

UNIVERSITÀ DEGLI STUDI DI NAPOLI

“FEDERICO II”

Scuola di Dottorato in Medicina Molecolare

Dottorato di Ricerca in Genetica e Medicina Molecolare



**Identificazione degli interattori molecolari di AF4, partner di
traslocazione di MLL nelle leucemie acute**

**Coordinatore:
Prof. Carmelo Bruno Bruni**

**Candidato:
Dott. Armando Cevenini**

Anno

2007

UNIVERSITÀ DEGLI STUDI DI NAPOLI

“FEDERICO II”

Dipartimento di Biochimica e Biotecnologie Mediche

Dottorato di Ricerca in Genetica e Medicina Molecolare

Coordinatore Prof. Carmelo Bruno Bruni

**Sede amministrativa:
Dipartimento di Biologia e Patologia Cellulare e Molecolare
“Luigi Califano”**

UNIVERSITÀ DEGLI STUDI DI NAPOLI

“FEDERICO II”

Dipartimento di Biochimica e Biotecnologie Mediche

**Tesi di Dottorato di Ricerca in Genetica e Medicina
Molecolare
XIX ciclo**

**Identificazione degli interattori molecolari di AF4, partner di
traslocazione di MLL nelle leucemie acute**

**Candidato:
Dott. Armando Cevenini**

**Docente guida e Relatore:
Chiar.mo Prof. Francesco Salvatore**

**Correlatore:
Chiar.ma Prof.ssa Gabriella Esposito**

Se ripenso a questi quattro anni di dottorato, mi viene spontaneo ricordare tutte le persone che mi sono state vicine nei momenti difficili, quelle che hanno gioito con me nei momenti felici, quelle che mi hanno dato consigli preziosi.

Tutte queste persone direttamente o indirettamente hanno contribuito a farmi giungere alla fine di questo iter formativo e al raggiungimento dei risultati scientifici riportati in questa tesi.

La lista dei ringraziamenti sarebbe così veramente lunga, per cui ho deciso di serbarla nel cuore anzi che esternarla per iscritto.

Tuttavia, prima che il lettore si addentri nelle problematiche scientifiche trattate in questa tesi, sento di dover ringraziare almeno quelle persone senza le quali non avrei avuto la possibilità di svolgere il progetto di ricerca di seguito riportato.

Ringrazio prima di tutto il Professor Francesco Salvatore, mio docente guida, per avermi accolto nei suoi laboratori, per avermi dato la possibilità di lavorare a questo progetto e per avermi sempre guidato, nonostante i suoi moltissimi impegni, lungo questo percorso di crescita professionale.

Ringrazio la Professoressa Gabriella Esposito per avermi accolto nel suo gruppo e per aver messo a mia disposizione costantemente e senza riserve il suo tempo, la sua competenza e la sua esperienza.

Ringrazio inoltre la Professoressa Margherita Ruoppolo ed il Dottor Alessandro Cuomo, per la loro preziosa collaborazione, senza la quale non sarebbe stato possibile giungere alla maggior parte dei risultati presentati in questa tesi e li ringrazio di cuore per il loro costante apporto umano, scientifico e morale.

INDICE

1. INTRODUZIONE	pag. 7
1.1. Le leucemie	pag. 8
1.2. La leucemia linfoblastica acuta (ALL)	pag. 10
1.3. Il ruolo di MLL nella patogenesi molecolare della ALL	pag. 11
1.4. Il gene e la proteina MLL	pag. 12
1.5. MLL e i geni Hox	pag. 16
1.6. Partners di fusione di MLL	pag. 17
1.7. La traslocazione t(4;11)(q21;q23) nella patogenesi molecolare della ALL	pag. 20
1.8. La proteina umana AF4	pag. 22
1.9. Funzione di AF4	pag. 25
1.10. Ruolo delle proteine SIAH nella proteolisi di AF4	pag. 25
1.11. Interattori di AF4	pag. 27
1.12. La chimera MLL-AF4	pag. 29
2. MATERIALI E METODI	pag. 31
2.1. Clonaggi	pag. 32
2.1.1. cDNA	pag. 32
2.1.2. PCR	pag. 32
2.1.3. Mutagenesi sito-specifica	pag. 32
2.1.4. Oligonucleotidi	pag. 33
2.1.5. Plasmidi	pag. 35
2.2. Espressione e purificazione di proteine ricombinanti in sistema procariotico	pag. 40
2.3. Colture cellulari e trasfezioni	pag. 41
2.3.1. Linee cellulari	pag. 41
2.3.2. Colture cellulari	pag. 44
2.3.3. Trasfezioni	pag. 44
2.4. Estratti proteici	pag. 45
2.5. Western blot	pag. 46
2.6. Immunoprecipitazioni	pag. 47
2.7. Analisi degli immunocomplessi mediante SDS-PAGE e spettrometria di massa	pag. 48
2.8. Analisi in silico	pag. 49

2.9. Microscopia a fluorescenza	pag. 50
3. RISULTATI	pag. 51
3.1. Identificazione degli interattori di AF4	pag. 52
3.1.1. Premesse	pag. 52
3.1.2. Espressione e purificazione di AF4 e suoi frammenti in sistema procariotico	pag. 52
3.1.3. Espressione di AF4 in sistemi eucariotici	pag. 58
3.1.4. Precipitazione per immunoaffinità di un frammento N-terminale di AF4	pag. 62
3.1.5. Identificazione degli interattori proteici del frammento N-terminale di AF4	pag. 63
3.1.6. Validazione dei dati di MS mediante co-immunoprecipitazione e Western blot	pag. 68
3.1.7. Analisi in silico degli interattori proteici di AF4	pag. 70
3.1.8. Espressione e immunoprecipitazione di un frammento C-terminale di AF4	pag. 71
3.1.9. Espressione di AF4 in cellule HEK 293 mediante trasfezione stabile	pag. 73
3.2. Analisi di domini funzionali della proteina AF4	pag. 75
3.2.1. Analisi dei residui fosforilati di AF4	pag. 75
3.2.2. Individuazione dei domini di AF4 responsabili dell'interazione con Cdk9 e 14-3-3 θ	pag. 79
3.2.3. Definizione dei segnali di localizzazione nucleare di AF4	pag. 81
4. DISCUSSIONE	pag. 85
5. BIBLIOGRAFIA	pag. 91

1. INTRODUZIONE

1.1. Le leucemie

Con il termine “leucemia” si indica un gruppo eterogeneo di patologie ematologiche caratterizzate dalla proliferazione neoplastica e clonale di cellule ematopoietiche.

Le leucemie si classificano in leucemie mieloidi, se il clone proliferante è un precursore della linea staminale mieloide, e leucemie linfoidei, se l’alterazione clonale è a carico di un precursore della linea staminale linfoide (1).

Parliamo, inoltre, di leucemie acute quando il clone leucemico è bloccato in una fase precoce di differenziazione; in questo caso la patologia è più aggressiva e la prognosi infausta (1).

Le leucemie croniche sono invece caratterizzate dalla proliferazione di precursori ematopoietici maggiormente differenziati che solitamente invadono il midollo e si riversano nel circolo sanguigno (1).

Distinguiamo, quindi, 4 tipi di leucemie:

Leucemia Mieloide Acuta (AML)

Leucemia Mieloide Cronica (CML)

Leucemia Linfatica Acuta (ALL)

Leucemia Linfatica Cronica (CLL).

La patogenesi clonale delle leucemie è determinata dall’alterazione di una singola cellula che, indifferenziata, prolifera, invade il midollo, e blocca la maturazione di altre linee ematopoietiche (1).

La sintomatologia leucemica è eterogenea e varia in base all’origine e al grado di differenziazione del clone espanso anche se solitamente l’invasione midollare causa anemia per insufficienza di produzione di globuli rossi, gravi infezioni per ridotta produzione di leucociti neutrofili ed emorragie per ridotta produzione di piastrine.

Analogamente, le strategie terapeutiche variano in base al tipo di leucemia; solitamente si tenta di arrestare la proliferazione del clone espanso utilizzando citostatici e chemioterapici; il trapianto di midollo, nonostante non sempre attuabile, rappresenta una valida alternativa anche al fine di ripopolare il midollo osseo distrutto dai chemioterapici (1).

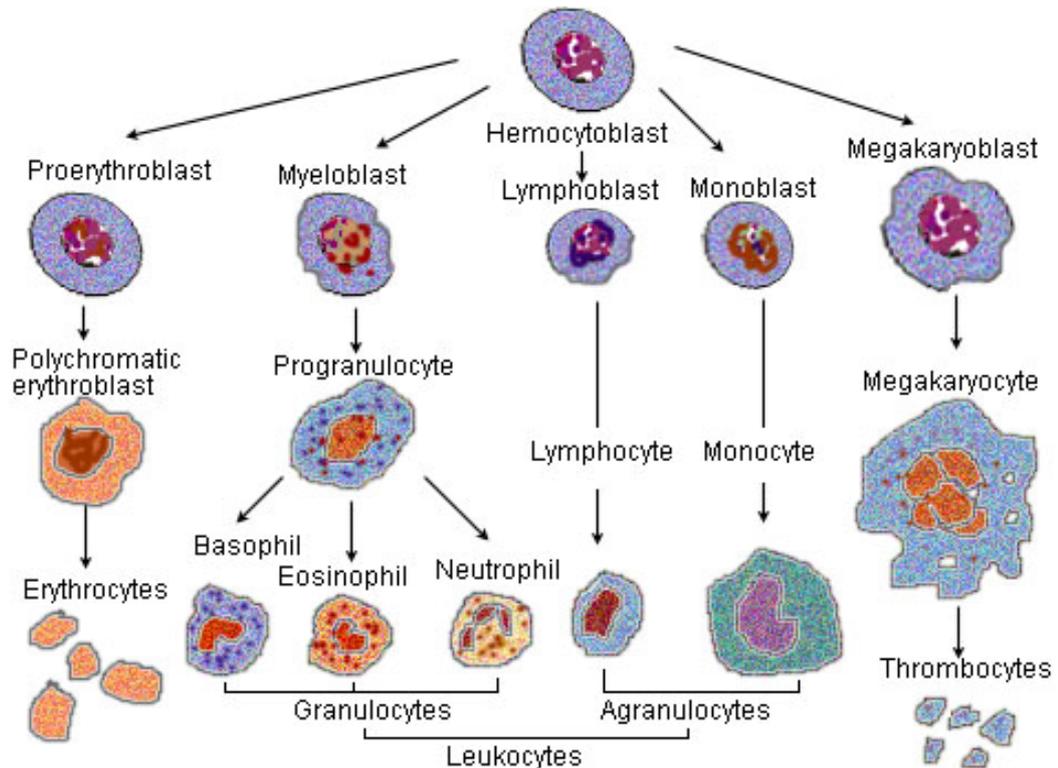


Figura 1. Ematopoiesi: produzione e differenziazione delle cellule ematiche. Da un'unica cellula staminale totipotente (emocitoblasto) originano tutte le cellule del sangue. Tale processo differenziativo può essere bloccato a diversi livelli dando origine alle varie forme leucemiche.

1.2. La leucemia linfoblastica acuta (ALL)

La leucemia linfoblastica acuta (ALL) è la forma leucemica maggiormente diffusa in età infantile. Essa rappresenta la quasi totalità (80%) delle leucemie osservabili prima dei 4 anni e solo raramente compare in individui che hanno superato i 30 anni di età. Le rare forme di ALL in età adulta rappresentano solitamente un'evoluzione acuta della leucemia cronica. Inoltre, l'esordio di tale patologia è solitamente acuto e non vi è una fase indolente e preleucemica. La proliferazione neoplastica alla base di tale leucemia riguarda un precursore molto immaturo (bastro) della linea linfoide. L'accumulo di linfociti immaturi si osserva in diverse sedi anatomiche (midollo osseo, sangue periferico, linfonodi, milza, testicoli e SNC) ed è solitamente accompagnato da un aumento del volume dei linfonodi e da epatosplenomegalia. Talvolta compare anche meningite leucemica. La prognosi di tale patologia è solitamente infausta a causa della rapida evoluzione e delle ridotte possibilità terapeutiche. Emorragie ed infezioni rappresentano le principali cause di morte. Secondo la classificazione FAB (Franco-Americano-Britannica), la ALL si suddivide in tre categorie, morfologicamente distinte, denominate L1, L2, L3, sulla base delle dimensioni dei blasti e sulla presenza della Tdt (deossinucleotidil trasferasi terminale), enzima caratteristico dei linfociti T e B immaturi, che partecipa al processo di rimaneggiamento dei geni delle immunoglobuline e del *T-cell receptor*. Le tre forme si differenziano anche per la prognosi: L1 è la forma a decorso migliore, L2 manifesta una prognosi intermedia, mentre L3 presenta la prognosi più infausta. Di conseguenza, diventa possibile pianificare una chemioterapia più o meno aggressiva a seconda della classe di appartenenza della forma leucemica (1).

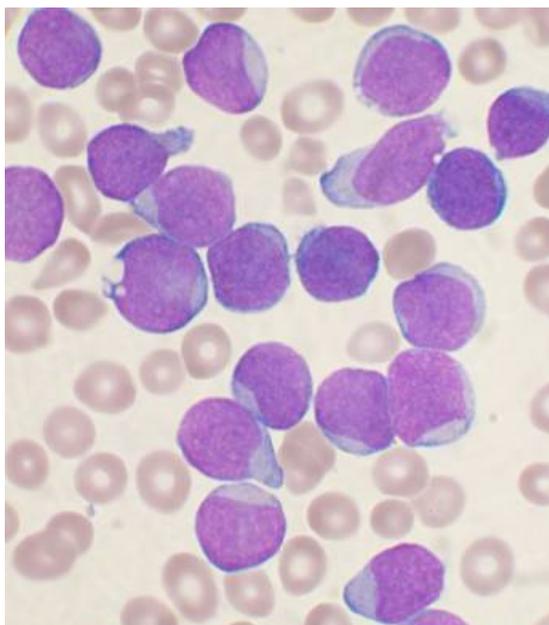


Figura 2. ALL: Striscio di sangue midollare. Le dimensioni e la tonalità delle cellule mostrano una leucemia linfatica acuta in avanzamento, inoltre la presenza di nucleoli e di cromatina fortemente immatura sono un chiaro segno di mancata differenziazione cellulare.

1.3. Il ruolo di MLL nella patogenesi molecolare della ALL

E' noto che più del 65 % delle leucemie acute sono causate a livello molecolare da alterazioni geniche acquisite a livello somatico (2). Tali alterazioni strutturali coinvolgono geni importanti per la regolazione dell'espressione genica e modificano i normali *pathways* di crescita e differenziazione cellulare. Il braccio lungo del cromosoma 11 è stato a lungo definito dagli oncoematologi come un *hot spot* per i riarrangiamenti genomici. In particolare, la banda citogenetica 11q23 è nota a causa del suo coinvolgimento nella patogenesi di alcune leucemie con decorso aggressivo. Oggi sappiamo che in questo *locus* è posizionato il gene MLL (*Mixed Lineage Leukemia*), conosciuto anche come HRX o ALL-1, e che alcune alterazioni in 11q23 convertono MLL in un oncogene attivo (3,4). Tali alterazioni sono per lo più rappresentate da traslocazioni cromosomiche che fondono il gene *MLL* con un gene partner che risiede

su un cromosoma differente. Sono ampiamente descritti anche casi di inserzione, duplicazione e inversione che coinvolgono lo stesso gene e che danno luogo all'insorgenza di patologie neoplastiche della linea emopoietica. Il meccanismo, però, più frequentemente riscontrato nelle leucemie rimane la traslocazione.

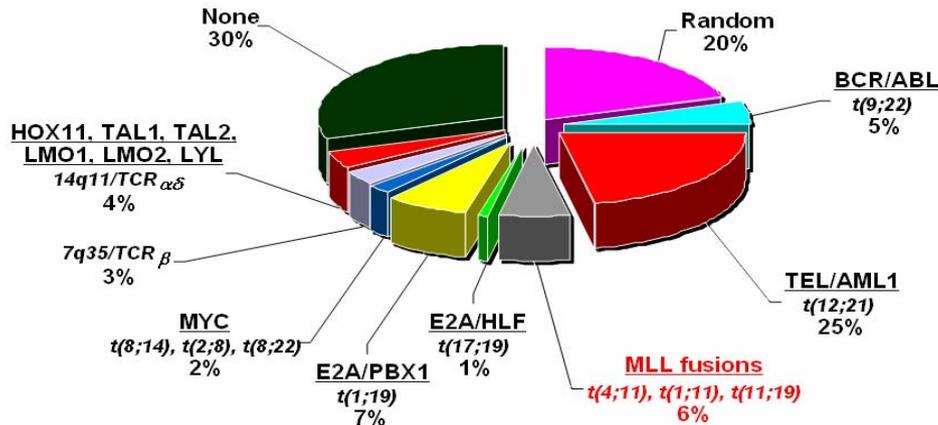


Figura 3. Rappresentazione schematica della distribuzione dei principali riarrangiamenti cromosomici causativi della ALL. In evidenza la percentuale rappresentata dalle traslocazioni che coinvolgono MLL.

1.4. Il gene e la proteina MLL

Il gene MLL è costituito da 36 esoni distribuiti su 100 kb e produce un mRNA di 12 kb che codifica per una proteina di circa 3969 amminoacidi, con un peso molecolare di circa 430 kDa (fig. 4). Questa proteina, che è largamente espressa nello sviluppo dell'embrione mentre lo è a bassi livelli in molti tessuti adulti, possiede vari domini di regolazione della trascrizione genica. La proteina matura grazie ad un taglio proteolitico citoplasmatico post-traduzionale a livello dei residui amminoacidici 2666-2667 e 2718-2719, dal quale originano un frammento N-terminale di ~320 kDa (MLL^N) e un altro C-terminale di ~180 kDa (MLL^C) per azione dell'enzima treonina-aspartasi, TASPASI1. I

due frammenti rimangono legati in maniera non covalente tramite il dominio FYRN (2023-2073), il dominio FYRC e parte del dominio SET (3666-3876), in un dimero che rappresenta la forma attiva della proteina. Bloccando la sintesi della proteina TASPAS1, si ha un accumulo della proteina MLL non processata proteoliticamente e quindi incapace di svolgere la sua funzione (3,4).

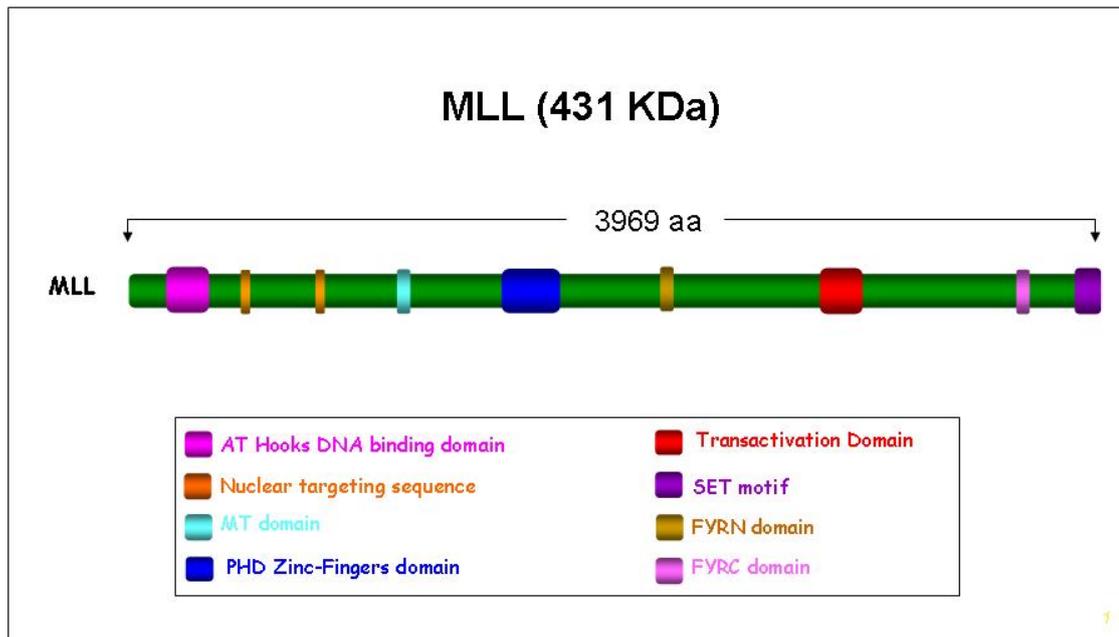


Figura 4. Rappresentazione schematica della proteina MLL e dei suoi domini funzionali.

MLL contiene una regione C-terminale ad alta omologia con la proteina TRITHORAX (TRX) di *Drosophila* ed è stata chiamata per questo motivo HRX, *Human Trithorax*. TRX è membro di un gruppo di proteine attivatrici, identificate da uno *screening* di mutanti che sopprimono il fenotipo causato dalla perdita dei repressori POLYCOMB-simili. Le proteine del gruppo Trithorax (Trx-G) e quelle del gruppo Polycomb (Pc-G) insieme costituiscono la cosiddetta “memoria cellulare”. Questo meccanismo è necessario nello sviluppo e nella differenziazione, ogni volta che il *pattern* trascrizionale deve essere “impresso”, per poi essere ereditato dalle future generazioni cellulari. Lo stesso processo permette alle cellule di mantenere un certo stato di

sviluppo e di *lineage-committment* anche in assenza dei fattori trascrizionali inizialmente determinanti per il differenziamento. In genere le Pc-G sono associate a modifiche che inducono repressione trascrizionale, mentre le Trx-G sono per lo più degli attivatori. I principali *target* di Pc-G e Trx-G sono i geni per i fattori trascrizionali *homeobox* (*Hox*). All'estremità N-terminale della porzione MLL^N, che con più probabilità collabora nella ricerca dei geni *target*, sono stati identificati tre domini AT-*hooks*, che mediano il legame al solco minore di una sequenza di DNA ricca in AT. Questi domini stabilizzano i complessi DNA-proteina innescando cambiamenti conformazionali o mediano le interazioni proteina-proteina attraverso il legame al DNA. Oltre agli AT-*hooks*, un secondo dominio di legame al DNA di MLL^N, riportato in letteratura col nome di *MT domain*, è stato rilevato per omologia con il dominio regolatorio della citosina metil-trasferasi. Questi domini contengono la sequenza specifica CXXC e legano dinucleotidi CpG non metilati, caratteristica delle isole CpG dei geni trascrizionalmente attivi (3).

Tra i domini AT-*hooks* ed il dominio MT (aa 1032-1395) ci sono due brevi regioni, NLS1 e NLS2 che rappresentano la localizzazione subnucleare della proteina MLL. Il dominio MT rimane in tutte le proteine di fusione di MLL che si formano in seguito alla traslocazione ed è richiesto per la capacità trasformante delle proteine di fusione. Esso oltre a contenere la regione di omologia alla DNA metiltrasferasi, DNMT1, che include il motivo *zinc-finger* CXXC, media la repressione trascrizionale attraverso il reclutamento delle istone-deacetilasi HDAC1 e HDAC2.

Tra i domini FYRN e FYRC è localizzato il dominio di transattivazione (TAD), non presente in TRITHORAX di *Drosophila*, ma conservato negli omologhi di MLL nei Vertebrati. Esso lega direttamente il coattivatore CBP (*CREB-binding protein*), un'acetiltrasferasi probabilmente coinvolta nell'acetilazione di H3 e H4, in prossimità dei promotori *HoxA9* e *HoxC8*. In seguito alla traslocazione cromosomica, la regione di

fusione si trova appena al di là degli esoni codificanti per il lato C-terminale del dominio di repressione, definendo il “*major break region*”. Il dominio SET in prossimità della porzione C-terminale di MLL è la sede dell’attività istone-metilasi, che va a modificare l’istone 3 in Lys4 (H3-K4) e si lega direttamente ai promotori *HoxA9* e *HoxC8*. La metilazione H3-K4 è collegata, dunque, ad uno stato di trascrizione attivo. E’ probabile che i promotori dei geni *Hox* siano H3-K4 metilati tramite il dominio SET di MLL, che giocherebbe, quindi, un ruolo diretto nell’attivazione trascrizionale dei geni *Hox* (5). MLL si è mostrato essere più efficiente nella metilazione di istoni H3 pre-acetilati suggerendo, quindi, una preferenza per le regioni non condensate della cromatina. Inoltre, il dominio SET di MLL recluta INI1, componente integrale del macchinario di decondensazione del nucleosoma SWI/SNF. Un dominio di 55 residui posizionato agli amminoacidi 2829-2883 di MLL ha una forte attività transattivante. La sostituzione sistematica degli amminoacidi in questo dominio indica che è essenziale un *core* di 9 amminoacidi, e che l’alterazione dei residui acidi ed idrofobici abolisce l’attività (6).

Si pensa che MLL serva come piattaforma per integrare i tre maggiori aspetti della biologia degli istoni, acetilazione, metilazione e rimodellamento, in modo da “imprimere” e rendere così ereditabile l’attivazione dei rispettivi geni *target*. Sfortunatamente si conosce ancora poco su come le modificazioni degli istoni indotte da MLL siano tradotte in attività trascrizionale. Due potenziali meccanismi, mutualmente non-esclusivi sono in discussione. In primo luogo, la metilazione degli istoni H3K4 potrebbe inibire sia l’attacco del complesso repressore NuRD (istone-deacetilasi) sia il complesso di rimodellamento del nucleosoma. In alternativa lo stato della cromatina indotto da MLL potrebbe reclutare coattivatori come il complesso istone-acetilasi SAGA (3).

1.5. MLL e i geni Hox

La famiglia dei geni *Hox*, codificanti per i fattori trascrizionali *homeobox* gioca un importante ruolo nell'ematopoiesi. Studi recenti indicano che una caratteristica comune delle leucemie acute a prognosi infausta è l'espressione deregolata di alcune di queste proteine (7).

Diversi studi hanno dimostrato che nei pazienti affetti da ALL c'è uno specifico *pattern* di espressione genica. Molti geni sono *over*-espressi nelle leucemie con aberrazioni di MLL e tra questi troviamo i geni dell'*homeobox* *HoxA7*, *HoxA9* e il cofattore MEIS1. Poichè i geni dell'*homeobox* sono *target* diretti di MLL *wild-type* e i domini che più probabilmente interagiscono con i geni *target* sono localizzati all'N-terminale, la porzione conservata nella fusione di MLL con altri geni, non sorprende trovare una potente attivazione dei geni *Hox* a valle di queste proteine chimeriche ad alta attività transattivante. Normalmente l'espressione dei geni *Hox* è alta nelle cellule staminali ematopoietiche e si spegne gradualmente durante il differenziamento; un errore nella *downregolazione* dei geni *Hox* inibisce la maturazione ematopoietica e può causare leucemia. Inoltre, molti di questi geni iperespressi nelle leucemie umane con proteine di fusione di MLL sono *upregolati* anche nelle cellule con amplificazioni di MLL (fig. 5). Comunque, sebbene molte evidenze identifichino i geni *homeobox* come gli effettori a valle di MLL più importanti, ci sono sicuramente anche altri geni che possono contribuire al processo di trasformazione leucemica (3).

1.6. Partners di fusione di MLL

Approssimativamente il 3-10% dei pazienti con leucemia mieloide acuta (AML) possiede una traslocazione del gene MLL; l'8-10% dei pazienti con leucemia linfoblastica acuta (ALL) che coinvolge il *lineage* B ha una traslocazione di MLL.

