule Panc1 con sonda -3500 wild type e mutata. Da ciascun canale sono state escisse le bande all'altezza indicata della freccia.

Contestualmente alla scansione del gel si è proceduto all'escissione delle bande dal gel. Queste sono state sminuzzate e sottoposte ad un opportuno protocollo di idrolisi in situ con tripsina, le miscele peptidiche ottenute sono state analizzate mediante nano LC-MS/MS. Dal momento che i complessi DNA-proteine dell'esperimento EMSA sono stati frazionati mediante una tecnica a bassa risoluzione, quale è l'elettroforesi monodimensionale, è normale aspettarsi che ciascuna banda del gel contenga più proteine. È stato necessario quindi l'utilizzo della spettrometria di massa tandem associata alla cromatografia liquida capillare a fase inversa per separare ed analizzare le miscele peptidiche complesse ottenute dall'idrolisi di ciascuna banda del gel. Man mano che la cromatografia procede, i peptidi vengono ionizzati, analizzati, isolati e frammentati. In base agli spettri di frammentazione ottenuti, si sono ricavate informazioni relative alle sequenze amminoacidiche che, unitamente ai pesi molecolari accurati dei peptidi intatti, sono state utilizzate per l'identificazione delle proteine.

I dati ottenuti, sia come massa dei peptidi intatti che in termini di spettri di frammentazione (peak lists), sono stati inseriti nel software Mascot, utilizzando l'opzione MS/MS Ion Search. Il programma di ricerca simula l'esperimento di proteomica, calcolando i possibili valori di teorici m/z ottenuti dalla frammentazione dei peptidi, originati dalla digestione proteolitica, specificamente utilizzata per tutte le proteine la cui sequenza è presente nella

banca dati. I valori di massa dei frammenti ritrovati nello spettro sono quindi paragonati con quelli teorici. Il report restituito in output dal software contiene l'elenco delle proteine identificate in ordine decrescente di score. Lo score di probabilità d'identificazione di una proteina è data dalla somma degli score che il software attribuisce a ciascun peptide assegnato alla proteina ed attribuito sulla base della qualità dello spettro di frammentazione. Lo score di identificazione di una proteina è una misura del numero e della bontà dei matches (corrispondenze), ossia della coincidenza dei valori sperimentali di massa con quelli teorici e risulta essere influenzato da parametri quali l'accuratezza della misura, il numero di peptidi appartenenti alla proteina con le relative intensità dei segnali negli spettri di frammentazione e la percentuale di sequenza proteica coperta dai peptidi identificati. La banca dati utilizzata per la ricerca è stata l'NCBI, in quanto è ricca e frequentemente aggiornata. A partire dal numero di accesso della proteina nella banca è possibile ottenere la sua sequenza, che viene utilizzata poi con un'altro software, ExPasy Blast, che allinea sequenze proteiche presenti in diverse banche dati. Questo programma consente di ricavare per ogni proteina analizzata il codice identificativo della stessa proteina relativa alla banca dati Uni-Prot, che sebbene meno aggiornata rispetto alla NCBI, ha il vantaggio di essere non ridondante e di contenere un maggior numero di informazioni, sia di carattere funzionale che strutturale, sulle proteine presenti. Gli allineamenti delle sequenze proteiche sono stati effettuati consultando banche dati di proteina in Homo Sapiens. Per tutti gli esperimenti EMSA, le proteine identificate nel canale del controllo sono state sottratte a quelle identificate nel campione, al fine di ottenere le liste dei putativi interattori. Come ipotizzato precedentemente per le bande ad alto peso molecolare, che non risentivano dello spiazzamento con la sonda fredda, gli interattori identificati risultano essere proteine che legano il DNA in maniera aspecifica, come ad esempio XRCC5 ed XRCC6 che legano le estremità libere del DNA. Di seguito sono riportate le tabelle, in cui sono elencati tutti gli interattori relativi alle bande a basso peso molecolare, dei vari esperimenti EMSA.

Tabella 2.1 Proteine presenti in cellule HepG2 ed interagenti con la sonda -1640 nella forma wild type.

Proteine	Score	Peptidi	SP code
14-3-3 protein gamma (YWHAG)	258	4	P61981
14-3-3 protein epsilon (YWHAE)	207	2	P62258
Annexin A5 (ANXA5)	191	3	P08758
DNA-(apurinic or apyrimidinic site) lyase (APEX 1)	134	3	P27695
Translationally-controlled tumor protein (TPT1)	103	3	P13693

Tabella 2.2 Proteine presenti in cellule HepG2 ed interagenti con la sonda -1640 nella forma mutata.

Proteine	Score	Peptidi	SP code
14-3-3 protein epsilon (YWHAE)	395	7	P62258
14-3-3 protein gamma (YWHAG)	216	2	P61981
14-3-3 protein beta/alpha (YWHAB)	212	4	P31946
Annexin A5 (ANXA5)	208	4	P08758
High mobility group protein B2 (HMGB2)	131	3	P26583
Heterog. Nucl. Ribonucleoprot. A1 (HNRNPA1)	81	2	P09651
Protein S100-A9 (S100A9)	69	2	P06702
DNA-(apurinic or apyrimidinic site) lyase (APEX 1)	65	2	P27695

Tabella 2.3 Proteine presenti in cellule Panc1 ed interagenti con la sonda -1640 nella forma wild type.

Proteine	Score	Peptidi	SP code
14-3-3 protein zeta/delta (YWHAZ)	90	1	P63104
14-3-3 protein beta/alpha (YWHAB)	54	1	P31946

Tabella 2.4 Proteine presenti in cellule Panc1 ed interagenti con la sonda -1640 nella forma mutata.

Proteine	Score	Peptidi	SP code
High mobility group protein B3 (HMGB3)	533	10	O15347
High mobility group protein B2 (HMGB2)	353	3	P26583
14-3-3 protein epsilon (YWHAE)	233	4	P62258
Acidic leucine-rich nucl. Phosphoprot. 32 family member A (ANP32A)	223	4	P39687
14-3-3 protein zeta/delta (YWHAZ)	220	3	P63104
14-3-3 protein gamma (YWHAG)	154	2	P61981
14-3-3 protein theta (YWHAQ)	106	2	P27348
Protein S100-A8 (S100 A8)	93	2	P05109
Annexin A5 (ANXA5)	82	2	P08758
DNA-(apurinic or apyrimidinic site) lyase (APEX 1)	47	1	P27695

Tabella 2.5 Proteine presenti in cellule HepG2 ed interagenti con la sonda -3500 nella forma wild type.

Proteine	Score	Peptidi	SP code
DNA-(apurinic or apyrimidinic site) lyase (APEX 1)	127	3	P27695

Tabella 2.6 Proteine presenti in cellule HepG2 ed interagenti con la sonda -3500 nella forma mutata.

Proteine	Score	Peptidi	SP code
High mobility group protein B2 (HMGB2)	195	5	P26583
DNA-(apurinic or apyrimidinic site) lyase (APEX 1)	182	4	P27695
Eukaryotic translation initiation factor 6 (EIF6)	136	2	P56537
Proliferating cell nuclear antigen (PCNA)	115	4	P12004
F-box only protein 2 (FBXO2)	55	1	Q9UK22

Tabella 2.7 Proteine presenti in cellule Panc1 ed interagenti con la sonda -3500 nella forma wild type.

Proteine	Score	Peptidi	SP code
14-3-3 protein zeta/delta (YWHAZ)	435	7	P63104
14-3-3 protein epsilon (YWHAE)	340	5	P62258
High mobility group protein B2 (HMGB2)	295	5	P26583
High mobility group protein B1 (HMGB1)	244	2	P09429
Single-stranded DNA-bind. Prot., mitoc. (SSBP1)	171	4	Q04837

Translationally-controlled tumor protein (TPT1)	161	4	P13693
Acidic leucine-rich nucl. Phosphoprot. 32 family member A (ANP32E)	149	3	Q9BTT0
DNA-(apurinic or apyrimidinic site) lyase (APEX 1)	78	2	P27695
DNA-3-methyladenine glycosylase (MPG)	71	3	P29372

Tabella 2.8 Proteine presenti in cellule Panc1 ed interagenti con la sonda -3500 nella forma mutata.

Proteine	Score	Peptidi	SP code
High mobility group protein B2 (HMGB2)	255	5	P26583
High mobility group protein B1 (HMGB1)	91	2	P09429
14-3-3 protein epsilon (YWHAE)	85	2	P62258
F-box only protein 2 (FBXO2)	133	2	Q9UK22
Acidic leucine-rich nucl. Phosphoprot. 32 family member A (ANP32E)	96	2	Q9BTT0
Translationally-controlled tumor protein (TPT1)	93	2	P13693
DNA-(apurinic or apyrimidinic site) lyase (APEX 1)	81	2	P27695
Annexin A1 (ANXA1)	81	2	P04083
Single-stranded DNA-bind. Prot., mitoc. (SSBP1)	80	2	Q04837

2.4 Discussione

Questo lavoro di tesi è stato volto all'isolamento e alla identificazione degli interattori di alcune regioni oligonucleotidiche situate a monte del promotore del gene *cftr*. In particolare, l'interesse per i complessi multiproteici interagenti con queste particolari sequenze geniche, che in teoria non dovrebbero svolgere alcuna funzione nella regolazione dell'espressione di *cftr*, nasce dal fatto che a seguito del sequenziamento genico in pazienti affetti da Fibrosi Cistica, i quali non presentavano alcuna mutazione all'interno delle canoniche regioni codificanti, sono state riscontrate mutazioni nelle 6000 coppie di basi a monte del sito di inizio della trascrizione. Negli ultimi anni i moderni studi proteomici e biochimici hanno dimostrato quanto sia importante, al fine di comprendere i fini meccanismi che sono alla base dell'espressione genica, definire come determinate proteine si leghino ad importanti sequenze di DNA e come mutazioni, anche puntiformi, possano alterare l'affinità di legame di fattori trascrizionali, che sono poi, probabilmente, alla base dell'insorgenza di gravi patologie. Pertanto questa attività di ricerca è stata indirizzata all'isolamento dei complessi proteici nucleari interagenti con le sequenze geniche di interesse, mediante un innovativo approccio di proteomica funzionale: l'EMSA/MS-MS.

La metodologia EMSA è largamente impiegata nell'ambito biologico al fine di validare ipotesi funzionali, volte a stabilire l'affinità di legame di una determinata proteina, la maggior parte delle volte ottenuta in maniera ricombinante, per una sonda oligonucleotidica. Il nostro approccio, invece, è stato focalizzato all'isolamento ed alla identificazione di interi complessi multiproteici, che in vitro vanno a legare regioni di DNA oggetto di studio, in modo da chiarire come eventuali mutazioni geniche possano alterare l'affinità di legame di alcune di queste proteine. E' stato dunque necessario apportare una serie di modifiche alle procedure EMSA convenzionali, alterando opportunamente vari passaggi della procedura trovando le condizioni ottimali per lo sviluppo di un approccio sperimentale in grado di portare all'identificazione di proteine specificamente legate ad una particolare sequenza oligonucleotidica. Al fine di limitare il gran numero di proteine che possono interagire in modo aspecifico con la sonda, e mascherarne i reali interattori, si è agito su diversi fronti. In primo luogo sono stati effettuati esperimenti volti all'ottenimento di una buona separazione dell'estratto nucleare di nostro interesse da quello citosolico, contenente verosimilmente proteine non coinvolte nel legame alla regione del promotore del gene di CFTR. In letteratura sono riportati diversi esperimenti di lisi cellulare frazionata, basati essenzialmente sull'utilizzo di due tamponi a diversa forza ionica. Tuttavia per la rottura della membrana nucleare è riportato l'utilizzo di tamponi a concentrazioni saline estremamente elevate. Nel

caso specifico un simile approccio avrebbe alterato notevolmente il legame della maggior parte delle proteine alla sequenze di DNA analizzata. Va ricordato, infatti, che le interazioni che si instaurano tra una proteina ed una sequenza oligonucleotidica sono interazioni deboli, in massima parte elettrostatiche, pertanto una elevata forza ionica sarebbe stata un notevole ostacolo. Si è giunti, in seguito a diverse prove, alla definizione di un protocollo di lisi soddisfacente la cui efficienza è stata validata mediante immunorivelazione di marcatori citosolici e nucleari. A questo punto si è passati alla definizione del corretto rapporto di tra la quantità di estratto proteico nucleare e la sonda oligonucleotidica utilizzata. Infatti un eccesso dell'uno o dell'altra avrebbe alterato il legame di specifiche proteine alla regione di DNA di interesse, andando ad aumentare notevolmente il numero di falsi positivi. Pertanto sono stati provati diversi rapporti proteine/sonda oligonucleotidica, prima di definire quello più idoneo, che è risultato essere:

10 µg estratto proteico/60 ng di sonda.

Ovviamente queste quantità sono sufficienti per visualizzare il ritardo della mobilità elettroforetica della sonda oligonucleotidica incubata con l'estratto proteico di interesse, rispetto alla medesima sonda libera, con metodi di immunorevelazione o per fluorescenza. La spettrometria di massa seppure nota per essere una tecnica estremamente sensibile, necessita di quantità superiori di estratto proteico, soprattutto quando si è interessati alla identificazione di fattori trascrizionali, i quali rispetto ad altre proteine sono di norma meno espressi. La definizione della quantità ottimale di estratto nucleare da usare negli esperimenti è stato uno degli step più delicati della fase di messa a punto. In genere i classici approcci di proteomica funzionale, a cui fanno seguito le analisi di spettrometria di massa, prevedono il frazionamento dei complessi proteici, isolati a seguito ad Affinity o Immunoprecipitazione, mediante elettroforesi monodimensionale in condizioni denaturanti. In questo caso questo approccio non è utilizzabile, infatti data la natura dell'analisi, volta ad isolare l'intero complesso proteico interagente con una sequenza di DNA, i campioni proteici devono essere separati su gel nativi, i quali consentono il caricamento in ciascun pozzetto di quantità di proteine più limitate rispetto ai gel denaturanti. Diverse sono state le corse elettroforetiche eseguite, volte a stabilire sia la massima quantità di estratto proteico da caricare in ciascun pozzetto che le condizioni di corsa più adatte all'ottenimento di una buona separazione tra la sonda libera e la sonda complessata dalle proteine. Alla fine si è deciso di utilizzare gel nativi di dimensioni 18 x 20 cm, e di caricare in totale 180 µg di estratto proteico incubato con la sonda di interesse, suddiviso in diversi pozzetti in modo che in ciascuno di essi vi fossero caricati 60 µg di proteine. Le prove preliminari in cui si era

caricato più materiale proteico per pozzetto, infatti, avevano dato problemi di diffusione durante la corsa elettroforetica. Definite le condizioni più opportune per la separazione dei complessi proteine-DNA, è stato affrontato il problema del modo migliore di visualizzare le bande di interesse sul gel per poterle successivamente escindere per l'identificazione delle proteine. In prima battuta sono state sintetizzate sonde di interesse marcate con biotina; si è poi proceduti al caricamento sul medesimo gel, sia dei campioni (estratto proteico + sonda marcata con biotina) destinati all'analisi di massa, che di una quantità inferiore degli stessi campioni, i quali sono poi stati sottoposti a western blot utilizzando l'anticorpo specifico contro la biotina. In questo modo, in seguito allo sviluppo del western blot, allineando la lastra autoradiografica al gel si possono identificare le bande relative alla sonda complessata dalle proteine. Questo approccio però è stato scartato per tre motivi fondamentali:

- 1) TEMPI SPERIMENTALI LUNGHI. Utilizzando questo approccio i tempi sperimentali si allungano notevolmente. Lo sviluppo del western blot richiede due giorni di lavoro, durante i quali il gel sarebbe stato conservato, seppure a 4°C, in acqua e ciò avrebbe potuto dare luogo a fenomeni di diffusione con conseguente perdita di materiale proteico.
- 2) BASSA SPECIFICITA' DI ANALISI: Le identificazioni delle bande relative ai complessi proteine-DNA, effettuata per confronto del gel con la lastra derivante da western blot, non è accurata e potrebbe condurre a notevoli errori durante la procedura di escissione della banda. Di conseguenza avremmo potuto o perdere materiale proteico utile ai fini della analisi di massa oppure escindere oltre alla banda elettroforetica di interesse anche un eccesso di gel contenente proteine contaminanti, che avrebbe comportato notevoli problemi, di aumento del background aspecifico.
- 3) COSTI ELEVATI. Questa tipologia di esperimenti comporta già di per sé dei costi notevoli, basti pensare sia alle elevate quantità di cellule da utilizzare per la lisi frazionata che alla produzione delle sonde oligonucleotidiche. Questi costi sarebbero risultati ancor più alti se per ogni esperimento preparativo si fosse reso necessario l'utilizzo di anticorpi e reattivi per la chemiluminescenza.

Per ovviare a queste problematiche si è pertanto deciso di marcare le sonde di interesse con un fluoroforo, nel caso specifico con la Cy5. In questo caso infatti appena terminata la corsa elettroforetica il gel viene sottoposto ad una scansione a 600 nm, lunghezza d'onda a cui assorbe il fluoroforo utilizzato. Questa analisi consente di visualizzare le bande di interesse ed escinderle istantaneamente con estrema precisione.

Per ciascun gel preparativo le bande di interesse sono state sottoposte ad un classico protocollo di idrolisi in situ con tripsina e le miscele peptidiche risultanti sono state analizzate mediante nano LC-MS/MS, consentendo l'identificazione degli interattori (Tabelle 2.1-2.8). Come si può notare il numero di interattori identificato, per ciascun esperimento, è particolarmente esiguo. Questo dato, se da un lato ha confermato l'elevata specificità di interazione tra proteine ed acidi nucleici, confermando che la procedura sviluppata è in grado di prevenire il legame aspecifico di proteine, tuttavia questo risultato ha indotto alcune considerazioni sulla natura delle sequenze di DNA utilizzate. Sonde oligonucleotidiche, costituite da 28-30 coppie di basi, forse hanno delle dimensioni troppo ridotte per consentire il legame di un intero complesso multiproteico. Pertanto è possibile che le proteine identificate siano in realtà solo una parte di un complesso proteico ben più ricco di proteine. Per ovviare a questo inconveniente si potrebbe procedere preventivamente alla multimerizzazione della sequenza oligonucleotidica di interesse, in modo da sintetizzare lunghe catene di DNA contenenti varie ripetizioni della stessa sequenza oligonucleotidica al fine di favorire sia i legami specifici rispetto a quelli aspecifici, sia l'isolamento di un intero complesso multiproteico. Ciò nonostante tra le proteine identificate con la procedura EMSA/MS-MS sviluppata sono presenti le proteine *S100-A9*, *S100-A8* [22-24] e le *Annexin 1 e 5* [25], che sono state identificate già in passato, attraverso classici esperimenti di proteomica funzionale, come interattori del gene *cftr*. Particolarmente interessante sembra essere la presenza di *Apex1* [26-28], che ritroviamo in quasi tutti gli esperimenti di EMSA/MS-MS condotti. *Apex1* è una importante proteina multifunzionale, la quale funge da "sensore" nella modulazione dell'attività di molti fattori trascrizionali. Questa proteina infatti ha attività ossidoreduttasica, e attraverso un meccanismo redox su determinati fattori trascrizionali ne aumenta o riduce l'affinità di legame al DNA.

