



UNIVERSITA' DEGLI STUDI DI TRIESTE

XXVI Ciclo del Dottorato di Ricerca
In Scienze della Riproduzione e dello Sviluppo
indirizzo Genetico Molecolare

Sviluppo di una strategia per la diagnosi molecolare dell'Anemia di Fanconi

MED/03 Genetica Medica

Dottoranda:	Roberta Bottega
Coordinatore:	Prof.ssa Giuliana Decorti
Supervisore di tesi:	Prof.ssa Anna Savoia

ANNO ACCADEMICO 2012-2013

RIASSUNTO

L'anemia di Fanconi (FA) è una malattia genetica rara caratterizzata da malformazioni congenite, pancitopenia, predisposizione al cancro e aumentata sensibilità ad agenti, quali diepossibutano e mitomicina C, che formano legami tra i due filamenti di DNA. La FA è causata da almeno 16 geni che costituiscono, insieme ad altri componenti, un pathway di riparazione del DNA. L'eterogeneità è uno dei principali motivi che complica la diagnosi molecolare della FA. E' pertanto necessario un processo a più livelli che implica lo screening di molti esoni o, in alternativa, l'allestimento di linee cellulari e l'analisi di complementazione per la caratterizzazione del gene candidato.

Gli scopi di questa tesi pertanto sono diretti a:

- Ridurre i tempi per l'identificazione del gene mutato sostituendo l'analisi di complementazione con quella di espressione delle proteine FA basandosi sul presupposto che prodotti mutati siano rapidamente degradati;
- Caratterizzare dal punto di vista molecolare gli effetti delle varianti identificate dall'analisi di sequenza.

Per quanto riguarda il primo obiettivo, ci siamo focalizzati sullo studio della proteina FANCA in 44 linee cellulari linfoblastoidi appartenenti ai diversi gruppi di complementazione. E' emerso che, fatta eccezione per FA-G, l'espressione di FANCA non è alterata da mutazioni nei geni *FANCB*, *FANCC* e *FANCD2*. Per quanto riguarda i pazienti con mutazioni in *FANCA*, invece, abbiamo osservato una correlazione tra il tipo di mutazione e il livello di espressione della proteina che può quelli essere paragonabile a quella dei controlli nel caso di mutazioni missenso o ampie delezioni *in frame*. In accordo con l'ipotesi invece, in presenza di mutazioni nonsense e frameshift in entrambi gli alleli del gene, non si ha produzione di proteina. Sulla base di questi dati possiamo concludere che l'analisi di FANCA non è soddisfacente per assegnare ai pazienti il corrispondente gruppo di complementazione. Tuttavia, da questo studio è emersa l'ipotesi di un'associazione tra l'espressione stabile delle proteine FANCA mutate e un

fenotipo meno grave nei pazienti. I dati preliminari dimostrano che queste proteine non sono traslocate nel nucleo e che quindi un'eventuale attività residua non sia da attribuire al processo di riparazione del DNA. Un potenziale ruolo andrebbe forse indagato a livello citoplasmatico dove, come sta emergendo dalla letteratura, almeno FANCG e FANCC, svolgono una funzione all'interno del mitocondrio tale da giustificare l'elevato grado di stress ossidativo delle cellule FA.

Per il secondo obiettivo, lo studio dei casi arruolati nell'ambito dell'AIEOP (Associazione Italiana Ematologia Oncologia Pediatrica) ha consentito l'identificazione delle mutazioni in 100 famiglie. Dall'analisi dei dati emerge che la maggior parte delle mutazioni colpisce il gene *FANCA* (85%), seguito da *FANCG* (9%), *FANCC* (3%), *FANCD2* (2%) e *FANCB* (1%). In assenza del dato di complementazione e/o in presenza di varianti alle quali non è sempre possibile attribuire un chiaro effetto patogenetico, sono state eseguite ulteriori indagini. Si citano a titolo di esempio la caratterizzazione delle ampie delezioni intrageniche mediante MLPA, l'analisi bioinformatica e a livello di RNA delle alterazioni di *splicing* che, qualora *in frame*, sono state ulteriormente confermate anche a livello proteico e, infine, lo studio bioinformatico di patogenicità delle sostituzioni aminoacidiche.

La formulazione di un algoritmo efficace e rapido per la diagnosi molecolare della FA, nonché la chiara definizione del significato patogenetico delle varianti identificate, è molto importante per corretta presa in carico del paziente e della famiglia sia per l'identificazione dei portatori che per la diagnosi prenatale.