Le tecniche di citogenetica e di biologia molecolare scoprono continuamente nuovi riarrangiamenti del gene MLL. In particolare, tutti i *partners* rari e non ancora conosciuti di MLL e la precisa localizzazione del *breakpoint* possono essere identificati o sfruttando la tecnologia della PCR *panhandle* su cDNA (8) o mediante "MLL *Fusion Chip*" (9). Numerosi studi hanno rivelato l'importanza di discriminare tra i vari tipi di traslocazione che coinvolgono il gene MLL per definire un appropriato trattamento terapeutico. Inoltre, che la prognosi delle leucemie con un riarrangiamento di MLL sembra essere differente, a seconda del gene *partner*, del *lineage* cellulare coinvolto nella trasformazione leucemica, dell'età del paziente e della terapia somministrata (10). Le traslocazioni di MLL, molto comuni anche nei bambini con AML, sono state identificate in oltre l'80% di tutti i casi di ALL infantile. Inoltre, sono state riscontrate in circa il 25% dei pazienti con leucemia collegata alla terapia per la cura di neoplasie pre-esistenti (*therapy-related leukemia*), in particolar modo quelli trattati con agenti chemioterapici che hanno come *target* le DNA topoisomerasi di tipo II (11).

Generalmente, una particolare traslocazione è associata ad un particolare sottotipo di leucemia ed il gene coinvolto nel riarrangiamento cromosomico presenta delle "preferenze" nella scelta del *partner*. Le traslocazioni che coinvolgono il gene MLL, invece, rappresentano un'eccezione, in quanto associate ad un largo spettro di patologie ematopoietiche maligne, incluse la AML, la ALL a cellule T e a cellule B, la sindrome mielodisplastica, il linfoma linfoblastico e il linfoma di Burkitt. Per di più,

MLL è considerato un oncogene “promiscuo”, in quanto sono stati trovati riarrangiamenti con più di 30 geni *partners* e 54 *breakpoints* differenti (12) (fig. 5).

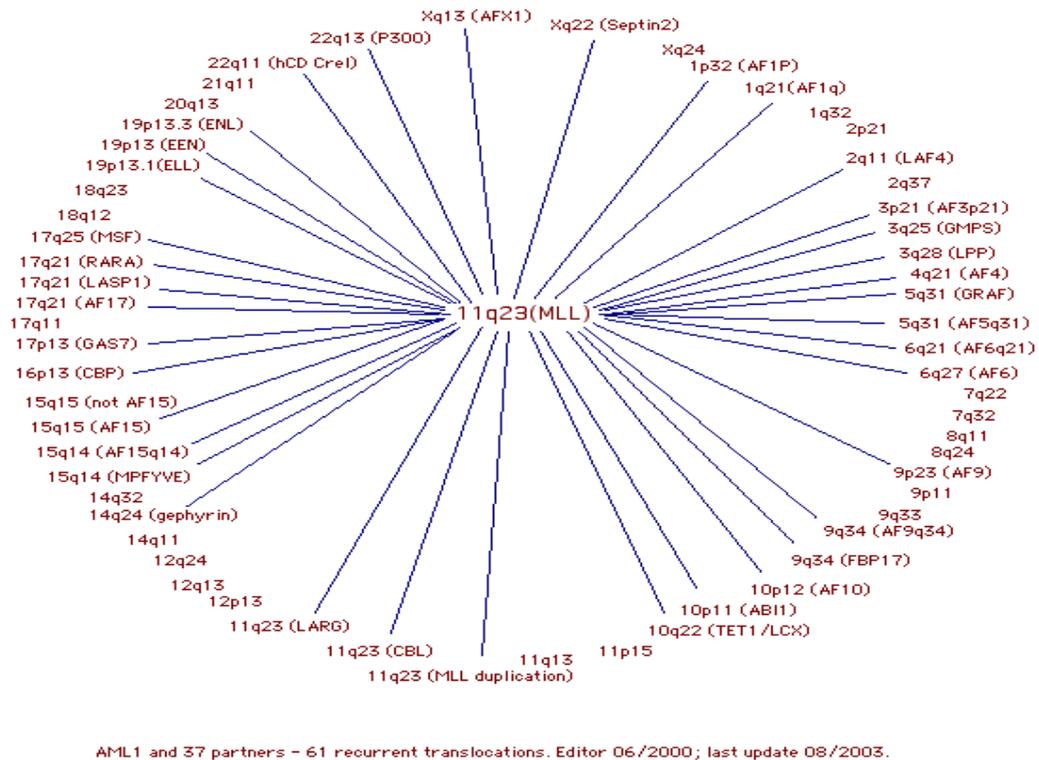


Figura 5. Le traslocazioni di MLL conosciute finora.

Il gene MLL può trovarsi fuso sia a proteine citoplasmatiche, sia a proteine localizzate nel nucleo. Le 5 traslocazioni, t4;11 (MLL-AF4), t9;11 (MLL-AF9), t11;19 (MLL-ENL), t10;11 (MLL-AF10) e t 11;19 (MLL-ELL) rendono conto da sole dei due terzi di tutti i casi in cui si ha un riarrangiamento di MLL (3).

Nonostante il largo spettro di *partners* interessati alla traslocazione, le fusioni rispettano tutte lo stesso schema: la sequenza C-terminale di MLL, incluso il dominio *zinc-finger* PHD, è sostituita *in frame* dalla proteina *partner*. Poichè le fusioni non mantengono il

motivo di interazione MLL^N/MLL^C, esse molto probabilmente non si associano con il C-terminale di MLL. Poichè le traslocazioni spesso sono bilanciate, può essere espresso anche il prodotto di fusione reciproco contenente il gene *partner* all’N-terminale legato alla sequenza C-terminale di MLL. In genere, ad eccezione dell’attività trasformante descritta per AF4-MLL, non ci sono evidenze che il prodotto reciproco sia importante per la leucemogenesi. Attualmente, si pensa che la proteina chimerica MLL trascritta dal cromosoma 11 derivativo agisce attraverso un meccanismo di *gain-of-function*. Lo scopo principale dei *partners* di fusione studiati finora sembra essere la conversione della proteina MLL troncata in un potente attivatore trascrizionale. Ciò può avvenire in diversi modi; spesso i *partners* contengono dei domini di attivazione, come ad esempio ENL, AF4 e AF9, per le quali è stato identificato un piccolo dominio transattivante C-terminale sufficiente a convertire MLL in un’oncoproteina attiva. Ulteriori dati a favore dell’attivazione come funzione chiave delle proteine *partners* di MLL derivano dai rari casi in cui MLL è fusa direttamente ad attivatori trascrizionali come p-300 e CBP. In alternativa, le proteine *partners* potrebbero reclutare degli attivatori come nel caso di ELL e ABI1. Una possibile spiegazione per le speciali proprietà transattivanti conferite dai *partners* di fusione della proteina MLL è la “MLL *web hypothesis*” (3). Secondo questa teoria, molti *partners* di MLL potrebbero essere membri dello stesso complesso multiproteico coinvolto nell’elongazione trascrizionale e nella modificazione della cromatina (13) (fig. 6).

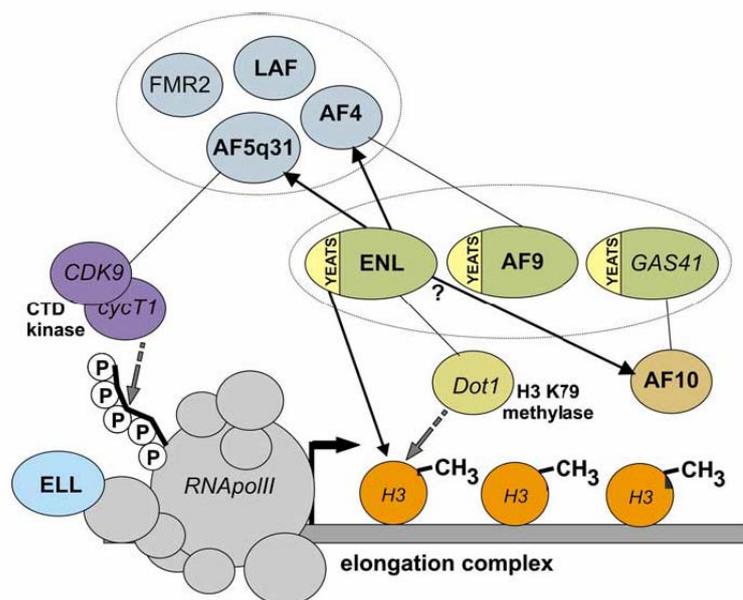


Figura 6. Molti dei *partners* di fusione di MLL sono connessi in un complesso associato con le modificazioni della cromatina che avvengono durante l'elongazione della trascrizione (13). La metilazione dell'istone H3 in Lisina K79 e la fosforilazione del dominio C-terminale della RNA polimerasi II sono eventi enzimatici necessari durante l'elongazione.

1.7. La traslocazione t(4;11)(q21;q23) nella patogenesi molecolare della ALL

Una delle più frequenti anomalie che coinvolgono il gene MLL è la traslocazione reciproca t(4;11)(q21;q23) che fonde la regione q23 del cromosoma 11, dove è situato MLL (11q23) con la regione q11 del cromosoma 4, dove è localizzato il gene AF4 (4q11) (14,15). AF4 si ritrova fuso ad MLL nel 50-70 % dei casi di leucemia linfoblastica acuta causata da traslocazioni cromosomiche che coinvolgono il *locus* 11q23. Tale traslocazione è stata riscontrata in più del 95% dei casi di leucemia linfoblastica acuta infantile (16). Questa forma leucemica è caratterizzata solitamente da una breve latenza (con insorgenza generalmente a 3 mesi), cui segue un decorso molto

rapido, con prognosi infausta e scarse possibilità terapeutiche (17). La traslocazione t(4;11)(q21;q23) causa la fusione *in frame* dei primi 8-11 esoni di MLL con gli ultimi 13-20 esoni di AF4; il gene di fusione viene trascritto in un mRNA di circa 12.5 kb che codifica una proteina chimerica di circa 240 kDa. Nonostante sia stato dimostrato che sia duplicazioni esoniche di MLL che il suo prodotto troncato sono sufficienti a indurre la trasformazione neoplastica di precursori emopoietici, il coinvolgimento di AF4 sembra influenzare sensibilmente il fenotipo leucemico giocando un ruolo importante nell'alterazione del differenziamento linfoide.

Il meccanismo mediante il quale la proteina di fusione induce neoplasia ed è responsabile della leucemigenesi non è completamente compreso. Sono stati proposti due modelli: il primo implica un'acquisizione di funzione da parte della proteina chimerica MLL/AF4 e prevede che questa sia in grado di indurre un'aberrante espressione di geni *target* compromettendo così la differenziazione dei linfoblasti (18); il secondo modello prevede che MLL/AF4 agisca come dominante negativo, bloccando la normale funzione fisiologica di AF4 e/o di MLL ed impedendo l'espressione di geni fondamentali per la differenziazione emopoietica. E' stato dedotto che la traslocazione t(4;11)(q21;q23) è in grado di trasformare progenitori staminali multipotenziali in quanto i blasti leucemici esprimono marker di cellule staminali (CD34 e HLA-DR), il marker CD19 della linea pro-B e il marker CD15 tipicamente espresso dai precursori mielomonocitici (19).

1.8. La proteina umana AF4

Il gene umano AF4 (chiamato anche FEL) è localizzato nella regione cromosomica 4q21 dove si estende per 134050 bp. E' costituito da 23 esoni (i primi 4 sono alternativamente presenti nel trascritto finale) che generano un trascritto maturo di 9390 bp caratterizzato da un "ORF" di 3633 bp. Il prodotto del gene AF4 umano è una proteina nucleare ricca in serina e prolina, di 1210 aa con un peso molecolare di circa 140 kDa. La funzione fisiologica di AF4 è attualmente ancora sconosciuta anche se la presenza di un dominio ad attività transattivante (aa 480-560) (6) e la sua localizzazione nucleare (20,16) mediata da almeno una sequenza di indirizzamento al nucleo (21) ne suggeriscono un ipotetico ruolo di regolatore trascrizionale (Fig.7). AF4 è una proteina ubiquitaria, ma è espressa a livelli più elevati in alcuni tessuti (22,23) quali placenta, rene, SNC, ed organi linfoidei. Nei topi, il gene AF4 è principalmente espresso nei timociti in via di maturazione e negli organi linfoidei adulti. In modelli murini di knockout genico di AF4 (topi AF4^{-/-}), lo sviluppo dei linfociti T e B risulta gravemente compromesso (24). Le analisi condotte su tali modelli hanno inoltre suggerito che la proteina AF4 gioca un ruolo importante nella linfoiesi soprattutto durante gli stadi precoci della proliferazione dei precursori, mentre non sembra essere richiesto per gli stadi terminali della differenziazione linfoitaria (24).

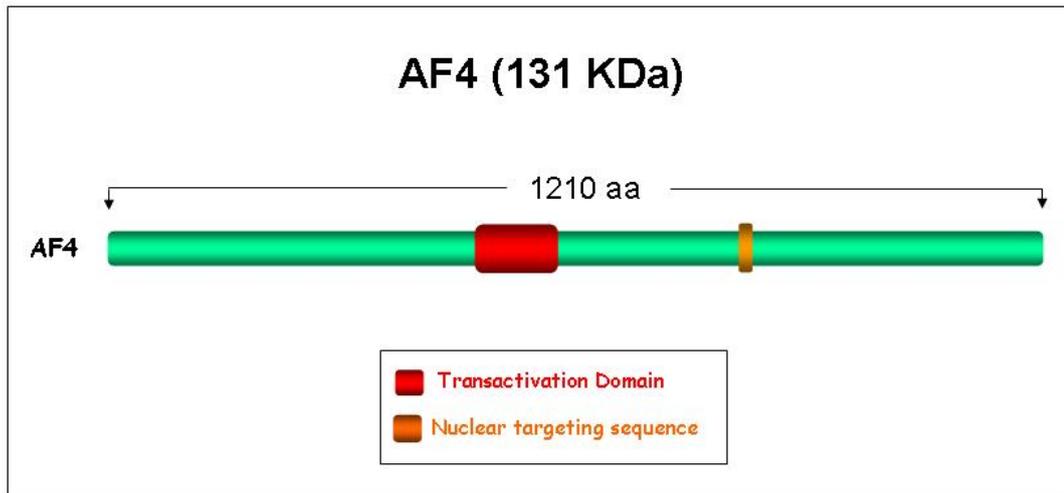


Figura 7. Proteina AF4. Rappresentazione schematica dei domini funzionali.

Insieme a LAF4, FMR2 e AF5Q31, AF4 appartiene alla famiglia di proteine ALF (AF4 LAF4 Family) (25-27). I membri della famiglia hanno in comune tre regioni con elevata similarità di sequenza, molto conservate anche dal punto di vista evolutivo (21): una regione N-terminale di circa 90 residui amminoacidici (aa 6-93) chiamata NHD (N-terminal Homology Domain); una regione, costituita da circa 180 residui (aa 467-522) e chiamata ALF domain, che comprende una sequenza ricca in serina e prolina (aa 444-480); una regione C-terminale di circa 260 aa (aa 954-1211) chiamata CHD (C-terminal Homology Domain) (Fig.8).

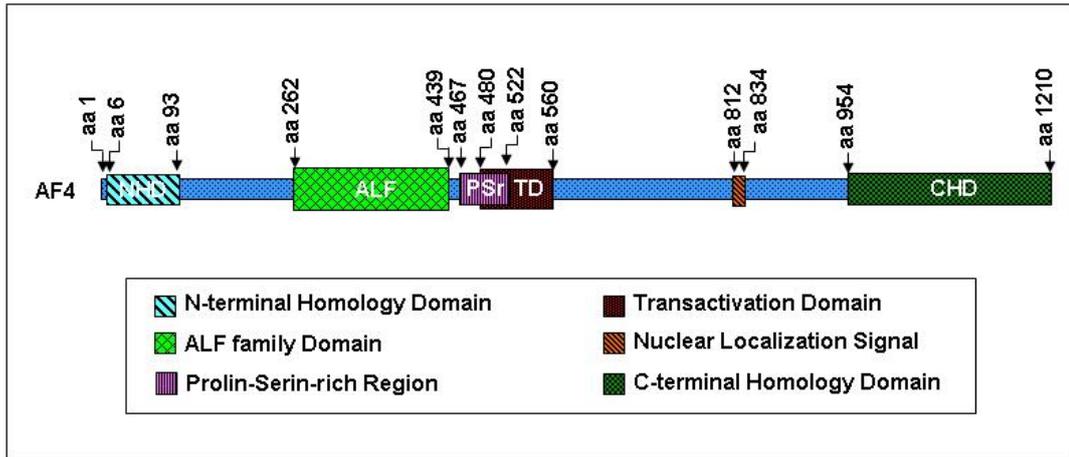


Figura 8. Proteina AF4/FEL. Rappresentazione schematica di tutti i domini di AF4. NHD (N-terminal Homology Domain) e CHD (C-terminal Homology Domain) sono molto conservati in tutti gli ortologhi di AF4. ALF è molto conservato in tutte le proteine della famiglia ALF ed in tutti i loro ortologhi. PSr (Prolin Serin rich region) è una regione ricca in serina e prolina. TD (Transactivation Domain) è un dominio ad elevata attività transattivante (6).

Il gene LAF4 è stato isolato dal linfoma di Burkitt e codifica per una proteina di 135 kDa (LAF4) che contiene una regione ricca in prolina e lisina e sequenze di localizzazione nucleari (NLS) (21). Al pari di AF4, la proteina LAF4 è un potenziale fattore di trascrizione con un dominio transattivante e capacità di legame al DNA in maniera non sequenza specifica (6). Tale proteina è probabilmente coinvolta nel differenziamento linfocitario, dati i suoi alti livelli d'espressione nei tessuti e nelle linee cellulari linfoidi B e T (24).

La proteina nucleare FMR2 è codificata, invece, da un gene che si estende nella regione q28 del cromosoma X. Alterazioni dell'espressione di tale gene, dovuti a meccanismi di espansioni di triplette sono associati ad una lieve forma di ritardo mentale (28).

L'ultimo componente della famiglia ALF ad essere stato identificato è AF5Q31, il cui gene, al pari di AF4, rappresenta un partner di fusione di MLL in traslocazioni cromosomiche coinvolte nella genesi del fenotipo leucemico (29).

1.9. Funzione di AF4

Recenti studi suggeriscono che la proteina AF4 insieme con AF9, ENL e AF10, funziona come un regolatore positivo della trascrizione genica, stimolando l'attività chinasi del fattore di elongazione della trascrizione della RNA polimerasi II, p-TEFb e mediando la metilazione di H3 in K79 dipendente da Dot1 (30). Ciò faciliterebbe la transizione della RNA polimerasi II lungo il template di DNA, in quanto si manterrebbe la struttura della cromatina in una conformazione aperta (fig.6). Risulta, inoltre, che p-TEFb fosforila la proteina AF4 e ne riduce la capacità di transattivazione, innescando un meccanismo di *feedback* negativo per il controllo dell'attività di elongazione di p-TEFb (30).

1.10. Ruolo delle proteine SIAH nella proteolisi di AF4

Lo studio del corrispettivo murino di AF4 (mAF4) ha evidenziato un'omologia del 64% con il gene umano (24). Tale vicinanza si rispecchia anche in un simile profilo di espressione genica: mAF4 è espresso ad alti livelli nei tessuti della linfopoiesi T (timo, linfonodi e rene) e nei compartimenti della linfopoiesi B (midollo osseo, fegato e polmoni) anche se a livelli minori rispetto al corrispettivo gene umano (24).

Altre evidenze sperimentali hanno mostrato come mAF4 sia altamente espresso anche a livello del SNC, soprattutto nelle cellule di Purkinje del cervelletto (31). Ciò suggerisce un ulteriore ruolo della proteina che potrebbe essere coinvolta in pathways fondamentali per la sopravvivenza neuronale.

Esperimenti di *doppio ibrido* e di coimmunoprecipitazione (28) hanno dimostrato una forte affinità tra AF4 e due proteine con funzione E3-Ubiquitina ligasi (SIAH1 e SIAH2) coinvolte nei pathways molecolari di diverse patologie neurodegenerative (29). I siti che mediano tale interazione sono il dominio SBD (SIAH Binding Domain) conservato in entrambe le proteine SIAH e una regione N-terminale di AF4 (aa 259-346) altamente conservata in tutte le proteine della famiglia ALF. In particolare, è stato identificato il dominio minimo di interazione in AF4 (*binding core motif*), rappresentato dalla sequenza amminoacidica “PXAXVXP” (28).

Il legame diretto tra AF4 e le proteine SIAH è fondamentale in quanto tali proteine sono in grado di mediare la rapida degradazione proteosoma-dipendente di AF4. L'importanza di tale pathway degradativo è stata messa in luce da studi condotti su un modello murino (32) ottenuto mediante mutagenesi chimica. Tale mutante manifestava atassia, cataratta e progressiva degenerazione delle cellule di Purkinje del cervelletto. Questo fenotipo dominante era caratterizzato da anomalie nei movimenti e nella deambulazione, da cui il nome di “topo robotico” attribuito a questo modello murino. La mappatura genetica e fisica del *locus* genico coinvolto in tale fenotipo ha permesso di identificare una mutazione *missenso* (V294A) nel gene AF4. Tale mutazione si trova nella regione di AF4 responsabile del legame con le proteine SIAH (all'interno del *binding core motif*) e ne riduce l'efficienza di binding di più del 70% (28): ne consegue una minore degradazione proteolitica e un accumulo della proteina mutante, probabilmente causativo della neurodegenerazione (28).

Questi dati fanno ipotizzare che la proteina “robotica” altera indirettamente le proprietà trascrizionali di AF4 prevenendo la sua fisiologica degradazione proteasoma-dipendente (turnover) mediata dall'interazione con le proteine SIAH. Inoltre, l'assenza di alterazioni neurologiche in topi AF4 *-/-* evidenzia chiaramente che la mutazione “robotica” induce un guadagno di funzione nel meccanismo fisiologico di AF4 dovuto

semplicemente alla sua sovraespressione. E' interessante notare che la proteina chimerica MLL-AF4, essendo priva di tale dominio di *binding*, è probabilmente protetta dalla degradazione SIAH-mediata, e sfugge quindi al controllo del *pathway* degradativo di AF4.

1.11. Interattori di AF4

Vari studi sperimentali stanno cercando di far luce sul ruolo biologico di AF4. Tali ricerche hanno portato tra l'altro all'identificazione di una serie di proteine che interagiscono con AF4. Tra i pochi interattori diretti di AF4 riportati in letteratura troviamo le proteine SIAH, la proteina AF9 e il suo omologo ENL, il fattore di elongazione della trascrizione ELL e la proteina SMAD9.

Le proteine SIAH, come già detto, mediano la degradazione proteasoma dipendente di AF4 e ne regolano il turnover. Le proteine AF9 e ENL fanno parte della stessa famiglia di proteine (ENL/AF9) e, come AF4, sono comuni partners di fusione di MLL coinvolti in diverse forme di AML e ALL. In particolare, il gene AF9, localizzato sul cromosoma 9, codifica per una proteina nucleare ricca in serina e prolina con capacità di attivare la trascrizione genica. L'alta identità di sequenza tra AF9 e ENL si evidenzia soprattutto all'estremità C-terminale dove è presente una sequenza di 90 amminoacidi che rappresenta il dominio minimo di transattivazione (16). Il sito di interazione tra AF4 e AF9 è localizzato in una regione C-terminale di AF4 altamente conservata in tutti gli altri membri della "ALF family". Inoltre, la regione di AF9 responsabile di tale interazione è presente anche nella proteina omologa ENL. Il legame tra AF4 e AF9 è

mantenuto inalterato nelle proteine di fusione MLL-AF4 e MLL-AF9 ed è stato dimostrato che MLL-AF4 è in grado di alterare la localizzazione subnucleare di AF9 (16).

Il gene ENL, invece, si estende nella regione 19p13.3 ed è coinvolto nella traslocazione t(11,19) anch'essa associata a diverse forme di AML e ALL (13,18). La proteina ENL mostra *in vitro* una debole attività trascrizionale promotore-specifica, che ne suggerisce un putativo ruolo di fattore di trascrizione. Tale tesi è avvalorata dal fatto che ENL mostra elevata identità di sequenza con la proteina di lievito *Anc1* che prende parte in due diversi complessi di trascrizione basale. Mediante l'interazione con la proteina ENL, AF4 interagisce indirettamente con altre proteine coinvolte nel processo trascrizionale, tra cui e AF10 e Dot1.

Un'altro interattore diretto di AF4 è ELL, l'unico partner di fusione di MLL di cui è nota la funzione biochimica (33,34), in quanto saggi *in vitro* ne evidenziano chiaramente il ruolo di fattore di elongazione trascrizionale (35). ELL agisce sopprimendo le pause che l'RNA polimerasi II effettua durante la trascrizione. Non è stato ancora dimostrato, tuttavia, se tale funzione sia richiesta per l'attività oncogenica della proteina MLL-ELL.. La proteina AF4 interagisce indirettamente con l'ELL Associated Factor 1 (EAF1) che lega l'estremità C-terminale di ELL e ne favorisce l'attività transattivante(36).

Altro interattore diretto di AF4 è la proteina SMAD9 (37), anche chiamata SMAD8A e/o SMAD8B. Le proteine R-Smads (SMAD1, SMAD5 e SMAD8) sono mediatori intracellulari delle proteine BMPs (bone morphogenetic proteins) coinvolte nel pathway differenziativo dei mammiferi. In particolare, le proteine R-Smads possono agire come regolatori positivi della trascrizione genica interagendo con il complesso basale di trascrizione RNA polimerasi II-dipendente (38).

1.12. La chimera MLL-AF4

In seguito alla traslocazione t(4;11)(q21;q23) sul cromosoma der(11) si forma un gene ibrido che comprende la regione 5' del gene MLL e la regione 3' del gene AF4. Benchè i breakpoint dei due geni possano essere variabili (introni 8-11 di MLL) (introni 12-19 di AF4), la fusione risulta sempre *in frame*, per cui il gene ibrido viene trascritto e maturato in un mRNA chimerico di circa 12.5 kb (la dimensione può variare leggermente a seconda dei *breakpoint*) codificante una proteina di circa 240 kDa (21) (fig. 9)

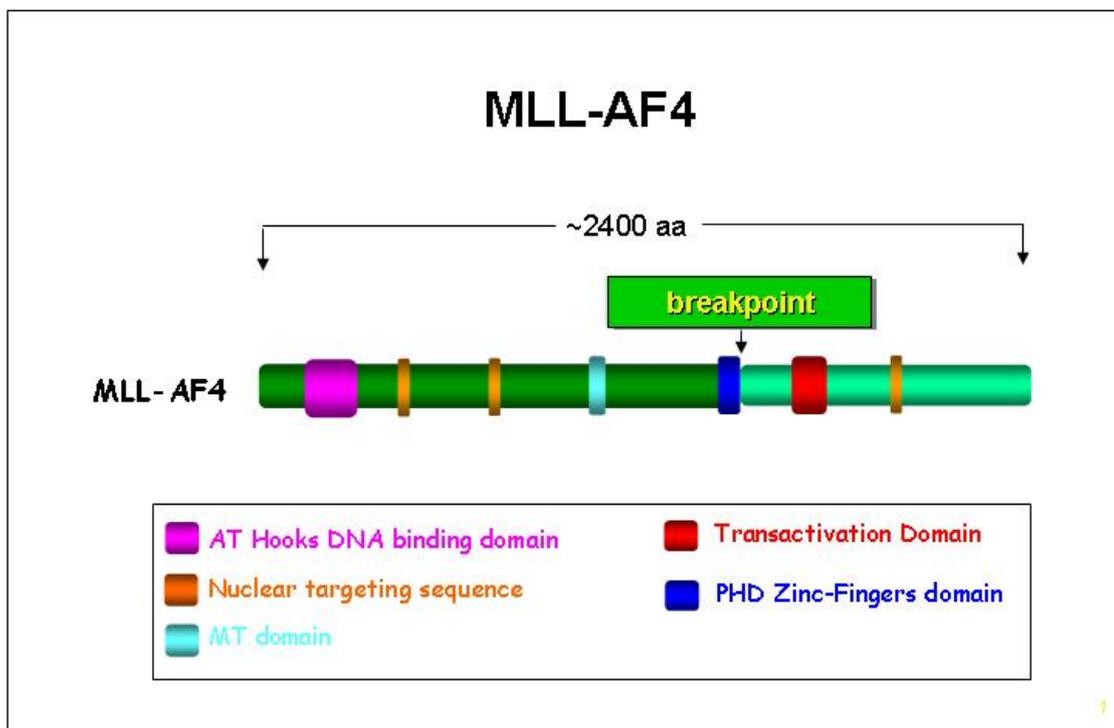


Figura 9. Rappresentazione grafica della proteina chimerica MLL-AF4, prodotta in seguito alla traslocazione cromosomica reciproca t(4;11).