Bibliografia

1. McIntosh I, Cutting GR. Cystic Fibrosis transmembrane conductance regulator and etiology and pathogenesis of cystic fibrosis. *FASEB J*, **6**(10), 2775-82, 1992.
2. Friedman KJ, Silverman LM. Cystic fibrosis syndrome: a new paradigm for inherited disorders and implications for molecular diagnostics. *Clin Chem*, **45**(7), 929-31, 1999.
3. Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science*, **245**(4922), 1066-73, 1989.
4. Gregory RJ, Cheng SH, Rich DP, Marshall J, Paul S, Hehir K, Ostedgaard L, Klinger KW, Welsh MJ, Smith AE. Expression and characterization of the cystic fibrosis transmembrane conductance regulator. *Nature*, **347**(6290), 382-86, 1990.
5. Rendine S et al. Genetic history of cystic fibrosis mutations in Italy. I. Regional distribution. *Ann Hum Genet* **61**(Pt.5), 411-24, 1997.
6. Bienvenue T, Les bases moléculaires de l'hétérogénéité phénotypique dans la mucoviscidose. *Am Biol Clin*, **55**, 113-21, 1997.
7. Audrézet MP, Chen JM, Raguénès O, Chuzhanova N, Giteau K, Le Maréchal C, Quéré I, Cooper DN, Férec C (2004), Genomic rearrangements in the CFTR gene: extensive allelic heterogeneity and diverse mutational mechanisms. *Hum Mutat* **23**(4), 343-57, 2004.
8. Salvatore F, Scudiero O, Castaldo G. Genotype-Phenotype Correlation in Cystic Fibrosis: the role of modifier genes. *Am J Med Genet*, **111**(1), 88-95, 2002.
9. Hantash FM, Redman JB, Starn K, Anderson B, Buller A, McGinniss MJ, Quan F, Peng M, Sun W, Strom CM. Novel and recurrent rearrangements in the CFTR gene: clinical and laboratory implications for cystic fibrosis screening. *Hum Genet*. **119**(1-2):126-36, 2006.
10. Pagnozzi D., Monti M., Orrù S., Pucci P. Functional proteomics. *Clin Chim Acta*, **357**(2), 140-50, 2005.
11. Alberts B. The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell*, **92**(3), 291-94, 1998.
12. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, **415**(6868), 141-47, 2002.
13. Godovac-Zimmermann J, Brown LR., Perspectives for mass spectrometry and functional proteomics. *Mass Spectrom Rev*, **20**(1), 1-57, 2001.

- 14.** Godovac-Zimmermann J, Brown LR. Proteomics approaches to elucidation of signal transduction pathways. *Curr Opin Mol Ther*; **5**(3), 241-9, 2003.
- 15.** Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, et al. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature*, **415**(6868), 180-83, 2002.
- 16.** Lewis TS, Hunt JB, Aveline LD, Jonscher KR, Louie DF, Yeh JM, et al. Identification of novel MAP kinase pathway signalling target by functional proteomics and mass spectrometry. *Mol Cell*, **6**(6), 1343-54, 2000.
- 17.** Zhang C, Zhang Z, Castle J, Sun S, Johnson J, Krainer AR, Zhang MQ. Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2 *Genes Dev*, **23**(18), 2550-63, 2008.
- 18.** A.V. Makeyev and S.A. Liebhaber. The poly(C)-binding proteins: A multiplicity of functions and a search for mechanisms *RNA*, **8**:265-278, 2002.
- 19.** Jian-Ping Yang, Thipparthi R Reddy, Ky T Truong, Modem Suhasini and Flossie Wong-Staal. Functional interaction of Sam68 and heterogeneous nuclear ribonucleoprotein K *Oncogene*, **21**, 7187-7194, 2002.
- 20.** Albert Jeltsch. On the enzymatic Properties of Dnmt1 *Landes Bioscience*, **1**:2, 63-66, 2006.
- 21.** Giordano S, Amato F, Elce A, Monti M, Iannone C, Pucci P, Seia M, Angioni A, Zarrilli, Castaldo G, Tomaiuolo R. Molecular and functional analysis of the large 5' promoter region of CFTR gene revealed pathogenic mutations in CF and CFTR-related disorders. *The Journal of Molecular Diagnostics*, 1525-1578, 2013.
- 22.** K. Hsu et al. Anti-infective protective properties of S100 calgranulins. *Antiinflamm Antiallergy Agents Med Chem.***8**(4): 290–305, 2009.
- 23.** G. Bode et al. Interaction between S100A8/A9 and Annexin A6 Is Involved in the Calcium-induced Cell Surface Exposition of S100A8/A9. *The journal of biological Chemistry*, **283**(6), 31776–31784, 2008.
- 24.** Atallah M. et al. Constitutive Neutrophil Apoptosis: Regulation by Cell Concentration via S100 A8/9 and the MEK – ERK Pathway. *PloS One*, **7**(2), 2012.
- 25.** Pupjalis D. et al. Annexin A1 released from apoptotic cells acts through formyl peptide receptors to dampen inflammatory monocyte activation via JAK/STAT/SOCS signaling. *EMBO Molecular medicine*, **3**, 102-114, 2011.
- 26.** Chen Li et al. Quantitative proteomics reveal up-regulated protein expression of the SET complex associated with Hepatocellular Carcinoma. *J.Proteome Res.* **11**, 871-885, 2012.

- 27.** Kishor K. et al. Transcriptional Regulatory Functions of Mammalian AP-Endonuclease (APE1/Ref-1) an Essential Multifunctional Protein. *Antioxidans e redox signaling*, 11(3), 2009.
- 28.** R. Abbotts et al. Human AP endonuclease 1 (APE1): from mechanistic insights to druggable target in cancer. *Cancer Treatment Reviews*, 36, 425-435, 2010.

3. Studio degli interattori del mRNA di ENPP1 mediante approcci di proteomica funzionale.

3.1 Resistenza insulinica

La resistenza all'azione biologica dell'insulina svolge un ruolo determinante nella predisposizione al diabete mellito tipo 2 (DT2) ed alle malattie cardiovascolari con un pesante impatto sulla qualità e durata della vita. Caratteristica fondamentale della sindrome metabolica è la ridotta capacità da parte dei tessuti periferici di rispondere all'azione dell'insulina; tale "resistenza" può essere considerata in larga misura responsabile del DT2; nel 2001 il National Institutes of Health ha suggerito i criteri diagnostici e classificativi del diabete mellito, ed ha stabilito che la presenza di almeno tre dei disordini inseriti nell'Adult Panel Treatment III (Tabella 3.1) siano sufficienti per identificare un quadro di sindrome metabolica [1].

Tabella 3.1 Criteri per la diagnosi del DT2.

Sindrome Metabolica: criteri ATP III

(la diagnosi è certa quando 3 o più fattori di rischio sono presenti)

Fattori di Rischio	Misura
• Obesità addominale (circonferenza vita)	
Uomini	> 102 cm
Donne	> 88 cm
• Trigliceridi	≥ 150 mg/dL
• Colesterolo HDL	
Uomini	< 40 mg/dL
Donne	< 50 mg/dL
• Pressione del sangue	≥ 130 / ≥ 85 mm hg
• Glicemia a digiuno	≥ 110 mg/dL

L'insulina rappresenta il principale regolatore del metabolismo glucidico, lipidico e proteico. Essa agisce attivando il trasporto di glucosio nei tessuti bersaglio, in particolare,

nel muscolo, nel tessuto adiposo ed inoltre agisce sulla produzione epatica di glucosio attraverso la inibizione della glicogenolisi e della glicogenogenesi. In maniera non diretta, l'insulina, regola anche il metabolismo dei globuli rossi e del cervello.

Nella sindrome da insulino-resistenza i tessuti bersaglio non rispondono più adeguatamente allo stimolo insulinico e si realizza, pertanto, uno stato di iperinsulinemia accompagnato da un aumento della glicemia; quest'ultima costituisce, a sua volta, uno stimolo ad una ulteriore secrezione di insulina, finalizzato al tentativo di mantenere l'omeostasi glucidica [2]. L'insulina esercita la sua azione sui tessuti bersaglio legandosi al proprio recettore di membrana (insulin receptor IR), una glicoproteina di membrana con attività tirosin-chinasica (IR-TK). Il IR è un tetramero costituito da due subunità alfa (α) extracellulari (regione del recettore che lega l'insulina) e due subunità beta (β) per lo più intracellulari (attività tirosin-chinasica del recettore). Le quattro subunità sono tenute insieme da ponti disolfuro. Le α -subunità del recettore inibiscono l'attività chinasi della porzione intracellulare. L'insulina legandosi al recettore, fa sì che il Connecting Domain (CD) possa indurre un cambio conformazionale del recettore tale da permettere l'autofosforilazione nei residui di tirosina 1158 / 1162 / 1163 del dominio citoplasmatico, e dei residui di tirosina 1328 / 1334 della porzione intramembrana. Il recettore così attivato fosforila una varietà di proteine cellulari implicate, nella trasduzione del segnale dell'insulina e nella molteplicità degli effetti dell'insulina sui tessuti (Fig.3.1). La fosforilazione in serina e treonina, invece, deprime l'attività chinasi del recettore e la sua capacità di trasmettere il segnale dell'insulina[3].

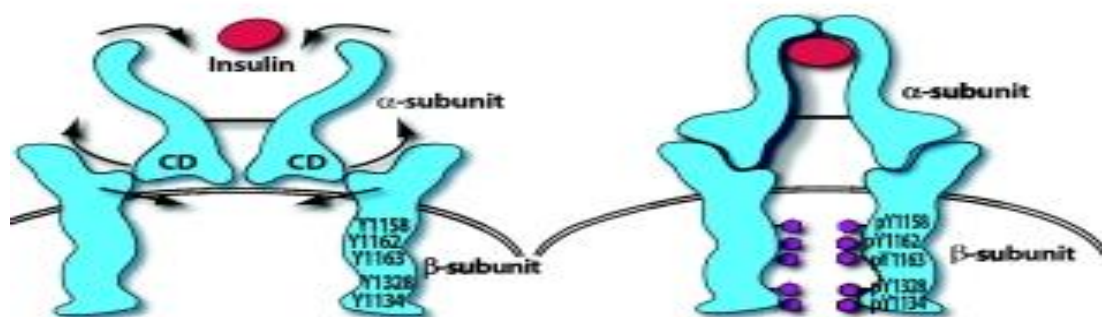


Figura 3.1 Legame dell'insulina al suo recettore di membrana.

Il meccanismo molecolare alla base della resistenza all'azione insulinica non è ancora del tutto spiegato. La cascata di eventi che permette all'insulina di esplicare la sua azione sui tessuti bersaglio, può essere alterata da varianti (polimorfismi o mutazioni) nei geni che codificano per proteine che inibiscono il segnale insulinico [4]. Tra le molecole che modulano l'azione insulinica e che, quindi, potrebbero giocare un ruolo nel determinare la resistenza insulinica (SRI), sono da annoverare gli inibitori dell'attività tirosin-chinasica del recettore dell'insulina (IR-TK); ciò è plausibile non solo per l'importanza di questa tappa nel segnale insulinico ma anche perché l'attività IR-TK risulta ridotta nei tessuti bersaglio dell'azione insulinica di soggetti con SRI. Recentemente, sono stati descritti diversi inibitori del IR-TK tra cui PC-1 [5].

PC-1 (conosciuta anche come ENPP1) è un ectoenzima di membrana [6]; il gene che codifica per questa proteina è costituito da 25 esoni ed è localizzato sul cromosoma 6q22-23 (Fig. 3.2).

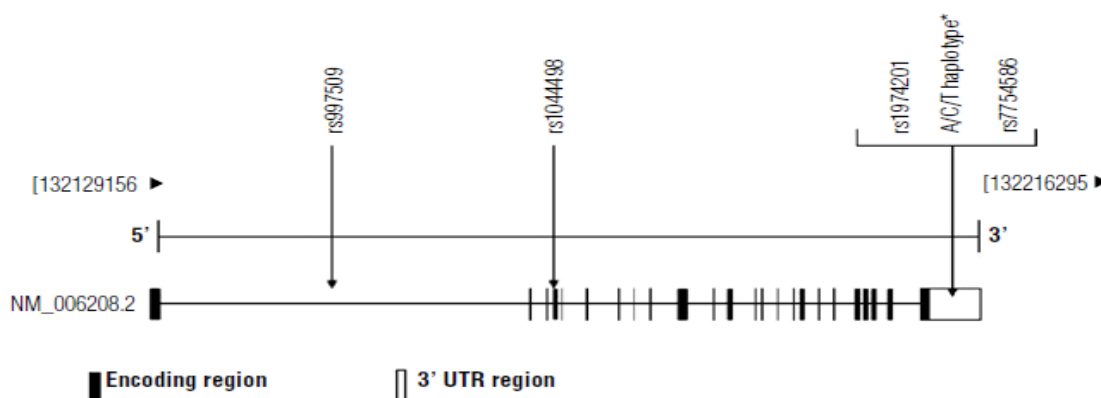


Figura 3.2: Schema del gene che codifica per ENPP1. In verticale sono mostrate le mutazioni causative dell'insorgenza del diabete mellito tipo 2.

ENPP1 è membro di una famiglia, costituita da cinque enzimi, in grado di regolare il metabolismo dei nucleotidi; questa proteina è espressa in vari tessuti tra cui anche i tessuti target per l'azione insulinica [7, 8]. Studi relativi alla struttura di ENPP1, hanno dimostrato che essa presenta un piccolo dominio N-terminale, attraverso il quale è ancorata alla membrana plasmatica, un dominio trans membrana, due domini somatomedin-B-like, ed un dominio C-terminale, più grande, che sporge nella matrice extracellulare (Fig.3.3). ENPP1 è

un omodimero di 230-260 kDa, oppure a seconda del tipo cellulare si presenta in una forma più ridotta, con un peso molecolare che varia tra i 115 ed i 135 kDa; nell'uomo ENPP1 è costituita da 873 amminoacidi. Il residuo di Thr, in posizione 204, svolge attività fosfodiesterasica e pirofosfatase, eventuali modifiche di questo residuo amminoacidico inibiscono l'attività enzimatica di ENPP1 [3].

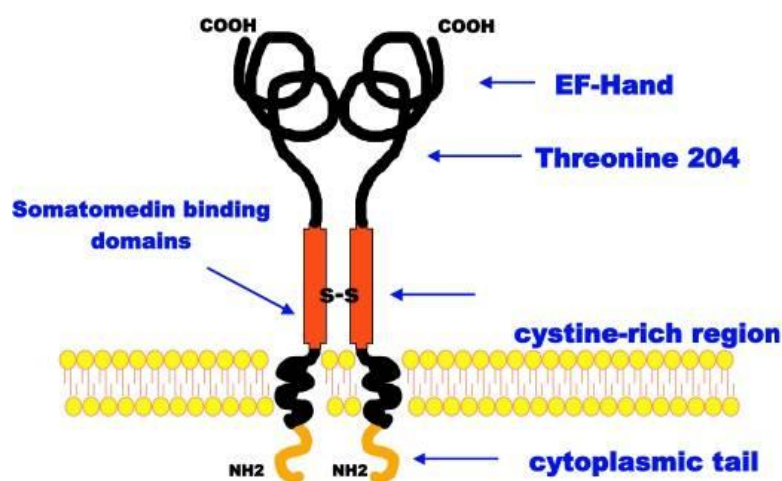


Figura 3.3 Struttura della proteina ENPP1.

Questa proteina se iper-espressa in tessuti bersaglio causa resistenza insulinica, in parecchi modelli cellulari ed animali [9, 10].

Nell'uomo l'over-espressione di ENPP1 gioca un ruolo nella patogenesi della insulino-resistenza in vivo come suggerito dagli aumentati contenuti proteici nel tessuto muscolare, nel tessuto adiposo, ed in fibroblasti di obesi o in pazienti con DT2[11, 12]. Questa over-espressione è stata osservata anche nei tessuti target di individui insulino-resistenti non diabetici e non obesi, indicando che l'alterata espressione di ENPP1 umana è un iniziale ed intrinseco difetto piuttosto che un effetto metabolico secondario dato dalla obesità o dal DT2. L'aumentato contenuto tissutale di ENPP1 in soggetti con SRI e/o DT2 e la riduzione del IR-TK in cellule che iper-esprimono ENPP1 hanno fortemente suggerito un ruolo del gene che codifica per ENPP1 nella fisiopatologia della SRI e del DT2. Inoltre è stata dimostrata l'elevata affinità di legame di ENPP1 per la regione amminoacidica sottostante al CD, pertanto quando ENPP1 è iper-espressa va a legare questa sequenza amminoacidica

impedendo il cambio conformazionale del recettore di membrana; ciò inibisce la trasduzione del segnale insulinico (Fig. 3.4).

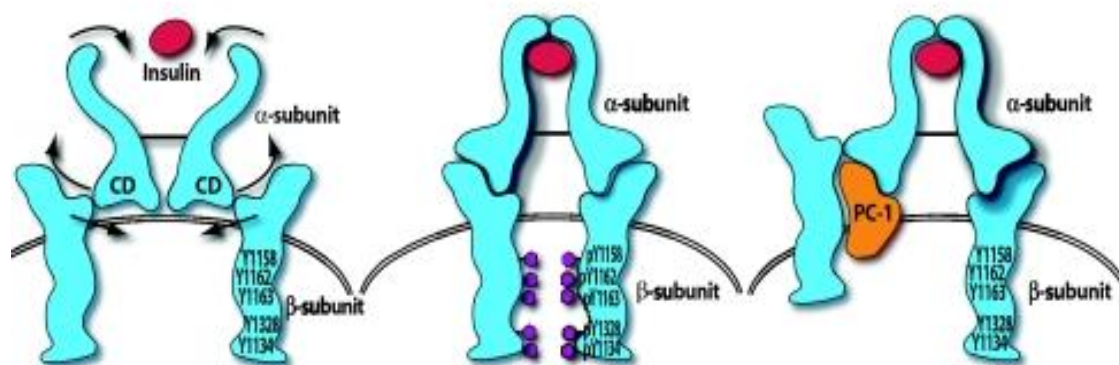


Figura 3.4 Legame di ENPP1 al recettore dell'insulina.

Nonostante il suo potenziale ruolo patogenetico nella insulino-resistenza poco si sa dei meccanismi molecolari che modulano i livelli di espressione di ENPP1. I processi che vanno dallo splicing del RNA messaggero (mRNA) alla sua traduzione sul ribosoma rappresentano un fondamentale stadio di controllo dell'espressione genica. Questi meccanismi di regolazione post-trascrizionale, di cui solo recentemente si è cominciato ad apprezzare l'importanza, sono modulati dall'interazione di fattori in trans con sequenze specifiche (elementi in cis) non tradotte (UTR) dell'mRNA. Queste regioni sono soprattutto caratterizzate da sequenze ricche in adenina e uracile [13] che legano proteine specifiche con la conseguente stabilizzazione o degradazione dell'mRNA, e la differente regolazione dell'espressione genica. Dati recenti individuano il 3'UTR di ENPP1 [14] come un potenziale sito regolatorio importante per la successiva espressione della proteina, suggerendo che fattori in trans interagiscono con l'mRNA di ENPP1 modulandone ed influenzandone la sua espressione.

Lo scopo di questo lavoro è stato quello di identificare le eventuali proteine che legano il 3'UTR di ENPP1 contribuendo alla modulazione del suo mRNA, e che potrebbero quindi essere la causa della sua over-espressione e del suo ruolo nell'insulino-resistenza e nelle complicanze correlate.

3.2 Materiali e Metodi

3.2.1 Materiali

Il detergente usato per l'estrazione proteica è il Triton X-100 della J.T. Baker.

I prodotti della SIGMA utilizzati sono: il ditiotreitolo (DTT); la tripsina, l'Hepes, l'Acido etilendiammintetraacetico (EDTA) e la resina Adipic Acid Dehydrazide Agarose Beads

Tutti i materiali utilizzati per le tecniche elettroforetiche: 30% Acrilammide/ Bis, N,N,N',N'-tetrametilendiammina (TEMED), sodio-dodecil solfato (SDS), ammoniopersolfato (APS), la miscela di proteine a peso molecolare noto "total blue" usata come standard nella migrazione elettroforetica sono della BIO-RAD.

La iodoacetamide (IAM), l'ammonio bicarbonato, il fenilmetilsulfonilfluoruro (PMSF) e l'EGTA sono della FLUKA.

Gli inibitori di proteasi "Complete mini EDTA" free sono della ROCHE.

Il Comassie Colloidale Blue è della PIERCE.