INDICE

1. INTRODUZIONE	1
1.1 CARATTERISTICHE CLINICHE	2
1.2 FENOTIPO CELLULARE	4
1.3 REVERSIONE DEL FENOTIPO	5
1.4 ETEROGENEITA' GENETICA.....	7
1.5 IL GENE <i>FANCA</i>	7
1.6 IL PATHWAY FANCONI.....	12
1.6.1 Evoluzione delle proteine Fanconi	12
1.6.2 Classificazione funzionale delle proteine Fanconi	14
1.6.3 Il pathway di riparazione del danno al DNA: i complessi I e II	16
1.6.4 Il pathway di riparazione del danno al DNA: il complesso III	19
1.6.5 Altre funzioni del pathway Fanconi	21
1.7 LA TERAPIA DELLA FA	23
2. SCOPO	25
3. MATERIALI E METODI	28
3.1 PAZIENTI, COLTURE CELLULARI, DNA E RNA	29
3.2 REAZIONI DI AMPLIFICAZIONE (PCR).....	30
3.3 SEQUENZIAMENTO E ANALISI DELLE SEQUENZE	33
3.4 REAZIONE DI RETROTRASCRIZIONE E RT-PCR	34
3.5 ANALISI BIOINFORMATICA DELLE MUTAZIONI MISSENSE E DEI SITI DI SPLICING	35
3.6 IMMUNOBLOTTING.....	36
3.6.1 Estrazione delle proteine e determinazione della concentrazione.....	36
3.6.2 Separazione delle proteine e trasferimento sul filtro di nitrocellulosa ..	36

3.6.3 Rilevazione delle proteine di interesse	37
3.7 MULTIPLEX LIGATION-DEPENDENT PROBE AMPLIFICATION (MLPA)	37
4. RISULTATI	39
4.1 STRATEGIA DIAGNOSTICA PER LA DIAGNOSI MOLECOLARE DELL'ANEMIA DI FANCONI	40
4.2 ANALISI MOLECOLARE DELLA CASISTICA ITALIANA.....	45
4.2.1 Mutazioni del gene <i>FANCA</i>	46
4.2.2 Caratterizzazione delle mutazioni.....	46
4.2.3 Effetto del fondatore delle mutazione in <i>FANCA</i>	53
4.2.4 Mutazioni in geni non- <i>FANCA</i>	54
4.3 ANALISI DELL'ESPRESSIONE PROTEICA PER LA DIAGNOSI DELL'ANEMIA DI FANCONI	59
4.3.1 Espressione della proteina <i>FANCA</i> e correlazione con il gruppo di complementazione.....	59
4.3.2 Gruppo di complementazione FA-A: correlazione tra genotipo ed espressione di <i>FANCA</i>	61
4.3.3 Espressione della proteina <i>FANCA</i> nei pazienti con ampie delezioni <i>in frame</i>	63
4.3.4 Localizzazione nucleare e citoplasmatica di <i>FANCA</i> mutata	64
5. DISCUSSIONE	66
6. CONCLUSIONI	73
7. BIBLIOGRAFIA.....	75

1. INTRODUZIONE

INTRODUZIONE

L'anemia di Fanconi (FA) è una rara malattia genetica descritta per la prima volta da Guido Fanconi nel 1927. Da allora, numerosi altri casi sono stati descritti e studiati rendendo possibile una chiara definizione della malattia anche a livello molecolare. La malattia si presenta con una frequenza di 1 caso ogni 160.000 nati con una frequenza del portatore stimata superiore a 1/200. In alcune popolazioni, tuttavia, la frequenza del portatore è aumentata a causa di effetti del fondatore. Fino ad oggi più di 2.000 casi sono stati riportati in letteratura.