L'aberrazione cromosomica t(4;11) è stata rilevata *in utero* e nel sangue neonatale suggerendo l'origine prenatale di tale anomalia genetica. Il fatto che nei bambini la ALL abbia un esordio molto precoce avvalora l'ipotesi che l'oncoproteina MLL-AF4 abbia un forte effetto tumorigenico e possa essere sufficiente ad instaurare la ALL. Questa ipotesi è ulteriormente corroborata dal fatto che i riarrangiamenti dei geni delle immunoglobuline e del *T cell receptor* sono rari nelle leucemie (5).

La proteina MLL-AF4 contiene la regione N-terminale di MLL e la regione C-terminale di AF4. In particolare, essa conserva il dominio di omologia con la DNA metiltrasferasi, gli *AT-hooks* e il segnale di localizzazione nucleare di MLL (39) e almeno il dominio transattivante di AF4. La composizione della chimera MLL-AF4 suggerisce che essa possa agire come oncoproteina tramite un meccanismo di *gain-of-function* o un meccanismo dominante negativo nella leucemogenesi (40). Sembra, inoltre, che entrambi i prodotti che si formano in seguito alla traslocazione 4;11, AF4-MLL e MLL-AF4 (rispettivamente sul cromosoma 4 e sul cromosoma 11), possano influenzare l'apoptosi, la crescita e il ciclo cellulare, in sistemi cellulari *in vitro* (41). In particolare, la proteina MLL-AF4 sembra essere coinvolta nella regolazione dei geni che codificano per gli inibitori delle chinasi ciclina dipendenti (CdKIs) tra cui p27kip1, oncosoppressore coinvolto nella regolazione della progressione del ciclo cellulare. Infatti, il decremento dell'attività di questa proteina, dovuto ad una diminuzione sia dell'RNA che della proteina, si traduce in progressione tumorale a livello di vari tessuti (42).

2. MATERIALI E METODI

2.1. Clonaggi

2.1.1. cDNA

Il cDNA full length di AF4 è stato preparato a partire da RNA totale estratto da cellule HEK293, utilizzando esameri a sequenza casuale come innesco e la trascrittasi inversa Super Script III (Invitrogen). Il cDNA così ottenuto è stato amplificato utilizzando oligonucleotidi specifici per AF4.

2.1.2. PCR

Tutte le PCR sono state condotte utilizzando la *Taq* DNAPolimerasi “Expand High Fidelity PCR System” (Roche) e oligonucleotidi di innesco dotati di siti di restrizione per l’inserimento dei frammenti amplificati negli opportuni vettori. Tutti i costrutti realizzati sono stati completamente sequenziati.

Il programma di cicli utilizzato è riportato nella seguente tabella:

1 ciclo	94°C	2 min
35 cicli	94°C	30 sec
	56°C	40 sec
	72°C	1 min/Kb
1 ciclo	72°C	7 min

2.1.3. Mutagenesi sito-specifica

La reazione di mutagenesi sito-specifica è stata eseguita con il kit di mutagenesi QuickChange (Stratagene), utilizzando come stampo il costrutto FLAG-AF4 ed una coppia di oligonucleotidi autocomplementari ed opportunamente mutati in posizione

centrale. In un volume di reazione di 50µl sono stati utilizzati 50 mg di DNA, 125 ng di ciascuno dei due primer, 1 µl di dNTP mix, 3.5 U/ µl di DNA polimerasi *PfuTurbo*.

Il programma di cicli utilizzato è riportato nella seguente tabella:

1 ciclo	95°C	30 sec
18 cicli	95°C	30 sec
	55°C	1 min
	68°C	17 min

Il prodotto di sintesi è stato digerito con l'enzima *Dpn I* per eliminare il DNA parentale metilato. 1 µl della reazione di digestione, contenente DNA plasmidico mutato di nuova sintesi, è stato utilizzato per trasformare cellule di *Epicurian Coli* del ceppo XL1-Blue chimicamente competenti.

2.1.4. Oligonucleotidi

Costrutto GST-AF4

Forward

GAF4-1F 5' -GCCGGGATCCGCAGCCCAGTCAAGTTTGTA-3'

GAF4-2F 5' -GCCGGGATCCGGCAGCGGCAGCAGGAC-3'

Reverse

GAF4-1R 5' -GCCGGTCTGACCTCGAGGGCTGGGGTATCCGAGAG-3'

GAF4-2R 5' -GCCGCTCGAGTTAAGGTGTTTTGGTTAATTCCTTG-3'

Costrutto GPA1

Forward

GAF4-1F 5' -GCCGGGATCCGCAGCCCAGTCAAGTTTGTA-3'

Reverse

GPA1-R 5' -GCCGCTCGAGTTAAGGGAGGGAGGTGGAAATGT-3'

Costrutto GPA2

Forward

GPA2-F 5' -GCCGGGATCCGGCAGCAGCAATAACAGTAA-3'

Reverse

GPA2-R 5' -GCCGCTCGAGTTACGTCCTTTTCGTCTTCACCT-3'

Costrutto GPA3

Forward

GPA3-F 5' -GCCGGGATCCCTGGACAACCTGGCTGACC-3'

Reverse
GPA3-R 5' -GCCGCTCGAGTTATGGCAAAGAAGGCACTGGA-3'

Costrutto GPA4

Forward
GPA4-F 5' -GCCGGGATCCTCACAGTCATCTTCATCTTCAT-3'
Reverse
GAF4-2R 5' -GCCGCTCGAGTTAAGGTGTTTTGGTTAATTCTTG-3'

Costrutto Flag-AF4

Forward
FAF4-1F 5' -GCCGGGTACCA GCAGCCCAGTCAAGTTTGTA-3'
FAF4-2F 5' -GCCGGGTACC CCACGAGTCAGGAGCATTCT-3'
Reverse
FAF4-1R 5' -GCCGGGATCCTTACCCTCTTCGTCTTCACCT-3'
FAF4-2R 5' -GCCGGGATCCTTAAGGTGTTTTGGTTAATTCTTG-3'

Costrutto Flag-AF4-1

Forward
FAF4-1F 5' -GCCGGGTACCA GCAGCCCAGTCAAGTTTGTA-3'
Reverse
FAF4-1R 5' -GCCGGGATCCCGTCTCTTCGTCTTCACCT-3'

Costrutto Flag-AF4-1.1

Forward
FAF4-1F 5' -GCCGGGTACCA GCAGCCCAGTCAAGTTTGTA-3'
Reverse
FA1.1-R 5' -GCCGGGATCCTTAGAGGGAGGTGGAAATGTCTG-3'

Costrutto Flag-AF4-1.2

Forward
FA1.2-F GCCGGGTACC CAGCAGCAATAACAGTAAAGG
Reverse
FA1.2-R GCCGGGATCCTTATTCTAGGGGCTCATTTTCTTC

Costrutto Flag-AF4-1.3

Forward
FA1.3-F GCCGGGTACCA GCATCAGCACATTCCAGCAG
Reverse
FAF4-1R 5' -GCCGGGATCCTTACCCTCTTCGTCTTCACCT-3'

Costrutto Flag-AF4-2

Forward
FA2-F 5' -GCCGGGTACC CCAGGGCTTCTTCCTATG-3'
Reverse
FAF4-2R 5' -GCCGGGATCCTTAAGGTGTTTTGGTTAATTCTTG-3'

Costrutto Flag-AF4S

Forward
GAF4-1F 5' -GCCGGGATCCGCAGCCCAGTCAAGTTTGTA-3'
3XFL-F 5' -GCCGGGTACC AATGGGCGGTAGGCGTGTA-3'
Reverse
GAF4-2R 5' -GCCGCTCGAGTTAAGGTGTTTTGGTTAATTCTTG-3'
3XFL-R 5' -GCCGGGATCCCTTGTTCATCGTCATCCTTGTA-3'

Mutagenesi AF4 S588A

Forward
AF4-S588A-F 5' -GGAGCTGTCAGAAGGCTCCGGCACAGCAG-3'
Reverse
S588A-R 5' -CTGCTGTGCCGGAGCCTTCTGACAGCTCC-3'
Tm: 82.3°C

Mutagenesi AF4 T766I

Forward
T766I-F 5' -CTTGATGGTGAAGATCATCCTAGACCTGCTCTCTC-3'
Reverse
T766I-R 5' -GAGAGAGCAGGTCTAGGATGATCTTCACCATCAAG-3'
Tm: 79.3°C

Costrutto Flag-AF4-cyt

Forward
FAF4-1F 5' -**GCCGGGTACCA**GCAGCCCAGTCAAGTTTGTA-3'
Reverse
FA-CYT-R 5' -**GCCGGGATCCTT**AGACTCAGAGAGCGAGAGCAGT-3'

Costrutto GFP-NLS

Forward
AF4NS-F 5' -**GCCGCTCGAGCT**CCCTCCAGTGAGAAGAAGAA-3'
Reverse
AF4NS-R 5' -**GCCGGAATTCTCA**CTCTACTCTTCTCTGCTTGG-3'

Costrutto GFP-NLS1

Forward
AF4NS-F 5' -**GCCGCTCGAGCT**CCCTCCAGTGAGAAGAAGAA-3'
Reverse
AF4NS1-R 5' -**GCCGGAATTCTCA**CTTGGTGTCTCTCAAAGGCA-3'

Costrutto GFP-NLS2

Forward
AF4NS2-F 5' -**GCCGCTCGAGCT**CTGCTCTCACCGCTCAGG-3'
Reverse
AF4NS2-R 5' -**GCCGGAATTCTCA**CTCCAGTCTGATTTTCTTGT-3'

Costrutto GFP-NLS3

Forward
AF4NS3-F 5' -**GCCGCTCGAGCT**AAGGAAATCAAATCACAGTCA-3'
Reverse
AF4NS-R 5' -**GCCGGAATTCTCA**CTCTACTCTTCTCTGCTT-3'

BamHI

XhoI

SalI

KpnI

Codone di stop

Estremità aggiuntiva

Spaziatore

2.1.5. Plasmidi

pGEX-4T1

Il vettore di espressione procariotico *pGEX-4T1* (Amersham) è un plasmide di circa 4.9 Kb, contiene un'origine di replicazione batterica *oriC*, il gene *amp^r*, che conferisce

resistenza all'antibiotico ampicillina necessario per la selezione positiva dei batteri trasformati, l'operone batterico *LacI^q* per l'induzione controllata, e un sito di clonaggio multiplo (MCS, Multiple Cloning Site). La peculiarità di questo plasmide sta nel possedere, a monte del sito di clonaggio, il gene che codifica per l'enzima glutatione S-transferasi (GST) che andrà ad inserirsi *in frame* all'estremo N-terminale della proteina di interesse codificata dall'inserto.

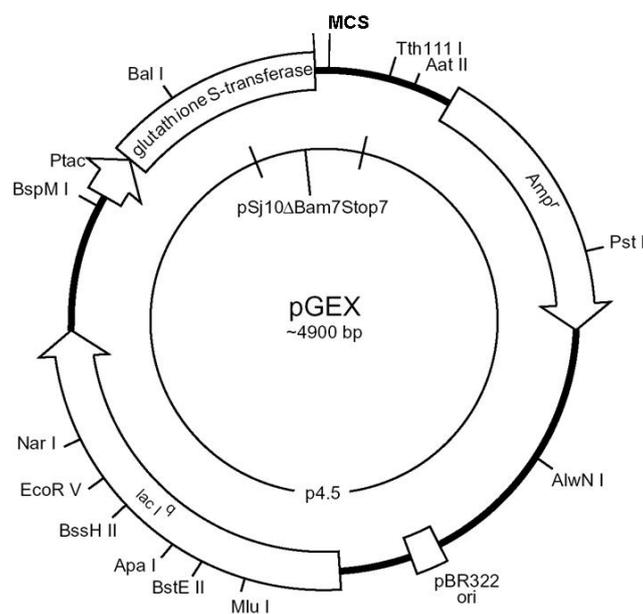


Figura 10. Rappresentazione schematica del vettore *pGEX-4T1*.

N-terminal-p3X-FLAG-CMV-7.1

Il vettore eucariotico *N-terminal-p3X-FLAG-CMV-7.1* (Sigma-Aldrich) è lungo 4717 bp e contiene un'origine di replicazione batterica *f1 ori*, il gene *amp^r*, e un MCS. La sequenza codificante per l'epitopo esapeptidico FLAG (Asp-Tyr-Lys-Xaa-Xaa-Asp) è presente in triplice copia ed è sotto il controllo di un promotore forte derivante dal Citomegalovirus umano (CMV), che assicura l'espressione ad alti livelli del transgene clonato. Inoltre, il terzo FLAG contiene una sequenza di riconoscimento per un

enterochinasi, enzima che permette di staccare, se necessario, il peptide 3xFLAG dalla proteina ricombinante.

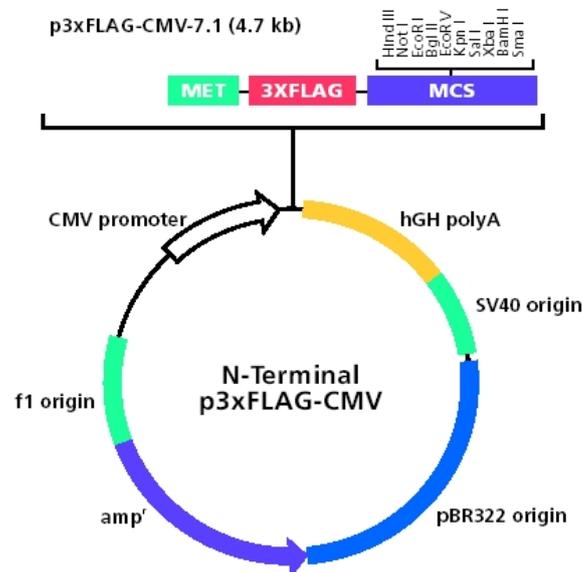


Figura 11. Rappresentazione schematica del vettore *p3X-FLAG*.

pcDNA3

pcDNA3 è un vettore di espressione per cellule eucariotiche che presenta le seguenti caratteristiche: un forte promotore quale CMV; l'origine di replicazione di SV40; il gene per la resistenza all'ampicillina per la selezione in *E.coli*; il segnale di poliadenilazione di BGH (Bovine Growth Hormone) per la corretta terminazione all'estremità 3' terminale dell'mRNA della proteina inserita nel vettore; il gene per la resistenza alla neomicina che consente di selezionare, usando G418, le cellule contenenti il plasmide stabilmente integrato nel genoma; l'MCS e a monte di quest'ultimo, le sequenze SP6 e T7 che possono essere utilizzate come *priming* sia per reazioni di sequenza sia per la trascrizione *in vitro*.

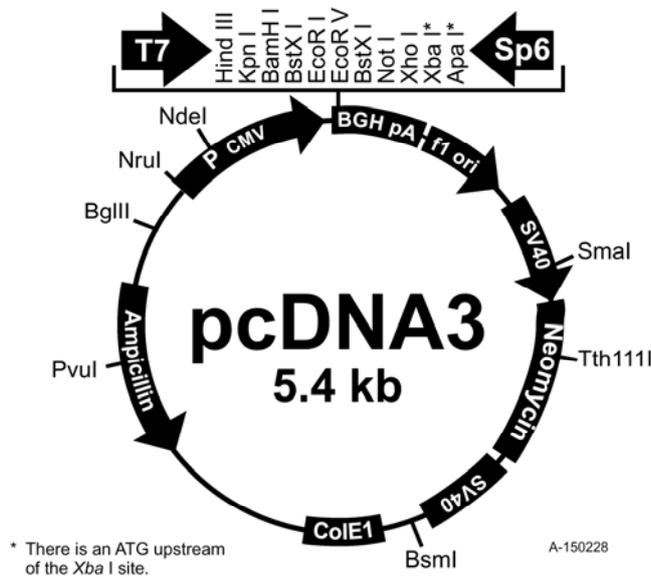


Figura 12. Rappresentazione schematica del vettore *pcDNA3*.

pEGFP-N1

pEGFP-N1 (Clontech) codifica per una variante della proteina selvatica GFP che è stata ottimizzata per una più brillante fluorescenza ed un'alta espressione in cellule di mammifero. In particolare, *pEGFP-N1* codifica per la variante GFPmut1 che contiene la doppia sostituzione Phe64Leu e Ser65Thr. A monte della sequenza codificante si trovano il promotore di CMV, per assicurare elevati livelli trascrizionali, e sequenze di tipo Kozak (sito di inizio della traduzione), per incrementare l'efficienza di traduzione nelle cellule eucariotiche e l'MCS. A valle della sequenza di EGFP si trova l'SV40 poli A. Il gene clonato nell'MCS, se nel corretto codice di lettura, sarà espresso in fusione all'N-terminale dell'EGFP. Il segnale di poliadenilazione di SV40, che si trova a valle del gene dell'EGFP, dirige la corretta terminazione dell'estremità 3' dell'mRNA di EGFP. Il vettore contiene inoltre l'origine di SV40 per la replicazione in cellule di mammifero che esprimano l'antigene T di SV40. Contiene inoltre una cassetta

per la resistenza alla neomicina (neo) che consente di selezionare usando G418 le cellule che contengono stabilmente inserito il plasmide nel proprio genoma; questa cassetta è costituita dal promotore di SV40, dal gene per la resistenza neomicina/kanamicina di Tn5, e dal segnale di poliadenilazione per il gene della timidin-kinasi di Herpes simplex. Inoltre, a valle di questa cassetta è espressa la resistenza alla Kanamicina in *E. coli*. pEGFP-N1 è provvisto, inoltre, dell'origine di replicazione pUC19, per la propagazione in *E. coli* ed una *f1 ori* per la produzione di DNA a singolo filamento.

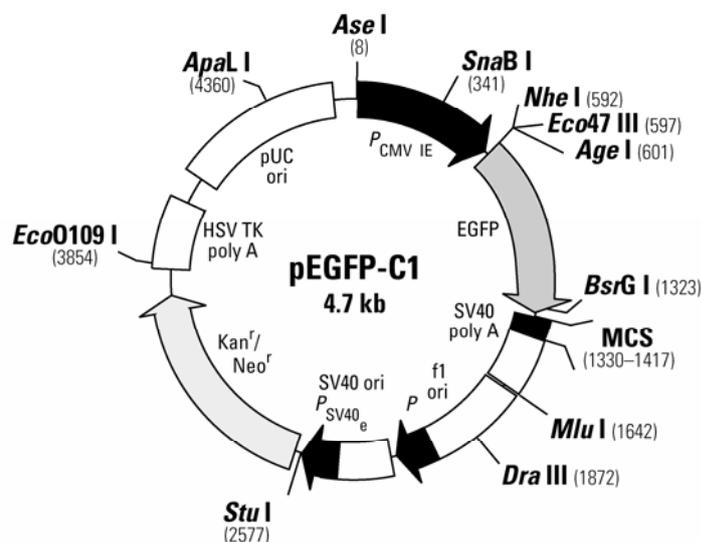


Figura 13. Rappresentazione schematica del vettore *pEGFP-N1*.

2.2. Espressione e purificazione di proteine ricombinanti in sistema procariotico

I costrutti ottenuti sfruttando il plasmide pGEX-4T1 sono stati utilizzati per trasformare cellule di *E. coli* del ceppo BL21/DE3 rese chimicamente competenti con il CaCl_2 . Queste sono state utilizzate per la realizzazione di colture batteriche in grado di esprimere le proteine ricombinanti fuse al loro N-terminale con la GST. Tali proteine sono state purificate mediante cromatografia di affinità utilizzando la resina “Glutathione-Sepharose-4B (Amersham Bioscience). Le condizioni sperimentali messe a punto sono le seguenti: sia crescita che induzione a 37°C per 160 min in LB, ampicillina $100 \mu\text{g/ml}$, IPTG 1 mM ; lisi mediante sonicazione in PBS con inibitori di proteasi e Triton 1%; “binding” per 30 min in agitazione a T.A.; lavaggi con PBS; eluizione con *buffer* Tris-HCl 50 mM pH 8 con inibitori di proteasi, DTT 1 mM , glutatione ridotto 10 mM .

2.3. Colture cellulari e trasfezioni

2.1.2 Linee cellulari

RS4;11

Precursori leucemici di linfociti B umani, stabiliti dal midollo osseo di una donna di 32 anni affetta da ALL. Crescono in sospensione singolarmente, in mezzo alpha-MEM 90%, FBS 10%, fino ad una densità massima di 4×10^6 cellule/ml, a 37°C, con CO₂ 5%, e presentano un ciclo replicativo di circa 50 h.

Jurkat

Linfociti T leucemici, stabiliti dal sangue periferico di un ragazzo di 14 anni affetto da ALL. Crescono in sospensione singolarmente o a piccoli gruppi in mezzo RPMI 90%, FBS 10%, L-glutamina 2 mM, fino ad una densità massima di 1.5×10^6 cellule/ml, a 37°C, con CO₂ 5%, e presentano un ciclo replicativo di circa 25-35 h.

697

Precursori leucemici di linfociti B umani, stabiliti dal midollo osseo di un ragazzo di 12 anni affetto da ALL. Crescono in sospensione, singolarmente o a piccoli gruppi fino ad una densità massima di 1×10^6 cellule/ml, in un mezzo costituito da RPMI 80-90%, FBS 10-20% a 37°C con CO₂ 5% e presentano un ciclo replicativo di circa 30-40 h.

CEM

Linfociti T leucemici, stabiliti dal sangue periferico di una bambina Caucasica di 3 anni affetta da ALL. Crescono in sospensione o formando un monostrato di cellule

epitelioidi aderenti, in mezzo RPMI 90%, FBS 10%, fino ad una densità massima di 2×10^6 cellule/ml o di $4-5 \times 10^6$ cellule/80 cm², a 37°C, con CO₂ 5%, e presentano un ciclo replicativo di circa 24-30 h.

REH

Precursori leucemici di linfociti B umani, stabiliti dal sangue periferico di una ragazza Nord-Africana di 15 anni affetta da ALL. Crescono in sospensione singolarmente, in mezzo RPMI 90%, FBS 10%, fino ad una densità massima di $3-5 \times 10^6$ cellule/ml, a 37°C, con CO₂ 5%, e presentano un ciclo replicativo di circa 50-70 h.

TOM-1

Precursori leucemici di linfociti B umani, stabiliti dal midollo osseo di una donna di 54 anni affetta da ALL, cromosoma Philadelphia-positiva, portatrice della variante ALL del gene di fusione BCR-ABL (e1-a2). Crescono in sospensione, singolarmente o a piccoli gruppi fino ad una densità massima di 2.5×10^6 cellule/ml, in un mezzo costituito da RPMI 80%, FBS 20% a 37°C con CO₂ 5% e presentano un ciclo replicativo di circa 60-70 h.

K562

Mieloblasti leucemici, stabiliti da effusione pleurale di una donna di 53 anni con leucemia mieloide cronica (MCL) in crisi blastica e contenenti il cromosoma Philadelphia con un gene di fusione BCR-ABL b3-a2. Crescono in sospensione singolarmente, in mezzo RPMI 90%, FBS 10%, fino ad una densità massima di $1-1.5 \times 10^6$ cellule/ml, a 37°C, con CO₂ 5%, e presentano un ciclo replicativo di circa 30-40 h.

HEK 293

Cellule di rene embrionale umano, stabilite da un rene embrionale primario trasformato con adenovirus di tipo 5. Crescono in monostrato come cellule aderenti fibroblastoidi in mezzo DMEM 85-90%, FBS 10-15%, fino ad una densità massima di 40×10^6 cellule/175 cm², a 37°C, con CO₂ 5%, e presentano un ciclo replicativo di circa 20-24 h.

SAOS

Cellule di osteosarcoma umano, stabilite da sarcoma osteogenico primario di una ragazza Caucasica di 11 anni. Crescono in monostrato come cellule epitelioidi aderenti, in mezzo DMEM 85-90%, FBS 10-15%, fino ad una densità massima di 10×10^6 cellule/175 cm², a 37°C, con CO₂ 5%, e presentano un ciclo replicativo di circa 43 h.

HeLa

Cellule di carcinoma della cervice uterina, stabilite dal carcinoma della cervice epitelioide di una donna di colore di 31 anni. Crescono in monostrato come cellule epitelioidi aderenti, in mezzo DMEM 90%, FBS 10%, L-glutammina 2 mM, fino ad una densità massima di $5-15 \times 10^6$ cellule/175 cm², a 37°C, con CO₂ 5%, e presentano un ciclo replicativo di circa 48 h.

COS-1

Cellule di rene di scimmia verde africana (*Cercopithecus aethiops*), stabilite da cellule CV-1 di scimmia, trasformate con un mutante difettivo di virus SV-40. Crescono in monostrato come cellule aderenti fibroblastoidi in mezzo DMEM 90%, FBS 10%, fino ad una densità massima di 5×10^6 cellule/80 cm², a 37°C, con CO₂ 5%, e presentano un ciclo replicativo di circa 48 h.

SK-N

Cellule di neuroblastoma umano, stabilite da metastasi sopraorbitali di un neuroblastoma di una ragazza di 14 anni. Crescono in monostrato come cellule aderenti fibroblastoidi in mezzo DMEM 90%, FBS 10%, L-glutammina 2 mM, fino ad una densità massima di 15×10^6 cellule/175 cm², a 37°C, con CO₂ 5%, e presentano un ciclo replicativo di circa 48 h.

2.3.1. Colture cellulari

Le cellule HEK293, HeLa, Cos-1, SAOS e SK-NB-E sono state tenute in coltura utilizzando come mezzo “DMEM” supplementato con FBS 10% finale. Per le K562, le RS4;11, le REH e le 697 è invece stato utilizzato il mezzo “RPMI”, FBS 10%, mentre per le TOM-1 il mezzo “RPMI” è stato supplementato con FBS 15%.

2.3.2. Trasfezioni

Gli esperimenti di trasfezione effettuati per verificare il livello di espressione dei costrutti Flag-AF4 e Flag-AF4-1 nelle varie linee cellulari sono stati condotti utilizzando il reagente Lipofectamine 2000 (Invitrogen) (secondo il protocollo suggerito dalla ditta) o il reagente Fugene (Roche) (secondo il protocollo suggerito dalla ditta).

Le trasfezioni per gli esperimenti di immunoprecipitazione sono state condotte nella linea cellulare HEK293 utilizzando la tecnica del Calcio Fosfato ed un tempo di trasfezione di 48h.

Le trasfezioni per gli esperimenti di immunofluorescenza sono state condotte per 12 h in cellule SAOS con la metodica del Calcio Fosfato.

Le trasfezioni per gli esperimenti di localizzazione subcellulare di proteine GFP sono stati condotti per 12 h in cellule HeLa con Lipofectamine 2000.

Le cellule che crescono in monostrato adese alla piastra (HEK293, HeLa, Cos-1, SAOS, SK-NB-E) sono state trasfettate ad confluenza di circa $1 \times 10^3/\text{mm}^2$, mentre le cellule che crescono in sospensione (K562, Jurkat, RS4;11, REH, 697, CEM, TOM-1) sono state trasfettate ad una concentrazione di circa $1 \times 10^3/\text{ml}$.

2.4. Estratti proteici

Gli estratti proteici per la valutazione del livello di espressione delle proteine ricombinanti in cellule eucariotiche sono stati preparati effettuando la lisi delle cellule in un *buffer* Tris-HCl 20 mM pH 8, NaCl 150 mM, NP-40 1%, PMSF 0,5 mM, EDTA 1 mM, Glicerolo 10%, cocktail di inibitori di proteasi (per estratti cellulari di mammifero) (Sigma-Aldrich) 1:100. Tale *buffer* è stato utilizzato in quantità di 40 μl per ogni milione di cellule.