3.2.2 Metodi

3.2.2.1 Colture cellulari

Le cellule utilizzate per effettuare gli esperimenti, fornite dal laboratorio della Prof.ssa Di Paola responsabile di un gruppo di ricerca presso il poliambulatorio "Casa Sollievo della Sofferenza" con il quale si è collaborato, sono le HepG2 mantenute nel mezzo di crescita Dulbecco's Modified Eagle's Nutrient Mixture F-12 Ham's (DMEM/F12) al 10% di siero fetale bovino in incubatore a 37°C e al 5% CO₂. Le cellule sono state tenute in coltura per 48 ore in fiasche di 75 cm² raccolte e sottoposte a lisi per la raccolta delle proteine citoplasmatiche.

3.2.2.2 Estrazione delle proteine

Dopo aver rimosso il mezzo di coltura dalle piastre, le cellule sono state lavate con PBS, sono state recuperate e la sospensione è stata quindi trasferita in un tubo per centrifuga sterile. Quest'ultimo è stato centrifugato per 5 min a 1000 rpm. Il precipitato di cellule così ottenuto è stato risospeso in tampone di lisi per l'estrazione di proteine citosoliche (Tris/HCl 50 mM, NaCl 150 mM, Glicerolo 10%, PMSF 0.4 mM, EDTA 1mM, cocktail di inibitori, detergente alla concentrazione desiderata nel caso specifico Triton X-100 1%) vortexato per

15 minuti ed infine posto ad agitare per 30 minuti a 4° C. La sospensione è stata quindi centrifugata per 20 minuti a 4° C a 3750 rpm; il sopranatante, costituente l'estratto proteico citosolico, è stato filtrato con filtri da 0.45 µm, trasferito in un nuovo tubo e conservato a – 80° C. Di tale estratto è stata determinata la concentrazione proteica mediante saggio di Bradford, utilizzando il reattivo “Bio Rad protein Assay”, e la proteina di riferimento per costruire la retta di taratura è l'albumina di siero bovino.

3.2.2.3 Isolamento di complessi proteici mediante tecniche di Affinity

Per poter isolare i complessi proteici interagenti con il 3' UTR di ENPP1 è stato eseguito un esperimento di Affinity, come descritto nel capitolo 1 al paragrafo 1.4.3.1, mediante la formazione di un legame covalente tra la sonda oligonucleotidica di interesse e le beads utilizzate.

- *Preparazione mRNA*

14 µg di ENPP1-3'UTR mRNA e il rispettivo mRNA di controllo, sono stati incubati con 400 µl di *reaction buffer* (100 mM CH₃COONa, pH 5.2, e 5 mM NaIO₄) per 1 h al buio e a temperatura ambiente. L'mRNA è stato poi precipitato con etanolo e successivamente risospeso in 100 µl di CH₃COONa 100 mM, pH 5.2.

- *Binding mRNA/beads*

Sono stati prelevati 100 µL di *Adipic Acid Dehydrazide Agarose Beads* (50 µL di resina secca) per ogni mg di estratto proteico. La resina è stata lavata in modo da eliminare la soluzione di glicerolo. I lavaggi sono stati effettuati aggiungendo una quantità di CH₃COONa pari a quella di resina per almeno cinque volte, lasciando la sospensione in ghiaccio per alcuni minuti ed infine centrifugando per 2 minuti a 3000 rpm. L'mRNA è stato poi incubato con la resina precedentemente lavata. L'incubazione è stata realizzata a 4°C su ruota per 12 h.

- *Precleaning*

Parallelamente al binding mRNA/beads, sono stati incubati, sia per il campione che per il controllo, 3 mg di estratto citosolico derivante da cellule HepG2 con la stessa quantità di *Adipic Acid Dehydrazide Agarose Beads* usate per il binding.

- *Immunoprecipitazione*

L'estratto è stato poi centrifugato per 5 minuti a 2000 rpm, recuperato il sopranatante il quale è stato incubato con mRNA/beads. L'incubazione è stata eseguita a temperatura ambiente, per 30 minuti, in un volume finale di 0,6 ml. E' stata aggiunta Heparina a concentrazione finale di 7 µg/µl. Dopo l'incubazione con la resina gli estratti sono stati centrifugati per 5 minuti a 2000 rpm, il sopranatante è stato allontanato, mentre la resina è stata lavata sei volte con 1mL di *washing buffer*.

- *Eluizione*

L'eluizione dei complessi legati alle beads è stata effettuata utilizzando *loading buffer 2 X* (100 mM Tris/Cl PH 6.8, 4% SDS, 0.2% blu di bromofenolo, 20% glicerolo) contenente 200 mM DTT che funge da agente riducente.

3.2.2.4 SDS-PAGE

L'SDS-PAGE (Sodio Dodecyl Sulfate-Polyacrilammide Gel Electrophoresis) è una metodica che consente di separare le proteine in base alle loro dimensioni ed al loro peso molecolare relativo. Tale tecnica viene eseguita in condizioni denaturanti mediante l'uso di un detergente anionico ossia l'SDS, sodio dodecil-solfato ($\text{CH}_3(\text{CH}_2)_{10}\text{-CH}_2\text{O-SO}_3^-$). Questo denaturante proteico conferisce una carica netta negativa, praticamente uniforme, alle catene proteiche da separare, la cui migrazione nel gel dipenderà di conseguenza esclusivamente dal peso molecolare. La miscela da separare in SDS-PAGE viene dapprima bollita per 5 minuti in un tampone loading buffer (100 mM Tris/Cl PH 6.8, 200 mM ditiotreitolo, 4% SDS, 0.2% blu di bromofenolo, 20% glicerolo) che contiene DTT che riduce eventuali ponti di solfuro che tengono insieme la struttura terziaria della proteina e l'SDS che si lega fortemente ad essa e la denatura; in media una molecola di SDS si lega ogni due residui di amminoacido. La carica nativa della proteina viene quindi definitivamente eliminata dall'SDS. Tutti i complessi proteina-SDS si muoveranno verso l'anodo e, grazie alle proprietà di setaccio molecolare del gel, le proteine si separeranno. Facendo correre insieme al campione degli standard di peso molecolare noto è possibile risalire al peso molecolare delle proteine della miscela proteica. La corsa elettroforetica viene seguita grazie ad un colorante tracciante ionizzabile, di solito il blu di bromofenolo, e il glicerolo presente nel loading buffer rende la densità della soluzione campione tale da stratificare facilmente il campione attraverso il tampone elettroforetico sul fondo, quando viene iniettato nel pozzetto di caricamento.

3.2.2.5 SDS-PAGE preparativa

Una volta preparati in *loading buffer 2x* i campioni sono stati denaturati a 98° C per 5 min e rapidamente caricati su un gel al 12,5%, 10x10 cm. La corsa elettroforetica è stata eseguita a voltaggio costante

3.2.2.6 Colorazione

Le proteine sono state visualizzate mediante una colorazione con Coomassie Colloidale blue (PIERCE).

La visualizzazione delle proteine avviene attraverso l'immersione del gel in una soluzione di Coomassie per un'intera notte; dopodichè il colorante in eccesso è stato rimosso mediante estensivi lavaggi con acqua deionizzata.

Idrolisi in situ

Capitolo 2, paragrafo 2.2.2.6.

Analisi mediante LC-MS/MS

Capitolo 2, paragrafo 2.2.2.7.2

Western blotting

Capitolo 2, paragrafo 2.2.2.3

3.3 Risultati

In prima battuta è stata effettuata una analisi in silico della regione del 3'UTR di ENPP1 per individuare potenziali siti regolatori presenti e già noti per la loro funzione.

Per poter individuare la regione del 3'UTR, più altamente conservata nell'evoluzione della specie, che poteva rappresentare un potenziale sito di legame per eventuali proteine citoplasmatiche con funzione regolatoria, è stata, prima di tutto, effettuata una ricerca per omologia di sequenza tra la regione del 3'UTR del gene di ENPP1 umano (numero di accesso M57736) e quella del gene ENPP1 di topo (numero di accesso J02700). Il 3'UTR di ENPP1 di topo è lungo 395 bp con il 60% di omologia con il 3'UTR di ENPP1 umano che è altre 306 bp più lungo (Fig. 3.3). Nella regione a più alta omologia è localizzato un aplotipo formato da 3 SNPs di cui è già stato descritto il ruolo nella modulazione della stabilità dell'mRNA di ENPP1. In questa regione altamente conservata, la sequenza che va dal nucleotide 2750 al nucleotide 3176 è stata utilizzata come DNA template per trascrivere l'RNA probe utilizzato negli esperimenti riportati di seguito.

mouse	CTACCAGGACCGACAAGAGTCAGTTTCAGAACTGCTGAGGTTGAAAAACACATTTGCCAAT	2810
human	<u>CTATCAACAAGAAAAGAGCCAGTTTCAGACATTTAAAGTGAAAAACACATTTGCCAAC</u>	2775
	*** ** * ** ***** ***** * * * ***** *****	
mouse	CTTCAGCCAAGAAGACTGATGTTTTTTTATTAAAAACAAAAGAAAACAAAACAC	2870
human	<u>CTTTAGCCAAGAAGACTGATATGTTTTTTATCCC</u> ----- <u>CAAACAC</u>	2816
	*** ***** ***** ***** *****	
mouse	CATAGATCCTTTTGAAGAGTCTTATATTTGATACAGTCCTCTA---CACTTTTGCATT	2926
human	<u>CATGAATCTTTTTGAGAGACCTTATATTTTATATAGTCCTCTAGCTACACTATTGCATT</u>	2876
	*** ** ***** ** ***** ** ***** ** *****	
mouse	GTTTGAAACGGTCGAGTGGAGTTACAACGGGACTCCCTGTGTGGTGTGATGTCCCTG	2986
human	<u>GTTCAGAACTGTCGACCAGAGTTAGAACGGAG</u> -- <u>CCCTCGGTGATCGGGACATCTCAG</u>	2933
	*** ***** ***** ***** * * * * * ***** * * * * *	
mouse	GCTGGGTTGTGTGACGACTCAGCACATC--TGCAGAGTGTTCCTGTCTGTGCCATGCAG	3044
human	<u>GGAACCTGCGT</u> --- <u>ACTCAGCACAGCAGTGGAGAGTGTTCCTGTTGAATCTGCACAT</u>	2989
	* ** ** ***** * * * ***** * * * *	
mouse	ACTTCC--TGTCTAAGAATTAGATGTGTCTACTAACGCACGGGG--AGTAAAGACA--CTTCA	3100
human	<u>ATTTGAATGTGTAAGCATTGTATACATTGATCAAGTTCGGGGGAATAAAGACAGACCACA</u>	3049
	* ** ** ***** ** * * * * * ***** * ***** * **	
mouse	CCTCACACC---TGAAGTGTTCCTTAAGGGACGAGGGGCAGGTACGTGGTCTGGGGAC	3156
human	<u>CCTAAACTGCCTTTCTGCTTCTCTTAAGGAGAAGTAGCTGTGAACATTGTCTGGATACT</u>	3109
	*** *	
mouse	CTGATGTTGAATCCT---ATTGTTGTTAATAAACTGACTAAAGGACTGGGGTAGCTC	3212
human	<u>CAGATATTGAATCTTTCTTACTATTGGTAATAAACCTTGAT</u> -- <u>GGCATTGGGCAAACAG</u>	3167
	* ** ** ***** * * * * * ***** ** * * * * * * * *	
mouse	ATGTCCCATTT-----	3224
human	<u>TAGACTTATAGTAGGGTTGGGGTAGCCCATGTTATGTGACTATCTTTATGAGAATTTTAA</u>	3227
	* * * * *	
mouse	-----	
human	<u>AGTGGTCTGGATATCTTTAACTTGGAGTTTCATTCTTTTATTGTAATCAAAAAAA</u>	3287
mouse	-----	
human	<u>AATTAACAGAAGCCAAAATACTTCTGAGACCTTGTTTCAATCTTTGCTGTATATCCCCTC</u>	3347
mouse	-----	
human	<u>AAAATCCAAGTTATTAATCTTATGTGTTTTCTTTTAAATTTTTTGATTGGATTCTTTAG</u>	3407
mouse	-----	
human	<u>ATTTAATGGTCAAATGAGTTCAACTTTGAGGGACGATCTTTGAATATACTTACCTATTA</u>	3467
mouse	-----	
human	<u>TAAAATCTTACTTTGTATTGTATTT</u> 3493	

Figura 3.3 Omologia di sequenza del 3'UTR di PC-1 tra uomo e topo. Sequenza nucleotidica di PC-1 umana (M57736) e di topo (J02700), presentata dalla posizione nucleotidica 2750 alla 3493. Nel paragonare le due sequenze i nucleotidi identici sono stati evidenziati con un asterisco. Sono stati sottolineati i primer utilizzati per trascrivere il probe ad RNA.

La sequenza oligonucleotidica di interesse ed una sequenza random di RNA di eguale lunghezza, utilizzata come controllo, sono state sottoposte a vari cicli di amplificazione mediante PCR.

Per quanto riguarda l'esperimento di Affinity sono stati utilizzati circa 90 milioni di cellule HepG2, suddivise in due aliquote da 45 milioni, in modo da poter utilizzare successivamente gli estratti per l'incubazione con la sonda di interesse e il suo controllo. Il primo punto del protocollo ha previsto una lisi citosolica mediante l'aggiunta di un tampone, in rapporto di 1:5 rispetto al pellet.

I pellets così trattati sono stati posti in ghiaccio e vi è stato, poi, aggiunto il detergente Triton X-100 al fine di ottenere una completa lisi citosolica. In queste condizioni è stato possibile isolare, sul fondo della provetta, il pellet molto compatto, il surnatante, invece, risultava essere costituito dall'estratto citosolico. Terminata la lisi si è proceduto a determinare la concentrazione proteica, del campione recuperato, mediante il saggio di Bradford. Per ciascuno dei due campioni nucleari sono stati ottenuti 300 μ l di estratto contenente circa 3 mg di proteine. La concentrazione proteica, stimata mediante saggio di Bradford, dell'estratto citosolico ottenuto dalla lisi cellulare, è risultata essere indispensabile per stabilire la quantità di esca oligonucleotidica e quindi di beads da utilizzare per isolare i complessi proteici interagenti con la sequenza di mRNA di interesse. Le condizioni utilizzate per la cromatografia di affinità sono riportate in Tabella 3.2.

Tabella 3.2 Rapporti utilizzati per l'esperimento di Affinity.

mg proteine /μg esca	3 mg proteina / 14 μ g esca
μg esca / μl resina	14 μ g esca / 300 μ l resina

Le sonde oligonucleotidiche, ottenute mediante PCR, sono state covalentemente legate alle beads. Si è proceduto, in primo luogo, con un pre-cleaning del campione proteico il quale è stato incubato con le beads non derivatizzate. In tal modo è stato possibile "pre-purificare" l'estratto citosolico, eliminando quei componenti proteici che mostravano avere affinità solo per il supporto solido. L'estratto "pre-cleared", a questo punto è stato nuovamente incubato con delle beads sulle quali erano state precedentemente immobilizzate le esche

oligonucleotidiche. L'eluizione dalle resine è stata effettuata aggiungendo del Loading Buffer, contenente DTT 200 mM.

I complessi così purificati sono stati frazionati mediante elettroforesi monodimensionale su gel di poliacrilammide al 12,5% in condizioni denaturanti (SDS-PAGE). Il gel ottenuto con la procedura appena descritta, colorato al Comassie per la visualizzazione delle bande proteiche, è mostrato in Figura 3.4.

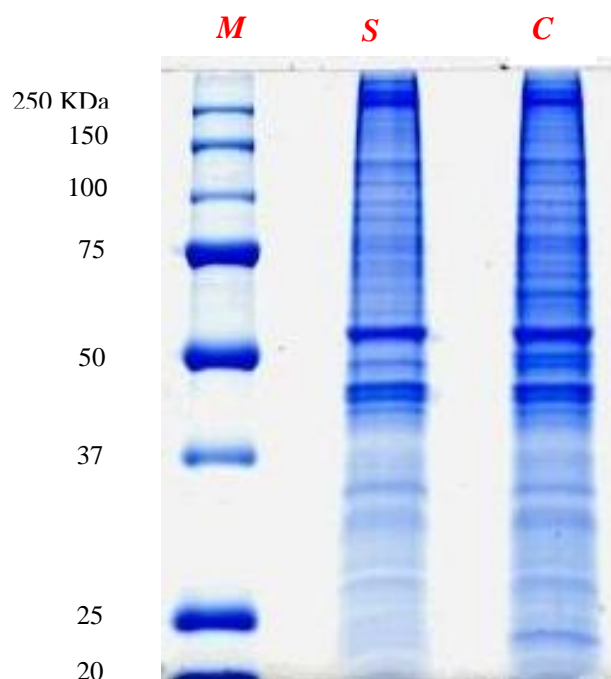


Figura 3.4 SDS- PAGE eluato proveniente dalla cromatografia di affinità su estratto citosolico per il campione(*S*) ed il controllo (*C*). *M* rappresenta una miscela di proteine a peso molecolare noto usata come standard.

A causa della bassa risoluzione bisogna considerare che una stessa banda proteica possa contenere molteplici proteine. Il gel è stato completamente tagliato, sono state ottenute 37 bande per entrambe le corsie, come mostrato in figura 3.5.

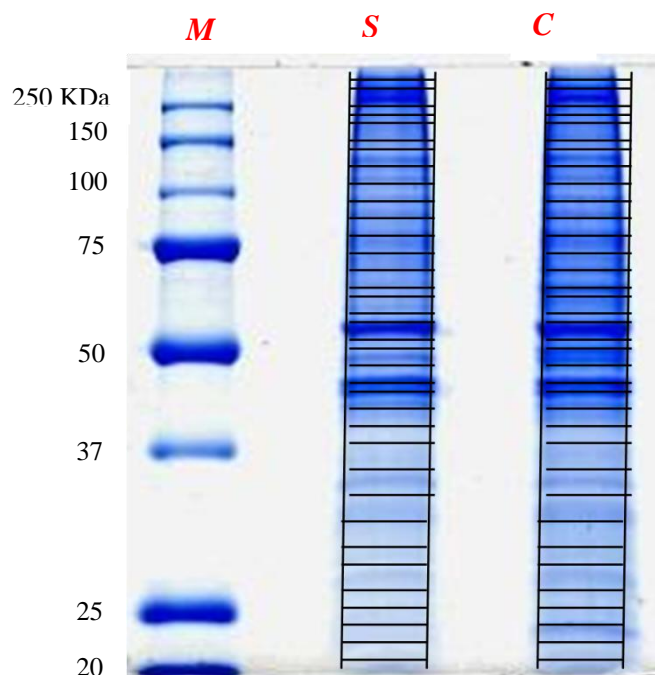


Figura 3.5 Bande escisse dal gel e sottoposte ad analici LC-MS/MS.

Le bande escisse dal gel, sono state sottoposte ad idrolisi in situ con tripsina e le miscele peptidiche ottenute sono state analizzate mediante LC-MS/MS.

La presenza di più proteine in una medesima banda richiedeva la necessità di una tecnica analitica particolarmente sensibile ed efficace. Si è reso necessario quindi l'utilizzo della spettrometria di massa tandem associata alla cromatografia liquida capillare a fase inversa, in quanto le miscele peptidiche da analizzare risultavano essere particolarmente complesse. I peptidi, quindi, sono stati frazionati mediante HPLC capillare ed introdotti nella sorgente dello spettrometro di massa man mano che eluivano dalla colonna cromatografica. Di seguito (Fig.3.6) è riportato il cromatogramma relativo all'analisi di una miscela peptidica derivante da una delle bande escisse dal gel dell'esperimento in esame.

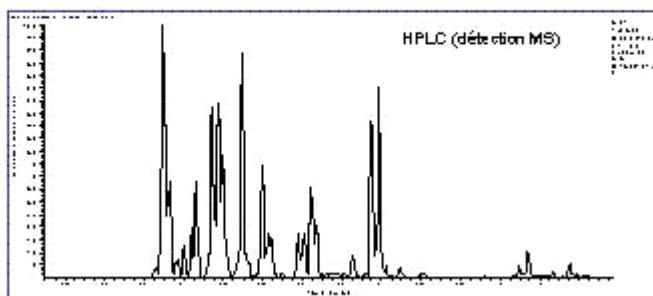


Figura 3.6 Cromatogramma relativo alla miscela peptidica.