1.1 CARATTERISTICHE CLINICHE

La FA è caratterizzata da una notevole eterogeneità fenotipica che si traduce con la possibile associazione di diverse anomalie riscontrabili nei pazienti. La più importante caratteristica clinica si manifesta a livello del sangue. La FA, infatti, è la più comune tra le sindromi ereditabili causate da anomalie dell'eritropoiesi. Alla nascita, la conta dei globuli rossi è generalmente normale. Il primo segno rilevabile della malattia è la macrocitosi seguita da trombocitopenia, neutropenia e anemia aplastica che evolve rapidamente in pancitopenia con esito fatale nella maggior parte dei casi. Le anomalie ematologiche si presentano generalmente all'età media di 7 anni e, l'incidenza totale a 40 anni è del 90% (Auerbach, 2009). La malattia è spesso accompagnata da malformazioni congenite. Nel 75% circa dei pazienti si riscontrano anomalie scheletriche tra cui ipoplasia di radio e pollice, deformazioni a livello vertebrale e scoliosi. E' interessante notare che le anomalie scheletriche spesso si presentano in maniera monolaterale o bialterale ma asimmetrica; tale comportamento è indice di una componente stocastica nello sviluppo delle malformazioni. In altre parole, un difetto nel

pathway FA non è sufficiente a causare il difetto di per sé ma aumenta notevolmente la probabilità che si manifesti.

Spesso i bambini affetti presentano zone di iper- e ipopigmentazione che si traducono in caratteristiche macchie cutanee cosiddette "color caffelatte". La FA è associata anche a problemi di ritardo della crescita dovuti a deficienze ormonali o ipotiroidismo sia in utero che dopo la nascita. Circa 1/4 dei pazienti presenta ritardo mentale, anomalie dei reni e del tratto urinario, anomalie oculari e sordità dovute ad anomalie strutturali; 1/3 dei pazienti presenta invece microcefalia.

Nei maschi spesso si riscontra ipogenitalia, ipospadia e difetti nella spermatogenesi che sono la causa della generale infertilità maschile. Anche le femmine possono presentare sviluppo incompleto dei genitali ma, se non trattate con androgeni, possono essere fertili sebbene le gravidanze siano spesso associate a complicanze.

Manifestazioni cliniche più rare comprendono malformazioni cardiache, del sistema nervoso e dell'apparato gastrointestinale (Tischkowitz and Hodgson, 2003).

Il notevole grado di eterogeneità fenotipica che caratterizza la FA sia per quanto riguarda le manifestazioni ematologiche che quelle malformative rende conto del fatto che non tutti i pazienti manifestino la totalità dei sintomi tipici della malattia e con la stessa severità. Addirittura si riscontra una notevole variabilità tra pazienti appartenenti alla stessa famiglia pur condividendo le stesse mutazioni. Per questo motivo molto probabilmente non è mai stato possibile stabilire una correlazione tra genotipo e gravità dei sintomi.

A complicare ulteriormente il quadro è importante ricordare che lo spettro delle manifestazioni cliniche è sovrapponibile con quello di altre patologie rendendo difficoltoso il riconoscimento dei pazienti FA da quelli affetti da altre sindromi quali l'anemia di Diamond-Blackfan, la sindrome Nijmegen (NBS), la sindrome di Seckel o la discheratosi congenita (Auerbach, 2009).

Un'altra caratteristica che accomuna i pazienti FA è l'elevata predisposizione allo sviluppo di neoplasie sia sotto forma di tumori solidi che leucemie in giovane età. I tumori solidi nei pazienti FA confrontati con quelli della popolazione generale seguono una distribuzione inusuale in quanto più della metà colpiscono testa, collo, esofago e vulva. I tumori al fegato, generalmente benigni, sono probabilmente correlati all'utilizzo degli androgeni in terapia. Tra i tumori ematologici, i più frequenti sono le leucemie non

linfocitiche come la leucemia mieloide acuta (LMA) e la sindrome mielodisplastica (MDS) che compaiono ad un'età media di 11,3 anni. L'incidenza totale dei tumori ematologici e non-ematologici nei pazienti FA prima dei 40 anni di età si assesta attorno al 33% e al 28% rispettivamente (Mathew, 2006).