Gli estratti proteici per gli esperimenti di immunoprecipitazione sono stati preparati utilizzando per la lisi un *buffer* Tris-HCl 50 mM pH 8, NaCl 150 mM, NP-40 0,1%, PMSF 0,5 mM, EDTA 1 mM, Glicerolo 10%, cocktail di inibitori di proteasi 1:100. Tale *buffer* è stato utilizzato in quantità di 200 μl per ogni milione di cellule.

In tutti i casi le cellule sono state incubate con il *buffer* di lisi per 20 min a 4°C, e gli estratti sono stati chiarificati mediante centrifugazione a 4°C per 30 min a 12000 RCF.

2.5. Western Blot

I Western Blot per la valutazione del livello di espressione delle proteine ricombinanti sono stati condotti utilizzando 40-60 μg di estratto proteico. Per l'analisi delle purificazioni con glutatione-sepharosio, invece sono stati utilizzati 5 μg di eluato. Per l'analisi degli esperimenti di immunoprecipitazione con resina anti-Flag, sono stati utilizzati 10-20 μl di eluato derivante da 2 mg di estratto oppure 10-20 μl di immunoprecipitato non eluito. Per l'analisi degli esperimenti di immunoprecipitazione con anticorpi specifici e proteina-A/G-sepharosio, sono stati utilizzati 20 μl di resina adoperata per 2 mg di estratto. In tutti i casi, dopo l'aggiunta del Laemly *Buffer* (Biorad; v/v 1:1), i campioni sono stati incubati per 10 min a 95° C e per 5 min a 4°C e sono poi stati caricati su SDS-PAGE all'8% o al 10%. Le proteine sono state trasferite dai gel su membrane di nitrocellulosa mediante elettroblotting a 25 Volts per 16 h a 4°C, in tampone Tris-Glicina. Dopo il trasferimento, i filtri sono stati sottoposti a *blocking* per 1 h a temperatura ambiente, in TBS 1X (Tris *Buffer* Saline), Tween 0.1%, proteine del latte (Biorad) 5% e poi incubati a temperatura ambiente con gli anticorpi primari per 90 min in TBS 1X, Tween 0.1%, proteine del latte 5%. Dopo 3 lavaggi di 5 min in TBS 1X, Tween 0.1%, i filtri sono stati incubati a temperatura ambiente con gli anticorpi secondari per 45 min TBS 1X, Tween 0.1%, proteine del latte 5% e poi di nuovo lavati 3 volte in TBS 1X, Tween 0.1%.

Gli anticorpi primari adoperati e le rispettive diluizioni di utilizzo sono: mouse monoclonal anti-Flag (Sigma-Aldrich) 1:5000, rabbit policlonal anti-Cdk9 (Santa Cruz) 1:1000, rabbit policlonal anti-14-3-3 θ (Santa Cruz) 1:500, goat policlonal anti-CRSP130 (Santa Cruz) 1:200, goat policlonal anti-CRSP33 (Santa Cruz) 1:200, rabbit policlonal anti-GST (Sigma-Aldrich) 1:5000.

Gli anticorpi secondari adoperati sono coniugati con enzima HRP e le rispettive diluizioni di utilizzo sono: anti-mouse 1:5000, anti-rabbit 1:5000, anti-goat 1:10000. Nel caso dei Western per l'analisi degli immunoprecipitati la rivelazione delle proteine sui filtri è avvenuta mediante l'utilizzo del kit "ECL plus" (Amersham). Negli altri casi è stato invece utilizzato il kit "ECL" (Amersham).

2.6. Immunoprecipitazioni

Per effettuare gli esperimenti di immunoprecipitazione sulle proteine ricombinanti dotate di triplo epitopo Flag, sono stati utilizzati 10 µl di anticorpi anti-Flag-M2-Agarose (Sigma-Aldrich) per ogni mg di estratto. L'incubazione è stata condotta per 3 h a 4°C in agitazione. Sono poi state effettuate, a 4°C "in batch", due serie di 3 lavaggi ognuna. La prima con un *buffer* Tris-HCl 50 mM pH 8, NaCl 300 mM, NP-40 0,1%, PMSF 0,5 mM, EDTA 1 mM, Glicerolo 10% e inibitori di proteasi 1:100. La seconda serie invece è stata effettuata con un *buffer* Tris-HCl 50 mM pH 8, NaCl 150 mM, NP-40 0,1%, PMSF 0,5 mM, EDTA 1 mM, Glicerolo 10% e inibitori di proteasi 1:100. I campioni sono stati poi trasferiti in microcolonne da centrifuga. L'eluizione è avvenuta in centrifuga a 4°C per 2 min a 5000 RCF utilizzando un *buffer* Tris-HCl 50 mM pH 8, NaCl 150 mM, NP-40 0,1%, PMSF 0,5 mM, EDTA 1 mM, glicerolo 10%, inibitori di proteasi 1:100, e p3xFlag-peptide (Sigma-Aldrich) 200 µg/ml. Le immunoprecipitazioni con anticorpi specifici sono state effettuate utilizzando 20 µl di anticorpo per ogni 10 mg di estratto, incubando per 16 h a 4°C in agitazione. Sono stati aggiunti 20 ml di proteina-A/G-sepharoso, incubando per 1 h a 4°C in agitazione e sono stati effettuati 3

lavaggi con un *buffer* Tris-HCl 50 mM pH 8, NaCl 150 mM, NP-40 0,1%, PMSF 0,5 mM, EDTA 1 mM, Glicerolo 10% e inibitori di proteasi 1:100.

2.7. Analisi degli immunocomplessi mediante SDS-PAGE e spettrometria di massa

Dopo la purificazione per immunoaffinità con anti-Flag-agarosio, 50 µg di proteine eluite sono stati precipitati in metanolo/cloroformio e risospesi in Laemly *Buffer*, incubati a 95° per 10 min, a 4°C per 5 min e poi caricati su SDS-PAGE 10%, utilizzando come marcatore di peso molecolare il Precision Plus (Sigma-Aldrich). I pattern di bande elettroforetiche sono stati visualizzati utilizzando il Gel Colloide Blue Stain Reagent (Pierce).

Le bande sul gel sono state tagliate, idrolizzate, e digerite tripticamente. Le miscele peptidiche risultanti sono state analizzate mediante un apparato nano Liquid Chromatography – ElectroSpray Ionizzazione – tandem Mass Spectrometry (nanoLC-ESI-MS/MS), con trappola LC/MSD XCT Ultra (Agilent Technologies) ed equipaggiato con un'interfaccia 1100 HPLC –Chip/MS Cube (Agilent Technologies). I dati grezzi ottenuti dalle analisi con l'apparato LC-ESI-MS/MS sono stati convertiti in formato di testo Mascot, per identificare le proteine mediante l'utilizzo del software Mascot versione 2.1 “in home” (MatrixScience) (45). La ricerca delle proteine è stata effettuata contro i database NCBI nr-20061017 e Sprot-50.9. Per ogni ione analizzato, il software Mascot calcola un punteggio di probabilità di *Mowse* (45) pari a $-10 \times \log(P)$, dove *P* è la probabilità che la corrispondenza sia un evento *random*. Singoli punteggi maggiori di 38 indicano identità o elevata similitudine ($p < 0.005$) in quanto mostrano un buon

rapporto segnale/rumore che porta ad un'interpretazione non ambigua del dato. Gli spettri MS/MS dei peptidi con punteggi Mascot pari a 38 sono stati ispezionati manualmente e inclusi nella lista delle identificazioni riuscite solo quando è stata osservata una serie di almeno quattro ioni y o b continui. Tutte le analisi di spettrometria di massa sono state effettuate nel laboratorio di proteomica del CEINGE Biotecnologie Avanzate s.c.a r.l..

2.8. Analisi *in silico*

Tutte le interazioni di AF4 sono state esaminate utilizzando i dati presenti nello *Human Protein Reference Database* (HPRD; www.hprd.org) e nel *Biomolecular Interaction Network Database* (BIND; <http://bond.unleashedinformatics.com/Action>), per costruire *in silico* una mappa dell'interattoma della proteina AF4. Il *clustering* delle proteine è stato eseguito utilizzando il *software* DAVID (www.david.abcc.ncifcrf.gov).

2.9. Microscopia a fluorescenza

Le cellule adese ai vetrini sono state fissate per 5 min a temperatura ambiente con PBS 1X, PFA (paraformaldeide) 4% e lavate con PBS 1X.

Le cellule esprimenti proteine ricombinanti con epitopo Flag, dopo il fissaggio, sono state permeabilizzate mediante incubazione a temperatura ambiente per 30 min in PBS 1X, Triton 0,2% e lavate in PBS 1X. Sono state sottoposte a blocking mediante incubazione per 1 h a temperatura ambiente in PBS 1X, Triton 0,1 %, FBS 10% e poi lavate in PBS 1X. Successivamente, sono state incubate con l'anticorpo primario anti-Flag monoclonale 1:500, in PBS 1X, Triton 0,1 %, FBS 10% e poi lavate in PBS 1X. Sono state incubate per 3 h a temperatura ambiente con l'anticorpo secondario anti-mouse coniugato con TRITC diluito 1:100 in PBS 1X, Triton 0,1 %, FBS 10% e poi lavate in PBS 1X. Sono state trattate con DAPI al buio a temperatura ambiente per 5 min e poi lavate con PBS 1X.

Le cellule esprimenti proteine ricombinanti GFP, dopo il fissaggio sono state trattate direttamente con DAPI al buio a temperatura ambiente per 5 min e poi lavate con PBS 1X.

In tutti i casi i vetrini sono stati montati con l'utilizzo di una soluzione contenente PBS 1X e Glicerolo 15% e sigillati con smalto trasparente. I vetrini sono stati poi osservati al microscopio a fluorescenza (Zeiss) dotato di fotocamera digitale ed interfaccia software, utilizzando un obiettivo 100X a immersione. Le immagini sono state acquisite grazie al software Axio Vision (Zeiss).

3. RISULTATI

3.1. Identificazione degli interattori di AF4

3.1.1. Premesse

Se i meccanismi molecolari, che sono alla base delle leucemie causate dalla traslocazione t(4;11), sono ancora poco compresi, questo è dovuto al fatto che poco o nulla si conosce della proteina AF4 e della sua funzione biologica. Trattandosi di un putativo fattore trascrizionale, che apparentemente non presenta motivi conosciuti di legame al DNA, una delle strategie migliori per risalire alla sua funzione era quella di identificare i suoi eventuali interattori proteici. Un possibile approccio alla identificazione di interattori proteici di AF4, premettendo che non esistevano sul mercato anticorpi specifici, era quello di esprimere la proteina AF4 ricombinante fusa ad un tag, ed utilizzarla come esca molecolare per isolare le proteine in grado di formare con essa complessi multi-proteici, provvedendo poi all'identificazione dei componenti del complesso mediante spettrometria di massa.

3.1.2. Espressione e purificazione di AF4 e suoi frammenti in sistema procariotico

La tecnica più semplice per isolare fattori proteici in grado di legarsi ad AF4 sembrava essere quella di esprimere questa proteina in un sistema eterologo batterico, di purificarla e di utilizzarla come esca icubandola con estratti proteici di linee cellulari o di tessuti umani. A tal scopo, abbiamo espresso in E.coli AF4 umana fusa con la glutatione S-transferasi (GST), per poi effettuare esperimenti di GST *pull-down*. Abbiamo pertanto effettuato il clonaggio del cDNA di AF4 nel plasmide *pGEX-4T1* (Amersham Biosciences), un vettore d'espressione procariotico in grado di esprimere la proteina codificata dall'inserito in fusione al suo N-terminale con GST (**Fig.14**).

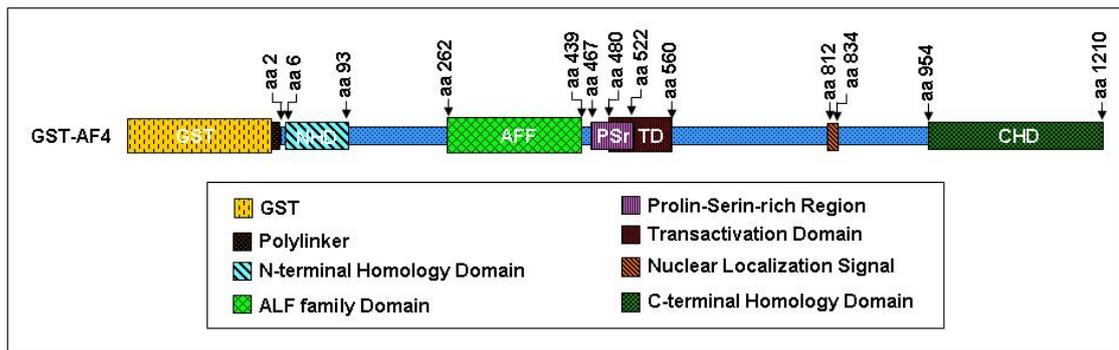


Figura 14. Rappresentazione schematica della proteina ricombinante espressa dal costrutto GST-AF4, in cui l'estremità N-terminale di AF4 è fusa con quella C-terminale dell'enzima Glutathione-S-transferasi. In figura sono schematizzati diversi domini di AF4. NHD (N-terminal Homology Domain) e CHD (C-terminal Homology Domain) sono molto conservati in tutti gli ortologhi di AF4. ALF è molto conservato in tutte le proteine della famiglia ALF ed in tutti i loro ortologhi. PSr (Prolin Serin rich region) è una regione ricca in serina e prolina. TD (Transactivation Domain) è un dominio ad elevata attività transattivante (6).

L'inserto è stato completamente sequenziato, per confermarne l'inserimento nella corretta cornice di lettura rispetto al GST-tag e la corrispondenza con la sequenza nucleotidica presente in banca dati. Il costrutto, denominato GST-AF4, è stato quindi utilizzato per trasformare cellule di *E. coli* del ceppo BL21/DE3, adatto all'espressione inducibile del gene esogeno. Sono state determinate le migliori condizioni (temperatura, densità batterica, concentrazione di IPTG) per l'espressione della proteina AF4, nonché della sua purificazione mediante cromatografia di affinità (resina Sepharose-Glutathione), resa possibile grazie alla presenza all'N-terminale del GST-tag. Le varie fasi del processo di purificazione sono state valutate mediante SDS-PAGE seguito da colorazione al Blue-Comassie e da Western Blot con anticorpi anti-GST (dati non riportati). Sebbene la proteina GST-AF4 sia prodotta in buona quantità e sia solubile in condizioni native, essa non mostra all'analisi su SDS-PAGE la massa molecolare attesa, bensì migra su gel come una proteina di circa 125 KDa (**Fig.15**).

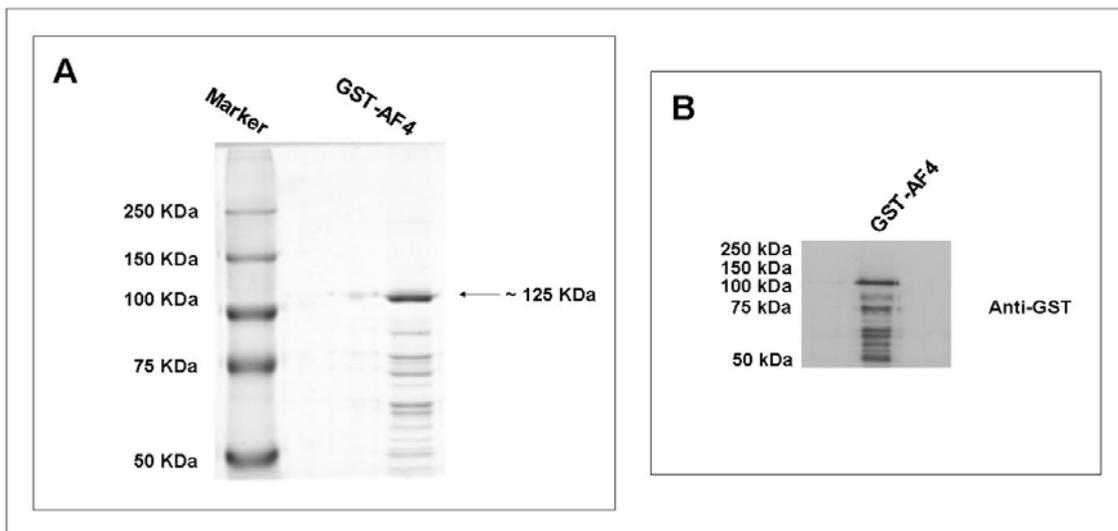


Figura 15. **A)** Analisi mediante SDS-PAGE all'8% colorato al Blue Comassie della proteina ricombinante GST-AF4 espressa in E.coli BL21/DE3 e purificata mediante cromatografia di affinità, con una resina Glutazione-Sepharosio. Sul gel sono stati caricati 4 μ g di eluato. Come mostrato in figura la proteina non migra sul gel all'altezza attesa. Infatti il peso molecolare atteso deriva dalla somma dei pesi della Glutazione-S-transferasi che è di circa 27 KDa e quello di AF4 che è di 140 KDa. Tuttavia, in base ai dati riportati in letteratura, AF4 migra come una proteina di circa 180 KDa, in quanto ha un punto isoelettrico teorico di circa 9.8 (proteina basica). L'osservazione che la proteina purificata non migri all'altezza attesa (circa 210) KDa può dipendere dall'incapacità del sistema di sintesi proteica batterico può non essere in grado di produrre in maniera efficace e completa una proteina eterologa di dimensioni molto grandi. In alternativa, si potrebbe ipotizzare una rapida degradazione della proteina a causa di un *fold*ing non corretto. I fenomeni di degradazione sono anche suggeriti dalle numerose bande di basso peso molecolare visualizzate sul gel. **B)** Analisi mediante Western Blot all'8% con anti-GST (Sigma-Aldrich) della proteina ricombinante GST-AF4 espressa in E.coli BL21/DE3 e purificata mediante cromatografia di affinità, con una resina Glutazione-Sepharosio. Sul gel sono stati caricati 0,5 μ g di eluato.

L'analisi mediante spettrometria di massa indica che il prodotto finale dei processi di espressione e purificazione in procarioti è una proteina che contiene, oltre alla GST, solo i primi 700 residui amminoacidici di AF4 (**Fig.16**).

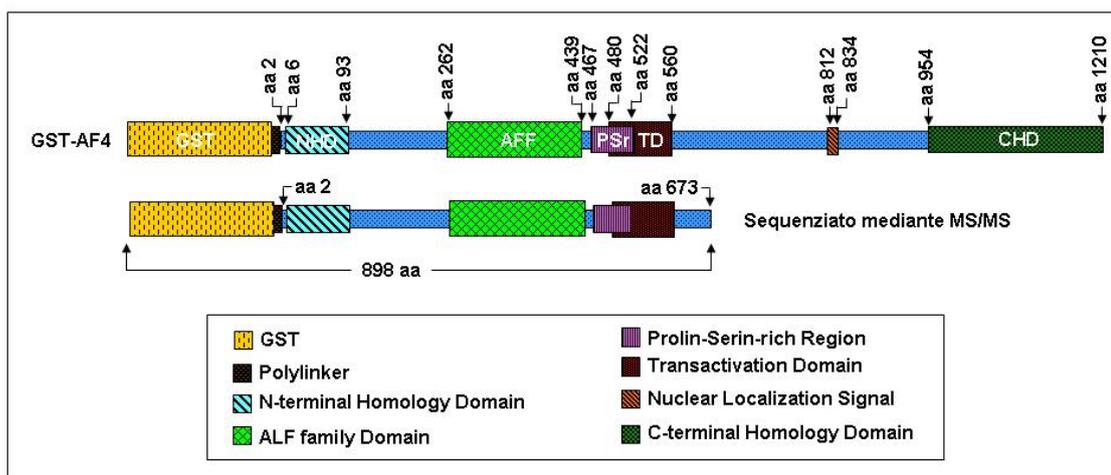


Figura 16. Rappresentazione schematica della porzione di AF4 prodotta a partire dal costrutto GST-AF4, così come stabilito dall'analisi della sua sequenza, eseguita nel laboratorio di proteomica del CEINGE Biotecnologie Avanzate s.c.a.r.l.. 100 μ g di proteina ricombinante GST-AF4 purificata mediante cromatografia per affinità sono stati separati mediante SDS-PAGE e quindi sequenziati. La proteina ricombinante purificata è costituita da 898 aa (contiene i primi 673 aa di AF4), e non contiene le regioni carbossiterminali di AF4. Ciò è in accordo con la massa molecolare di 125 KDa valutata mediante SDS-PAGE (vedi Fig.2).

Inoltre, le analisi mediante SDS-PAGE (**Fig.15**) e Western blot di AF4 ricombinante purificata per affinità mostrano la presenza, oltre della proteina di interesse, di molte bande di peso molecolare inferiore, costituite probabilmente da prodotti di degradazione che ne rendevano sconsigliabile l'utilizzo negli esperimenti successivi. Abbiamo quindi suddiviso arbitrariamente la sequenza del gene AF4 in quattro regioni parzialmente sovrapposte e abbiamo sfruttato ancora il vettore *pGEX-4T1* per esprimere i quattro peptidi corrispondenti, che sono stati denominati rispettivamente GPA1, GPA2, GPA3, GPA4 (**Fig.17**).

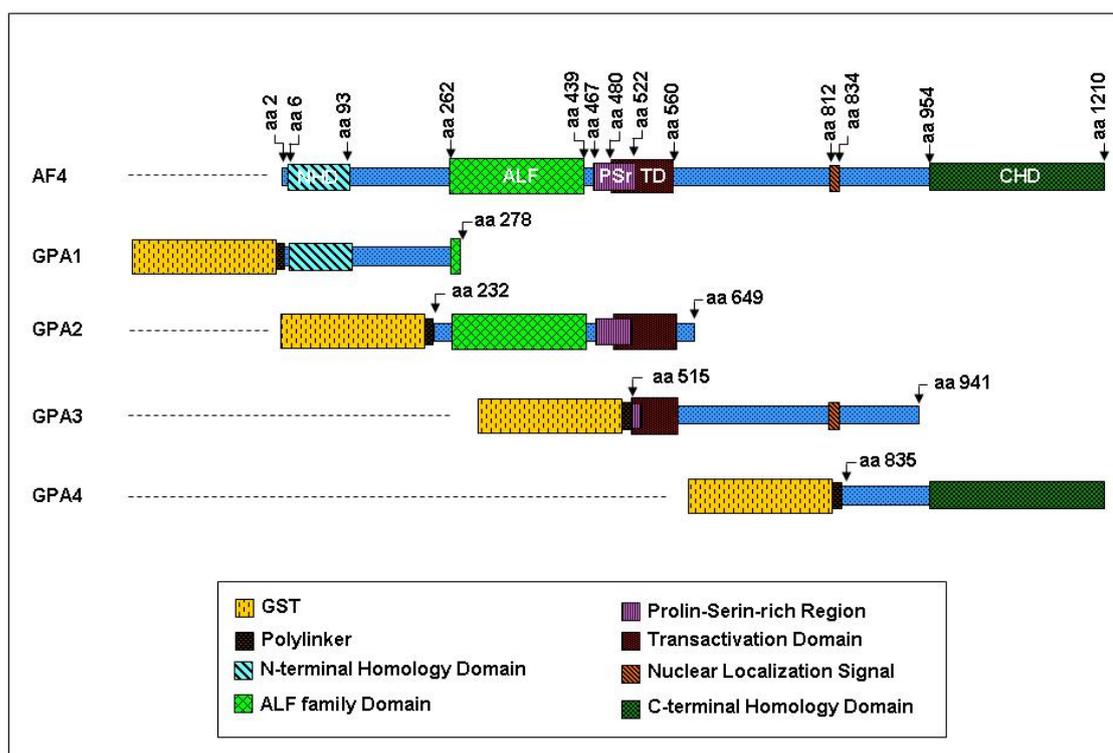


Figura 17. Rappresentazione schematica delle proteina ricombinanti espresse dai costrutti GPA1, GPA2, GPA3, GPA4, in cui le estremità N-terminali dei frammenti di AF4 sono fuse con quella C-terminale dell'enzima Glutazione-S-transferasi. Come si vede in figura, i frammenti di AF4 sono stati scelti in modo da avere estremità leggermente sovrapposte l'uno con l'altro. Il primo costrutto contiene completamente il dominio NHD, il secondo contiene completamente i domini AFF, PSr e TD. Il terzo frammento include buona parte del dominio TD ed il segnale di localizzazione nucleare. L'ultimo costrutto contiene il dominio CHD.

Il dominio transattivante e il dominio di localizzazione nucleare di AF4 sono stati inclusi per intero nel secondo e nel terzo costrutto rispettivamente. I costrutti sono stati interamente sequenziati ed utilizzati per trasformare cellule di *E. coli* BL21/DE3. Sono state stabilite le condizioni più opportune di crescita batterica, induzione nonché di purificazione delle proteine ricombinanti fuse al tag GST. Con l'analisi mediante SDS-PAGE e Western Blot, abbiamo verificato che i 4 quattro frammenti proteici si esprimono a buoni livelli e hanno la dimensione attesa. Tuttavia anche in questo caso abbiamo dovuto constatare che la metodologia utilizzata non ci consentiva di ottenere frazioni purificate sufficientemente pulite (**Fig.18**) da poter essere utilizzate per

eventuali successivi esperimenti di “GST pull-down”. Inoltre le procedure di purificazione utilizzate ci consentivano di ottenere una resa accettabile solo nel caso dei peptidi GPA1 e GPA2.

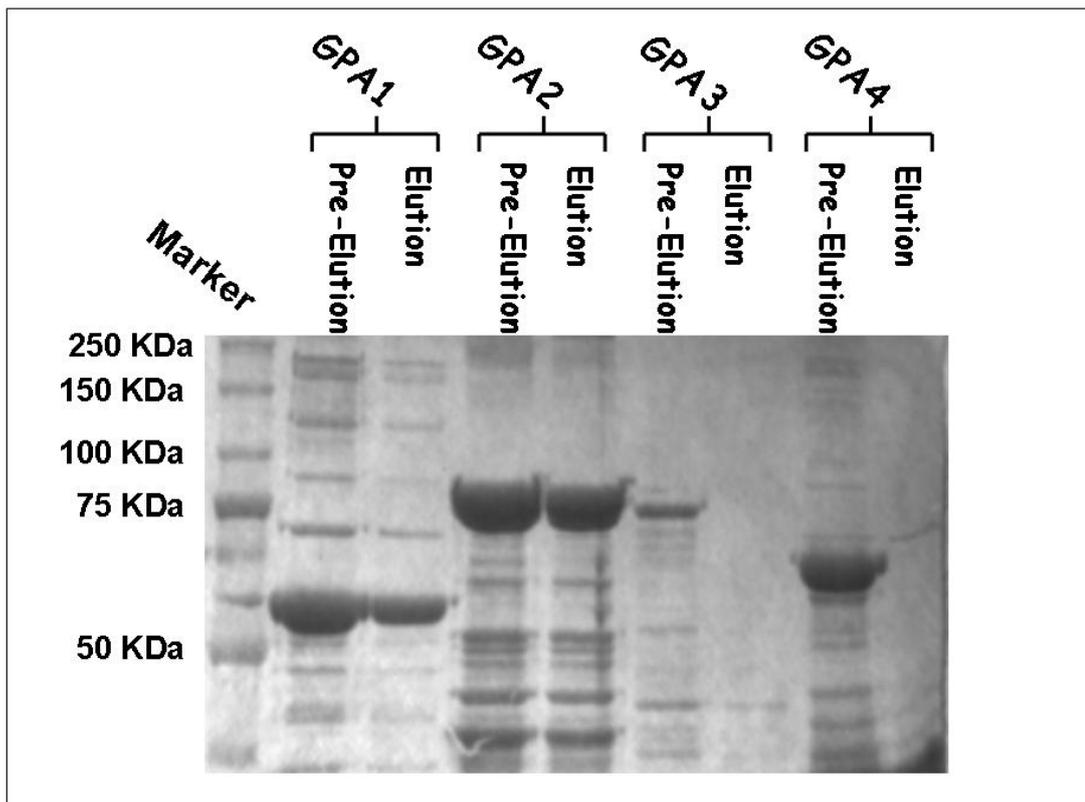


Figura 18. Analisi mediante SDS-PAGE al 10% colorato al Blue Comassie delle proteine ricombinanti GPA1, GPA2, GPA3, GPA4 espresse in E.coli BL21/DE3 e purificate per affinità, con una resina Glutazione-Sepharosio. Si tenga presente che le quattro proteine nelle condizioni utilizzate (vedi Par.XXX) vengono espresse in maniera quantitativamente identica. Come descritto nella sezione materiali e metodi (**vedi Par.XXX**) è stato utilizzato 1 μ l di resina al 50% per ogni ml di coltura batterica, ed 1 μ l di buffer di eluizione per ogni μ l di resina al 50%. A ciascuna delle proteine sono state riservati due pozzetti del gel. Nel primo pozzetto sono stati caricati 20 μ l di resina pre-eluzione, mentre nel secondo sono stati caricati 20 μ l di eluato. Come si vede in figura, solo le prime due proteine ricombinanti vengono purificate con efficienza. Per quanto riguarda GPA3, essa non si lega alla resina in maniera efficace. Nel caso invece della proteina GPA4, essa si lega in maniera efficiente alla resina, ma non se ne distacca.