Nella sorgente i peptidi sono stati ionizzati e ciò ha permesso la determinazione del loro peso molecolare accurato. Contemporaneamente gli ioni peptidici prodotti sono stati isolati uno alla volta e frammentati nella cella di collisione producendo uno spettro di frammentazione da cui è stato possibile ottenere informazioni di sequenza.

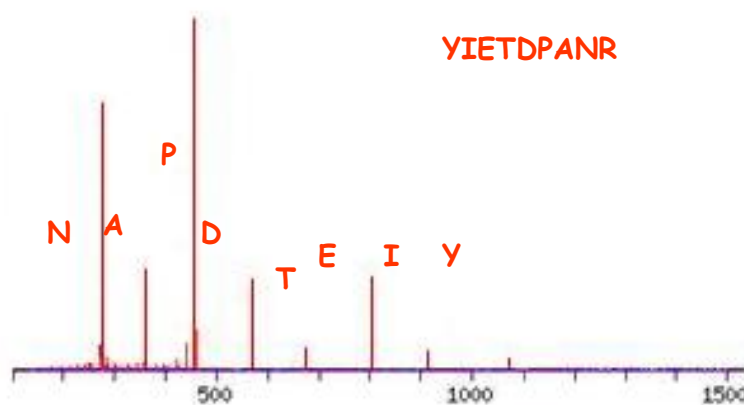


Figura 3.7 Frammentazione di un peptide.

Le analisi del campione di interesse e del controllo sono state svolte in parallelo in modo da poter confrontare gli interattori di ciascuno di essi e da poter sottrarre ai campioni le proteine identificate anche nel controllo.

Nella Tabella 3.3 si riportano le identificazioni relative al 3'UTR mRNA di ENPP1, si può osservare che alcune bande non hanno fornito risultati positivi poiché tutte le proteine identificate sono state eliminate in quanto presenti in modo rilevante anche nel controllo.

Tabella 3.3 Proteine identificate nell'esperimento di Affinity.

BANDA	PROTEIN ID	SCORE	PEPTIDI	MW (kDa)	SP CODE
2	gi 10436768 U5 small nuclear ribonucleoprotein 200 kDa helicase	88	2	208	<u>Q75643</u>
4	gi 533202 Enhancer of mRNA-decapping protein 4	187	6	151	Q6P2E9
	gi 136398 KIAA0169 protein Nucleoporin NUP188 homolog	84	2	198	Q5SRE5
6	gi 13399322 N-acetyltransferase 10 (GCN5-related)	145	4	117	Q9H0A0
	gi 18539467 nucleolar RNA-associated protein beta	59	3	113	Q9H6R4
	gi 1575536 regulator of nonsense transcript stability	51	2	124	Q92900
7	gi 12803113 Elongation factor Tu GTP binding domain containing 2	129	4	110	Q15029
	gi 5881961 Dhm1-like protein (5'-3' exoribonuclease 2)	79	2	109	Q9H0D6
8	gi 19550955 ligase III, DNA, ATP-dependent	110	2	97	P49916
	gi 4557469 adaptor-related protein complex 2, beta 1 subunit isoform b AP-2 complex subunit beta	112	3	105	P63010
	gi 4557761 mutS homolog 2	53	2	105	P43246
9	gi 3641674 gamma1-adaptin	47	2	92	O43747
10	gi 5031755 heterogeneous nuclear ribonucleoprotein R isoform 2	219	4	72	O43390
	gi 24497453 nucleoporin 88kDa	121	3	85	Q99567
	gi 7657013 apoptosis antagonizing transcription factor	82	2	63	Q9NY61
11	gi 4557495 cleavage stimulation factor subunit 3 isoform 1	204	3	83	Q12996
	gi 157743245 WD repeat domain 43	109	3	77	Q15061
	gi 5802970 AFG3 ATPase family gene 3-like 2	59	4	89	Q9Y4W6
12	gi 23308697 signal recognition particle receptor	74	3	70	P08240
13	gi 3668141 PBK1 protein	88	2	58	O76021
14	gi 5730027 KH domain containing, RNA binding, signal transduction associated 1	129	3	48	Q07666
	gi 4758412 polypeptide N-acetylgalactosaminyltransferase 2	96	2	65	Q10471
	gi 1575609 FUSE binding protein 3	86	2	64	Q96124
15	gi 1710248 protein disulfide isomerase-related protein 5	176	2	46	Q15084

	gi 1809248 siah binding protein 1 (Poly(U)-binding-splicing factor PUF60)	90	3	58	Q9UHX1
	gi 2352902 cell cycle progression 2 protein	81	2	56	Q969Z0
16	gi 4759276 RNA, U3 small nucleolar interacting protein 2	63	2	52	O43818
	gi 7657381 PRP19/PSO4 pre-mRNA processing factor 19 homolog	62	1	56	Q9UMS4
	gi 26051237 nucleoporin 54kDa	60	2	55	Q7Z3B4
17	gi 6912588 pseudoautosomal GTP-binding protein-like protein	63	2	45	O43824
19	gi 194375558 Cleavage stimulation factor 50 kDa subunit (CstF-50)	166	4	48	Q05048
	gi 7637838 programmed cell death 9 28S ribosomal protein S30 mit (Programmed cell death P 9)	73	3	51	Q9NP92
	gi 4506221 26S proteasome non-ATPase regulatory subunit 12	49	2	53	O00232
23	gi 4505763 Phosphoglycerate kinase 1	443	8	50	P00558
	gi 62088002 TIA1 protein variant Nucleolysin TIA-1 isoform p40 (RNA-binding protein TIA-1)	150	4	52	P31483
25	gi 15680185 NOL1/NOP2/Sun domain family, member 4 Putative methyltransferase NSUN4	99	3	43	Q96CB9
	gi 116283278 TFB1M protein Dimethyladenosine transferase 1 mit (mtTFB1)	80	3	39	Q8WVM0
	gi 33188463 mitochondrial ribosomal protein S9 precursor 28S ribosomal protein S9 mit	79	2	46	P82933
	gi 15342081 39S ribosomal protein L38 mit	78	2	41	Q96DV4
26	gi 13183088 Inactive hydroxysteroid dehydrogenase-like protein 1	110	3	35	Q3SXM5
	gi 34335134 Protein SEC13 homolog	78	2	36	P55735
	gi 1125770 AA tRNA synth complex-interacting multifunctional P 2	78	2	37	Q13155
27	gi 38327502 hnRNP D-like (AU-rich element RNA-binding factor)	224	5	33	O14979
	gi 45006986 AA tRNA synth complex-interacting multifunctional P 1	163	3	35	Q12904
	gi 15529982 U3 small nucleolar ribonucleoprotein protein IMP4	148	2	34	Q96G21
	gi 14277912 UDP-glucose 4-epimerase	96	2	38	Q14376
	gi 1905874 PDZ and LIM domain protein 1 (Elfin)	73	2	36	O00151

Tra gli interattori identificati hanno suscitato particolare interesse la *Polypeptide N-acetylgalactosaminyltransferase 2* (GALNT2) e *Sam 68*. Pertanto per GALNT2 sono stati condotti diversi esperimenti funzionali.

3.3.1 Influenza di GALNT2 sull'espressione di ENPP1

L'identificazione di GALNT2 ha avuto una particolare rilevanza. In prima battuta, è stata ulteriormente validata l'interazione tra questa proteina ed il 3'UTR mRNA di ENPP1. A questo proposito è stato effettuato un esperimento di RNA pull down, che ha previsto l'incubazione delle medesime sonde oligonucleotidiche utilizzate nell'esperimento preparativo con un estratto proteico totale derivante da cellule HepG2. In seguito è stato condotto un esperimento di western blotting, utilizzando uno specifico anticorpo per GALNT2 (Fig. 3.9).

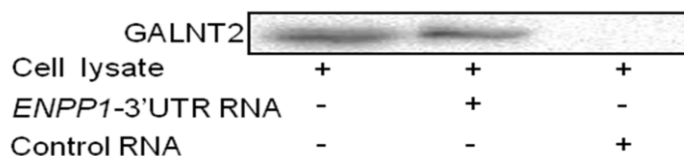


Figura 3.9 Western Blotting per confermare l'interazione tra GALNT2 ed il 3'UTR mRNA di ENPP1.

Alla luce di quanto ottenuto dall'esperimento di western blotting è stato possibile confermare l'interazione diretta tra GALANT2 ed il 3'UTR di ENPP1. Per comprendere il ruolo esercitato da GALNT2 nella modulazione dell'mRNA di ENPP1 sono stati condotti diversi esperimenti funzionali, tra i quali anche esperimenti di silenziamento genico. La linea cellulare HepG2 è stata trasfettata con il siRNA di GALNT2 inducendo in questo modo il silenziamento del gene; come controllo la stessa linea cellulare è stata trasfettata con un siRNA scrambled. Dopo 48 e 96h dal silenziamento, è stato valutata la bontà dell'esperimento mediante western blotting, utilizzando l'anticorpo diretto contro GALNT2; la normalizzazione è stata condotta andando a rivelare, nel medesimo esperimento, la presenza di GAPDH (Fig. 3.10)

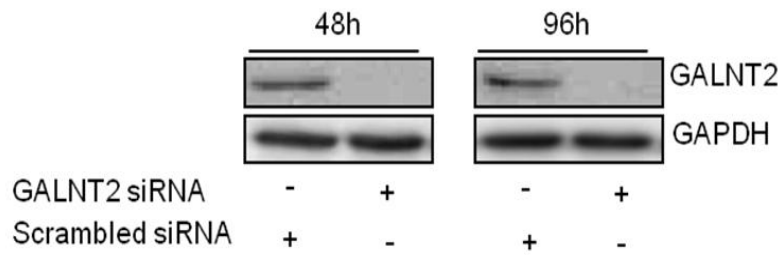


Figura 3.10 Western Blotting per confermare l'avvenuto silenziamento di GALNT2 nella linea cellulare HepG2.

Al silenziamento genico di GALNT2 hanno fatto seguito degli esperimenti di Real Time-PCR volti alla valutazione di eventuali alterazioni dell'espressione del mRNA di ENPP1 in relazione all'assenza di GALNT2 (Fig. 3.11)

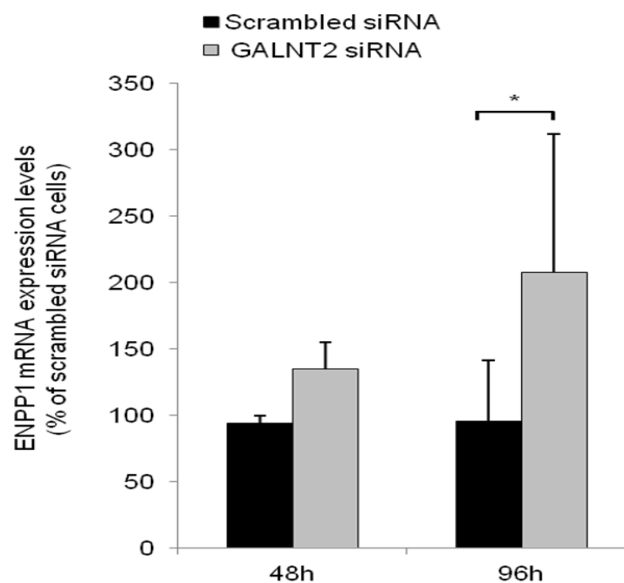


Figura 3.11 Istogramma relativo alla Real Time-PCR per la valutazione dell'espressione dell'mRNA di ENPP1 in seguito al silenziamento di GALNT2 (grigio) e di un RNA scrambled (nero).

Appare evidente come in seguito al silenziamento di GALNT2 i livelli di espressione del mRNA di ENPP1 aumentano sensibilmente, l'efficienza trascrizionale del gene codificante per ENPP1 sembra essere inversamente proporzionale all'espressione di GALNT2. Si è deciso di non limitarsi solo alle analisi del trascritto di ENPP1, ma di andare a valutare, mediante western blotting, anche l'influenza di GALNT2 sull'espressione della proteina ENPP1 (Fig. 3.12).

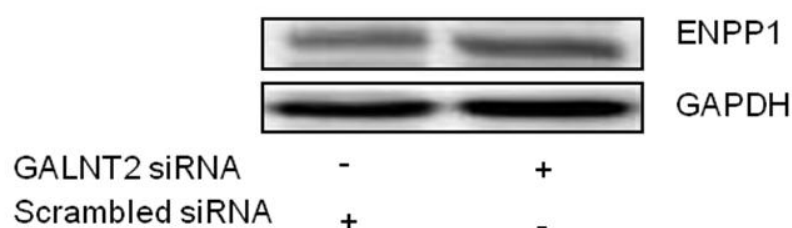


Figura 3.12 Western blot per la valutazione dell'espressione di ENPP1 in seguito al silenziamento di GALNT2. In particolare una stessa quantità di estratto proteico proveniente da cellule HepG2 è stato incubato con il siRNA di GALNT2 e un siRNA scrambled. Le proteine sono state frazionate mediante SDS-PAGE e si è valutata la presenza di ENPP1 in entrambe le condizioni utilizzando l'anticorpo specifico contro la proteina di interesse. La normalizzazione è stata effettuata valutando la presenza di GAPDH.

Anche in questo caso i dati sono in accordo con quelli derivanti dagli esperimenti volti alla valutazione del mRNA di ENPP1 in presenza ed in assenza di GALNT2. È evidente infatti che in presenza di GALNT2 la quantità di proteina ENPP1 espressa è notevolmente inferiore rispetto a quando GALNT2 è silenziato [15].

3.4 Discussione

Molti processi cellulari sono svolti dalle proteine non come entità individuali ma in combinazione con altre componenti proteiche stabilendo interazioni specifiche, che ne determinano la funzione biologica. La comprensione delle funzioni di una proteina, così come la definizione dei meccanismi molecolari all'interno di una cellula è deducibile quindi dall'identificazione dei partners con cui essa interagisce. L'identificazione e lo studio dei componenti dei complessi multiproteici rappresenta un ideale campo di applicazione per approcci di proteomica funzionale, focalizzati alla definizione della funzione biologica di proteine sconosciute ed alla descrizione dei meccanismi cellulari a livello molecolare mediante identificazione delle interazioni proteina-proteina in vivo.

Una delle tecniche comunemente usata per la purificazione dei complessi multiproteici coinvolgente la proteina d'interesse (esca) dall'intero estratto cellulare, si basa sulla cromatografia di affinità, la quale trova ampia applicazione anche nella caratterizzazione di complessi proteici che legano DNA e/o RNA.

Questo lavoro è stato volto all'isolamento e alla identificazione degli interattori del 3'UTR dell'mRNA di ENPP1, coinvolto nella resistenza insulinica. In particolare, negli ultimi anni è aumentato l'interesse per lo studio di complessi proteici in grado di controllare fenomeni post trascrizionali, come ad esempio la regolazione dell'espressione genica, attraverso la regolazione del pre mRNA e del mRNA. L'interesse per lo studio del 3'UTR mRNA di ENPP1 nasce dal fatto che in seguito al sequenziamento di questa regione sono emersi numerosi polimorfismi causativi dell'insorgenza del DT2. Inoltre, solo negli ultimi anni si è compreso che questa regione dell'mRNA di ENPP1, particolarmente ricca in uracile ed adenina, è in grado di legare proteine che modulano l'attività post trascrizionale del gene. Il 3'UTR pertanto risulta essere un potenziale sito di regolazione genica. La conoscenza degli interattori di questa particolare regione risulta quindi uno strumento estremamente importante per spiegare l'insorgenza il perché dell'iper espressione di ENPP1 in paziente che mostrano resistenza insulinica. In particolare, questo lavoro di tesi, ha previsto l'isolamento di complessi proteici, interagenti sia con la sequenza oligonucleotidica di interesse, mediante cromatografia di affinità. Sono stati utilizzati a tal proposito specifici costrutti oligonucleotidici, ottenuti attraverso diversi cicli di PCR, come esca per la cromatografia di affinità. Il ligando è stato immobilizzato covalentemente sulla resina. I componenti proteici specificamente legati all'esca, costituenti i complessi purificati, sono stati eluiti, frazionati mediante elettroforesi mono-dimensionale SDS-PAGE, sottoposti ad idrolisi triptica, analizzati mediante tecniche di spettrometria di massa tandem. La necessità

di utilizzare una tecnica analitica così efficace è stata dettata dalla complessità delle miscele peptidiche da analizzare, derivanti da più proteine presenti sotto una stessa banda da gel. Servendoci di LCMSMS è stato possibile ottenere sia informazioni sul peso accurato dei singoli peptidi sia informazioni sulla sequenza; attraverso ricerche in banche dati utilizzando il programma Mascot si è giunti all'identificazione dei putativi interattori, che sono stati opportunamente raggruppati in classi funzionali, come mostrato in figura 3.8.

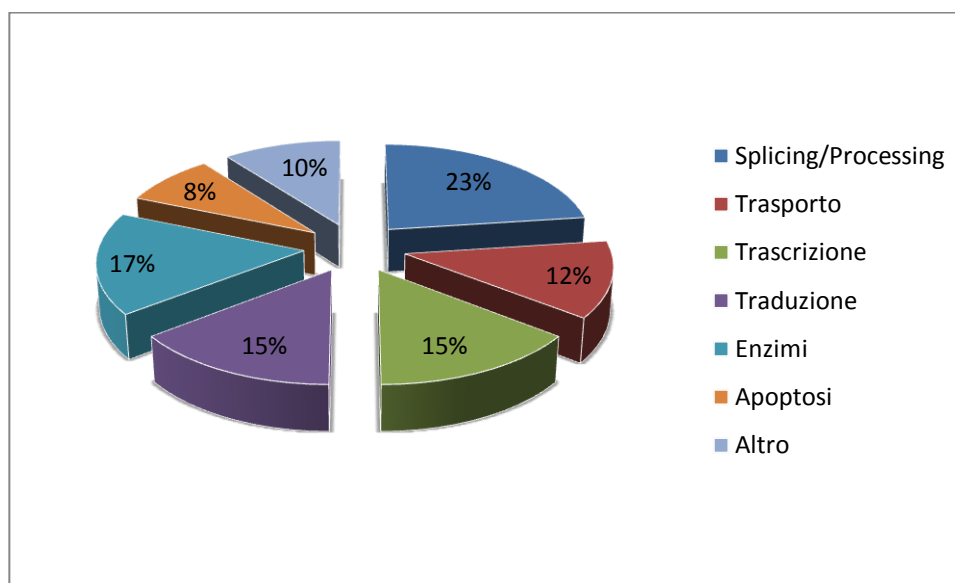


Figura 3.8 Classificazione funzionale dei putativi interattori del 3'UTR mRNA di ENPP1.

La maggioranza delle proteine identificate risulta appartenere alla classe dei *fattori di Splicing* e di *Processing* dell'mRNA confermando la correttezza dell'esperimento. Oltre ai fattori di splicing sono state identificate anche proteine coinvolte in processi di trascrizione e traduzione. Oggi è diffusamente accertato ed accettato che i processi di sintesi (trascrizione) e di maturazione (splicing) del mRNA sono strettamente coordinati, sia in termini di spazio che di tempo, in maniera molto più stringente di quanto si potesse immaginare fino a pochi anni fa. Tra gli interattori identificati ritroviamo la *Nucleolysin TIA-1 isoform p40* (TIA1) e la *Poly(U) binding splicing factor PUF60* (PUF60)[16, 17]. Entrambe queste proteine giocano un ruolo fondamentale nel processo di splicing, sia in vivo che in vitro. In particolare PUF60 e TIA1 facilitano la "ricognizione" del sito di splicing al 3' del mRNA, favorendo quindi l'assemblaggio del complesso spliceosomale [18]. Inoltre è stato dimostrato che TIA1 ha elevata affinità di binding per regioni ricche di adenina ed uracile

(AU-rich elements, ARE), ed è noto che il 3'UTR del mRNA è particolarmente ricco di queste sequenze; TIA1 quindi lega con elevata specificità il 3'UTR del mRNA di ENPP1[16].