1.2 FENOTIPO CELLULARE

Le cellule FA sono caratterizzate da un elevato grado di instabilità cromosomica con prolungamento o arresto in fase G2 del ciclo cellulare dovuto al fatto che le cellule non riescono a superare il *check-point* di controllo di integrità del genoma prima della divisione. Questa caratteristica è condivisa da altre patologie quali l'ataxia teleangiectasia e la sindrome di Bloom che, pur presentando sintomi clinici diversi, hanno in comune la predisposizione alle neoplasie trattandosi di malattie causate da difetti nella riparazione del DNA (Jacquemont and Taniguchi, 2007). A differenza di queste malattie, le cellule FA presentano un elevato grado di instabilità genomica indotta da esposizione ad agenti cross-linkanti quali diepossibutano (DEB), mitomicina C (MMC) e cisplatino. Tali sostanze creano legami tra i filamenti del DNA (Figura 1). Se esposte a queste sostanze, le cellule FA manifestano un'elevata fragilità cromosomica che porta ad un aumento delle rotture cromatidiche e alla formazione di strutture triradiali e quadriradiali d'interscambio tra cromosomi non omologhi (Joenje et al., 1995). Il danno che ne consegue (ICLs, *interstrand e intrastrand DNA crosslinks*) inibisce processi essenziali quali la replicazione del DNA e la trascrizione e deve necessariamente essere riparato o bypassato per permettere alle cellule di sopravvivere.

Questa caratteristica cellulare è stata sfruttata per lo sviluppo del cosiddetto "Test al DEB". Si tratta di un'analisi citogenetica che permette di diagnosticare in modo corretto i pazienti FA. Tale test è stato effettuato anche in diagnosi prenatale sulle cellule amniotiche, su quelle dei villi corionici o sul sangue fetale. E' un test molto delicato non sempre di facile interpretazione, al quale non si ricorre mai se sono note le mutazioni nella famiglia.

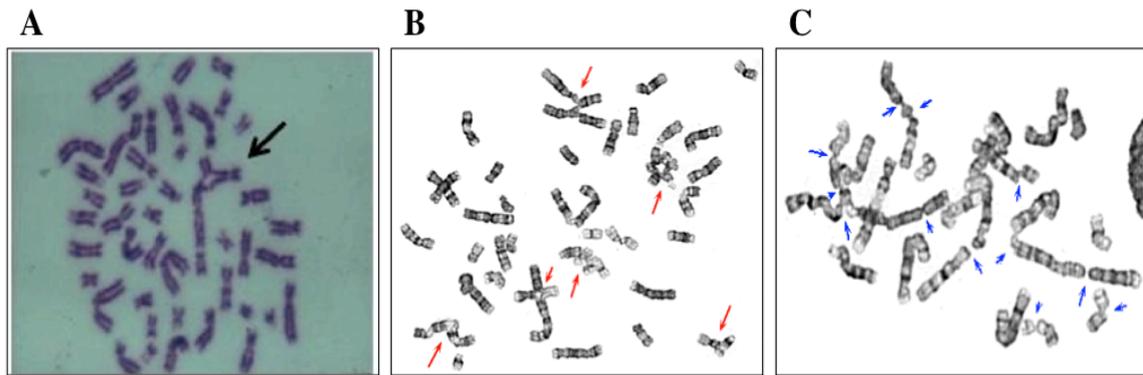


Figura 1. Analisi del cariotipo in un paziente con FA. (A) Alterazioni cromosomiche spontanee con formazione di una struttura triradiale complessa (freccia). (B) (C) Alterazioni cromosomiche evidenziate in un cariotipo dopo il trattamento con agenti alchilanti: sono visibili strutture complesse triradiali e quadriradiali (freccie rosse) e rotture su uno o entrambi i cromatidi (freccie blu) tipiche delle cellule FA.

In alternativa o in combinazione con il test al DEB molti laboratori utilizzano la citometria a flusso per rilevare, dopo esposizione ad agenti cross-linking, il numero di cellule in fase G2 (Auerbach, 2009).

Recenti studi hanno identificato nelle aldeidi, intermedi di reazione di numerosi pathway cellulari, gli agenti fisiologici che producono i *cross-link* dannosi per il doppio filamento del DNA (Langevin et al., 2011).

1.3 REVERSIONE DEL FENOTIPO

Numerosi pazienti FA sono considerati mosaici, almeno a livello ematopoietico. Il mosaicismo, in questo caso, è definito dalla presenza di due popolazioni cellulari distinte, una con instabilità cromosomica e quindi con fenotipo FA e l'altra con fenotipo normale. Tale aspetto è stato evidenziato analizzando centinaia di linee linfoblastoidi allestite da pazienti FA con un test al DEB positivo e osservando che il 20%, dopo un breve periodo in coltura, manifestavano resistenza al DEB e/o alla MMC (Joenje et al., 1995).