3.1.3. Espressione di AF4 in sistemi eucariotici

Un altro modo per isolare un eventuale complesso proteico utilizzando come esca molecolare AF4, era quella di esprimere la proteina-esca in un sistema cellulare eucariotico. Abbiamo pertanto clonato l'intero cDNA di AF4 nel vettore eucariotico *N-terminal-CMV-p3X-FLAG 7.1* (Sigma-Aldrich), per esprimere la proteina ricombinante fusa con un triplo epitopo eptapeptidico, denominato "Flag", al suo N-terminale. Il costrutto così ottenuto è stato denominato Flag-AF4. Inoltre, nella fase intermedia della strategia di clonaggio, abbiamo ottenuto un costrutto contenente le prime 2 Kb di AF4 che è stato denominato Flag-AF4-1 (**Fig.19**).

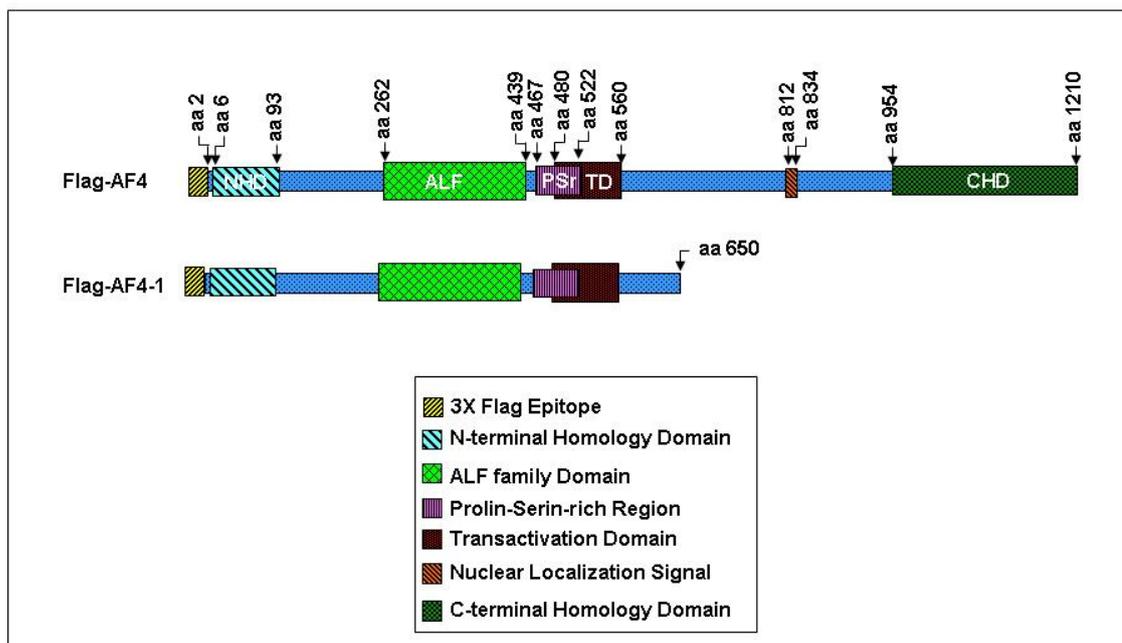


Figura 19. Rappresentazione schematica delle proteine ricombinanti espresse dai costrutti Flag-AF4 e Flag-AF4-1, in cui le estremità N-terminali di AF4 e del frammento AF4-1 sono fuse con quella C-terminale del triplo epitopo eptapeptidico Flag (3xFlag). Come si vede dalla figura, il frammento AF4-1 contiene i domini NHD, AFF, PSr e TD.

Entrambi i costrutti sono stati completamente sequenziati e utilizzati in esperimenti di trasfezione transiente nelle seguenti linee cellulari: Cos-1, HEK293, HeLa, SK-NB-E, SAOS, K562, RS4;11, TOM-1, REH, CEM e 697. Abbiamo saggiato diverse condizioni di trasfezione, diversi tempi, nonché diversi reagenti quali Lipofectamine 2000 (Invitrogen) e Fugene (Roche). Abbiamo inoltre sperimentato diverse strategie per impedire o limitare la possibile degradazione della proteina. Sono pertanto state utilizzate quantità scalari di inibitori del proteasoma nel mezzo di crescita e anche quantità diverse e tipi differenti di inibitori di proteasi nei *buffer* di lisi. Come dimostrato mediante Western Blot con anticorpi Anti-Flag, Flag-AF4 sembra esprimersi solo nelle cellule della linea HEK293 e SAOS (**Fig.20**), ma non in quantità sufficiente da poter realizzare una “immuno-affinity precipitation“ preparativa. Flag-AF4-1 si esprime invece a buoni livelli nelle linee Cos-1, SK-NB-E, HeLa e SAOS, ma soprattutto nelle HEK293 (**Fig.20**).

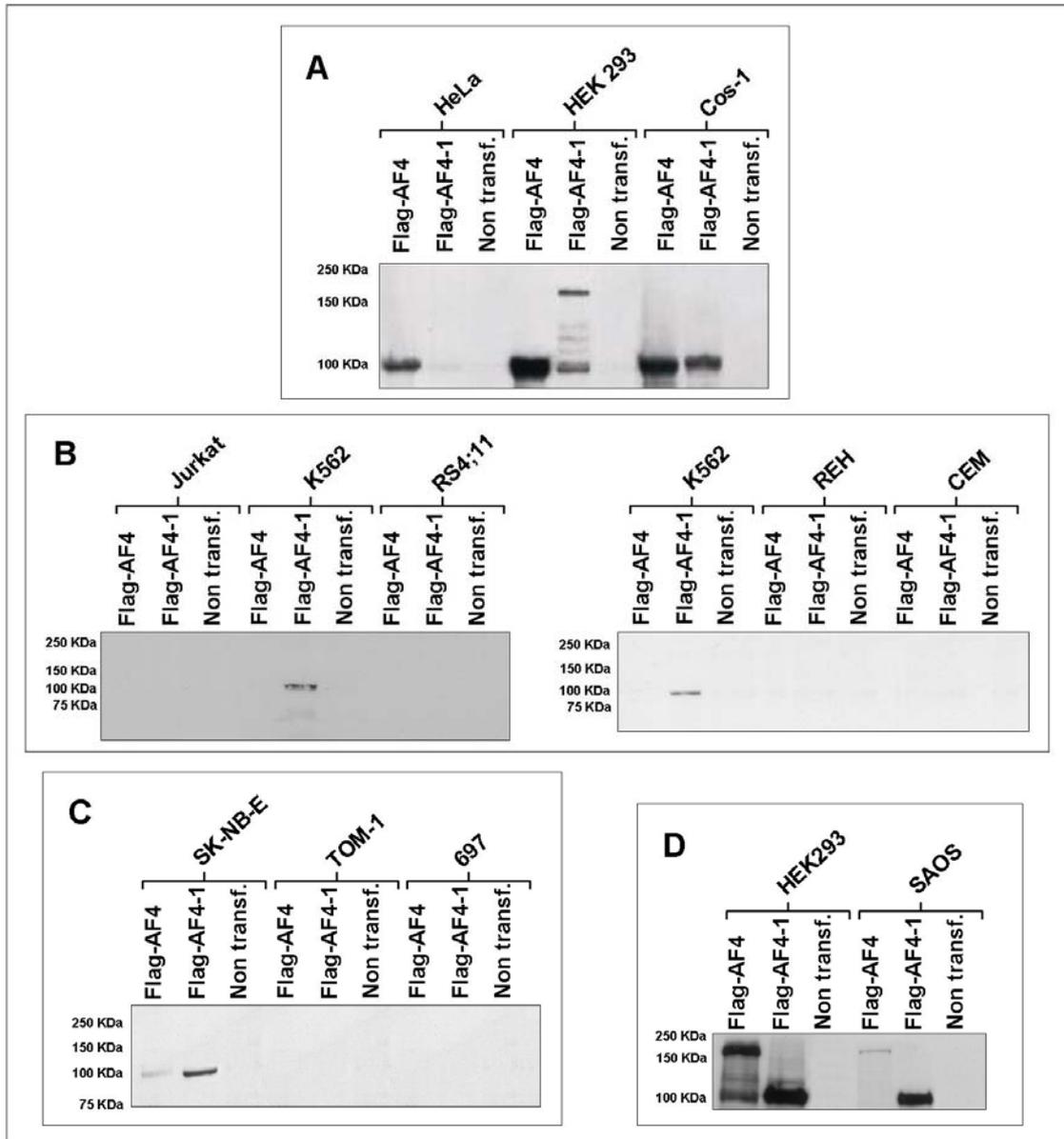


Figura 20. Analisi mediante Western Blot con anticorpi anti-Flag-M2 (Sigma-Aldrich) per la valutazione dell'espressione delle proteine ricombinanti Flag-AF4 e Flag-AF4-1 in linee cellulari di mammifero. Si può notare che Flag-AF4 si esprime solo nelle linee cellulari HEK293 e SAOS. Il costrutto Flag-AF4 si esprime in diverse linee, ma presenta un maggior livello di espressione nelle HEK293. **A)** Espressione di Flag-AF4 e Flag-AF4-1 in cellule HeLa, HEK293 e Cos-1, dopo trasfezione transiente. Le cellule ad una confluenza di circa $1 \times 10^3/\text{mm}^2$ sono state trasfettate con ciascuno dei due costrutti per 48 h con Lipofectamine 2000 (Invitrogen) (secondo il protocollo suggerito dalla ditta). **B)** Espressione di Flag-AF4 e Flag-AF4-1 in cellule K562, RS4;11, Jurkat, REH, e CEM dopo trasfezione transiente. Le cellule ad una concentrazione di circa $1 \times 10^3/\text{ml}$ sono state trasfettate con ciascuno dei due costrutti per 48 h con Fugene (Roche) (secondo il protocollo suggerito dalla ditta). **C)** Espressione di Flag-AF4 e Flag-AF4-1 in cellule SK-N-BE, TOM-1 e 697. mediante trasfezione transiente. Le SK-N-BE ad una confluenza di circa $1 \times 10^3/\text{mm}^2$ sono state trasfettate per 48 h con Lipofectamine 2000 mentre le TOM-1 e le 697 ad

una concentrazione di circa 1×10^3 /ml sono state trasfettate per 48 h con Fugene. **D)** Espressione di Flag-AF4 e Flag-AF4-1 in cellule HeLa, HEK293 e Cos-1, dopo trasfezione transiente. Le cellule ad una confluenza di circa 1×10^3 /mm² sono state trasfettate con ciascuno dei due costrutti per 48 h con Lipofectamine 2000 (Invitrogen) (secondo il protocollo suggerito dalla ditta).

Per le analisi dei punti A, B e C, 40 µg di ciascun estratto proteico sono stati caricati su gel all'8% ed analizzati mediante Western blot.

Per le analisi del punto D, 60 µg di ciascun estratto proteico sono stati caricati su gel all'8% ed analizzati mediante Western blot.

Si può notare che nessuno dei due costrutti si esprime nelle linee cellulari RS4;11, Jurkat, REH, CEM, TOM-1 e 697 e nelle K562 è appena rilevabile solo il segnale relativo a Flag-AF4-1. Tuttavia il fenomeno è in gran parte spiegabile con la scarsa efficienza di trasfezione caratteristica dei linfoblasti e dei mieloblasti.

Abbiamo quindi ottimizzato le condizioni per introdurre con buona efficienza questo costrutto nelle HEK293, utilizzando la metodica di trasfezione del fosfato di calcio (**Fig.21**), più economica rispetto alla lipofectamina, in vista di esperimenti su un maggior numero di cellule.

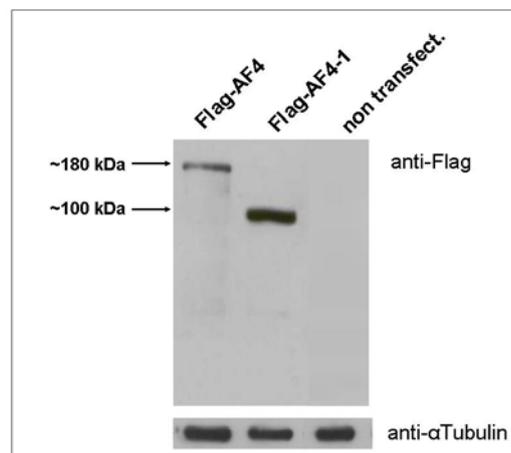


Figura 21. Analisi mediante Western Blot delle proteine ricombinanti Flag-AF4 e Flag-AF4-1 espresse in cellule HEK293 tranfettate in maniera transiente con calcio fosfato. In particolare le cellule ad una confluenza di circa 1×10^3 /mm² sono state trasfettate per 48 h. 40 µg di ciascun estratto proteico è stato caricato su gel al 10% ed analizzato mediante Western blot con anticorpi anti-Flag-M2 (Sigma-Aldrich). La normalizzazione è stata effettuata mediante Western Blot con anticorpi anti-αTubulina. Come si può vedere, a parità di condizioni, la quantità di proteina Flag-AF4-1 è circa 5-6 volte maggiore di quella di Flag-AF4.

3.1.4. Precipitazione per immunoaffinità di un frammento N-terminale di AF4

Dalle cellule HEK293 trasfettate con il costrutto Flag-AF4-1, che esprime la regione N-terminale di AF4 e che contiene il dominio di transattivazione, abbiamo preparato estratti cellulari totali utilizzando bassa concentrazione di detergenti per favorire le interazioni proteiche (vedi Cap.2). Abbiamo isolato da questi estratti le proteine di interesse mediante immunoprecipitazione per affinità con anticorpi anti-FLAG legati a microsfere di agarosio e successiva eluizione dell'immunocomplesso con un *buffer* contenente alte concentrazioni di peptide 3xFLAG (vedi Cap.2).

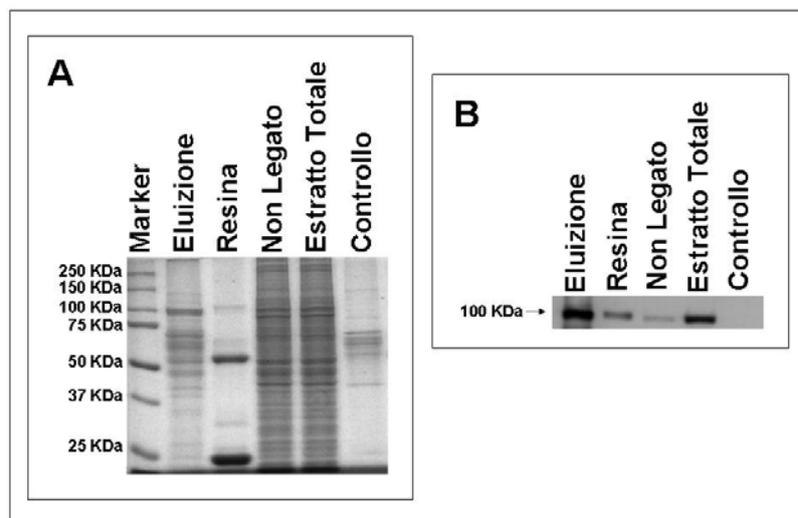


Figura 22. A) Analisi mediante SDS-PAGE colorato al comassie colloidale della proteina ricombinante Flag-AF4-1 espressa in HEK293 e purificata per immunoaffinità con la resina anti-Flag-M2-Agarosio (Sigma-Aldrich). In particolare sono stati utilizzati 10 μ l di resina al 50% e 10 μ l di buffer di eluizione per ogni mg di estratto proteico totale. Sono state analizzate tutte fasi principali del processo di purificazione. Gli eluati del campione e del controllo sono stati precipitati in metanolo/cloroformio, liofilizzati e risospesi nel buffer di caricamento. Sul gel sono stati caricati nel secondo pozzetto e nell'ultimo pozzetto rispettivamente 25 μ g di eluato del campione e una pari quantità in volume del controllo. Negli altri pozzetti sono stati caricati rispettivamente 20 μ l di resina post-eluizione, 50 μ g di estratto post-precipitazione e 50 μ g di estratto totale. B) Analisi mediante Western blot con anti-Flag, della proteina ricombinante Flag-AF4 1 espressa in HEK293 e purificata per immunoaffinità con la resina anti-Flag-M2-Agarosio (Sigma-Aldrich). L'eluato quantizzato spettrofotometricamente è concentrato circa 0,1 μ g/ μ l. Sono state analizzati tutte le fasi principali del processo di purificazione. Sul gel per il western blot sono stati caricati in

ciascun pozzetto rispettivamente 10 μ l di eluato, 10 μ l di resina post-eluzione, 50 μ g di estratto post-precipitazione, 50 μ g di estratto totale e 10 μ l di eluato del controllo derivante da cellule non trasfettate.

Per verificare l'efficienza del processo di purificazione per immuno-affinità abbiamo rivelato la presenza dell'esca nell'eluato mediante Western Blot con anticorpi anti-Flag (Fig.22). Inoltre la quantizzazione spettrofotometrica mediante BioRad Protein Assay ci ha permesso di stimare che la resa del processo era tale da consentirci di eluire una quantità di proteine, sufficiente per le analisi successive.

3.1.5. Identificazione degli interattori proteici del frammento N-terminale di AF4

Le proteine isolate in seguito alla purificazione per immunoaffinità, sono state da noi separate mediante gel 10% SDS-PAGE monodimensionale colorato al Comassie colloidale (Fig.23).

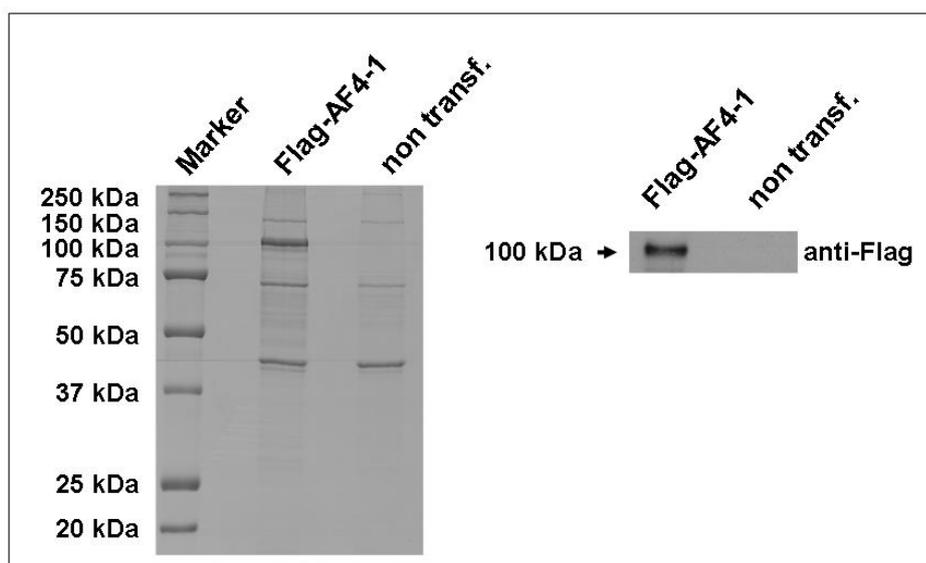


Figura 23. Separazione su SDS-PAGE al 10%, colorato al comassie colloidale, delle proteine isolate in seguito a purificazione per immunoaffinità con anti-Flag-agarosio. In particolare sono state analizzate le proteine precipitate da un estratto cellulare di HEK293 trasfettate con

Flag-AF4-1 e di un estratto da cellule non trasfettate. La proteina Flag-AF4-1 è stata utilizzata come esca per legare ed isolare tutte le proteine cellulari in grado di interagire con il frammento proteico AF4-1. In particolare sono stati caricati su gel 50 µg di eluato del campione ed una quantità corrispondente quantità in volume del controllo. La presenza dell'esca è stata verificata mediante Western blot con anti-Flag.

Come controllo negativo è stato analizzato un immunoprecipitato da HEK293 non trasfettate. Le bande contenenti le diverse proteine sono state escisse dal gel, digerite *in situ* con opportuni enzimi proteolitici (vedi Cap.2) e la miscela risultante è stata analizzata mediante tecniche di spettrometria di massa (vedi Cap.2) nel laboratorio di proteomica del CEINGE Biotecnologie Avanzate s.c.a.r.l.. In particolare, i campioni sono stati analizzati mediante un apparato LC/ESI-MS/MS (Liquid Cromatography / ElectroSpray Ionizzazione – tandem Mass Spectrometry) (44), per determinare i valori di massa dei peptidi ed ottenere informazioni, anche se parziali, sulla sequenza dei frammenti proteolitici (47). L'identificazione delle proteine è stata effettuata mediante programmi bioinformatici specifici presenti in rete (ProFound, Mascot, MS-Fit, etc.) che utilizzano informazioni quali i valori di massa accurati dei peptidi, il tipo di proteasi usata per l'idrolisi, il peso molecolare della proteina o il punto isoelettrico. Per l'identificazione delle proteine "pescate", i valori di massa registrati dagli spettri sono paragonati con quelli provenienti dalla digestione teorica di tutte le proteine presenti nelle banche dati (48). Per selezionare le proteine che interagiscono specificamente con AF4-1, abbiamo sottratto le specie presenti anche nel controllo negativo. Grazie alle tecniche appena descritte abbiamo potuto identificare 60 interattori proteici di AF4 (Tab.1).

Tabella 1. Identificazione degli interattori proteici di AF4-1 mediante LC/ESI-MS/MS

Gene Name	Accession Number	Protein
Transcription regulatory protein:		
PPARBP	gi 2765322	Activator recruited cofactor 205kDa component
MED12	gi 4827042	Mediator of RNA polymerase II transcription, subunit 12 homolog
CRSP2	gi 4580326	Cofactor required for Sp1 transcriptional activation, subunit 2, 150kDa
CRSP3	gi 28558969	Cofactor required for Sp1 transcriptional activation, subunit 3, 130kDa
THRAP4	gi 8699628	Vitamin D receptor-interacting protein complex component DRIP100
CRSP6	gi 28558975	Cofactor required for Sp1 transcriptional activation, subunit 6, 77kDa
CRSP7	gi 28558977	Cofactor required for Sp1 transcriptional activation, subunit 7, 70kDa
PSIP1	gi 19923653	Transcriptional coactivator p52/p75 (PC4 and SFRS1 interacting protein 2)
RBBP4	gi 13111851	Retinoblastoma binding protein 4
RUVBL2	gi 13111851	48 kDa TATA box binding protein interacting protein
EAF1	gi 27370592	ELL associated factor 1
MED4	gi 7141320	p36 TRAP/SMCC/PC2 subunit (Mediator of RNA polymerase II transcription, subunit 4 homolog)
CRSP8	gi 7141322	Cofactor required for Sp1 transcriptional activation, subunit 8, 34kDa
CRSP9	gi 13528909	Cofactor required for Sp1 transcriptional activation, subunit 9
MED6	gi 3329506	RNA polymerase transcriptional regulation mediator
MED8	gi 33988564	Mediator of RNA polymerase II transcription subunit 8 Activator-recruited cofactor 32 kDa component
TRFP	gi 4323033	Trf (TATA binding protein-related factor)-proximal homolog
SURB7	gi 1515377	RNA polymerase II holoenzyme component SRB7
ELL3	gi 10130023	RNA polymerase II elongation factor ELL3
TCEA1	gi 313223	Transcription elongation factor A (SII), 1
DNA-directed RNA polymerase Protein:		
POLR2A	gi 36124	RNA polymerase II largest subunit
POLR2C	gi 2920711	Polymerase (RNA) II (DNA directed) polypeptide C, 33kDa
POLR2E	gi 1060912	Polymerase II, RNA, subunit E
DNA binding protein:		
MCM3	gi 1552242	DNA polymerase alpha holoenzyme-associated protein P1 (p102 protein)
RPA1	gi 17390283	replication protein A1, 70kDa
RUVBL1	gi 15277588	49-kDa TATA box-binding protein-interacting protein
HIST1H2BO	gi 31979	histone H2A.2
RNA binding protein:		
NCL	gi 189306	nucleolin
PRPF31	gi 40254869	PRP31 pre-mRNA processing factor 31 homolog
SERBP1	gi 4929579	Chromodomain helicase DNA binding protein 3 interacting protein (SERPINE1 mRNA binding protein 1)
SSB	gi 34209	La. Protein

SNRPD2	gi 29294624	Small nuclear ribonucleoprotein D2 polypeptide 16.5kDa
SRP14	gi 48735296	Signal recognition particle 14kDa
Serine/threonine kinase protein:		
CDK8	gi 4502745	CDK8 protein kinase
CDK9	gi 12805029	Cyclin-dependent kinase 9 (CDC2-related kinase)
Serine/threonine phosphatase protein:		
PPP2R1A	gi 178663	protein phosphatase 2 (formerly 2A) regulatory subunit A (PR 65) alpha isoform
Ribonucleoprotein:		
HNRPK	gi 460789	heterogeneous nuclear ribonucleoprotein K
TRIM21	gi 338490	Tripartite motif containing protein 21 (52 KDa ribonucleoprotein autoantigen R0/SS-A)
Translation regulator protein:		
EEF1B2	gi 12652911	Eukaryotic translation elongation factor 1 beta 1
Auxiliary transport protein:		
GDI2	gi 285975	Rab GDP dissociation inhibitor beta
Receptor signaling complex scaffold protein:		
YWHAE	gi 12655169	14-3-3 protein, epsilon isoform (Protein kinase C inhibitor protein 1)
SPIN	gi 5730065	SPINL
FGF11	gi 21618445	Fibroblast growth factor 11
YWHAQ	gi 55594676	14-3-3 protein theta (14-3-3 protein T-cell)
AP3S1	gi 4502861	Adapter-related protein complex 3 sigma 1 subunit
Transmembrane receptor protein tyrosine kinase protein:		
FGFR2	gi 29432	fibroblast growth factor (FGR) receptor
Lipid Kinase protein:		
PIP5K2B	gi 1857637	phosphatidylinositol-4-phosphate 5-kinase type II beta
PIP5K2C	gi 21322230	phosphatidylinositol-4-phosphate 5-kinase, type II, gamma
Heterotrimeric G-protein GTPase protein:		
GNB4	gi 12654119	Guanine nucleotide binding protein (G protein), beta polypeptide 4
Guanyl-nucleotide exchange factor protein:		
ARHGEF4	gi 8809845	Rho guanine nucleotide exchange factor (GEF) 4
Calcium binding protein, (Signal transduction)		
S100A8	gi 21614544	S100 calcium-binding protein A8
S100A9	gi 4506773	S100 calcium-binding protein A9
Enzyme: Dehydrogenase		
GAPDH	gi 31645	glyceraldehyde-3-phosphate dehydrogenase
MDH2	gi 12804929	malate dehydrogenase 2, NAD
Enzyme: Synthase		
PTS	gi 4506331	6-pyruvoyltetrahydropterin synthase
Unknown:		
SMAD9	gi 13959539	MAD mothers against decapentaplegic homolog 9
LSM14	gi 16877144	LSM14 homolog A (RNA-associated protein 55)
DRG1	gi 17939479	Developmentally regulated GTP binding protein 1 (Neural precursor cell expressed developmentally down regulated 3)
HCCA2	gi 55249549	HCCA2 protein (MOB2)
C20orf11	gi 21594655	Chromosome 20 open reading frame 11

Tali proteine umane sono state da noi classificate, mediante le informazioni del database Gene Ontology, in base alla localizzazione subcellulare e ai processi biologici in cui sono coinvolte. Da questa classificazione è risultato che il 62% delle proteine identificate è nucleare, ed il 59% è coinvolto nella trascrizione (**Fig.24**).