Gli esperimenti funzionali eseguiti per comprendere la relazione tra l'mRNA di ENPP1 e GALNT2, hanno dato risultati concordanti sia per la valutazione dell'espressione del trascritto di ENPP1 in assenza di GALNT2, che per l'espressione della proteina. In particolare si evince chiaramente che i livelli di espressione di ENPP1 risultano essere inversamente proporzionali all'espressione di GALNT2. Pertanto gli studi condotti sulle cellule HepG2 hanno dimostrato che GALNT2 lega in maniera specifica il 3'UTR mRNA di ENPP1 diminuendone sensibilmente sia la trascrizione che l'espressione (Figura 3.9).

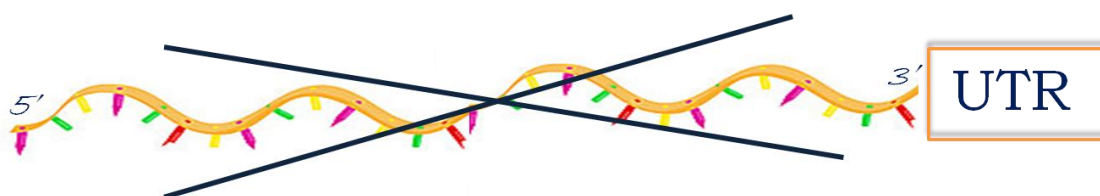


Figura 3.9: Legame di GALNT2 al 3'UTR mRNA di ENPP1.

I dati finora ottenuti suggeriscono chiaramente che la regione del 3'UTR è direttamente coinvolta nella regolazione dell'espressione genica di ENPP1 e che GALNT2 è un nuovo potenziale modulatore del signaling dell'insulina [15].

Un ruolo completamente opposto a GALNT2 è quello di Sam68. Questa proteina è nota per essere essenzialmente un importante fattore di regolazione dello splicing alternativo del mRNA [19], a cui si lega in maniera diretta, soprattutto durante l'adipogenesi; tuttavia dati recenti di letteratura hanno dimostrato che Sam68 svolge un ruolo fondamentale nello sviluppo della resistenza insulinica e dell'obesità. In particolare è stato osservato che in topi di laboratorio in cui Sam68 è silenziato, anche in presenza di un regime alimentare ipercalorico, vi è immunità dallo sviluppo di obesità e di resistenza insulinica [20]. Questi dati quindi lasciano ipotizzare che Sam 68 promuova in qualche modo l'espressione di ENPP1. In conclusione possiamo affermare che un unico esperimento di RNA pull down ha evidenziato due fenomeni opposti. In un caso è stato possibile identificare GALNT2, potente inibitore dell'attività di ENPP1 e di conseguenza in sua presenza vi è una notevole riduzione del fenomeno di resistenza insulinica. Dall'altra parte, invece, la presenza di Sam68 indica

chiaramente il suo potenziale ruolo di promotore dell'attività di ENPP1 con conseguente sviluppo di patologie legate al fenomeno di resistenza insulinica.

Bibliografia

1. Reaven G.M. Role of insulin resistance in human disease. *Diabetes* 37,1595-1607, 1988.
2. Pfeiffer EF, Kerner W. The artificial endocrine pancreas: its impact on the pathophysiology and treatment of diabetes mellitus. *Diabetes Care*. Jan-Feb;4(1):11-26, 1981.
3. Goldfine I.D: et al. The Role of Membrane Glycoprotein Plasma Cell Antigen 1/Ectonucleotide Pyrophosphatase Phosphodiesterase 1 in the Pathogenesis of Insulin Resistance and Related Abnormalities. *Endocrine reviews*, 29(1), 69-75, 2008.
4. Virkamaki A., Ueki K., Khan C.R. Protein-protein interaction in insulin signalling and the molecular mechanism of insulin resistance. *J. Clin. Invest.* 103, 931-943, 1999.
5. Almind K, Doria A., Khan C.R. Putting the genes for type II diabetes on the map. *Nat. Med* 7:277-279, 2001.
6. Maddux B.A. et al. Membrane glycoprotein PC-1 and insulin resistance in non-insulin-dependent diabetes mellitus. *Nature* 373, 448-51, 1995.
7. Maddux B.A. et al. Overexpression of the insulin receptor inhibitor PC-1/ENPP1 induces insulin resistance and hyperglycemia. *Am J Physiol Endocrinol Metab* 290, 746-749, 2006.
8. Chandalia M. et al. Adipose tissue dysfunction in humans: a potential role for the transmembrane protein ENPP1. *Clin. Endocrinol. Metab.*, 97, 2012.
9. Frittitta, L. et al. Elevated PC-1 content in cultured skin fibroblasts correlates with decreased in vivo and in vitro insulin action in nondiabetic subjects: evidence that PC-1 may be an intrinsic factor in impaired insulin receptor signaling. *Diabetes* 47, 1095-100, 1998.
10. Frittitta, L. et al. Increased adipose tissue PC-1 protein content, but not tumour necrosis factor-alpha gene expression, is associated with a reduction of both whole body insulin sensitivity and insulin receptor tyrosine-kinase activity. *Diabetologia* 40, 282-9, 1997.
11. Chin C.N. et al. Evidence that inhibition of insulin receptor signaling activity by PC1/ENPP1 is dependent on its enzyme activity. *Eur. J.Pharmacol.* 606, 17-24, 2009.
12. Costanzo B.V. et al. The Q allele variant(GLN 121) of membrane glycoprotein PC-1 interacts with the insulin receptor and inhibits insulin signaling more effectively than the common K allele variant (LYS 121). *Diabetes* 50, 831-836, 2001.
13. Lopez de Silanes I. et al. Identification of a target RNA motif for RNA-binding protein HuR. *PNAS* 101, 2987-2992, 2004.

- 14.** Frittitta, L. et al. A cluster of three single nucleotide polymorphisms in the 3'-untranslated region of human glycoprotein PC-1 gene stabilizes PC-1 mRNA and is associated with increased PC-1 protein content and insulin resistance-related abnormalities. *Diabetes* 50, 1952-1955, 2001.
- 15.** Marucci A et al. Role of GALNT2 in the modulation of ENPP1 expression and insulin signaling and action GALNT2: a novel modulator of insulin signaling. *Biochimica et biophysica Acta*, 2013.
- 16.** Gal-Mark N. et al. The pivotal role of TIA proteins in 5' splice-site selection of alu exons and across evolution. *PLoS Genet*, 5(11), 2009.
- 17.** Page-McCaw P.S. et al. PUF60: A novel U2AF65- related splicing activity. *RNA*, 5(12), 1548-60, 1999.
- 18.** Hastings M.L. et al. Control of pre-mRNA splicing by the general splicing factors PUF60 and U2AF(65). *PLoS One*, 2(6), 2007.
- 19.** Lijima T. et al. SAM68 regulates neuronal activity-dependent alternative splicing of neurexin-1. *Cell*. 147(7):1601-14, 2011.
- 20.** Huot M.E. et al. The Sam68 STAR RNA-binding protein regulates mTOR alternative splicing during adipogenesis. *Mol. Cell*.;46(2):187-99, 2012.

4. Messa a punto della metodologia ChIP/MS-MS per lo studio della proteina CBX7.

4.1 CBX7 e la carcinogenesi tiroidea

La tiroide è la più grande ghiandola endocrina umana [1]. E' situata nella regione del collo, sulla superficie anteriore della trachea, ed è costituita da due differenti tipi di cellule, le follicolari (TFC) e le parafollicolari o cellule C [2]. Più del 95% dei carcinomi tiroidei deriva dalle cellule follicolari [1] e sono le più comuni neoplasie endocrine, con circa 2500 casi stimati ogni anno negli Stati Uniti. Di contro, solo il 3% dei tumori tiroidei, definiti carcinomi tiroidei midollari, è di origine parafollicolare. I tumori tiroidei derivano dall'accumulo di diverse modificazioni nei geni critici coinvolti nel controllo della proliferazione cellulare. Le neoplasie tiroidee rappresentano un buon modello per lo studio degli eventi multistep coinvolti nella carcinogenesi, poiché esse comprendono un ampio spettro di lesioni con differenti gradi di malignità, che sono diagnosticate sulla base di parametri istologici e clinici. I tumori tiroidei derivanti da cellule follicolari includono: 1) adenomi benigni, che non sono invasivi e molto bene differenziati; 2) carcinomi, che sono divisi in ben differenziati, poco differenziati (PDTC) e non differenziati (ATC). I carcinomi tiroidei ben differenziati sono papillari (PTC) che rappresentano più del 70% dei tumori tiroidei maligni e follicolari (FTC) che rappresentano il 10% dei carcinomi tiroidei. Questi tumori, essendo differenziati, hanno una buona prognosi. [1,3] PDTC e ATC sembrano derivare dalla progressione di carcinomi differenziati [4]. Sebbene gli ATC rappresentano il 2-5% dei tumori tiroidei maligni, essi sono tra le neoplasie umane più letali che crescono rapidamente, sono molto aggressivi e sempre fatali. I PDTC invece rappresentano il 7% dei tumori tiroidei, sono forme intermedie tra i ben differenziati e i non differenziati, sia morfologicamente che dal punto di vista comportamentale. Sebbene sono stati sviluppati vari approcci terapeutici nella pratica clinica, la maggior parte di essi sono fatali. Quindi, la scoperta di nuovi approcci per la diagnosi del cancro a stadi precoci e per stabilire terapie migliori è un problema critico e urgente. Per arrivare a questo scopo, l'identificazione e la caratterizzazione di molecole chiave che partecipano nella carcinogenesi sono steps essenziali. Studi di microarray sono ampiamente usati per definire situazioni diagnostiche e prognostiche nel cancro e hanno condotto all'identificazione di una lunga lista di geni regolati nei carcinomi [5]. Anche per identificare i geni coinvolti nel processo di carcinogenesi tiroideo è stato usato questo approccio, che ha condotto all'identificazione di

geni la cui espressione è up- o down- regolata in linee cellulari di carcinomi paragonata con le culture cellulari primarie di origine tiroidea normale [6]. Tra i geni down-regolati è stato identificato CBX7. Questa proteina appartiene alla famiglia chromobox, codifica per una nuova proteina polycomb e la sua espressione mostra una riduzione progressiva, ben correlata con il grado di malignità delle neoplasie tiroidee. Inoltre, i livelli della proteina diminuiscono in una percentuale crescente di casi andando dagli adenomi benigni ai PTC, FTC e gli ATC. La correlazione tra la bassa espressione di CBX7 e gli istotipi più aggressivi sembra poter essere applicata anche ai carcinomi al seno, ovaie e intestino e sembra ridurre la sopravvivenza nel carcinoma al colon [7].

4.1.2 CBX7 e i geni Polycomb

Il gene *cbx7* (Chromobox homolog 7) è down-regolato nei carcinomi tiroidei e codifica per la proteina CBX7 che fa parte del complesso Polycomb [8]. I geni del gruppo Polycomb (PcG) collegano le modifiche degli istoni con la repressione trascrizionale [9,10]. Le proteine PcG sono parte di due complessi ad alto peso molecolare chiamati Polycomb repressive complexes 1 e 2, PRC1 e PRC2 (Figura 4.1).

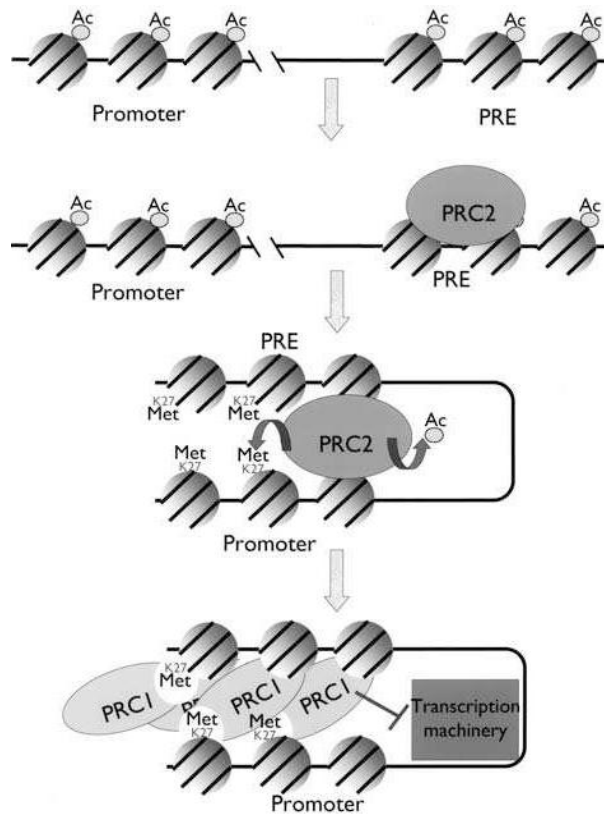


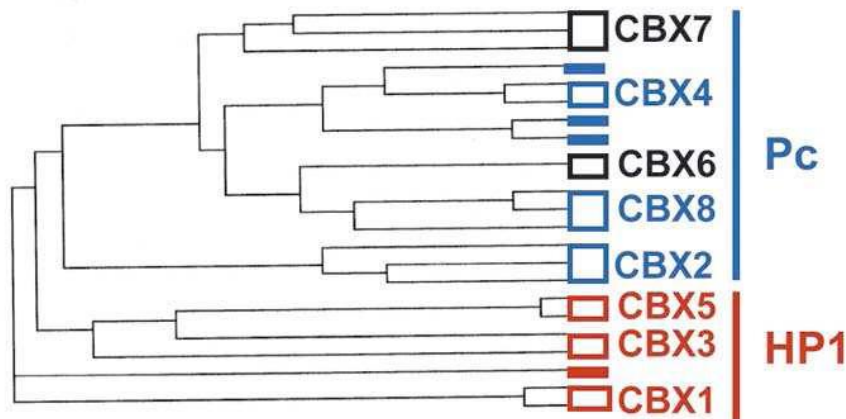
Figura 4.1 Meccanismo di repressione trascrizionale operato dalle proteine del gruppo Polycomb.

Il legame del Polycomb repressive complex 2 (PRC2) al Polycomb responsive elements (PRE) attiva la deacetilazione dell'istone e la metilazione sulla Lisina 27 dell'istone H3, entrambi nella regione PRE e del promotore dei geni silenziati. Questa modifica costituisce uno dei segnali per il reclutamento del Polycomb repressive complex 1 (PRC1), o complesso di mantenimento, che reprime l'espressione genica mediante l'inibizione dell'inizio della trascrizione.

Il complesso PRC1 riconosce e lega la lisina trimetilata in posizione 27 dell'istone H3 mediante le proteine Polycomb e contribuisce alla repressione genica inibendo l'inizio della trascrizione [11]. I membri del *core* di questo complesso nei mammiferi sono gli omologhi di *Drosophila* delle proteine Polycomb (Pc), Posterior sex combs (Psc), Sex combs extra (Sce) e Polyhomeiotic (Ph). Il complesso PRC2, invece, è associato alle attività istone deacetilasi e istone metiltransferasi (specifiche per la lisina 27 dell'istone H3) [12,13]. Tra i membri di questo secondo complesso ci sono Enhancer of Zeste (EZH2), Early embryonic deficient (EED), Suppressor of Zeste (SUZ12) e altre proteine associate. È ben noto che la deregolazione delle proteine PcG è coinvolta nella tumorigenesi. Molti studi infatti dimostrano che, ad esempio, Bmi-1 risulta amplificato nei tumori del sangue e nei medulloblastomi [14-17]; EZH2 è coinvolto nei tumori del sangue [14-18], e la sua espressione risulta incrementata nelle metastasi dei tumori alla prostata ed al seno [19,20];

alterazioni di Rae28/Ph1 sono osservate nella leucemia linfoblastica acuta [21] e SUZ12 risulta up-regolata nei tumori al colon[22]. Evidenze recenti mostrano che i geni PcG sono una nuova classe di oncogeni e antioncogeni, che potrebbero in futuro diventare centrali per lo sviluppo di nuove terapie del cancro basate su silenziamento genico[23,24]. I geni PcG come Bmi-1 ed EZH2 sono capaci di trasformazione cellulare, ma sono anche vitali per la sopravvivenza cellulare. Ad oggi, l'espressione anomala dei geni PcG è stata descritta nella maggior parte dei tumori umani. Inoltre, la correlazione tra l'espressione dei geni PcG e il comportamento biologico di sottotipi clinicamente definiti di cancro suggerisce che questi geni svolgono un ruolo centrale nell'oncogenesi, e possono essere presi in considerazione per la scoperta di nuovi marcatori diagnostici[25].

CBX7 e le altre proteine PcG contengono un chromodominio all'estremità N-terminale (CBX7 tra gli aminoacidi 10 e 46). Questo dominio fu originariamente identificato in *Drosophila melanogaster* come una regione di 37 aminoacidi omologa ad heterochromatin protein 1 (HP1) and Polycomb (Pc) proteins[26]. Analisi filogenetiche e di sequenza del chromodominio (Figura 4.2) hanno rivelato che CBX7 mostra una grande similitudine ad altre proteine note o putative del gruppo Pc come CBX2 (Pc1), CBX4 (Pc2), CBX6 and CBX8 (Pc3). Questa similitudine è meno pronunciata rispetto alle proteine HP1, come CBX1 (HP1 β), CBX3 (HP1 γ), CBX5 (HP1 α) e, inoltre, CBX7 non contiene il “*chromo-shadow domain*” che è una caratteristica delle proteine HP1. CBX7 contiene anche altri specifici residui molto conservati che permettono di distinguere i chromodomini Pc e HP1: questi residui specifici delle proteine Pc sono necessari sia per la dimerizzazione delle proteine Pc che per il riconoscimento della Lys 27 trimetilata sull'istone H3 [27]. Oltre ad avere un'alta omologia entro il chromodominio, CBX7 è anche omologa alle proteine Pc nella regione carbossi-terminale (Figura 4C) definito Pc-box o C-box [28]. E' noto che il Pc box (aminoacidi 231-243) e il chromodominio sono entrambi necessari per la funzione di CBX7. E' stato visto che la proteina di fusione tra CBX7 e il DNA-binding domain di Gal4 è capace di reprimere, in modo dose-dipendente, la trascrizione del gene reporter 4xGal4-tk-luc, suggerendo che CBX7 funziona come repressore della trascrizione [8]. Così, CBX7 sembra agire come altre proteine Pc esercitando il suo effetto mediante la repressione trascrizionale.

A**B**

```

hcbx7  9  QVFAVESIRKKRVRKGGKVEYLVKWKGWPPKYSTWEPEEHI  48
hcbx4 10 -VFAVESIEKKRIRKGRVEYLVKWRGWSPKYNTWEPEENI  48
hcbx6  9  RVFAAESIIKRRIRKGRIEYLVKWKGWAIKYSTWEPEENI  48
hcbx8  9  RVFAAEALLKRRIRKGRMEYLVKWKGWSQKYSTWEPEENI  48
hcbx2 10  QVFAAECILSKRLRKGKLEYLVKWRGWSSKHNSWEPEENI  49

```

C

```

hCBX8  364 TDVTSNFLTVTIKE  377
hCBX6  371 TDVTSNLLTVTIKE  384
hCBX2  269 TDVTANLITVTVKE  292
hCBX4  541 TDVTANCLTVTFKE  554
hCBX7  231 TDITANSITVTFRE  244

```

Figura 4.2 CBX7 e chromodominio

Identificazione di CBX7 come proteina appartenente alla famiglia Polycomb. (A) Analisi filogenetica delle proteine Chromobox (CBX). Le linee raggruppate mediante rettangoli aperti denotano ortologi della stessa protein CBX (nome a destra). I rettangoli pieni si riferiscono a proteine che non possono essere considerate puramente ortologi di una qualsiasi delle proteine CBX. Le proteine Pc sono mostrate in blu, mentre le proteine HP1 in nero. (B) Allineamento del chromodominio di CBX7 con quello di altre Polycomb umane note e putative. Residui identici sono mostrati in giallo, mentre residui simili in azzurro. (C) Allineamento della porzione Polycomb di CBX7 con quello di altre Polycomb umane note e putative. I colori sono descritti in (B).