Una spiegazione a questo fenomeno, definito "reversione del fenotipo", si basa sull'ipotesi che le cellule linfoblastoidi resistenti derivino da linfociti "normali" del paziente. Tali cellule, se immortalizzate, acquisiscono un vantaggio selettivo dovuto ad

una maggiore capacità di replicazione rispetto alle cellule FA. Questa ipotesi è stata confermata a livello molecolare in diversi studi che hanno messo in evidenza come, eventi di ricombinazione mitotica, mutazioni somatiche (*de novo* e “back mutation”) o conversione genica, abbiano generato alleli *wild type* del gene mutato creando una condizione di eterozigosi sufficiente a ripristinare il fenotipo normale (Hirschhorn, 2003).

La reversione del fenotipo è stata finora dimostrata solo in cellule ematopoietiche ma non si può escludere che possa interessare anche altri tessuti e che rappresenti una delle cause di estrema variabilità clinica dei pazienti FA. Interessando un numero consistente di pazienti (circa il 20%), le cause di questo fenomeno potrebbero essere direttamente associate ai difetti di riparazione del DNA che sono alla base della malattia. Questo perché un elevato tasso di errori potrebbe inserire, anche se raramente, alterazioni in grado di correggere il fenotipo con importanti implicazioni cliniche. L’espansione clonale di una cellula staminale emopoietica con reversione potrebbe portare ad una correzione del fenotipo ematologico o perlomeno al blocco dell’evoluzione verso la pancitopenia conferendo ai mosaici un fenotipo più lieve. In questo contesto quindi, potrebbe essere giustificato parlare di “terapia genica naturale”. In realtà, il mosaicismo, pur rappresentando un “trampolino” verso l’attenuazione del quadro clinico, pone grossi interrogativi dal punto di vista clinico; innanzitutto in caso di trapianto e di valutazione dei rischi associati all’insorgenza di neoplasie la coesistenza di popolazioni di cellule diversamente sensibili ad agenti citotossici, è un aspetto da tenere in forte considerazione quando si progettano le modalità di condizionamento (Gross et al., 2002). Inoltre, nei pazienti con mosaicismo emopoietico il test al DEB non è sempre di facile interpretazione. In questi casi, per confermare o escludere il sospetto diagnostico, si deve ricorrere all’analisi di altre cellule come per esempio i fibroblasti (Auerbach, 2009).

1.4 ETEROGENEITA' GENETICA

La FA è una malattia geneticamente eterogenea con ben 16 gruppi di complementazione ad oggi identificati, a ciascuno dei quali è associato un gene (Tabella 1).

Il gene *FANCC* è stato il primo ad essere identificato grazie allo sviluppo di una strategia di clonaggio funzionale che sfrutta l'analisi di complementazione (Joenje et al., 1995). Negli anni successivi, lo stesso approccio ha portato alla scoperta di altri quattro geni: *FANCA*, *FANCG*, *FANCF* e *FANCE* (Wang, 2007). In particolare, *FANCA* è stato l'unico gene isolato anche mediante un approccio di "clonaggio per posizione", una strategia che sfruttando la disponibilità di famiglie informative, ha permesso prima di localizzare il gene sul cromosoma 16 e successivamente di ricercare le mutazioni nei geni candidati. Gli altri undici geni sono stati identificati utilizzando una serie complessa di strategie, tra cui quella di maggiore successo è stata l'identificazione di geni candidati selezionati in quanto subunità di un complesso costituito da proteine FA note (Tabella 1) (Kim and D'Andrea, 2012).

Dei 16 diversi geni FA, *FANCA* è quello più frequentemente mutato nei pazienti (65%). Seguono *FANCC* e *FANCG* che, insieme, sono implicati in un altro 20% dei casi. Tutte le mutazioni identificate nei geni Fanconi sono elencate nel "Fanconi Anemia Mutation Database" gestito dalla Rockefeller University (<http://www.rockefeller.edu/fanconi/>). E' da notare che, alcuni geni, sono mutati in un numero veramente limitato di pazienti e, pertanto, rappresentano forme molto rare di FA (Ameziane et al., 2008). Per questo motivo, sarà brevemente descritto solo il gene *FANCA*.