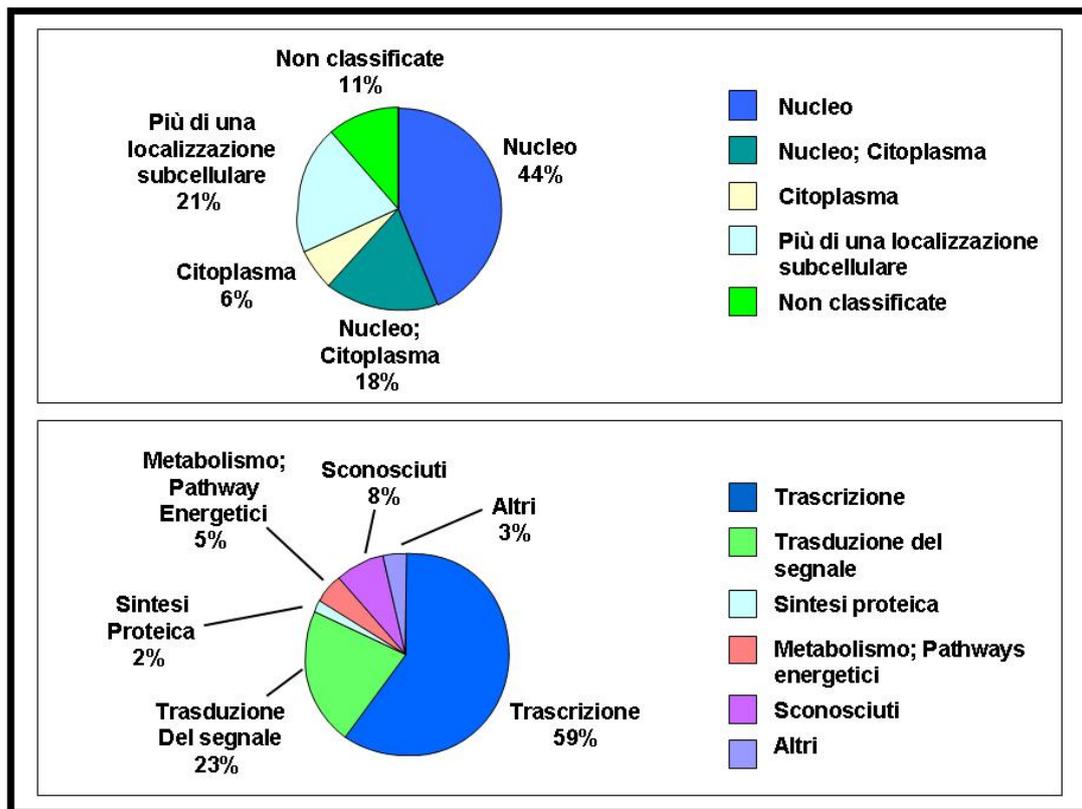


Figura 24. Classificazione grafica degli interattori proteici di AF4-1, identificati con le nostre analisi, effettuata secondo il database *Gene Ontology*, che distingue le varie proteine in base rispettivamente alla localizzazione cellulare e ai processi biologici in cui esse sono coinvolte. Come mostrato in figura il 44% delle proteine identificate è localizzato nel nucleo mentre il 18% sembra presentare una localizzazione mista nucleo/cytosol. Quindi il 62% degli interattori di AF4 identificati in questo lavoro sono proteine nucleari. Inoltre il 59% delle proteine individuate è coinvolto nel processo della trascrizione.

3.1.6. Validazione dei dati di MS mediante co-immunoprecipitazione e Western Blot

Per corroborare l'autenticità dei risultati ottenuti mediante le tecniche proteomiche appena descritte, abbiamo effettuato esperimenti di immunoprecipitazione e Western blot, utilizzando anticorpi specifici diretti contro alcuni dei putativi partner proteici di AF4. In questi esperimenti è stato utilizzato come controllo un ulteriore costrutto denominato Flag-Rep1, costituito dal vettore *N-terminal-CMV-p3X-FLAG 7.1* in cui è inserito il cDNA codificante la proteina umana Rep1. Gli estratti provenienti da HEK293 trasfettate rispettivamente con Flag-AF4, Flag-AF4-1, Flag-Rep1, e da HEK293 non trasfettate, sono stati utilizzati per esperimenti di immunoprecipitazione utilizzando anticorpi anti-Flag-Agarosio. Gli immunoprecipitati sono stati analizzati mediante Western Blot con anticorpi specifici per gli interattori identificati. Gli stessi estratti sono stati parallelamente utilizzati per esperimenti di co-immunoprecipitazione utilizzando gli anticorpi specifici per gli interattori di AF4 e una resina Proteina-A/G-Agarosio. Tali immunoprecipitati sono poi stati analizzati mediante Western Blot con anticorpi Anti-Flag per rivelare la presenza delle proteine ricombinanti dotate dell'epitopo 3xFlag. I dati di MS sono stati confermati in tutti i casi in cui gli anticorpi a nostra disposizione si sono dimostrati efficienti per la rivelazione su Western Blot. In tali casi nelle IP con anti-Flag-Sepharosio le bande rivelate con gli anticorpi specifici erano visibili solo nelle cellule trasfettate con Flag-AF4 e Flag-AF4-1, e non in quelle trasfettate con Flag-Rep1 o nel controllo non trasfettato (**Fig.25**). Nelle IP con anticorpi specifici e Proteina-A/G-Sepharosio le bande rivelate con l'anti-Flag erano visibili solo nelle cellule trasfettate con Flag-AF4 e Flag-AF4-1, e non in quelle trasfettate con Flag-Rep1 o nel controllo non trasfettato (**Fig.25**). Con questi esperimenti abbiamo potuto confermare finora l'interazione specifica tra AF4 e le proteine CRSP3, MED7, Cdk9 e 14-3-3-θ (**Fig.25**).

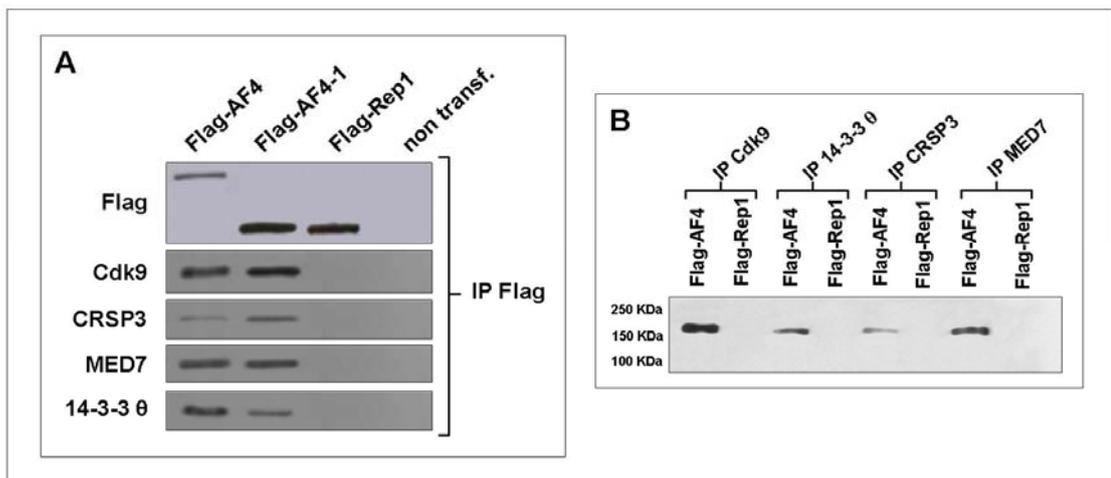


Figura 25. Validazione mediante Western Blot con anticorpi specifici di alcuni interattori di AF4, isolati per immunoprecipitazione di affinità e identificati dall'analisi di spettrometria di massa. Gli esperimenti di immunoprecipitazione sono stati condotti come precedentemente illustrato utilizzando la linea cellulare HEK293. Oltre al controllo non trasfettato è stato utilizzato un ulteriore controllo costituito da cellule in cui è stata espressa, mediante trasfezione transiente, la proteina ricombinante Flag-Rep1, costituita dalla proteina umana Rep-1 fusa al suo estremo amminoterminale con l'epitopo 3xFlag. Tale proteina è coinvolta in processi biologici molto diversi da quelli in cui potrebbe essere coinvolta AF4 ed il costrutto Flag-Rep1 era già disponibile nel laboratorio in cui è stato svolto questo lavoro. A) Analisi mediante Western blot degli eluati ottenuti da esperimenti di immunoprecipitazione effettuati utilizzando 10 μ l di resina anti-Flag-agarosio e 10 μ l di buffer di eluizione per ogni mg di estratto proteico totale. Il Western blot con anticorpi anti-Flag è stato condotto caricando in ciascun pozzetto del gel 10 μ l di eluato. Mentre i Western blot con gli anticorpi specifici per gli interattori di AF4 sono stati condotti caricando 20 μ l di eluato. Come si vede in figura, le proteine identificate per MS si legano oltre che al frammento AF4-1 anche alla proteina AF4 intera, mentre non si legano a Rep1. B) Analisi mediante Western blot degli immunoprecipitati ottenuti utilizzando gli anticorpi specifici per ciascun interattore di AF4 da cellule HEK293 trasfettate con Flag-AF4 e con Flag-Rep1. Sono stati utilizzati 1 μ g di anticorpo specifico e 5 μ l di resina Proteina-A/G-agarosio (Santa Cruz Biotecnologies) per ogni mg di estratto proteico totale. In ogni pozzetto del gel sono stati caricati 20 μ l di immunoprecipitato e il legame con AF4 è stato rivelato mediante Western blot con anticorpi anti-Flag. Come si vede in figura le proteine identificate per MS legano AF4 mentre non legano Rep1.

3.1.7. Analisi *in silico* degli interattori proteici di AF4

Abbiamo eseguito un'analisi proteomica *in silico* delle proteine che interagiscono direttamente o indirettamente con AF4. Utilizzando il database HPRD ed il software DAVID NIH abbiamo potuto effettuare un dettagliato clustering, che ci ha permesso di costruire un network di interazioni tra proteine coinvolte con AF4 nello stesso scenario funzionale (Fig.26). Questa analisi ha confermato ed ha esteso i dati contenuti in un recente report (29) secondo il quale AF4 fa parte del complesso di proteine che costituiscono il macchinario della RNA polimerasi II e che regolano positivamente il processo trascrizionale.

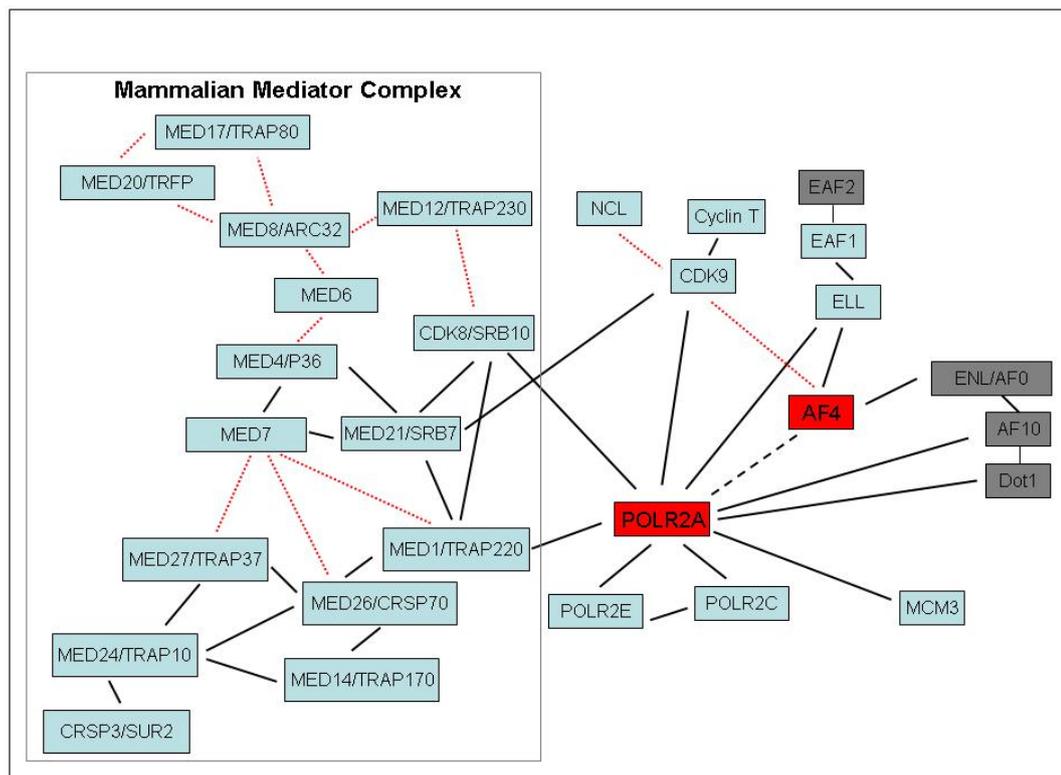


Figura 26. Rappresentazione schematica del *network* di interazioni tra i *partners* molecolari di AF4 coinvolti nello stesso scenario funzionale, la trascrizione mediata dalla RNAPol II. Questa mappa di interazioni è stata realizzata grazie al database HPRD (Human Protein Reference Database) e al software DAVID NIH. Le proteine nelle caselle celesti sono quelle identificate in questo lavoro. Le proteine in grigio sono quelle di cui già si conosceva l'interazione con AF4. POLR2A e AF4 sono in rosso perchè considerate il cuore del *network*. Le linee continue rappresentano le interazioni dirette. Le linee tratteggiate sono invece le interazioni di cui non si è dimostrato sperimentalmente se fossero dirette o mediate.

3.1.8. Espressione e immunoprecipitazione di un frammento C-terminale di AF4

Per completare l'identificazione degli interattori proteici di AF4, abbiamo clonato nel vettore *N-terminal-CMV-p3X-FLAG 7.1* il cDNA corrispondente circa agli ultimi 600 residui della sequenza amminoacidica di AF4 (vedi Cap.2). Abbiamo così ottenuto un costrutto denominato Flag-AF4-2 (**Fig.27**), che è stato da noi utilizzato per esperimenti di trasfezione transiente in cellule HEK293 con la metodica del calcio fosfato.

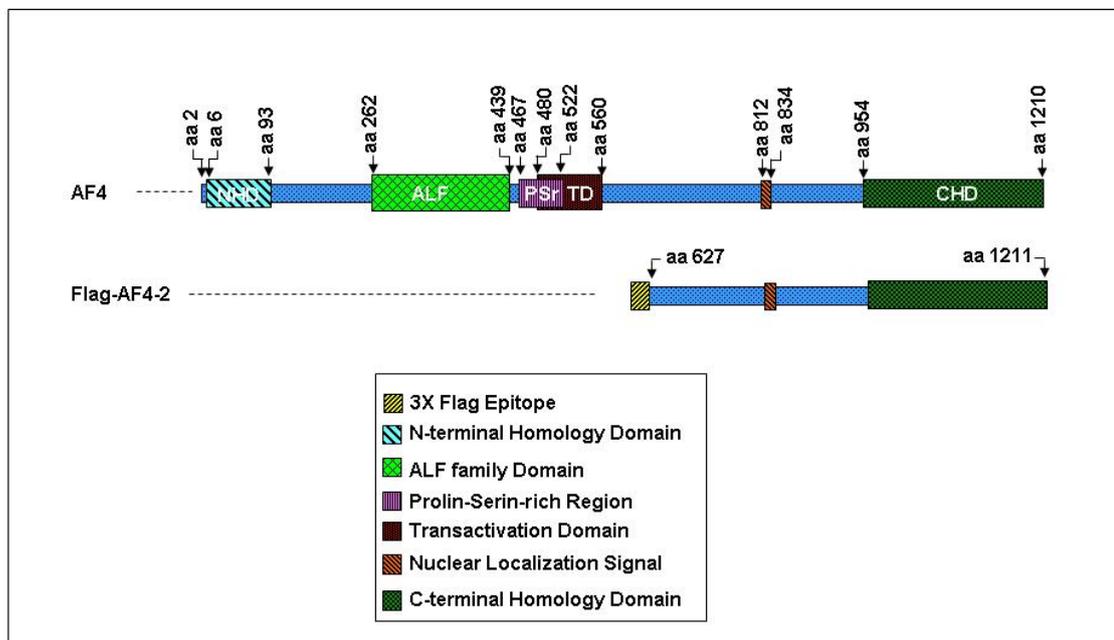


Figura 27. Rappresentazione schematica della proteina ricombinante espressa dal costrutto Flag-AF4-2, in cui l'estremità N-terminale del frammento AF4-2 è fusa con quella C-terminale del triplo epitopo eptapeptidico Flag (3xFlag). Come si vede dalla figura il frammento AF4-2 contiene il dominio CHD e la sequenza di localizzazione nucleare.

Mediante Western Blot con anticorpo anti-Flag, abbiamo potuto osservare che il frammento proteico ricombinante Flag-AF4-2 si esprime bene nella linea cellulare testata (**Fig.28**). Sono quindi stati effettuati esperimenti di immunoprecipitazione con l'anticorpo anti-FLAG-Sepharosio. Le proteine legate sono state quindi eluite con il peptide FLAG. Oltre all'eluato, tutte le fasi intermedie del procedimento di

immunoprecipitazione sono state analizzate mediante Western Blot con anti-Flag (Fig.28). Tale analisi mostrava che apparentemente l'esca Flag-AF4-2 si lega alla resina e viene eluita con buona efficienza. Le proteine eluite dall'affinità sono state poi analizzate mediante gel 10% SDS/PAGE monodimensionale colorato al Comassie colloidale (Fig.28). Tuttavia sul gel colorato non è visibile la banda dell'esca Flag-AF4-2. Anche l'analisi mediante MS non ha permesso di rilevare la presenza dell'esca. E' possibile che, nelle condizioni sperimentali finora utilizzate, la quantità di proteina eluita dalla resina è molto scarsa, benchè visibile per Western Blot.

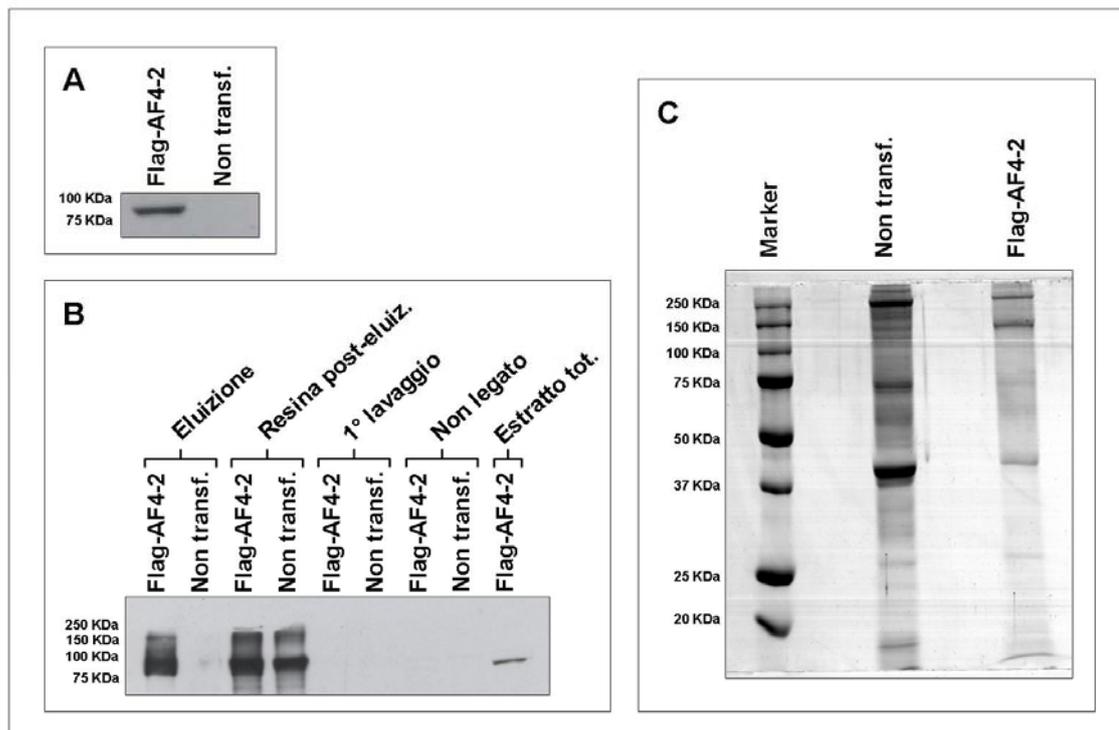


Figura 28. A) Analisi mediante Western blot con anti-Flag della proteina ricombinante Flag-AF4-2 espressa in seguito a trasfezione transiente in cellule HEK293. B) Analisi mediante Western blot con anti-Flag della proteina ricombinante Flag-AF4-2 purificata per immunoaffinità con la resina anti-Flag-M2-Agarosio (Sigma-Aldrich). In particolare sono stati utilizzati 10 μ l di resina al 50% e 10 μ l di buffer di eluizione per ogni mg di estratto totale. Gli eluati del campione e del controllo sono stati precipitati in metanolo/clorofornio, liofilizzati e risospesi nel buffer di caricamento. Sono state analizzate tutte le fasi principali del processo di purificazione. Sul gel per il western sono stati caricati in ciascun pozzetto rispettivamente, 4 μ g di eluato del campione e una pari quantità in volume del controllo non

trasfettato. Sono inoltre stati caricati rispettivamente sia per il campione che per il controllo 20 µl di resina post-eluzione, 40 µl del primo lavaggio e 40 µg di estratto post-precipitazione. Inoltre sono stati caricati 40 µg di estratto totale del campione. C) Separazione mediante SDS-PAGE colorato al comassie colloidale della proteina ricombinante Flag-AF4-2 purificata per immunoaffinità con la resina anti-Flag-M2-Agarosio (Sigma-Aldrich). Aliquote degli stessi eluati, precipitati in metanolo/cloroformio e analizzati mediante *western blot*, sono stati risospesi nel buffer di caricamento. Sul gel colorato al comassie colloidale sono stati caricati 50 µg di eluato del campione e una pari quantità in volume del controllo.

Come si può notare in figura, la *lane* del controllo contiene molte bande, e anche piuttosto marcate, mentre normalmente, con le metodiche messe a punto l'eluato di controllo pur derivando da una pari quantità di estratto totale rispetto al campione è molto poco concentrato e mostra sul gel molte bande. Infatti le interazioni aspecifiche della resina anti-Flag-agarosio sono molto poche. Tuttavia il problema principale che emerge da questo esperimento è che nella *lane* del campione non è visibile la banda dell'esca Flag-AF4-2. Anche l'analisi di MS non ha permesso di rilevare la presenza dell'esca. Evidentemente la proteina ricombinante Flag-AF4-2 per motivi dovuti alla sua struttura terziaria pur legandosi efficacemente alla resina non viene eluita in maniera efficace. Quindi benchè sia visibile analizzando l'eluato mediante *western blot* non è presente in quantità sufficiente a renderla utilizzabile per le analisi di proteomica funzionale.

3.1.9. Espressione di AF4 in cellule HEK293 mediante trasfezione stabile

Per ottenere un costrutto che ci consentisse di esprimere stabilmente e a livelli elevati in cellule eucariotiche la proteina ricombinante Flag-AF4, abbiamo clonato l'intero cDNA codificante AF4 nel vettore *pcDNA3* (Invitrogen) (vedi Cap.2). Tale vettore presenta il vantaggio rispetto ad *N-terminal-CMV-p3X-FLAG 7.1*, di contenere il gene per la resistenza alla Neomicina, utile per selezionare i cloni cellulari che in seguito ad eventi casuali molto rari integrino nel loro genoma il costrutto, consentendone l'espressione in maniera stabile. Tuttavia, il vettore *pcDNA3* non presenta al suo interno la sequenza codificante il triplo epitopo Flag. Ci siamo quindi preoccupati di inserire nel vettore, a monte della sequenza di AF4 ed *in frame* con questa, la breve sequenza codificante il *3xFlag*, avendo cura di inserire, al 5' di questo anche una sequenza *Kozak* per la

traduzione. L'intero costrutto è stato completamente sequenziato ed è stato denominato Flag-AF4S.

Abbiamo realizzato su cellule HEK293 esperimenti di trasfezione con la metodica del Calcio Fosfato selezionando i cloni stabili mediante aggiunta di Neomicina nel mezzo di coltura. Alla fine del processo di selezione, sono stati prelevati e fatti crescere separatamente 15 cloni. Per ciascuno dei cloni l'espressione della proteina ricombinante è stata analizzata mediante Western Blot con anti-Flag (**Fig.29**).

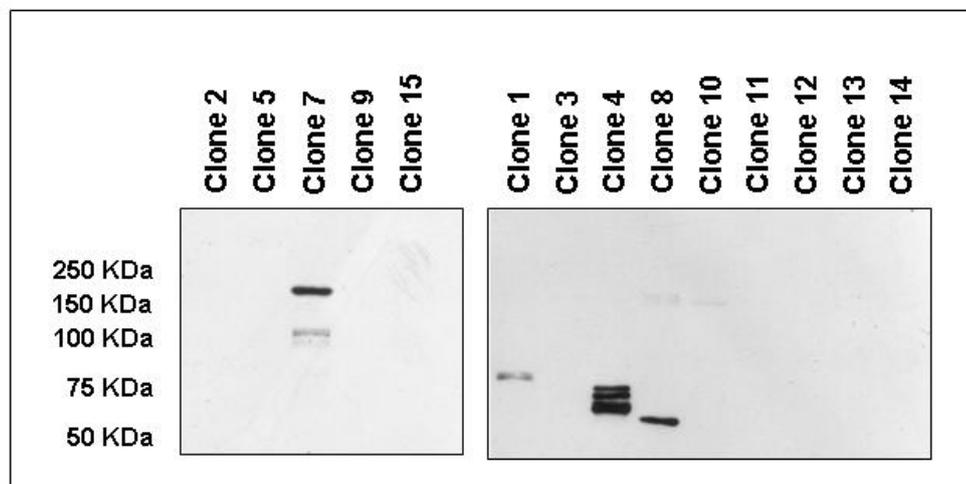


Figura 29. Analisi mediante Western blot con anti-Flag dell'espressione della proteina ricombinante AF4 fusa al suo N-terminale con il 3xFlag, in cloni HEK293 trasfettati stabilmente con il costrutto Flag-AF4S. Tale costrutto è stato ottenuto clonando separatamente AF4 ed il 3xFlag in un vettore contenente il gene per la resistenza alla neomicina. I cloni positivi sono stati selezionati in mezzo contenente neomicina. 15 cloni sono stati isolati e ne è stata analizzata la capacità di esprimere la proteina ricombinante. Come si vede in figura, solo il clone 7 denominato HFA7 esprime la proteina intera. Gli altri cloni hanno acquisito la resistenza alla neomicina e quindi devono aver integrato nel genoma il plasmide. Tuttavia l'evento di integrazione nel genoma della cellula del vettore è casuale sia come posizione sul genoma che come posizione del sito in cui il plasmide si apre generando le estremità ricombinogene. Quindi è possibile che nel caso della trasfezione stabile di un costrutto di grandi dimensioni quale è Flag-AF4S (l'inserto è circa 1300 bp), molti cloni non esprimano la proteina ricombinante o esprimano solo parti di essa.