CBX7 è anche capace di interagire con vari membri del gruppo Pc. Infatti, come altre proteine Pc, CBX7 è capace di interagire con se stessa e con la proteina ring-finger Ring1 [8], come riportato da esperimenti di Glutathione S-transferase (GST) pull-down [29,30]. Inoltre, nel nucleo di diverse linee cellulari, CBX7 localizza con Ring1 in strutture distinte chiamate Pc-bodies [31], come dimostrato da studi di immunofluorescenza. In topo è stato

dimostrato che il chromodominio di CBX7 è essenziale per il legame a H3K9me3 e H3K27me3, in alcuni casi associato con lo stato repressivo. E' stato anche trovato che, *in vitro*, il cromodominio di CBX7 può legare RNA e che, *in vivo*, l'interazione di CBX7 con la cromatina, in particolare con il cromosoma X inattivo, dipende parzialmente dalla sua associazione con l'RNA [32]. CBX7 è altamente espresso in parecchi tessuti, quali il cervello, cuore, muscoli scheletrici e reni [8]. Uno studio recente ha dimostrato che il gene *cbx7* è drasticamente ridotto in sei linee cellulari di carcinoma tiroideo rispetto a normali cellule tiroidee. Il ripristino dell'espressione di CBX7 nelle cellule cancerose tiroidee riduce la loro velocità di crescita, indicando che CBX7 gioca un ruolo critico nella regolazione della proliferazione delle cellule tiroidee trasformate [7]. Il presente progetto di tesi ha avuto come obiettivo lo studio funzionale del fattore trascrizionale CBX7. Si è pensato di disegnare una strategia sperimentale che consentisse, in un'unica procedura, di identificare i complessi proteici di CBX7 operanti sul DNA e contestualmente le regioni del DNA coinvolte in tali interazioni. A tale scopo è stato messo a punto un opportuno protocollo di ChIP opportunamente modificato che ha consentito in ultima analisi di identificare sia le proteine legate al DNA che le porzioni di DNA coinvolte nell'interazione.

4.2 Materiali e Metodi

4.2.1 Materiali

Il detergente usato per l'estrazione proteica il Triton X-100 della BIORAD.

Le beads utilizzate per la Chromatin Immunoprecipitation sono le Dynabeads PROTEIN A, della INVITROGEN.

L'anticorpo V5 tag Antibody è della ABCAM; l'anticorpo Anti-Acetyl-Histone H3 è della Millipore.

I prodotti della SIGMA utilizzati sono:

il ditiotreitolo (DTT), la tripsina, il Na-deoxycholate, il N-lauroylsarcosine, l'Hepes l'Acido etilendiammintetraacetico (EDTA) e l'NP-40:

I materiali utilizzati per le tecniche elettroforetiche, Mini-Protean TGX 4-12% e la miscela di proteine a peso molecolare noto "total blue" usata come standard nella migrazione elettroforetica sono della BIO-RAD.

La iodoacetamide (IAM), l'ammonio bicarbonato, la fenilmetilsulfonilfluoruro (PMSF) e l'Acido etilenglicol bis (2 amminoetil) N,N,N',N'-tetraacetico (EGTA) sono della FLUKA.

Gli inibitori di proteasi sono della ROCHE.

4.2.2 Metodi

4.2.2.1 Colture cellulare

Le cellule utilizzate negli esperimenti eseguiti durante questo lavoro di tesi sono state fornite dal gruppo del prof. Alfredo Fusco, con il quale si è collaborato per lo sviluppo del progetto.

Le cellule utilizzate sono cellule HEK 293 sia tal quali che trasfettate col costrutto di espressione del gene CBX7 recante all'estremità amminotermine il peptide v5 (necessario per purificare i complessi dei partners molecolari di CBX7).

4.2.2.2 Crosslink proteine nucleari/cromatina

Al fine stabilizzare i legami nativi proteina-DNA, le cellule HEK 293 sia trasfettate con CBX7 (campione) che non (controllo), sono state trattate con *Formaldehyde stock solution* (50 mM HEPES pH 7.5, 100 mM NaCl, 1mM EDTA, 0.5 mM EGTA, 11% formaldeide w/v). La *Formaldehyde stock solution* è stata aggiunta alle cellule in modo che la concentrazione finale di formaldeide fosse pari allo 0,75%. Le cellule sono state tenute sotto leggera agitazione su pedana basculante per 20 minuti a temperatura ambiente. Per bloccare

la fase di crosslink, ad opera della formaldeide, al termine dei 20 minuti alle cellule è stata aggiunta, in rapporto 1:20, della glicina 2,5 M. A questo punto le cellule sono state lavate, per 3 volte, con PBS freddo e centrifugate per 5 min a 1350 rpm ed a 4°C. Al termine della centrifugata è stato rimosso il surnatante.

4.2.2.3 Lisi cellulare frazionata

Per ottenere l'estratto nucleare, le cellule sono risospese, in rapporto di 10 ml per 100 milioni di cellule, nel *Lysis Buffer 1* (50 mM HEPES pH 7.5, 140 mM NaCl, 1mM EDTA, 10% glicerolo, 0.5% NP-40, 0.25% TritonX 100, 1X inibitori di proteasi). Le cellule sono state tenute in agitazione su pedana basculante per 10 minuti a 4°C e poi centrifugate per 5 min a 4°C a 1350 rpm. Al termine è stato recuperato il surnatante costituito dalle sole proteine citosoliche. A questo punto per allontanare dal pellet parte dei detergenti utilizzati nel primo step della lisi, è stato aggiunto, in rapporto di 10 ml per 100 milioni di cellule, il *Lysis Buffer 2* (10 mM Tris-HCl pH 8, 200 mM NaCl, 1mM EDTA, 0.5 mM EGTA, 1X inibitori di proteasi). Le cellule sono state tenute in agitazione su pedana basculante per 10 minuti a 4°C e poi centrifugate per 5 min a 4°C a 1350 rpm. Al termine si è proceduto con l'allontanamento del surnatante. Il pellet così ottenuto, costituito essenzialmente da cromatina e proteine nucleari, è stato risospeso, in rapporto di 3ml per 100 milioni di cellule, nel *Lysis Buffer 3* (10 mM Tris-HCl pH 8, 100 mM NaCl, 1mM EDTA, 0.5 mM EGTA, 0.1% Na-Deoxycholate, 0.5% N-lauroylsarcosine, 1X inibitori di proteasi).

4.2.2.4 Isolamento di complessi proteici mediante tecniche di Chromatin Immunoprecipitation

Per isolare i componenti proteici, coinvolgenti CBX7, presenti nel complesso proteina-DNA, è stato eseguito un esperimento di Chromatin Immunoprecipitation (ChIP), sfruttando la presenza, all'estremità amminoterminale della proteina esca, del peptide v5, per il quale sono disponibili commercialmente anticorpi che permettono di purificare i complessi.

- ***Sonicazione***

Al fine di ottenere frammenti di cromatina di 400-600 bp, l'estratto nucleare, sia del campione che del controllo, contenuto nel *Lysis Buffer 3*, è stato sottoposto a diversi cicli di sonicazione a 4°C. Il protocollo ha previsto che ciascun ciclo di sonicazione, della durata di 10 min, fosse condotto 30 sec ON / 1 min OFF. Al termine di ogni ciclo è stato prelevato un

volume pari ad 1/15 del campione, si è proceduto dapprima con il de-crosslink delle proteine dalla cromatina (1 h a 65°C) ed in seguito con l'estrazione e purificazione del DNA; quest'ultimo è stato caricato su gel di agarosio all'1.3% al fine di valutare la lunghezza dei frammenti ottenuti. I frammenti oligonucleotidici della lunghezza desiderata sono stati ottenuti dopo tre cicli di sonicazione. A questo punto sia al campione che al controllo è stato aggiunto del Triton X-100, con concentrazione finale pari all'1%. I campioni sono stati centrifugati per 10 min a 4°C ed a 13000 rpm. Al termine è stato recuperato il surnatante ed è stata determinata la concentrazione proteica mediante saggio Bradford, utilizzando il reattivo "Bio Rad Protein Assay". Come proteina di riferimento per costruire la retta di taratura è stata utilizzata l'Albumina di siero bovino. Prima di procedere con l'immunoprecipitazione il 3% in volume della cromatina, sia del campione che del controllo, è stato tenuto da parte; ciò ha consentito, in seguito, di normalizzare i risultati derivanti dalla Real Time PCR.

- ***Lavaggi beads magnetiche coniugate con la proteina A***

Sono stati eseguiti 3 lavaggi delle beads magnetiche coniugate con la proteina A, con la *Blocking Solution* (1x PBS, 0.5% BSA (W/V)) per allontanare la soluzione in cui sono di norma conservate. Terminati i lavaggi le beads sono state risospese in 600 µL di *blocking solution* e l'incubazione è stata condotta O.N. su ruota a 4°C.

- ***Incubazione Anticorpo anti V5 con complessi cromatina-proteine***

Parallelamente ai lavaggi delle beads, si è proceduto all'incubazione dell'estratto nucleare, sia del campione che del controllo, con l'anticorpo anti v5, utilizzando il seguente rapporto: 300 milioni di cellule/20 µg di anticorpo. L'incubazione è stata condotta O.N. su ruota a 4°C.

- ***Immunoprecipitazione della cromatina***

Gli immunocomplessi sono stati incubati con la resina coniugata con la proteina A, dalla quale era stata precedentemente allontanata la *Blocking Solution*. L'incubazione è stata eseguita a 4° C su ruota per 3 ore. Al termine il soprannatante è stato allontanato, mentre la resina è stata lavata tre volte con 1 ml di *Washing Buffer A* (1% TritonX-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl pH 7.6); ciascun lavaggio è stato condotto a 4°C su ruota per 5 minuti. Infine la resina è stata lavata un'ultima volta con 500 µL di *Washing Buffer B* (1% TritonX-100, 2 mM EDTA, 300 mM NaCl, 20 mM Tris-HCl pH 7.6). Il lavaggio è stato condotto a 4°C su ruota per 7 minuti, al termine si è proceduto al recupero del

surnatante. A questo punto 1/5 del surnatante è stato utilizzato per l'estrazione del DNA immunoprecipitato, dai restanti 4/5 sono stati eluiti i complessi proteici.

- ***Eluizione delle proteine immunoprecipitate***

L'eluizione dei complessi proteici immunoprecipitati in seguito all'esperimento di ChIP è avvenuta sui 400 µL di beads+cromatina/complessi proteici, in *Washing Buffer B*. L'eluato è stato posto su di un apposito supporto per beads magnetiche in modo da poter eliminare il surnatante. Sulle beads sono stati aggiunti 35 µL di *de crosslinking/elution buffer 1X* (250mM Tris-HCl pH 8.8, 2% SDS, 0.5 M β mercaptoetanol). La reazione è stata condotta a 95°C per 20 minuti. Al termine si è proceduto con una centrifugazione di 1 min a 13000 rpm ed a 4°C, è stato recuperato il surnatante e vi è stato aggiunto del *Loading buffer 2 X* (100 mM Tris/Cl PH 6.8, 4% SDS, 0.2% blu di bromofenolo, 20% glicerolo) contenente 200 mM DTT che funge da agente riducente.

- ***Purificazione del DNA immunoprecipitato***

La purificazione del DNA immunoprecipitato è avvenuta sui 100 µL di beads+cromatina/complessi proteici, in *Washing Buffer B*. L'eluato è stato posto su di un apposito supporto per beads magnetiche in modo da poter eliminare il surnatante. Alle beads è stato aggiunto 1 ml di TE 1X contenente 50 mM di NaCl. Le beads sono state dapprima risospese e poi centrifugate per 3 minuti a 200 rpm ed a 4°C; il surnatante è stato allontanato. A questo punto alle beads sono stati aggiunti 250 µL di *Elution Buffer* (TE 1X, 2% SDS), la reazione è stata condotta sotto agitazione per 30 minuti ed a 65°C. Al termine si è proceduto con una centrifugazione di 1 min a 13000 rpm ed a temperatura ambiente. L'eluato è stato recuperato e posto a 65°C O.N: in modo da decrosslinkare le proteine dalla cromatina. Il giorno seguente si è proceduto con l'estrazione del DNA con Fenolo/Cloroformio/ alcol isoamilico.

- ***Real Time-PCR***

Il DNA ottenuto è stato utilizzato per eseguire delle Real Time-PCR utilizzando i primer di geni target, ovvero di geni di cui è nota o si presuppone l'interazione con CBX7. In particolare i geni target contro cui è stata effettuata la Real Time-PCR sono: E-caderina, Ciclina E, P16. Come controllo negativo è stata eseguita una Real Time-PCR utilizzando il primer di GAPDH. Inoltre i valori derivanti dalla Real Time-PCR della cromatina INPUT

sono stati utilizzati per effettuare la normalizzazione dei valori di ottenuti attraverso la seguente formula:

$$\% \text{ Input} = 2^{\text{DCt}} \times 3; \text{DCt} = \text{Ct}(\text{input}) - \text{Ct}(\text{cIP}).$$

SDS-PAGE

Capitolo 3, paragrafo 3.2.2.4

Colorazione

Capitolo 3, paragrafo 3.2.2.6

Idrolisi in situ

Capitolo 2, paragrafo 2.2.2.6

Analisi mediante LC-MS/MS

Capitolo 2, paragrafo 2.2.2.7.2

Western blotting

Capitolo 2, paragrafo 2.2.2.3

4.3 Risultati

4.3.1 Messa a punto del protocollo di ChIP/MS-MS

La ChIP è una metodologia largamente utilizzata in studi di biologia molecolare relativi alla struttura e alla regolazione della cromatina attraverso l'identificazione delle regioni del DNA legate *in vivo* da particolari fattori proteici, fattori trascrizionali, istoni, ecc. Il protocollo classico di ChIP prevede la stabilizzazione dei legami nativi proteina-DNA mediante l'impiego di reattivi cross-link seguita dall'immunoprecipitazione dell'intero complesso proteina-DNA utilizzando un anticorpo specifico per la proteina di interesse. Le regioni specifiche di DNA così purificate vengono identificate solitamente mediante amplificazione con PCR e/o sequenziamento genico. Sebbene una forte spinta all'applicazione di questa strategia è derivata dalla disponibilità commerciale di una vasta gamma di anticorpi impiegabili con buoni risultati nell'immunoprecipitazione della cromatina, fino ad oggi essa è stata applicata quasi esclusivamente all'analisi del DNA. Il limite intrinseco di questa metodologia è costituito dal fatto che nonostante via sia una notevole complessità sperimentale il numero di informazioni che è possibile ricavare da un esperimento di questo tipo è estremamente esiguo. Infatti solo alcune regioni di DNA, legate *in vivo* dal fattore proteico che viene immunoprecipitato, possono essere amplificate per PCR, in relazione alla disponibilità dei primers di cui è possibile usufruire. Oltretutto nella maggior parte dei casi l'amplificazione è effettuata per quei geni su cui sono state già elaborate delle ipotesi funzionali, tali per cui è possibile presupporre un legame con il fattore trascrizionale di interesse. Inoltre, è necessario che il fattore trascrizionale che lega specificamente il DNA sia noto per provvedere alla fase di immunoprecipitazione con un opportuno anticorpo. Quindi nella sua accezione classica questa strategia sperimentale può rispondere alla domanda: quale regione del DNA è legata da uno specifico fattore proteico mentre non può delucidare il fattore o i fattori proteici in grado di legare una specifica sequenza di DNA. Inoltre la ChIP classica non è in grado di investigare l'intero complesso multiproteico che viene immunoprecipitato insieme alla regione di DNA legata dalla proteina di interesse. Soltanto in questi ultimi anni, con l'avvento della proteomica funzionale, si è cominciato ad impiegare la ChIP come strumento per l'identificazione delle proteine che *in vivo* sono legate funzionalmente ad un tratto di DNA, e che quindi agiscono in maniera concertata alla proteina di interesse. Una prima applicazione in tal senso è stata realizzata nello studio delle modificazioni chimiche covalenti delle proteine istoniche (acetilazione, metilazione, fosforilazione, etc.) e la loro correlazione con alterazioni della

struttura cromatinica e dell'attività trascrizionale (epigenomica). Nell'ambito di questo lavoro di tesi, questa metodologia è stata opportunamente modificata al fine di avere non solo informazioni circa le specifiche sequenze oligonucleotidiche a cui si lega la proteina di interesse, ma di identificare i componenti proteici presenti nel complesso proteine-DNA immunoprecipitato, mediante l'impiego di metodologie di spettrometria di massa. In questo modo è possibile identificare tutte le proteine che *in vivo*, legandosi funzionalmente ad un tratto di DNA, agiscono in maniera concertata alla proteina di interesse. Questa innovativa strategia sperimentale consente di avere uno scenario completo, sia da un punto di vista genomico che proteomico, del sistema che si sta analizzando. Inoltre la lista dei potenziali interattori della proteina esca unitamente alla conoscenza delle regioni del DNA su cui complessi proteici identificati svolgono il proprio ruolo funzionale consente di effettuare delle ipotesi funzionali sui diversi ruoli svolti dalla proteina di interesse nell'ambito della regolazione genica. Queste ipotesi successivamente possono essere validate effettuando opportuni esperimenti funzionali. La ChIP e l'MS quindi possono essere viste come due strategie complementari per investigare i componenti epigenetici della cromatina.

4.3.2 Ottimizzazione della strategia e prove di eluizione

Al fine di definire le condizioni più opportune che consentissero di ottenere non solo le classiche informazioni genomiche derivanti da un esperimento ChIP, ma contestualmente anche informazioni proteomiche è stato effettuato un esperimento preliminare per verificare l'effettiva immunoprecipitazione di una proteina di interesse. In particolare poiché la ChIP, nella sua accezione classica, è utilizzata per lo studio di modifiche istoniche, la proteina da immunoprecipitare in questa fase di messa a punto della strategia è stato l'istone H3, nella sua forma acetilata. Per questo esperimento sono stati utilizzati 50 milioni di cellule umane HEK293. Si è proceduto in prima battuta con la stabilizzazione dei legami nativi proteine-DNA mediante l'utilizzo di formaldeide. Le cellule così trattate sono state poi sottoposte ad un particolare protocollo di lisi cellulare frazionata, costituito dall'utilizzo di tre tamponi di lisi differenti, che consentissero di allontanare le proteine citosoliche da quella nucleari stabilmente ancorate alla cromatina. Terminata la lisi si è passati ad una delle fasi più delicate dell'intero esperimento: la "frammentazione" della cromatina mediante sonicazione. In genere i frammenti di cromatina da ottenere in seguito alla frammentazione devono essere costituiti da 400-600 bp, frammenti di dimensioni più ridotte sarebbero troppo piccoli per essere cross linkati ad un complesso multiproteico. Pertanto dopo ogni ciclo di sonicazione, condotto a 4°C secondo il programma 30 sec ON / 1 min OFF, sono stati prelevati pochi μ L

di campione, successivamente è stato effettuato un de-crosslink delle proteine dalla cromatina seguito da uno step di purificazione del DNA. In questo caso per ottenere i frammenti della lunghezza desiderata sono stati eseguiti 6 cicli di sonicazione. Terminata questa fase il campione è stato centrifugato al fine di recuperare le proteine nucleari crosslinkate alla cromatina. A questo punto è stata prelevata una quantità pari al 3% in volume dell'estratto proteine-cromatina, ed è stato tenuto da parte; questo campione (input) consentirà di verificare l'effettiva presenza della proteina di interesse all'interno dell'estratto. Al contrario di quanto accade generalmente per gli esperimenti di *Affinity* ed *Immunoprecipitazione*, in questo caso la quantità di anticorpo da utilizzare è stata determinata sulla base dei milioni di cellule adoperate per l'esperimento. Il rapporto utilizzato è stato il seguente:

100 milioni di cellule/5 μ g anticorpo

Pertanto l'estratto nucleare è stato incubato O.N. con l'anticorpo diretto contro l'istone H3 acetilato. Parallelamente si è proceduto anche con il lavaggio ed il trattamento delle beads magnetiche coniugate con la proteina A. Si è scelto di utilizzare questa resina in quanto essa ha affinità per il tipo di anticorpo utilizzato per la ChIP, consentendo così l'isolamento del complesso multiproteico di H3. Il rapporto utilizzato è stato il seguente:

10 μ g anticorpo/100 μ L beads in soluzione

In questo caso sono stati prelevati 25 μ L di beads e sono stati incubati con la *Blocking Solution* O.N.. Successivamente si è proceduto all'allontanamento della *Blocking Solution* dalle beads ed al loro trasferimento nella provetta contenente l'estratto nucleare incubato con l'anticorpo. Terminata l'incubazione la provetta è stata posta sull'apposito supporto per beads magnetiche e si è proceduto con l'allontanamento dell'unbound. A questo punto la resina è stata lavata per tre volte *Washing Buffer A* ed un'ultima volta con *Washing Buffer B*. A questo punto sono stati utilizzati due metodi di eluizione. Il primo ha previsto l'aggiunta 20 μ L di *de crosslinking/elution buffer*. Terminata questa prima eluizione, si è proceduto con una seconda eluizione che ha previsto l'aggiunta 20 μ L di *Loading buffer 2 X* contenente 200 mM DTT. Entrambi gli eluati sono stati frazionati mediante elettroforesi monodimensionale per eseguire poi un esperimento di western blotting. Le proteine presenti sul gel sono state quindi trasferite su membrana di nitrocellulosa. Il segnale relativo alla proteina di interesse è stato rivelato mediante metodi chemiluminescenti. La lastra autoradiografica impressionata è mostrata in figura 4.3.