1.5 IL GENE FANCA

Il gene *FANCA*, localizzato sul cromosoma 16q24.3, è contenuto in una regione di 80 Kb e suddiviso in 43 esoni (Figura 2, A); esso produce un trascritto di 4365 bp che codifica per una proteina di 1455 amminoacidi. Lo spettro delle mutazioni identificate in *FANCA* è molto ampio: circa un quarto del totale sono mutazioni del tipo *missense*,

seguono *nonsense*, microdelezioni/inserzioni e alterazioni dei siti di *splicing* che possono interessare qualsiasi punto del gene nonostante esistano degli “hot-spot” mutazionali (Castella et al., 2011, Savino et al., 1997). Inoltre, nel 20% circa dei casi la mutazione causativa è un’ampia delezione intragenica (Figura 2, B). Meccanismi di crossing-over ineguali tra ripetizioni *Alu* di cui sono particolarmente ricche le regioni introniche di *FANCA* sono alla base di questo fenomeno relativamente frequente (Castella et al., 2011, Morgan et al., 1999). Le piccole delezioni/inserzioni invece, sono probabilmente associate ad uno slittamento della polimerasi durante la replicazione in quanto spesso si localizzano in prossimità di ripetizioni di brevi stringhe di basi o di tratti omonucleotidici. Le mutazioni identificate in *FANCA*, come pure negli altri geni FA, sono elencate nel database delle mutazioni FA (<http://www.rockefeller.edu/fanconi/genes/>).

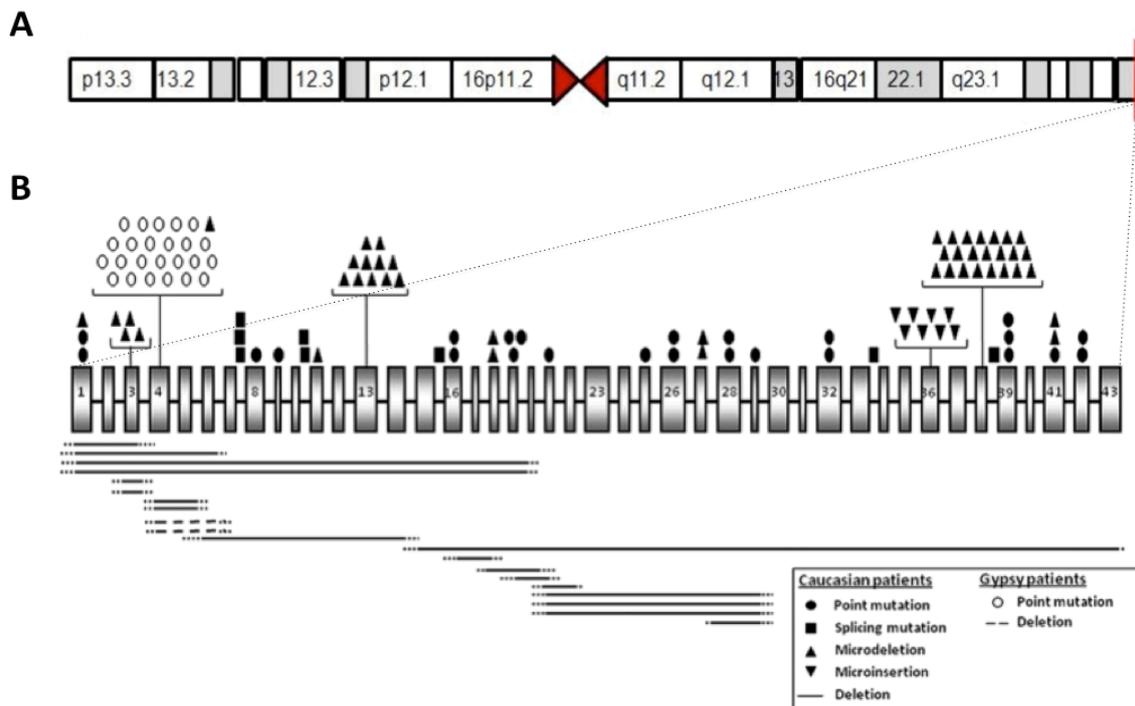


Figura 2. (A) Rappresentazione schematica del cromosoma 16 e localizzazione del gene *FANCA* in posizione 16q24.3 (barra rossa). (B) Localizzazione delle mutazioni identificate nella popolazione spagnola. Sono rappresentati tutti i 43 esoni del gene *FANCA* e le mutazioni identificate all’interno del gene (Adattato da Castella et al., 2011).