Solo uno dei cloni esprime in maniera rilevabile la proteina intera, mentre per gli altri il Western Blot non fornisce segnali apprezzabili o mostra bande di peso molecolare inferiore a quello atteso. Questo clone, da noi denominato HFA7 (Hek Flag Af4 clone 7) rappresenta un utile strumento per tentare di realizzare esperimenti di immunoprecipitazione su larga scala utilizzando come esca la proteina ricombinante AF4 *full length* fusa all'epitopo 3xFlag. Infatti il problema della scarsità di proteina Flag-AF4 ricavata mediante trasfezione transiente può essere superato mettendo semplicemente in coltura un grande numero di cellule HFA7, recuperando così grandi quantità di estratto. Inoltre il clone HFA7 costituisce un modello cellulare di overespressione di AF4, e può quindi risultare un utile strumento per futuri esperimenti di proteomica o di trascrittomica differenziale che possano condurre all'individuazione di geni regolati da AF4.

3.2. Analisi di domini funzionali della proteina AF4

3.2.1. Analisi dei residui fosforilati di AF4

Analisi bioinformatiche da noi condotte mostrano che AF4 possiede numerosi putativi siti di fosforilazione di cui alcuni contenuti in probabili sequenze consensus di legame con le 14-3-3. D'altra parte recenti studi (30) hanno dimostrato che Cdk9 fosforila AF4 regolandone negativamente l'attività transattivante. In letteratura sono però noti solo due residui fosforilati di AF4 (49), la Ser 588 e la Thr 766 (**Fig.30**).

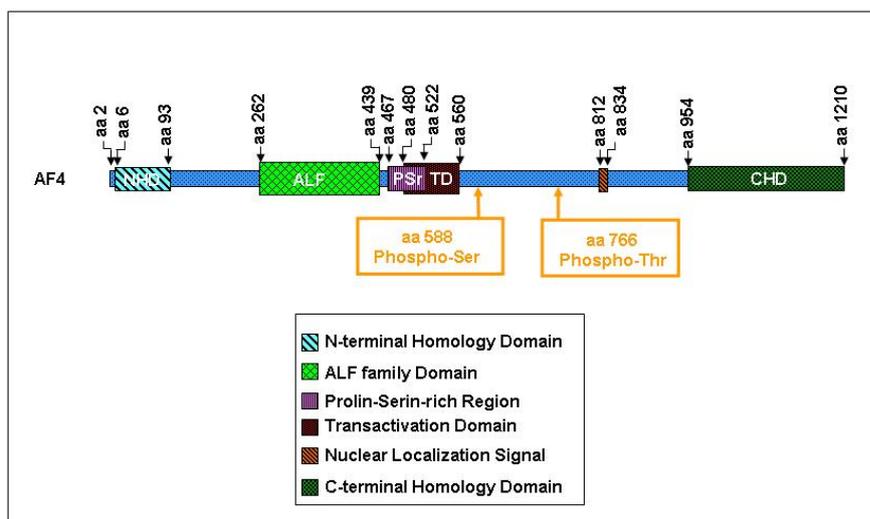


Figura 30. Rappresentazione schematica della proteina umana AF4, in cui sono riportate le posizioni dei due residui fosforilabili noti: la serina 588 e la treonina 766.

Per verificare se tali siti fossero importanti per il legame di AF4 con 14-3-3 θ , abbiamo realizzato mediante mutagenesi sito-specifica costrutti Flag-AF4, mutati nei siti di fosforilazione: Flag-AF4 S588A, Flag-AF4 T766I, Flag-AF4 S588A/T766I. Abbiamo espresso tali costrutti in HEK293 e abbiamo effettuato esperimenti di immunoprecipitazione per indagare sulla possibilità che qualcuno dei mutanti non legasse 14-3-3 θ . Tutti i mutanti immunoprecipitano con 14-3-3 θ in maniera paragonabile alla proteina WT (**Fig.31**).

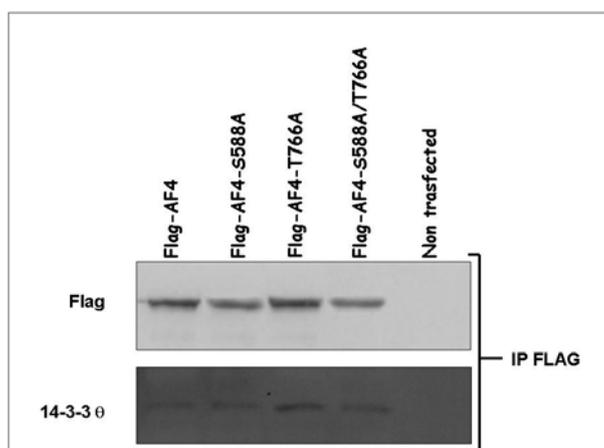


Figura 31. Analisi mediante Western blot dell'interazione tra AF4 e 14-3-3 θ . Esperimenti di immunoprecipitazione con resina anti-Flag-agarosio, sono

stati condotti su estratti di HEK293 in cui sono state espresse mediante trasfezione transiente le proteine ricombinante Flag-AF4 (*wild type*) e le proteine mutanti (nei siti fosforilabili) Flag-AF4 S588A, FlagAF4 T766I e Flag-AF4 S588A/T766I. In ciascun pozzetto sono stati caricati 20 μ l di immunoprecipitato. L'interazione con 14-3-3 è stata rivelata con anticorpi specifici. Come si vede in figura, tutti i mutanti di AF4 legano 14-3-3 θ in maniera paragonabile alla proteina *wild type*.

Contemporaneamente abbiamo indagato sulla possibilità che 14-3-3 θ possa regolare la localizzazione subcellulare di AF4 ed indirettamente il suo turnover. Per verificare se i mutanti presentassero una diversa localizzazione cellulare rispetto a quella nucleare del WT, abbiamo effettuato esperimenti di immunofluorescenza, in diverse linee cellulari, comprese le HEK293 (dati non riportati). Tuttavia abbiamo ottenuto buoni risultati solo nella linea cellulare SAOS (osteosarcoma umano) (**Fig.32**). Queste cellule esprimono Flag-AF4 ad un livello paragonabile a quello delle HEK293, ma aderiscono meglio ai vetrini e hanno caratteristiche morfologiche che ne rendono più facile l'osservazione al microscopio. Per questi esperimenti abbiamo utilizzato come controllo di localizzazione citosolica un costrutto, denominato Flag-AF4-cyt, ottenuto nel plasmide *N-terminal-CMV-p3X-FLAG 7.1*. Tale costrutto esprime fusa all'epitopo 3xFlag la regione aminotermine di AF4 che contiene i primi 480 residui della proteina. Tale costrutto non contiene né il dominio transattivante né la regione contenente i putativi segnali di localizzazione nucleare. Per rivelare nelle cellule le proteine ricombinanti Flag-AF4 (WT e mutanti), abbiamo utilizzato un anticorpo primario anti-Flag ed un secondario anti-mouse marcato con TRITC. Per marcare i nuclei abbiamo invece usato il DAPI. Avendo parallelamente osservato che i mutanti immunoprecipitavano con la 14-3-3 θ , non ci ha sorpreso constatare che anche la localizzazione subcellulare di AF4 mutante era nucleare come quella del WT (**Fig.32**). Resta comunque valida l'ipotesi che l'interazione con 14-3-3 θ possa avere un ruolo nella localizzazione cellulare di AF4 e quindi sulla sua regolazione.

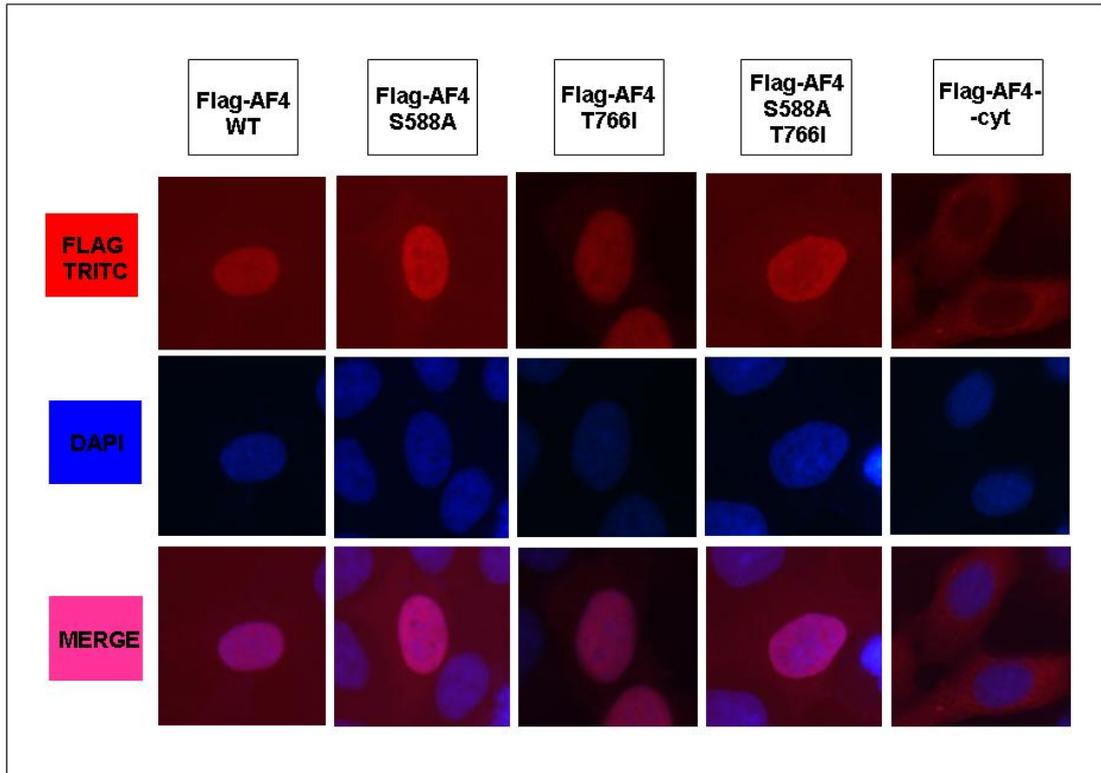


Figura 32. Analisi mediante microscopia a fluorescenza della localizzazione subcellulare della proteina umana AF4 e di tre suoi mutanti ottenuti mediante mutagenesi sito specifica nei siti di fosforilazione. In particolare le analisi sono state condotte su cellule SAOS in cui sono state espresse mediante trasfezione transiente la proteina ricombinante Flag-AF4 (*wild type*), le proteine mutanti (nei siti fosforilabili) Flag-AF4 S588A, FlagAF4 T766I e Flag-AF4 S588A/T766I, e una proteina di controllo, costituita dai primi 480 residui di AF4 fusi al 3xFlag. Tale proteina mancando della regione contenente il segnale di localizzazione nucleare di AF4 e del dominio TD, costituisce un controllo di localizzazione citosolica. Le proteine ricombinanti sono state legate con anticorpi primari anti-Flag e secondari anti-mouse marcati con TRITC, mentre i nuclei sono stati colorati con DAPI. Come si vede in figura la localizzazione delle proteine AF4 mutanti è nucleare come quella del *wild type*.

3.2.2. Individuazione dei domini di AF4 responsabili dell'interazione con Cdk9 e con 14-3-3 θ

Per individuare le regioni di AF4 responsabili dell'interazione con Cdk9 e con 14-3-3 θ , abbiamo diviso la sequenza di AF4-1 in tre parti parzialmente sovrapposte di circa 400 aa (vedi Cap.2) e abbiamo clonato i cDNA corrispondenti ai tre frammenti nel vettore *N-terminal-CMV-p3X-FLAG 7.1*. I tre costrutti sono stati denominati Flag-AF4-1.1, Flag-AF4-1.2, Flag-AF4-1.3 (**Fig.33**).

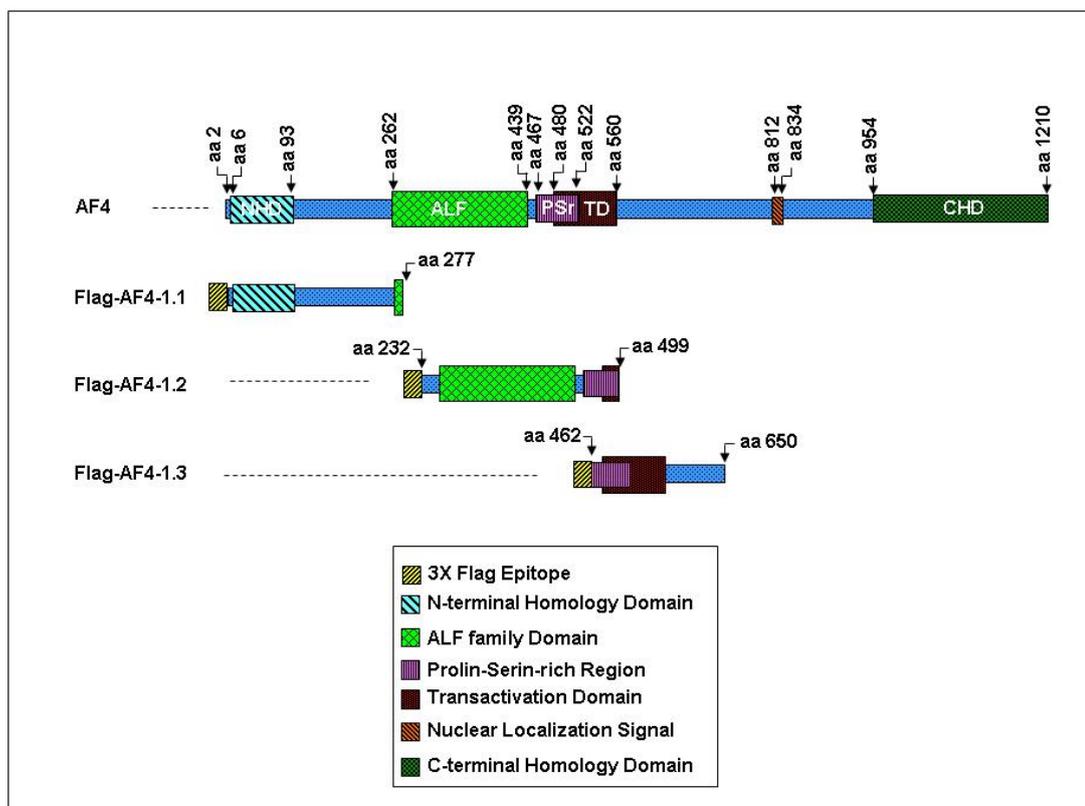


Figura 33. Rappresentazione schematica delle proteina ricombinanti codificate dai costrutti Flag-AF4-1.1, Flag-AF4-1.2 e Flag-AF4-1.3, costituite da frammenti di AF4 fusi con il 3xFlag. I tre peptidi insieme coprono tutta la sequenza del frammento AF4-1 di cui è stato identificato l'interattoma presentato in questa tesi. Come si vede in figura, i frammenti di AF4 sono stati scelti in modo da avere estremità leggermente sovrapposte l'uno con l'altro. Il primo costrutto contiene completamente il dominio NHD, il secondo contiene completamente il dominio AFF, e parte del TD. Il terzo frammento include completamente il dominio TD.

Dopo aver espresso tali costrutti in cellule HEK293, abbiamo effettuato esperimenti di immunoprecipitazione e Western Blot (**Fig.34**) secondo le modalità già descritte in precedenza.

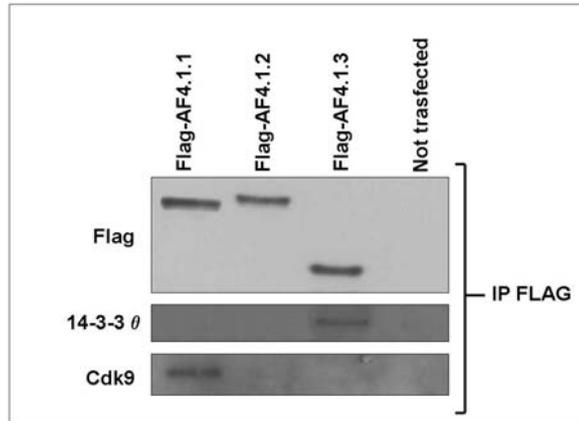


Figura 34. Analisi mediante Western blot degli immunoprecipitati ottenuti con resina anti-Flag-agarosio da estratti di HEK293 in cui sono state espresse, mediante trasfezione transiente, le proteine ricombinanti Flag-AF4-1.1, Flag-AF4-1.2, Flag-AF4-1.3. Questa analisi è stata effettuata per identificare le regioni minime di AF4 in grado di interagire con le proteine Cdk9 e 14-3-3 θ . Sono stati caricati 20 μ l di immunoprecipitato per ogni campione che sono stati analizzati con anti-Flag, anti-14-3-3 θ e anti-Cdk9. Come si vede in figura, solo il frammento AF4-1.3 (~210 aa) lega 14-3-3 θ , mentre Cdk9 interagisce solo con il frammento AF4-1.1 (~300 aa).

Abbiamo così osservato che solo Flag-AF4-1.1 lega Cdk9 e che solo Flag-AF4-1.3 lega 14-3-3 θ . Questi risultati rivestono particolare importanza, perchè ci consentono di affermare che la proteina chimerica oncogenica MLL-AF4 potrebbe non legare Cdk9, mentre potrebbe legare 14-3-3 θ . Inoltre, aver minimizzato la regione candidata per l'interazione con 14-3-3 θ ci permette di restringere a questa zona della proteina la ricerca di siti fosforilabili e di sequenze consensus per il legame alle 14-3-3 (**Fig.35**)

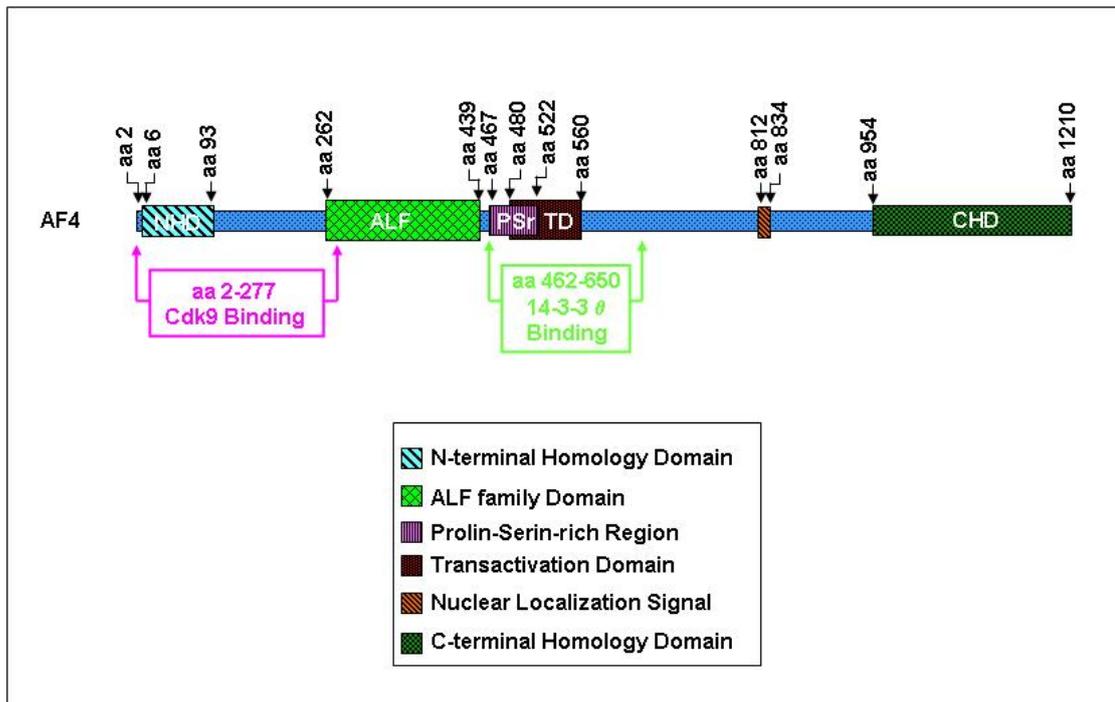


Figura 35. Rappresentazione schematica della proteina umana AF4, in cui sono riportate le posizioni e l'estensione delle regioni responsabili dell'interazione rispettivamente con Cdk9 e con 14-3-3 θ .

3.2.3. Definizione dei segnali di localizzazione nucleare di AF4

Benchè già da tempo siano stati mappati sperimentalmente i segnali di localizzazione nucleare dell'ortologo murino di AF4 (16), tali segnali sono finora stati individuati sulla sequenza di AF4 umana solo mediante predizioni bioinformatiche basate sull'omologia con i segnali murini o mediante ricerca delle *consensus* con matrici dei pesi. Abbiamo pertanto effettuato un'analisi più approfondita della regione di AF4 contenente le putative sequenze di indirizzamento al nucleo. Innanzitutto abbiamo effettuato una predizione dei putativi segnali con il Software online PSORT II (**Fig.36**). Abbiamo così

potuto individuare quattro sequenze predette di “nuclear targeting” in una regione che si estende dal residuo 665 al residuo 919 di AF4 (**Fig.36**). Abbiamo poi clonato tale regione nel vettore *pEGFP-C1* (Clontech) (vedi Cap.2) che consente l’espressione in sistemi eucariotici del frammento proteico codificato dall’inserito in fusione al suo N-terminale con la “Enhanced Green Fluorescent Protein” (EGFP). Tale costrutto è stato denominato GFP-NLS ed è stato utilizzato per esperimenti di trasfezione transiente utilizzando la Lipofectamine 2000 (Invitrogen). Abbiamo espresso il peptide codificato da tale costrutto in cellule HeLa, che sono tra le più adatte ad esperimenti di microscopia a fluorescenza per le loro caratteristiche morfologiche e perchè aderiscono bene ai vetrini. Come controllo negativo sono state utilizzate cellule trasfettate con il vettore vuoto, che esprimono quindi la sola EGFP, priva di segnali di indirizzamento. Dopo il fissaggio, le cellule sono state trattate con il colorante nucleare DAPI. L’osservazione dei vetrini al microscopio a fluorescenza (**Fig.37**) ha mostrato che mentre nelle cellule trasfettate con il vettore vuoto la fluorescenza della EGFP era diffusa in tutto il volume cellulare, nelle cellule che esprimevano GFP-NLS la fluorescenza nel verde era visibile solo nel nucleo ed era perfettamente sovrapponibile con quella del DAPI. Abbiamo suddiviso la sequenza della regione contenente gli NLS (Nuclear Localization Signals) in tre parti (vedi Cap.2) di cui le prime due contenessero rispettivamente il primo ed il secondo NLS e la terza contenesse gli ultimi due segnali che sono tra loro molto vicini. Abbiamo clonato i corrispondenti cDNA nel vettore *pEGFP-C1* e i costrutti così ottenuti sono stati denominati GFP-NLS1, GFP-NLS2 e GFP-NLS3 (**Fig.36**).

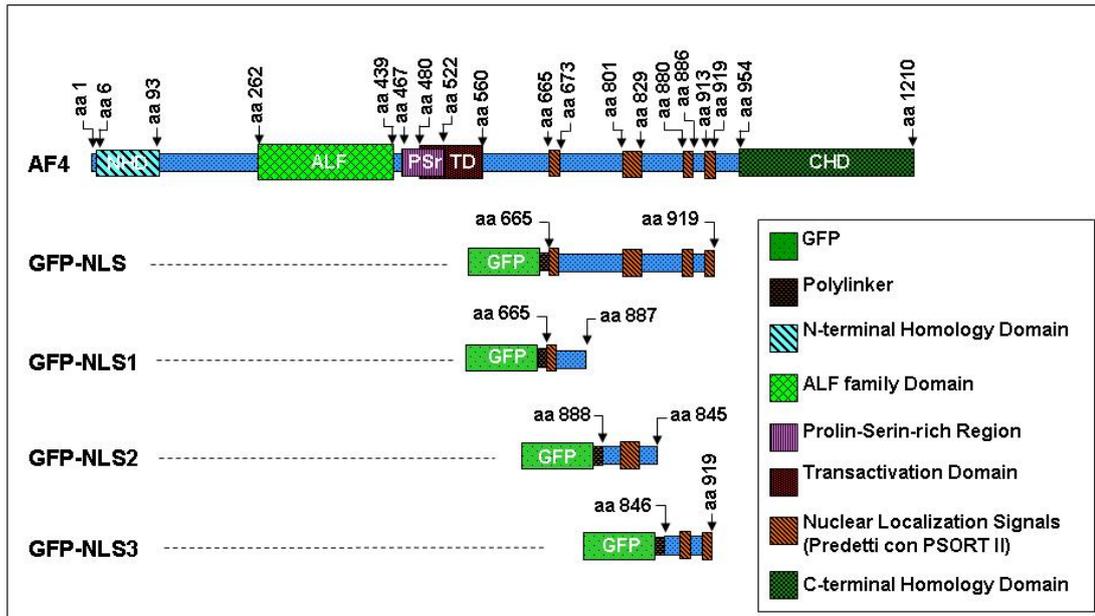


Figura 36. Rappresentazione schematica delle proteine ricombinanti espresse dai costrutti GFP-NLS, GFP-NLS1, GFP-NLS2, GFP-NLS2. Il costrutto GFP-NLS contiene la sequenza codificante la regione di AF4 umana che comprende i putativi segnali di localizzazione nucleare (NLS) predetti bioinformaticamente. I costrutti GFP-NLS1, GFP-NLS2, GFP-NLS2 sono stati ottenuti suddividendo in tre frammenti la regione contenente i vari NLS, in maniera tale che il primo frammento contenga il primo NLS, che il secondo frammento contenga il secondo NLS e che il terzo frammento contenga il terzo ed il quarto NLS (tra loro molto vicini). Tali costrutti sono in grado di esprimere i frammento di AF4 fusi all' N-terminale con la proteina GFP.

I peptidi codificati da tali costrutti sono stati espressi nelle HeLa e le cellule osservate al microscopio (**Fig.37**) con le modalità già descritte. La proteina GFP-NLS1 mostra la stessa localizzazione diffusa della EGFP espressa dal vettore vuoto. Invece GFP-NLS2 e GFP-NLS3 sono localizzate solo nel nucleo. Grazie a tali esperimenti abbiamo quindi potuto stabilire quali sono le sequenze importanti per la localizzazione nucleare della proteina umana AF4. Tali regioni si trovano in una regione della proteina che è presente in tutte le chimere conosciute MLL-AF4 e hanno pertanto un ruolo fondamentale nell'indirizzamento al nucleo di tali proteine chimeriche e quindi nel meccanismo patogenico della leucemia.