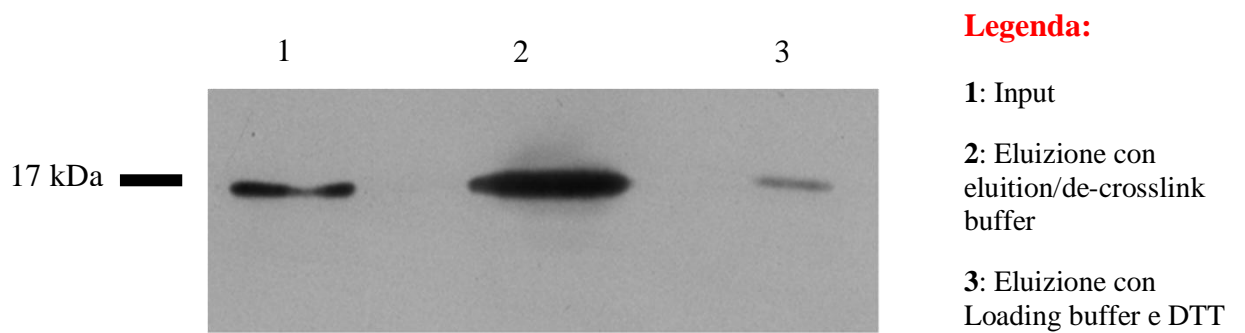


Figura 4.3 Western Blotting dell'esperimento analitico.

Il Western Blotting ha mostrato chiaramente non solo la presenza della proteina di interesse, dimostrando quindi che i rapporti sia di baeds che di anticorpo scelti per questo tipo di esperimento erano corretti, ma anche che l'eluizione migliore è stata ottenuta utilizzando *l'eluition/de cross-link buffer*.

4.3.3 ChIP/MS-MS Preparativa

L'esperimento preparativo di Chromatin Immunoprecipitation è stato effettuato utilizzando cellule umane HEK293 trasfettate in maniera transiente sia con un vettore di espressione codificante per la proteina CBX7 coniugata all'epitopo v5, sia con un vettore vuoto come controllo; in particolare per entrambe le linee cellulari sono stati utilizzati 300 milioni di cellule. La quantità di cellule è stata notevolmente incrementata rispetto alle quantità utilizzate nella fase di messa a punto, per consentire l'identificazione del complesso multiproteico interagente con CBX7 attraverso spettrometria di massa. In questo caso sono stati necessari tre cicli di sonicazione al fine di ottenere i frammenti oligonucleotidici della lunghezza desiderata. Come si può osservare dalla figura 4.4 il focus, sia per il campione (S) che per il controllo (C), è intorno alle 500 bp.

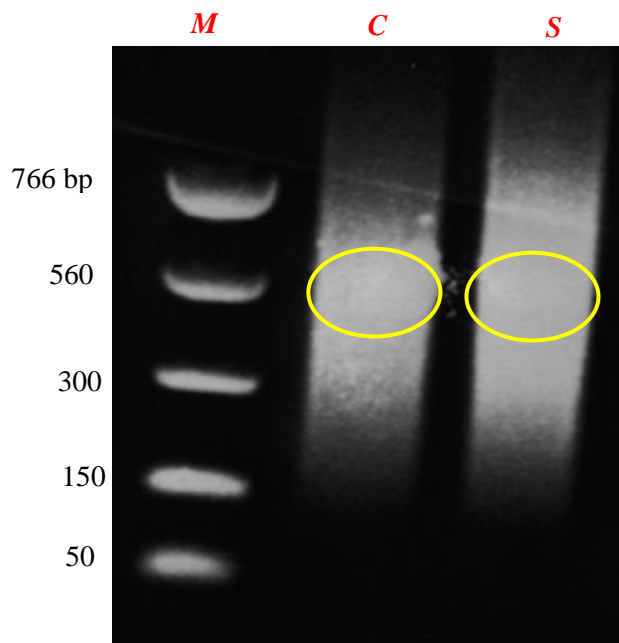


Figura 4.4 Gel d'agarosio per valutare la lunghezza dei frammenti cromatinici derivanti dalla sonicazione.

Terminata la fase di sonicazione sia il campione che il controllo sono stati centrifugati al fine di separare il pellet insolubile, costituito essenzialmente da membrane cellulari, dal surnatante contenente le proteine nucleari crosslinkate alla cromatina. Si è proceduto pertanto alla determinazione della concentrazione proteica mediante saggio Bradford. Per entrambi i campioni nucleari sono stati ottenuti 6 mL di estratto contenente circa 36 mg di proteine. Gli estratti nucleari, sia del campione che del controllo, sono stati incubati secondo il rapporto sopra citato, O.N. con l'anticorpo diretto contro l'epitopo v5. Parallelamente si è proceduto anche con il lavaggio ed il trattamento delle beads magnetiche coniugate con la proteina A. In questo caso sono stati prelevati 150 μ L di beads per ciascun campione, in accordo con il rapporto utilizzato durante la fase di messa a punto, e sono state incubate con la *Blocking Solution* O.N.. Successivamente si è proceduto all'allontanamento della *Blocking Solution* dalle beads ed al loro trasferimento nelle provette contenenti gli estratti nucleari incubati con l'anticorpo. Terminata l'incubazione le provette sono state poste sull'apposito supporto per beads magnetiche e si è proceduto con l'allontanamento dell'unbound. Le beads alle quali erano ancora immobilizzati i complessi DNA-proteine sono state risospese in *Washing Buffer B*; 1/5 del volume è stato utilizzato per l'estrazione del DNA da utilizzare per RT-PCR e i restanti 4/5 sono stati utilizzati per la purificazione dei complessi proteici immunoprecipitati. Sia i campioni destinati al recupero del DNA che quelli destinati all'eluizione dei complessi proteici sono stati posti sull'apposito supporto per

beads magnetiche; ciò ha consentito l'allontanamento dalle beads del buffer in cui erano risospese. A questo punto per l'estrazione del DNA le beads sono state lavate con TE 1X. Eliminato il tampone di lavaggio, le beads sono state risospese in *Elution Buffer*. Al termine si è proceduto al recupero del surnatante, costituito dai complessi cromatina/proteine. Per de-crosslinkare le proteine dal DNA il campione è stato posto O.N. a 65°C. In seguito si è proceduto all'estrazione del DNA mediante Fenolo/Cloroformio/Metanolo. Il DNA così ottenuto è stato utilizzato per eseguire degli esperimenti basati su amplificazioni con PCR e sequenze per l'eventuale identificazione di sequenze di DNA specificamente riconosciute da CBX7. I risultati sono riportati in figura 4.5. La procedura ha portato alla identificazione di sequenze appartenenti ai promotori dei geni E-caderina, p16 e Ciclina E.

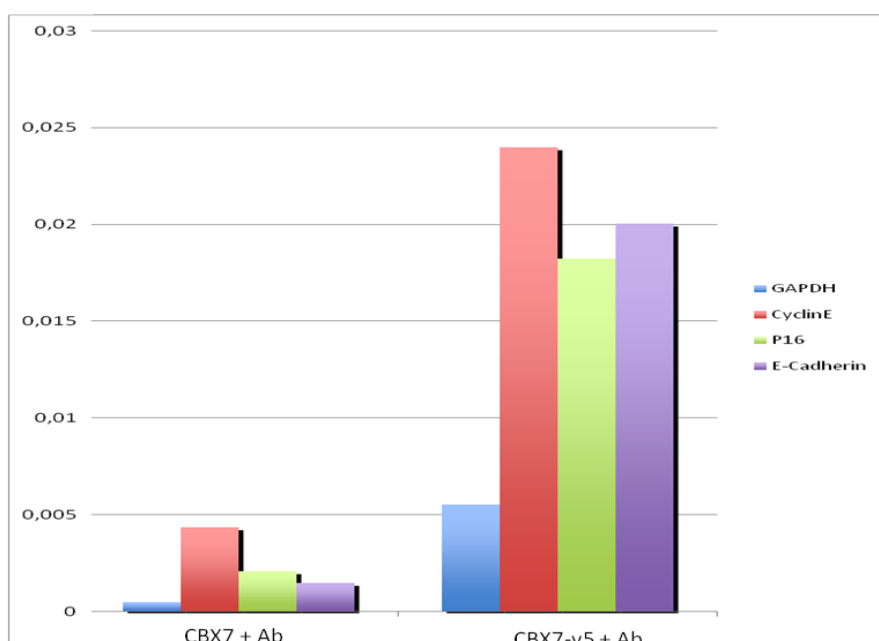


Figura 4.5: Istogrammi relativi alle Real-Time PCR eseguite sui frammenti cromatinici eluiti.

Per quanto concerne invece l'eluizione dei complessi multiproteici le beads sono state risospese in *De-crosslink/elution buffer*. Al termine si è proceduto al recupero del surnatante al quale è stato aggiunto del *Loading buffer* contenete DTT. L'avvenuta immunoprecipitazione di CBX7 è stata verificata mediante un esperimento preliminare di Western Blot,. Sono stati caricati, su gel di poliaccrilammide al 10%, 70µg di estratto cellulare nucleare (Input) sia del campione che del controllo ed una eguale quantità di entrambi gli estratti dopo la ChIP.

Il risultato ottenuto è riportato in Figura 4.6, che mostra chiaramente nel canale del campione immunoprecipitato, la presenza di un'intensa banda in corrispondenza del peso molecolare atteso per CBX7, una banda analoga, ma molto meno abbondante è presente anche nell'Input. L'arricchimento della proteina CBX7 nella ChIP verificata mediante western blotting ha confermato la riuscita dell'esperimento.

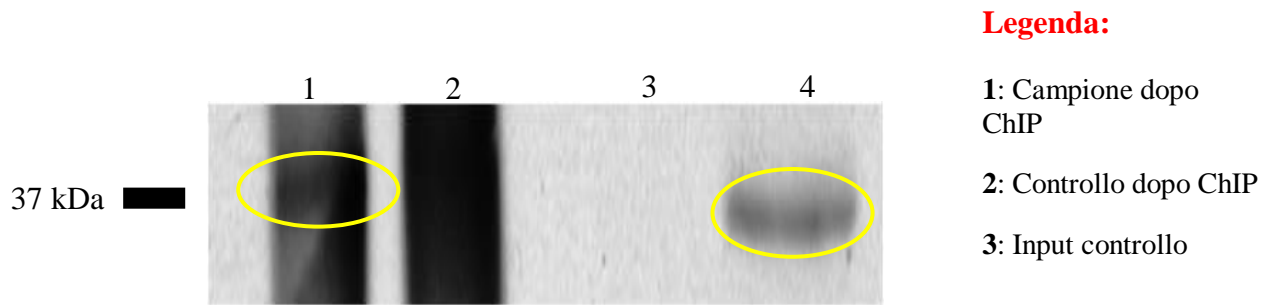


Figura 4.6 Western Blot per testare l'avvenuta immunoprecipitazione di CBX7.

Una volta effettuata la verifica analitica dell'esperimento di ChIP, la restante parte del campione e del controllo è stata frazionata mediante elettroforesi monodimensionale preparativa su gel di poliacrilammide al 4-12% in condizioni denaturanti (SDS-PAGE). Il gel ottenuto è stato poi colorato al Coomassie Colloidale Blue per la visualizzazione delle bande proteiche (Fig. 4.7).

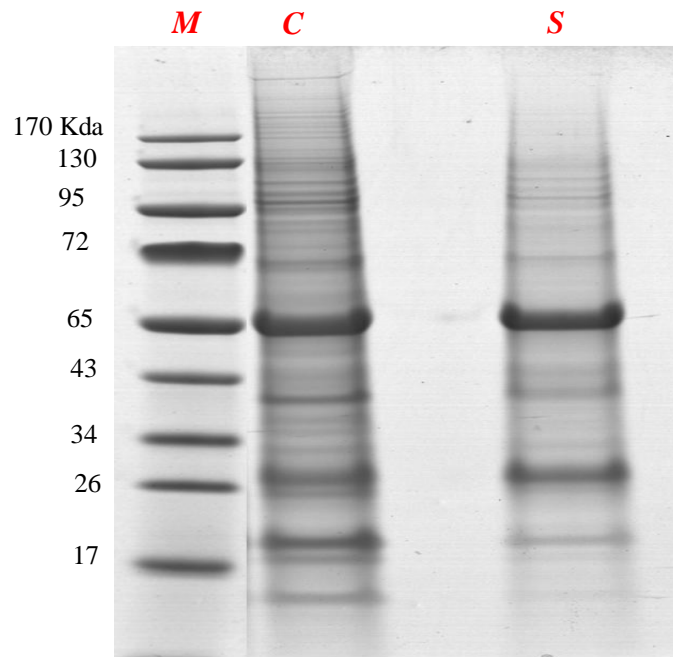


Figura 4.7 SDS-PAGE del campione immunoprecipitato (S) con il rispettivo controllo (C). M rappresenta una miscela di proteine a peso molecolare noto usata come standard.

Il gel è stato completamente tagliato, sono state ottenute 40 bande per ogni corsia, come mostrato in figura 4.8.

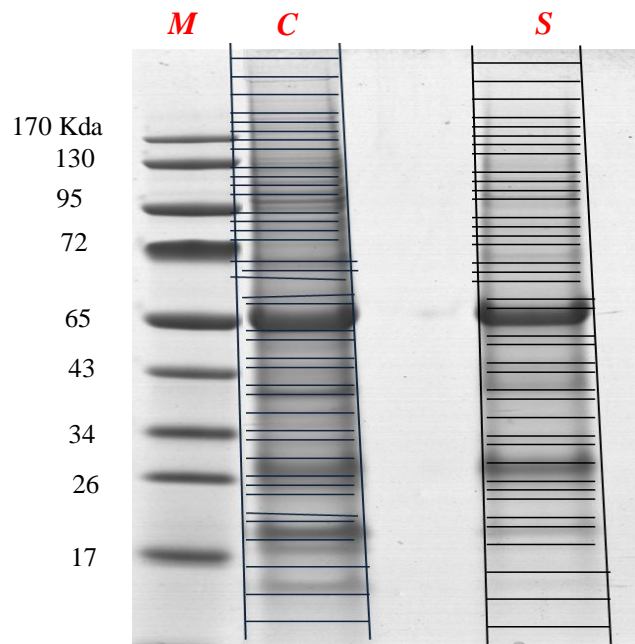


Figura 4.8: Bande escisse dal gel e sottoposte ad analici LC-MS/MS.

Le bande escisse dal gel, sono state sottoposte ad idrolisi in situ con tripsina e le miscele peptidiche ottenute sono state analizzate mediante LC-MS/MS. Le analisi del campione e del controllo sono state svolte in parallelo in modo da poter sottrarre ai campioni le proteine identificate anche nel controllo. Nella tabella 4.1 si riportano le proteine identificate.

Tabella 4.1 Proteine identificate in seguito all'esperimento di ChIP/MS-MS.

BANDA	PROTEIN ID	SCORE	PEPTIDI	MW (kDa)	SP code
14	gi 5174669 sex comb on midleg-like protein 2	101	3	78	Q9UQR0
19	gi 46411195 PIG48	270	7	61	P49368
	gi 179287 autoantigen	237	5	40	Q06265
	gi 1809248 siah binding protein 1	214	5	58	Q9UHX1
	gi 5729730 apoptosis inhibitor 5 isoform b	150	2	57	Q9BZZ5
	gi 71051977 HDAC2 protein	148	3	65	Q92769
	gi 5453607 T-complex protein 1 subunit eta isoform a	106	2	60	Q99832
	gi 4507947 tyrosyl-tRNA synthetase, cytoplasmic	74	2	59	P54577
20	gi 48762932 T-complex protein 1 subunit theta	234	5	60	P50990
	gi 7657381 pre-mRNA-processing factor 19	158	3	56	Q9UMS4
	gi 4504897 importin subunit alpha-2	99	4	58	P52292
	gi 4507115 fascin	80	2	55	Q16658
	gi 73760405 thymopoietin isoform beta	69	2	50	P42167
25	gi 7453577 protein arginine N-methyltransferase 1-variant 1 (PRMT1)	541	11	40	Q99873-3
	gi 460771 hnRNP-E1 (poly(rC)-binding protein 1	283	7	38	Q15365
	gi 181914 DNA-binding protein	105	3	36	Q14103-3
26	gi 28372505 chromobox protein homolog 7	348	9	28	O95931
	gi 19923973 BTB/POZ domain-containing protein KCTD12	293	8	36	Q96CX2
	gi 825671 B23 nucleophosmin (280 AA)	121	3	31	P06748
	gi 31645 glyceraldehyde-3-phosphate dehydrogenase	84	2	36	P04406
	gi 14141157 heterogeneous nuclear ribonucleoprotein H3 isoform a	90	3	37	P31942
	gi 55956919 heterogeneous nuclear ribonucleoprotein A/B isoform a	86	2	36	Q99729
	gi 14141166 poly(rC)-binding protein 2 isoform b	84	2	38	Q15366

Tra i vari interattori di CBX7 sono stati identificati la *Protein arginine N-methyltransferase 1* (PRMT1) nota per metilare i residui di arginina di numerosi target proteici, ed uno dei suoi substrati, l'*Heterogeneous nuclear ribonucleoprotein D0*. La presenza di PRMT1 è particolarmente significativa in quanto già in passato era stata identificata come interattore di CBX7, a seguito di diversi esperimenti di proteomica funzionale (Immunoprecipitazione, Co-immunoprecipitazione e separazione dei complessi proteici con Gradiente di glicerolo). Averla identificata anche stavolta è una ulteriore conferma dell'esito positivo di questa nuova strategia sperimentale; inoltre va ricordato che la metilazione del DNA è fondamentale nella regolazione della trascrizione. Uno dei principali substrati di PRMT1 è il residuo di arginina in posizione 3 dell'istone H4, in particolare PRMT1 dimetila in maniera asimmetrica questo residuo [33-36]. Per comprendere se in qualche modo CBX7 potesse

influenzare l'attività di PRMT1 la linea cellulare HEK 293, utilizzata nell'esperimento di ChIP/MS-MS è trasfettata in un caso con il siRNA di CBX7, in un altro caso con il siRNA di PRMT1 ed infine con il siRNA per entrambe le proteine, in modo da impedire l'espressione di una sola o di entrambe le proteine. Come controllo è stata utilizzata la medesima linea cellulare non trasfettata (nt). Le cellule così trattate sono state sottoposte a lisi cellulare frazionata, e sui tre differenti estratti nucleari è stato condotto un esperimento di wester blot, andando a verificare se la metilazione dell' istone H4 sul residuo di arginina 3 fosse influenzata dall'assenza di CBX7.