Nella cellula, la proteina *FANCA* si localizza sia nel citoplasma che nel nucleo con migrazioni che seguono le varie fasi del ciclo cellulare. In particolare, l’entrata e l’uscita di *FANCA* dal nucleo sono processi finemente regolati che prevedono l’interazione tra

specifici domini della stessa e proteine di trasporto. La porzione N-terminale di FANCA, infatti, contiene un segnale di localizzazione nucleare (NLS) costituito prevalentemente da amminoacidi basici indispensabile per l'entrata nel nucleo (Figura 3) e, pazienti con mutazioni in queste regioni, presentano la proteina FANCA solamente nel compartimento citoplasmatico (Lightfoot et al., 1999). Oltre a questo dominio, la presenza di cinque sequenze funzionali ricche in leucina (NESs) distribuite su tutta la lunghezza della proteina (Figura 3), assicurano l'interazione con il recettore nucleare CRM-1 permettendo alla proteina di attraversare il poro nucleare e tornare nel citoplasma (Ferrer et al., 2005).

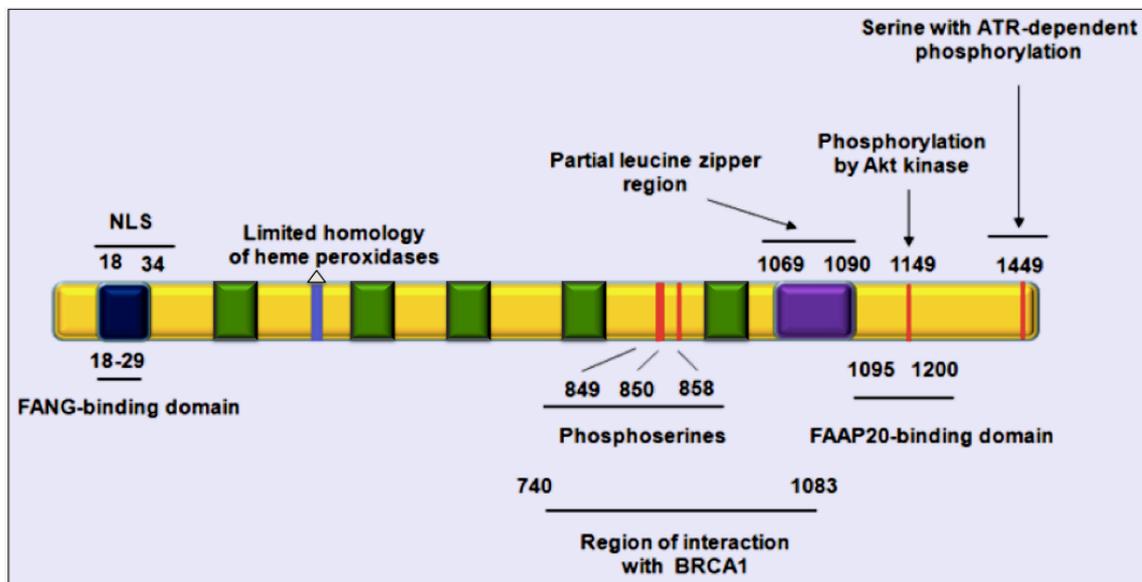


Figura 3. Rappresentazione schematica della proteina FANCA. Sono indicate le diverse regioni funzionali fino ad oggi identificate tra cui il segnale di localizzazione nucleare (NLS), il sito di legame con la proteina FANCG, i siti di fosforilazione e le regioni ricche in leucina (NESs, box verdi). Le proporzioni dei diversi domini non sono in scala. (Adattata da <http://atlasgeneticsoncology.org/Genes/FA1ID102.html>).

Altre regioni di FANCA non ancora ben caratterizzate, permettono l'interazione tra FANCA e FANCG portando alla formazione di un complesso tra le due proteine nel citoplasma. Studi di co-immunoprecipitazione e traduzione in vitro hanno indicato i residui compresi tra il 18 e il 29 di FANCA quali responsabili del legame con FANCG (Hodson and Walden, 2012). Le funzioni di tale complesso non sono ancora chiarite

tuttavia, si ritiene che svolga un ruolo importante nella localizzazione subcellulare di FANCA. Infatti, cellule di pazienti difettivi della proteina FANCG, non presentano FANCA a livello nucleare, difetto recuperabile con la successiva aggiunta della proteina mancante (Garcia-Higuera et al., 2000).

Per quanto riguarda le modificazioni post-traduzionali, l'unica ad oggi descritta per la proteina FANCA riguarda la fosforilazione a livello delle serine 1149 e 1449 (Collins et al., 2009) (Figura 3).