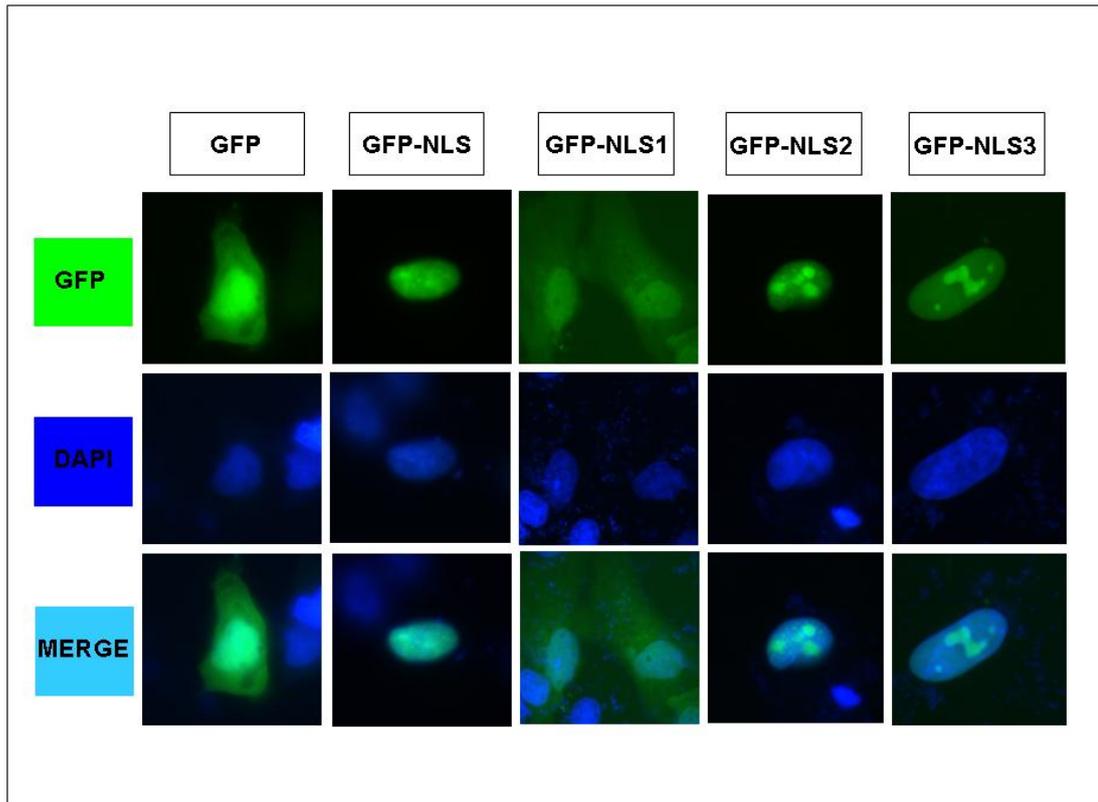


Figura 37. Analisi mediante microscopia a fluorescenza della localizzazione subcellulare delle proteine ricombinanti GFP-NLS, GFP-NLS1, GFP-NLS2, GFP-NLS2. In particolare le analisi sono state condotte su cellule HeLa in cui sono state espresse in seguito a trasfezione transiente le proteine ricombinanti. Tali analisi sono state condotte allo scopo di verificare l'efficienza dei segnali di localizzazione nucleare di AF4 umana predetti bioinformaticamente. Come controllo sono state utilizzate cellule in cui è stata espressa la GFP da sola. Come si vede in figura il primo dei segnali predetti non sembra mediare in maniera efficace la localizzazione della GFP nel nucleo.

4. DISCUSSIONE

Il lavoro presentato in questa tesi rappresenta il primo solido approccio di proteomica funzionale per l'identificazione delle proteine che interagiscono con la proteina umana AF4 coinvolta nella leucemia linfoblastica acuta t(4;11). Le nostre analisi ci hanno permesso di individuare 60 nuovi interattori di AF4. Più della metà di queste proteine sono nucleari e hanno un ruolo nella trascrizione mediata dall'RNA polimerasi II. In particolare, tra i fattori da noi trovati in complesso con AF4 sono presenti proprio tre subunità della RNAPol II: POLR2A, POLR2C e POLR2E (50). Inoltre le caratteristiche di molte delle proteine identificate ci consentono non solo di collocare AF4 tra i fattori che formano il sofisticato macchinario trascrizionale ma di attribuirgli una funzione di attivatore trascrizionale.

Uno dei primi elementi che ci spingono a tale considerazione è l'aver confermato nel corso del nostro studio l'interazione con due suoi partner proteici già noti: ELL ed EAF1. La prima di queste due proteine è un fattore positivo di elongazione trascrizionale (34,35), che sembrerebbe essere in grado di sopprimere le pause transitorie dell'RNA polimerasi lungo il DNA, aumentando così la velocità del processo trascrizionale durante la fase di elongazione dei trascritti. La cosa che rende ancor più interessante l'aver individuato questa proteina tra gli interattori di AF4 è il fatto che ELL si trova frequentemente fusa ad MLL nelle leucemie infantili (35,36). Per quanto riguarda EAF1 (ELL Associated Factor 1), è riportato che tale fattore attiva la trascrizione mediante interazione con il dominio carbossiterminale di ELL (36) ed è stato provato che tale dominio è essenziale per la leucemiogenesi indotta da MLL-ELL (51,52). Questi dati sembrano voler suggerire che la perturbazione del *pathway* di regolazione trascrizionale in cui AF4 è coinvolto riveste un ruolo primario nel meccanismo leucemiogenico.

Un'altro dato a favore dell'attività transattivante di AF4 è costituito dal fatto che dei 60 interattori di AF4 da noi individuati, ben 15 sono componenti del *Mammalian Mediator*

Complex, un complesso multiproteico composto da più di 25 proteine (53-55). Tale macrocomplesso noto anche come *CRSP* (Cofactor Required for SP1 activation) (56), costituisce un sistema modulare in grado di trasdurre al macchinario basale dell'RNA pol II segnali provenienti da attivatori e repressori trascrizionali. Infatti il *Mediator* non è in grado di legare specificamente il DNA ma interagisce fisicamente con l'RNA Polimerasi II e funziona da interfaccia tra fattori trascrizionali specifici ed il macchinario basale della trascrizione(53,54). La maggior parte delle proteine umane che compongono *CRSP* sono omologhe di proteine presenti in molte altre specie (dal lievito al topo). Ma la cosa forse più interessante è la grande plasticità del complesso *CRSP*, che è in grado di assumere configurazioni differenti a seconda del tipo di regolazione trascrizionale che deve essere messa in atto. Infatti le proteine che costituiscono il *Mammalian Mediator Complex* non prendono parte tutte insieme al complesso, ma sono organizzate in subunità che si assemblano in modo differente in risposta a diversi segnali, in modo da modulare finemente e specificamente i livelli di trascrizione(53,54). Alcune delle configurazioni che il complesso può assumere sono in grado di reprimere la trascrizione, mentre altre sono in grado di attivarla. Il dato forse più rilevante è che tutte le subunità di *CRSP* da noi identificate tra gli interattori di AF4 sono costituenti di configurazioni transattivanti del complesso.

Un'ulteriore prova del coinvolgimento di AF4 in un'importante via di attivazione della trascrizione è costituita dall'interazione da noi individuata e validata tra AF4 e Cdk9. Quest'ultima proteina forma, insieme alla ciclina T1 un complesso noto come p-TEFb, (57) che è un fattore positivo di regolazione della trascrizione mediata dalla RNAPol II. In particolare p-TEFb promuove la fase di elongazione del processo trascrizionale, fosforilando il dominio carbossiterminale della subunità maggiore della RNAPol II e il fattore negativo di elongazione DSIF. L'interazione tra AF4 e Cdk9 è stata riportata anche in un recente articolo (30), in cui gli autori tra l'altro dimostrano che il legame

con AF4 promuove l'attività chinasi di Cdk9 sul CTD (dominio carbossiterminale) della RNAPol, attivando così ulteriormente la trascrizione.

A questo punto il modello più plausibile per spiegare il meccanismo patogenico della leucemia linfoblastica acuta t(4;11) sembrerebbe quello dell'acquisizione di funzione da parte dell'ibrido MLL-AF4. Tale proteina chimerica potrebbe essere in grado di legarsi ai promotori dei geni regolati da MLL e di promuoverne la trascrizione secondo le modalità dipendenti dalla via di attivazione trascrizionale in cui è coinvolto AF4.

Era importante a questo punto determinare con maggior precisione quale fosse la regione di AF4 responsabile dell'interazione con Cdk9. Infatti se si fosse dimostrato che tale regione era conservata nell'ibrido MLL-AF4, si poteva ipotizzare per la leucemioinesi un meccanismo in cui MLL-AF4 mediante il legame con Cdk9 sarebbe stato in grado autoattivarsi in maniera costitutiva promuovendo la trascrizione dei geni bersaglio di MLL a livelli più elevati del normale. Tuttavia i nostri risultati dimostrano che i primi 300 residui dell'estremità amminotermine di AF4 che mancano nella chimera sono responsabili del legame con Cdk9, escludendo così che la proteina chimerica possa interagire con pTEF-b almeno mediante la porzione di AF4.

Ovviamente dopo aver messo in luce il coinvolgimento di AF4 in un'importante via di attivazione della trascrizione, viene da chiedersi in che maniera l'attività transattivante di questa proteina venga regolata. Una prima risposta in tal senso la si può trovare da alcuni dati recentemente pubblicati (30) che mostrano che Cdk9 oltre a fosforilare il dominio carbossiterminale dell'RNAPol II fosforila anche AF4, regolandone negativamente l'attività di attivazione trascrizionale. Tuttavia viene dimostrato che la fosforilazione di AF4 non ne influenza la stabilità. Il fatto che AF4 sia una proteina soggetta a fosforilazione emerge anche da un altro recente lavoro (49) di proteomica funzionale in cui vengono identificati sperimentalmente due siti fosforilabili di AF4, a

tutt'oggi individuati: la Ser588 e la Thr766. Questi ultimi dati hanno portato la nostra attenzione su altri due degli interattori da noi identificati: 14-3-3 ϵ e 14-3-3 θ . Quella delle 14-3-3 è una famiglia di proteine coinvolte in una grande varietà di processi biologici, inclusa l'oncogenesi (58). Le 14-3-3 legano proteine contenenti residui di serina e treonina fosforilati e ne modulano l'attività o la localizzazione subcellulare (58). Ci siamo chiesti se il legame con le proteine 14-3-3 possa influenzare la localizzazione subcellulare di AF4, che è normalmente nucleare. La localizzazione cellulare di AF4 potrebbe infatti essere un punto cruciale per il controllo del suo turnover e quindi per la regolazione della sua attività. Infatti la degradazione di AF4 è proteasoma dipendente ed è mediata dalla sua interazione con le proteine citosoliche SIAH1 e SIAH2. Inoltre la proteina chimerica MLL-AF4 sfugge a tale controllo, in quanto non contiene la regione di interazione con le proteine SIAH. Analisi bioinformatiche hanno mostrato che AF4 presenta diverse putative sequenze consensus per il legame alle 14-3-3 θ , di cui una contenente il primo dei due siti di fosforilati (Ser 588). Inoltre abbiamo mappato la regione di AF4 responsabile dell'interazione osservando che comprende la Ser 588 ed è contenuta nell'ibrido MLL-AF4.

Quindi se l'interazione con 14-3-3 θ gioca un ruolo nella regolazione di AF4, anche MLL-AF4 potrebbe rispondere a tale meccanismo regolativo. Comunque, mutando i due siti di fosforilazione (Ser 588 e Thr 766), la localizzazione di AF4 non varia così come non varia la capacità di legare 14-3-3 θ . Questo potrebbe dipendere dal fatto che i siti mutagenizzati non sono quelli responsabili del legame a 14-3-3 θ . Infatti, AF4 contiene diversi altri putativi siti fosforilabili in probabili consensus per le 14-3-3. Il fatto di aver ristretto la zona di interazione sulla sequenza di AF4 ci consentirà di concentrare le ricerche solo in questa regione.

Avendo focalizzato la nostra attenzione anche sulla regolazione della localizzazione subcellulare di AF4, abbiamo analizzato le regioni di AF4 responsabili

dell'indirizzamento al nucleo. I dati in letteratura in merito ai segnali di *nuclear targeting* di AF4 umana non erano molto chiari. Diversi autori hanno riportato sulla sequenza di AF4 umana uno o più putativi segnali di localizzazione (15,21,28,59,60), tuttavia l'identificazione delle sequenze di localizzazione nucleare è stata effettuata sperimentalmente solo su AF4 murina (16). Abbiamo pertanto individuato funzionalmente i segnali di localizzazione nucleare di AF4 umana mediante un approccio combinato di bioinformatica e di microscopia a fluorescenza. La regolazione della localizzazione subcellulare di AF4 e quindi l'influenza delle regioni di AF4 conservate nella chimera sulla localizzazione di MLL-AF4, ci sembra di primaria importanza. Infatti l'indirizzamento al nucleo di MLL-AF4 potrebbe dipendere dalle regioni di AF4 quanto da quelle di MLL. Se la localizzazione nucleare della chimera dovesse risultare indispensabile per il processo neoplastico e per la sopravvivenza delle cellule leucemiche, si potrebbe indagare sull'effetto del mancato indirizzamento al nucleo di AF4 nei tessuti emopoietici. Tale effetto, soprattutto se di durata limitata, potrebbe anche non essere disastroso per le cellule non leucemiche in quanto che il modello murino di knockout genico di AF4 è vitale. Quindi si potrebbero ipotizzare approcci terapeutici mirati sulle regioni di AF4 e della chimera MLL-AF4, responsabili della localizzazione nucleare.

In conclusione i risultati presentati in questa tesi contribuiscono all'avanzamento sulla comprensione delle funzioni della proteina umana AF4 e forniscono alcuni indizi sul meccanismo patogenico della leucemia linfoblastica acuta t(4;11). Tali risultati potranno costituire la base per successivi futuri esperimenti tesi alla ricerca di nuovi *target* terapeutici.

5. BIBLIOGRAFIA

1. Widmann. *Interpretazione clinica degli esami di laboratorio*. McGraw Hill, 172-177.
2. Look AT. *Oncogenic transcription factors in the human acute leukemias*. Science, 1997; 278: 1059-1064. Review.
3. Slany RK. *When epigenetics kills: MLL fusion proteins in leukemia*. Hematol Oncol. 2005 Mar;23(1):1-9. Review.
4. Z-Y Li, D-P Liu and C-C Liang. *New insight into the molecular mechanisms of MLL-associated leukemia*. Leukemia (2005) 19, 183–190
5. Daser A, Rabbitts TH. *The versatile mixed lineage leukaemia gene MLL and its many associations in leukaemogenesis*. Semin Cancer Biol. 2005 Jun;15(3):175-88. Review.
6. Prasad R, Yano T, Sorio C, Nakamura T, Rallapalli R, Gu Y, Leshkowitz D, Croce CM, Canaani E. *Domains with transcriptional regulatory activity within the ALL1 and AF4 proteins involved in acute leukemia*. Proc Natl Acad Sci U S A. 1995 Dec 19; 92(26):12160-4.
7. Eklund EA. *The role of HOX genes in malignant myeloid disease*. Curr Opin Hematol. 2007 Mar;14(2):85-9. Review.
8. Felix CA, Kim CS, Megonigal MD, Slater DJ, Jones DH, Spinner NB, Stump T, Hosler MR, Nowell PC, Lange BJ, Rappaport EF. *Panhandle polymerase chain reaction amplifies MLL genomic translocation breakpoint involving unknown partner gene*. Blood. 1997 Dec 15; 90(12): 4679-86.
9. Harrison CJ, Griffiths M, Moorman F, Schnittger S, Cayuela JM, Shurtleff S, Gottardi E, Mitterbauer G, Colomer D, Delabesse E, Casteras V. *A multicenter evaluation of comprehensive analysis of MLL translocations and fusion gene partners in acute leukemia using the MLL FusionChip device*. Cancer Genet Cytogenet. 2007 Feb; 173(1):17-22.
10. Pui CH and Campana D. *Age-related differences in leukemia biology and prognosis: the paradigm of MLL-AF4-positive acute lymphoblastic leukemia*. Leukemia. 2007 Apr;21(4):593-4.
11. Aplan PD. *Chromosomal translocations involving the MLL gene: molecular mechanisms*. DNA Repair (Amst). 2006 Sep 8;5(9-10):1265-72. Review.
12. Huret JL, Dessen P, Bernheim A. *An atlas of chromosomes in hematological malignancies. Example: 11q23 and MLL partners*. Leukemia. 2001 Jun; 15(6):987-9.
13. Zeisig, D., T., Bittner, C., B., Zeisig, B., B., Garcia-Cuellar, M., P., Hess, J., L., Slany, R., K.. *The eleven-nineteen-leukemia protein ENL connects nuclear MLL fusion partners with chromatin*. Oncogene. (2005) 24, 5525-32.

14. Hess, JL. *MLL: a histone methyltransferase disrupted in leukemia*. Trends Mol. Med. (2004) 10, 500-7. Review.
15. Morrissey, J., Tkachuk, D., C., Milatovich, A., Francke, U., Link, M., Cleary, M., L.. *A serine/proline-rich protein is fused to HRX in t(4;11) acute leukemias*. Blood. (1993) 81, 1124-31.
16. Erfurth F, Hemenway CS, de Erkenez AC, Domer PH. *MLL fusion partners AF4 and AF9 interact at subnuclear foci*. Leukemia. 2004 Jan; 18(1):92-102.
17. Zeleznic-Le NJ, Harden A.M, Rowley JD. *11q23 translocations split the "AT-hook" cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the mixed-lineage leukemia (MLL) gene*. Proc. Nadl. Acad. Sci. USA, 1994; 91: 10610-10614
18. Lavau C, Szilvassy SJ, Slany R, Cleary ML. *Immortalization and leukemic transformation of a myelomonocytic precursor by retrovirally transduced XRX-ENL*. EMBO J, 1997; 16: 4226-4237
19. Raimondi SC, Frestedt JL, Pui CH, Downing JR, Head DR, Kersey JH, Behm FG. *Acute lymphoblastic leukemias with deletion of 11q23 or a novel inversion (11)(p13q23) lack MLL gene rearrangements and have favorable clinical features*. Blood. 1995; 86(5):1881-6.
20. Nilson, I., Reichel, M., Ennas, M., G., Greim, R., Knorr, C., Siegler, G., Greil, J., Fey, G., H., Marschalek, R.. *Exon/intron structure of the human AF-4 gene, a member of the AF-4/LAF-4/FMR-2 gene family coding for a nuclear protein with structural alterations in acute leukaemia*. Br. J. Haematol. (1997) 98,157-69.
21. Li Q, Frestedt JL, Kersey JH. *AF4 encodes a ubiquitous protein that in both native and MLL-AF4 fusion types localizes to subnuclear compartments*. Blood, 1998; 92: 3841-3847.
22. Chen, C., S., Hilden, J., M., Frestedt, J., Domer, P., H., Moore, R., Korsmeyer, S., J., Kersey, J., H.. *The chromosome 4q21 gene (AF-4/FEL) is widely expressed in normal tissues and shows breakpoint diversity in t(4;11)(q21;q23) acute leukemia*. Blood (1993) 82. 1080-5.
23. Frestedt, J., L., Hilden, J., M., Moore, R.,O., Kersey, J., H.,. *Differential expression of AF4/FEL mRNA in human tissues*. Genet Anal. (1996) 12,147-9.
24. Isnard, P., Core, N., Naquet, P., Djabali, M.. *Altered lymphoid development in mice deficient for the mAF4 proto-oncogene*. (2000) Blood. 96, 705-10.
25. Gecz, J., Gedeon, A., K., Sutherland, G., R., Mulley, J., C.. *Identification of the gene FMR2, associated with FRAXE mental retardation*. Nat. Genet. (1996) 13, 105-8.

26. Gu, Y., Shen, Y., Gibbs, R., A., Nelson, D., L.. *Identification of FMR2, a novel gene associated with the FRAXE CCG repeat and CpG island*. Nat. Genet. (1996) 13, 109-13.
27. Ma, C., Staudt, L., M.. *LAF-4 encodes a lymphoid nuclear protein with transactivation potential that is homologous to AF-4, the gene fused to MLL in t(4;11) leukemias*. Blood. (1996) 87, 734-45.
28. Bursen, A., Moritz, S., Gaussmann, A., Moritz, S., Dingermann, T., Marschalek, R. *Interaction of AF4 wild-type and AF4.MLL fusion protein with SIAH proteins: indication for t(4;11) pathobiology?* Oncogene (2004) 23, 6237-6249.
29. Oliver. P., L., Bitoun, E., Clark, J., Jones, E., L., Davies, K., E.. *Mediation of Af4 protein function in the cerebellum by Siah proteins*. Proc. Natl. Acad. Sci. U S A. (2004) 101, 14901-6.
30. Bitoun E, Oliver PL, Davies KE. *The mixed-lineage leukemia fusion partner AF4 stimulates RNA polymerase II transcriptional elongation and mediates coordinated chromatin remodelling*. Human Molecular Genetics, 2007; 16:92-106
31. Isaacs, A., M., Oliver, P., L., Jones, E., L., Jeans, A., Potter, A., Hovik, B., H., Nolan, P., M., Vizor, L., Glenister, P., Simon, A., K., Gray, I., C., Spurr, N., K., Brown, S., D., Hunter, A., J., Davies, K., E.. *A mutation in Af4 is predicted to cause cerebellar ataxia and cataracts in the robotic mouse*. J Neurosci. (2003) 23, 1631-7.
32. Bitoun, E., Davies, K., E.. *The robotic mouse: unravelling the function of AF4 in the cerebellum*. Cerebellum. (2005) 4, 250-60.
33. Thirman, M., J., Levitan, D., A., Kobayashi, H., Simon, M., C., Rowley, J., D.. *Cloning of ELL, a gene that fuses to MLL in a t(11;19)(q23;p13.1) in acute myeloid leukemia*. Proc. Natl. Acad. Sci. USA. (1994) 91, 12110-4.
34. Conaway, J., W., Shilatifard, A., Dvir, A., Conaway, R., C., (2000) *Control of elongation by RNA polymerase II*. Trends Biochem. Sci. 25, 375-80.
35. Shilatifard, A, Lane W., S., Jackson, K., W., Conaway, R., C., Conaway J., W., (1996) *An RNA polymerase II elongation factor encoded by the human ELL gene*. Science 271, 1873-6.
36. Lavau, C., Luo, R., T., Du, C., Thirman, M., J., (2000) *Retrovirus-mediated gene transfer of MLL-ELL transforms primary myeloid progenitors and causes acute myeloid leukemias in mice*. Proc. Natl. Acad. Sci U S A. 97, 10984-9.
37. Colland, F., Jacq, X., Trouplin, V., Mouglin, C., Groizelau, C., Hamburger, A., Meil, A., Wojcik, J., Legrain, P., Gauthier, J-M., (2006) *Functional proteomics mapping of a human signalling pathway*. Genome Res. 14, 1324-1332

38. Binato R, Alvarez Martinez CE, Pizzatti L, Robert B, Abdelhay E *SMAD 8 binding to mice Msx1 basal promoter is required for transcriptional activation.* Biochem J, 2006 ; 393: 141-150.
39. Divoky V, Trka JM, Watzinger F, Lion T, *Cryptic splice site activation during RNA processing of MLL/AF4 chimeric transcripts in infants with t(4;11) positive ALL,* Gene. 2000 Apr 18;247(1-2):111-8.
40. Domer PH, Fakharzadeh SS, Chen CS, Jockel J, Johansen L, Silverman GA, Kersey JH, Korsmeyer SJ, *Acute mixed-lineage leukemia t(4;11)(q21;q23) generates an MLL-AF4 fusion product,* Proc Natl Acad Sci U S A. 1993 Aug 15;90(16):7884-8.
41. Gaussmann A, Wenger T, Eberle I, Bursen A, Bracharz S, Herr I, Dingermann T, Marshalek R, *Combined effects of the two reciprocal t(4;11) fusion proteins MLL.AF4 and AF4.MLL confer resistance to apoptosis, cell cycling capacity and growth transformation,* Oncogene. 2007 May 17;26(23):3352-63.
42. Xia ZB, Popovic R, Chen J, Theisler C, Stuart T, Santillan DA, Erfurth, Diaz MO, Zeleznik-Le NJ, *The MLL fusion gene, MLL-AF4, regulates cyclin-dependent kinase inhibitor CDKN1B (p27kip1) expression,* Proc Natl Acad Sci U S A. 2005 Sep 27;102(39):14028-33.
43. Jiang, L., He, L., Fountoulakis, M., (2004) *Comparison of protein precipitation methods for sample preparation prior to proteomic analysis.* J. Chromatogr. A. 1023, 317-20.
44. Yin, H., Killeen, K., Brennen, R., Sobek, D., Werlich, M., van de Goor, T., (2005) *Microfluidic chip for peptide analysis with an integrated HPLC column, sample enrichment column, and nanoelectrospray tip.* Anal. Chem. 77, 527-33.
45. Perkins, D., N., Pappin, D., J., Creasy, D., M., Cottrell, J., S., (1999) *Probability-based protein identification by searching sequence databases using mass spectrometry data.* Electrophoresis. 20, 3551-67.
46. Mathivanan, S., Periaswamy, B., Gandhi, T., Kandasamy, K., Suresh, S., Mohmood, R., Ramachandra, Y., Pandey, A., (2006) *An evaluation of human protein-protein interaction data in the public domain.* BMC Bioinformatic. 7 Suppl 5:S19.
47. Shen TL, Noon KR. *Liquid chromatography-mass spectrometry and tandem mass spectrometry of peptides and proteins.* Methods Mol Biol. 2004;251:111-40.
48. Moulder R, Filén JJ, Salmi J, Katajamaa M, Nevalainen OS, Oresic M, Aittokallio T, Lahesmaa R, Nyman TA. *A comparative evaluation of software for the analysis of liquid chromatography-tandem mass spectrometry data from isotope coded affinity tag experiments.* Proteomics. 2005 Jul;5(11):2748-60.

49. Beausoleil, S., A., Jedrychowski, M., Schwartz, D., Elias, J., E., Villen, J., Li, J., Cohn, M., A., Cantley, L., C., Gygi, S., P., (2004) *Large-scale characterization of HeLa cell nuclear phosphoproteins*. Proc. Natl. Acad. Sci. U S A. 101, 12130-5.
50. Acker J, de Graaff M, Cheynel I, Khazak V, Kedinger C, Vigneron M. *Interactions between the human RNA polymerase II subunits*. J Biol Chem. 1997 Jul 4;272(27):16815-21.
51. Luo, R.,T., Lavau, C., Du, C., Simone, F., Polak, P., E., Kawamata, S., Thirman, M., J., (2001) *The elongation domain of ELL is dispensable but its ELL-associated factor 1 interaction domain is essential for MLL-ELL-induced leukemogenesis*. Mol. Cell. Biol. 21, 5678-87.
52. Simone, F., Polak, P., E, Kaberlein, J., J., Luo, R., T., Levitan, D., A., Thirman, M., J., (2001) *EAF1, a novel ELL-associated factor that is delocalized by expression of the MLL-ELL fusion protein*. Blood. 98, 201-9.
53. Malik, S., Roeder, R., G., (2005) *Dynamic regulation of pol II transcription by the mammalian Mediator complex*. Trends Biochem. Sci. 30, 256-63.
54. Conaway, R., C., Sato, S., Tomomori-Sato, C., Yao, T., Conaway, J., W., (2005) *The mammalian Mediator complex and its role in transcriptional regulation*. Trends Biochem. Sci. 30, 250-5.
55. Woychik NA, Hampsey M. *The RNA polymerase II machinery: structure illuminates function*. Cell. 2002 Feb 22;108(4):453-63. Review.
56. Ryu S, Zhou S, Ladurner AG, Tjian R, *The transcriptional cofactor complex CRSP is required for activity of the enhancer-binding protein Sp1*, Nature. 1999 Feb 4; 397 (6718): 446-50.
57. Zhu, Y., Pe'ery, T., Peng, J., Ramanathan, Y., Marshall, N., Marshall, T., Amendt, B., Mathews. M., B., Price, D., H., (1997) *Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation in vitro*. Genes Dev. 11, 2622-32.
58. Mackintosh, C., (2004) *Dynamic interactions between 14-3-3 proteins and phosphoproteins regulate diverse cellular processes*. Biochem. J. 15, 329-42. Review.
59. Isnard P, Depetris D, Mattei MG, Ferrier P, Djabali M. *cDNA cloning, expression and chromosomal localization of the murine AF-4 gene involved in human leukemia*. Mamm Genome. 1998 Dec;9(12):1065-8. No abstract available.
60. Nakamura T, Alder H, Gu Y, Prasad R, Canaani O, Kamada N, Gale RP, Lange B, Crist WM, Nowell PC, et al. *Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in acute leukemia share sequence homology and/or common motifs*. Proc Natl Acad Sci U S A. 1993 May 15;90(10):4631-5.