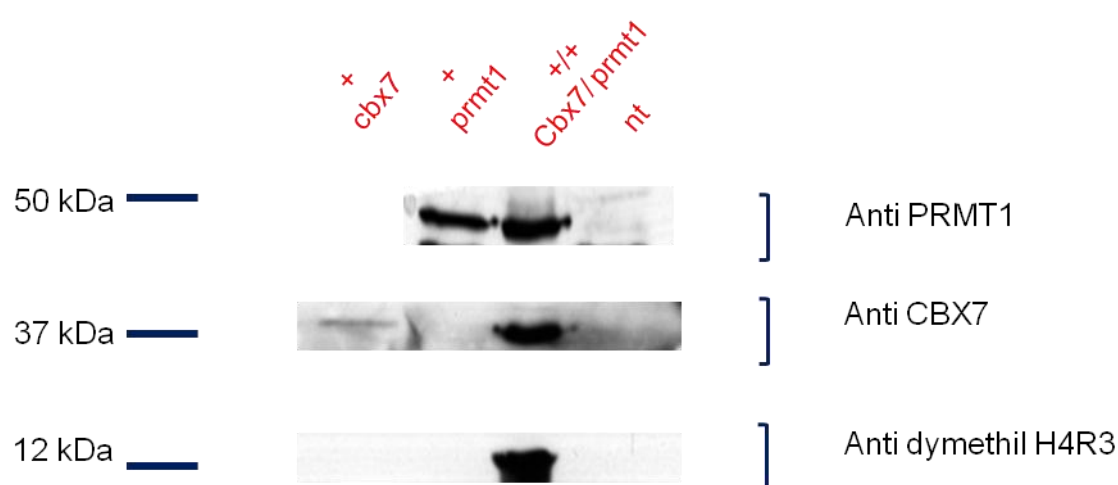


Figura 4.9 Western blot volto a validare l'effetto di CBX7 sull'attività di PRMT1.

Il dato mostrato in figura 4.9 dimostra che PRMT1 esplica la sua attività metiltrasferasica solo in presenza di CBX7.

Tra i risultati ottenuti l'altro dato particolarmente interessante è l'interazione tra CBX7 e l'*Histone deacetylase 2* (HDAC2), un noto repressore della trascrizione di vari geni importanti tra cui E-cadherin.

Infatti, la perdita di espressione di CBX7 potrebbe essere correlata con l'acquisizione dell'invasività accompagnata dalla perdita delle caratteristiche epiteliali e il guadagno del fenotipo mesenchimale, un processo conosciuto come transizione epitelio-mesenchimale. L'E-cadherin è il principale elemento dei nodi di adesione cellula-cellula, indispensabile per la normale morfologia epiteliale della cellula, emergendo perciò come caratteristica del fenotipo epiteliale. Nella maggior parte dei tumori, la down-regolazione dell'E-cadherin durante la progressione neoplastica avviene mediante meccanismi epigenetici, che includono la repressione trascrizionale, in alcuni casi mediata dall'attività di HDAC [37], e

l'ipermetilazione del promotore [38]. Solo in pochi casi sono state trovate mutazioni nel gene di E-cadherin che hanno portato all'assenza o all'espressione di una proteina non funzionale [39]. Per questo motivo è stata seguita l'espressione di E-cadherin in carcinomi tiroidei di differenti istotipi: è stata riscontrata un'evidente correlazione tra i livelli di espressione nei carcinomi tiroidei umani, tale per cui entrambi sono drasticamente down-regolate nei carcinomi tiroidei anaplastici paragonati al normale tessuto tiroideo. D'altra parte nei carcinomi tiroidei non è stata riscontrata ipermetilazione del promotore dell'E-cadherin. In questo lavoro di tesi è stato dimostrato, mediante Real Time-PCR condotta sui frammenti di cromatina immunoprecipitata, che *in vivo* CBX7 lega il promotore del gene dell'E-cadherin. Inoltre da studi condotti è stato verificato l'effetto dell'espressione di CBX7 sulla trascrizione di E-cadherin; cellule HEK 293 sono state transientemente cotrasfettate con un vettore di espressione codificante CBX7 e con un vettore reporter che contiene il gene della luciferasi sotto il controllo del promotore dell'E-cadherin. Come mostrato in figura 4.10 CBX7 aumenta l'attività trascrizionale del promotore dell'E-cadherin in modo dose-dipendente [40].

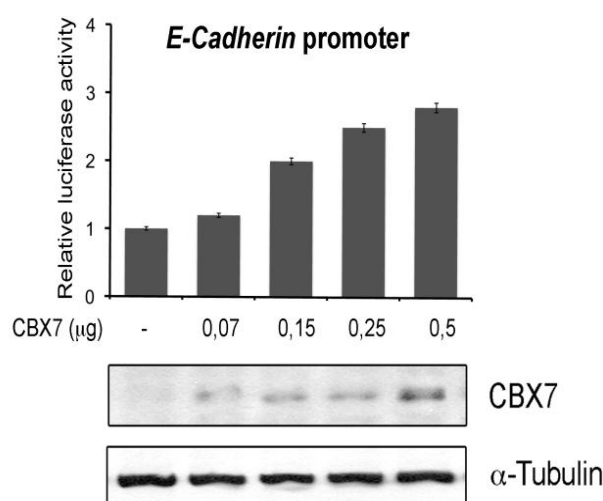


Figura 4.10 Analisi della risposta avuta incrementando la quantità di CBX7 sul vettore reporter dell'E-cadherin sotto il controllo della luciferase transientemente trasfettato in cellule HEK 293. L'analisi Western blot conferma l'incremento dell'espressione di CBX7. L'espressione seguita con α -Tubulin è servita come controllo per la quantità di proteina caricata.

4.4 Discussione

I tumori tiroidei derivano dall'accumulo di varie modifiche nei geni critici coinvolti nel controllo della proliferazione cellulare. Sebbene nella pratica clinica siano seguiti vari approcci terapeutici per la cura di questi, la maggior parte di essi non raggiunge esiti positivi. Così, la scoperta di nuovi approcci per diagnosticare il cancro in stadi precoci e per stabilire terapie efficaci rappresenta una questione critica e urgente. Per raggiungere questo scopo, l'identificazione e la caratterizzazione di molecole chiave che partecipano alla carcinogenesi sono punti essenziali. A tal proposito ci si è soffermati sul ruolo di CBX7, che potrebbe essere coinvolto nel processo di carcinogenesi tiroidea. CBX7 è espressa abbondantemente nella ghiandola tiroidea normale, ma la sua espressione diminuisce con la malignità e lo stadio neoplastico. Infatti, l'espressione di CBX7 è comparabile al tessuto tiroideo normale negli adenomi follicolari benigni, leggermente ridotta nei carcinomi tiroidei papillari e drasticamente ridotta, e in molti casi assente, nei carcinomi tiroidei follicolari. Si potrebbe così pensare che i livelli di CBX7 potrebbero servire per differenziare le neoplasie tiroidee benigne e maligne.

Allo scopo di studiare la funzione di CBX7 nelle cellule normali e capire perché è drasticamente down-regolato nelle cellule tumorali, abbiamo deciso di investigare le proteine che interagiscono con CBX7, mediante un innovativo approccio di proteomica funzionale. In particolare la metodologia da noi messa a punto, ChIP-MS/MS, ha consentito di ottenere, *in vivo*, non solo informazioni sui geni con cui CBX7 interagisce ma ci ha permesso di avere informazioni anche sulle proteine che interagiscono con essa. Tali approcci sono molto validi perché l'associazione di CBX7 con dei partners appartenenti a uno specifico complesso proteico coinvolto in un particolare meccanismo può dare suggerimenti sulla funzione biologica della proteina [41].

Alla luce dei risultati ottenuti dagli esperimenti funzionali volti alla validazione della diretta interazione di CBX7 e PRMT1 (Figura 4,9), è possibile ipotizzare che CBX7 possa esercitare la sua funzione sulla progressione del cancro alla tiroide anche attraverso l'interazione con PRMT1, modulando la metilazione di specifici istoni e quindi la trascrizione di geni che potrebbero essere coinvolti in fenomeni tumorali.

Questi dati lasciano ipotizzare, infatti, che l'interazione diretta delle due proteine ed in particolare la capacità di CBX7 di promuovere l'attività di PRMT1, sia uno dei meccanismi con cui probabilmente il suddetto fattore agisce da repressore trascrizionale. In questo scenario infatti la metilazione a carico dei residui di Arginina e Lisina degli istoni [33-36], va a stabilizzare ulteriormente la carica positiva di questi ultimi, aumentando

significativamente la loro interazione al DNA carico negativamente. Pertanto ciò impedisce la trascrizione genica. Infine in questo lavoro si è dimostrato non solo che CBX7 interagisce con la proteina HDAC2 inibendo la sua attività ma che questa interazione avviene sul promotore dell'E-cadherin. E' stata anche dimostrata la capacità di CBX7 di regolare positivamente l'espressione di E-cadherin interagendo con HDAC2 e inibendo la sua attività sul promotore dell'E-cadherin.

Tra gli altri geni legati *in vivo* da CBX7, la Real Time-PCR ha evidenziato anche la presenza del gene della *Ciclina E*. Recentemente è stato dimostrato che un'over-espressione di Ciclina E regola negativamente Cbx7, diminuendo l'espressione della proteina, in accordo con quanto avviene in presenza di carcinomi [42].

Infine particolarmente significativo è stato identificare nuove proteine coinvolte nella progressione tumorale, come ad esempio *Nucleophosmin* e *Sex comb on midleg-like protein 2* (SMCL2). Per quanto riguarda *Nucleophosmin*, altri esperimenti di CHIP, riportati in letteratura, hanno confermato l'elevata affinità di legame di questa proteina per il *core* istonico; inoltre studi recenti indicano che *Nucleophosmin* è in grado di stimolare *in vitro* la trascrizione [43, 44]. Particolarmente significativo è il fatto che questa proteina risulta essere over-espressa in cellule tumorali ed è in grado di esercitare un effetto oncogenico andando a legare ed inibendo l'attività di numerosi soppressori tumorali. Infine per quanto concerne SCML2 [45, 46], la sua identificazione risulta essere particolarmente interessante in quanto essa fa parte, insieme a CBX7, della famiglia delle chromobox ed è membro del polycomb repressive complex 1 che, insieme con il polycomb repressive complex 2, inibisce l'espressione di alcuni geni regolatori.

Bibliografia

1. Kondo T, Ezzat S, Asa SL. Pathogenetic mechanisms in thyroid follicular cell neoplasia. *Nat Rev Cancer* 2006;6:292–306.
2. De Felice M, Di Lauro R. Thyroid Development and Its Disorders: Genetics and Molecular Mechanisms. *Endocrine Reviews* 2004; 25:722–46.
3. Saltman B, Singh B, Hedvat CV, Wreesmann VB, and Ghossein R. Patterns of expression of cell cycle/apoptosis genes along the spectrum of thyroid carcinoma progression. *Surgery* 2006;140:899-905.
4. Van der Laan BF et al. The association of well-differentiated thyroid carcinoma with insular or anaplastic thyroid carcinoma: evidence for dedifferentiation in tumor progression. *Endocr Pathol* 1993;4:215-221.
5. Delys L, Detours V, Franc B, et al. Gene expression and the biological phenotype of papillary thyroid carcinomas. *Oncogene* 2007;26:7894–903.
6. Pallante P, Berlingieri, MT, Troncone, G, Kruhoffer M, Orntoft TF, Viglietto G, Caleo A, Migliaccio I, Decaussin-Petrucci, M, Santoro M, et al. . UbcH10 overexpression may represent a marker of anaplastic thyroid carcinomas. . *Br J Cancer* 2005;93: 464-71.
7. Pallante P, Federico A, Berlingieri MT, Bianco M, Ferraro A, Forzati F, Iaccarino A, Russo M, Pierantoni GM, Leone V, Sacchetti S, Troncone G, Santoro M, Fusco A. Loss of the CBX7 gene expression correlates with a highly malignant phenotype in thyroid cancer. *Cancer Res* 2008;68(16):6770-8.
8. Gil J, Bernard D, Martínez D, Beach D. Polycomb CBX7 has a unifying role in cellular lifespan. *Nat Cell Biol* 2004;6:67-72.
9. Otte AP, Kwaks TH. Gene repression by Polycomb group protein complexes: a distinct complex for every occasion? *Curr Opin Genet Dev* 2003;13:448–54.
10. Lund AH, van Lohuizen M. Polycomb complexes and silencing mechanisms. *Curr Opin Cell Biol* 2004;16:239–46.
11. Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y. Role of histone H3 lysine 27 methylation in Polycombgroup silencing. *Science* 2002;298:1039–43.
12. Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P, Reinberg D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev* 2002; 16: 2893-905.

13. Kirmizis A, Bartley SM, Kuzmichev A, Margueron R, Reinberg D, Green R, Farnham PJ. Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. *Genes Dev* 2004; 18: 1592-605.
14. Raaphorst FM, van Kemenade FJ, Blokzijl T, Fieret E, Hamer KM, Satijn DP, Otte AP, Meijer CJ. Coexpression of BMI-1 and EZH2 polycomb group genes in Reed-Sternberg cells of Hodgkin's disease. *Am J Pathol* 2000; 157: 709-15.
15. Beà S, Tort F, Pinyol M, Puig X, Hernández L, Hernández S, Fernandez PL, van Lohuizen M, Colomer D, Campo E. BMI-1 gene amplification and overexpression in hematological malignancies occur mainly in mantle cell lymphomas. *Cancer Res* 2001; 61: 2409-12.
16. Van Kemenade FJ, Raaphorst FM, Blokzijl T, Fieret E, Hamer KM, Satijn DP, Otte AP, Meijer CJ. Coexpression of BMI-1 and EZH2 polycombgroup proteins is associated with cycling cells and degree of malignancy in B-cell non-Hodgkin lymphoma. *Blood* 2001;97:3896–901.
17. Lessard J, Sauvageau G. Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. *Nature* 2003;423:255–60.
18. Visser HP, Gunster MJ, Kluin-Nelemans HC, Manders EM, Raaphorst FM, Meijer CJ, Willemze R, Otte AP. The Polycomb group protein EZH2 is upregulated in proliferating, cultured human mantle cell lymphoma. *Br J Haematol* 2001;112:950–8.
19. Varambally S, Dhanasekaran SM, Zhou M, Barrette TR, Kumar-Sinha C, N Sanda MG, Ghosh D, Pienta KJ, Sewalt RGAB, Otte AP, Rubin MA, Chinnaiyan AM. The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* 2002;419:624–9.
20. Klier CG, Cao Q, Varambally S, Shen R, Ota I, Tomlins SA, Ghosh D, Sewalt RGAB, Otte AP, Hayes DF, Sabel MS, Livant D, Weiss SJ, Rubin MA, Chinnaiyan AM. EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc Natl Acad Sci U S A* 2003;100:11606–11.
21. Tokimasa S, Ohta H, Sawada A, Matsuda Y, Kim JY, Nishiguchi S, Hara J, Takihara Y. Lack of the Polycomb-group gene *rae28* causes maturation arrest at the early B-cell developmental stage. *Exp Hematol* 2001;29:93–103.
22. Kirmizis A, Bartley SM, Farnham PJ. Identification of the polycomb group protein SU(Z)12 as a potential molecular target for human cancer therapy. *Mol Cancer Ther* 2003;2:113–21.
23. Nakao M, Minami T, Ueda Y, Sakamoto Y, Ichimura T. Epigenetic system: a pathway to malignancies and a therapeutic target. *Int J Hematol* 2004;80:103–107.

24. Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* 2004;429:457–63.
25. Raaphorst FM, Vermeer M, Fieret E, Blokzijl T, Dukers D, Sewalt RG, Otte AP, Willemze R, Meijer CJ. Site-specific expression of Polycombgroup genes encoding the HPC-HPH/PRC1 complex in clinically defined primary nodal and cutaneous large B-cell lymphomas. *Am J Pathol* 2004;164:533–42.
26. Paro R, Hogness DS. The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of *Drosophila*. *Proc Natl Acad Sci U S A* 1991;88:263–7.
27. Min J, Zhang Y, Xu R-M. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev* 2003;17:1823–8.
28. Jones DO, Cowell IG, Singh PB. Mammalian chromodomain proteins: their role in genome organisation and expression. *Bioessays* 2000;22:124–37.
29. Satijn DP, Otte AP. Polycomb group protein complexes: do different complexes regulate distinct target genes? *Biochim Biophys Acta* 1999;1447:1–16.
30. Satijn DP, Gunster MJ, van der Vlag J, Hamer KM, Schul W, Alkema MJ, Saurin AJ, Freemont PS, van Driel R, Otte AP. RING1 is associated with the polycomb group protein complex and acts as a transcriptional repressor. *Mol Cell Biol* 1997;17:4105–13.
31. Saurin AJ, Shiels C, Williamson J, Satijn DP, Otte AP, Sheer D, Freemont PS. The human polycomb group complex associates with pericentromeric heterochromatin to form a novel nuclear domain. *J Cell Biol* 1998;142:887–98.
32. Bernstein E, Duncan EM, Masui O, Gil J, Heard E, Allis CD. Mouse Polycomb Proteins Bind Differentially to Methylated Histone H3 and RNA and Are Enriched in Facultative Heterochromatin. *Mol Cell Biol* 2006;26:2560-9.
33. Di Lorenzo A, Bedford M.T. Histone Arginine methylation. *FEBS Lett*, 2011, 585(13): 2024-2031.
34. Yost M.J. et al. Targets in Epigenetics: Inhibiting the Methyl Writers of the Histone Code. *Current Clinical Genomics*, 2011, 5: 72-84.
35. Bedford M.T, Clarke S.G. Protein Arginine Methylation in Mammals: Who, What and Why. *Mol. Cell*, 2009.
36. Stallcup M.R. Role of protein methylation in chromatin remodeling and transcriptional regulation. *Oncogene*, 2001, 20:3014-3020.
37. Peinado H, Portillo F, Cano A. Transcriptional regulation of cadherins during development and carcinogenesis. *Int J Dev Biol* 2004; 48: 365-75.

38. Mick R, Ziober BL. c-Met expression in tall cell variant papillary carcinoma of the thyroid. *Cancer* 2003;98:1386–93.
39. Thiery JP, Sleeman J. Complex networks orchestrate epithelial-mesenchymal transitions. *Nat Rev Mol Cell Biol* 2006;7:131–42.
40. Ou JN, Torrisani J, Unterberger A, et al. Histone deacetylase inhibitor trichostatin A induces global and gene-specific DNA demethylation in human cancer cell lines. *Biochem Pharmacol* 2007; 73:1297–307.
41. Monti M, Orrù S, Pagnozzi D, Pucci P. Interaction proteomics. *Biosci Rep.* 2005; 25: 45-56.
42. Forzati F. et al. Tumor suppressor activity of CBX7 in lung carcinogenesis. *Cell Cycle* 2012, 11(10):1888-91
43. Murano K. et al. Transcription regulation of the rRNA gene by a multifunctional nucleolar protein, B23/nucleophosmin, through its histone chaperone activity. *Mol cel Biol*, 2008, 28(10):3114-26.
44. Ye K. Nucleophosmin/B23, a multifunctional protein that can regulate apoptosis. *Cancer Biol Ther*, 2005, 4(9): 918-23.
45. Montini E. et al. Identification of SCML2, a Second Human Gene Homologous to the *Drosophila* Sex comb on midleg (*Scm*): A New Gene Cluster on Xp22. *Genomics*, 1999, 58: 65-72.
46. Santiveri C.M: et al. The malignant brain tumor repeats of human SCML2 bind to peptides containing monomethylated lysine. *J. Mol. Biol*, 2008, 382(5): 1107-12.