Infine, il gene FANCA, pur essendo trascritto in tutti i tessuti, mostra alcune variazioni nel livello di espressione, molto elevato, per esempio, nel testicolo (Taniguchi and D'Andrea, 2006).

Gruppo di complementazione	Gene	Localizzazione cromosomica	Proteina (kda)	Ruolo in FANCD2Ub	Funzione della proteina	Pazienti FA (%)
A	FANCA	16q24.3	163	+	Complesso FA	65%
B	FANCB	Xp22.2	95	+	Complesso FA	circa 2%
C	FANCC	9q22.3	63	+	Complesso FA	10%
D1	FANCD1/BRCA2	13q12.3	380	-	Reclutamento RAD51	circa 2%
D2	FANCD2	3p26	155,162(Ub)	+	Proteina mono-Ub	circa 2%
E	FANCE	6p21-p22	60	+	Complesso FA	circa 2%
F	FANCF	11p15	42	+	Complesso FA	circa 2%
G	FANCG/XRCC9	9p13	68	+	Complesso FA	9%
I	FANCI	15q26.1	140,147(Ub)	+	Proteina mono-Ub	circa 2%
J	FANCI/BACH1/BRIP1	17q22.2	130	-	5'→3' DNA elicasi/ATPasi	circa 2%
L	FANCL/PHF9/POG	2p16.1	43	+	Complesso FA, E3 Ub ligasi	<0,2%
M	FANCM/Hef	14q21.3	250	+	Complesso FA, E3 ATPasi/translocasi	<0,2%
N	FANCN/PALB2	16p12.1	130	-	Regolazione localizzazione di BRCA2	circa 2%
O	RAD51C	17q25.1	42	-	Promozione formazione dei foci di RAD51	<0,2%
P	SLX4	16p13.3	200	-	Scaffold proteico per numerose proteine	<0,2%
Q	ERCC4	16p13.3	103	-	Endonucleasi specifica della riparazione	<0,2%

Tabella 1. Geni FA finora identificati. Ub = ubiquitinazione

1.6 IL PATHWAY FANCONI

1.6.1 Evoluzione delle proteine Fanconi

Numerosi sono gli studi condotti per cercare di capire il ruolo dei geni FA ed i meccanismi attraverso i quali le mutazioni causino la malattia. Tuttavia, nonostante i notevoli sforzi, i risultati ottenuti non hanno ancora chiarito del tutto la loro funzione per diversi motivi, primo fra tutti l'estrema complessità del pathway FA, sia dal punto di vista del numero di interattori che ne fanno parte, sia a causa del complicato meccanismo di regolazione necessario al corretto funzionamento dello stesso. I primi passi in questo senso si sono rivolti verso l'individuazione di domini importanti via via che i geni FA venivano identificati. Per questo tipo di lavoro, a partire dal 1992, con l'identificazione del primo gene del pathway, *FANCC*, sono stati adottati numerosi approcci: studi di omologia delle proteine con domini funzionali noti o con proteine già funzionalmente caratterizzate, esperimenti di localizzazione cellulare ed espressione tissutale, analisi di mutazioni per identificare domini critici, analisi delle interazioni proteiche con partner molecolari e creazione di modelli animali.

Sebbene molte proteine FA presentino motivi per l'interazione con altri componenti, solo alcune presentano qualche omologia con domini funzionali ben definiti come, ad esempio, le proteine *FANCM*, *FANCI*, *FANCO*, *FANCI* e *FANCP* che possiedono domini endonucleasici e/o elicasici associati ad un'attività sul DNA e la proteina *FANCL* che possiede un'attività E3 ubiquitina ligasica (Wang, 2007). Per molti anni, questo ha reso difficile la formulazione di ipotesi relative al funzionamento del pathway.

Anche gli studi evolutivi non si sono rivelati risolutivi: i primi geni studiati, *FANCA* e *FANCG*, presentavano ortologhi solo nei vertebrati avvalorando l'ipotesi che le proteine FA appartenessero ad un pathway evolutivamente recente. *FANCD2* ha cambiato questa concezione perché i rispettivi geni ortologhi sono presenti anche nei vermi, nei lieviti, nelle muffe e nelle piante (Patel and Joenje, 2007). Infatti, recenti studi hanno confermato il ruolo di *FANCD2* in risposta al danno al DNA a seguito della sua ubiquitinazione anche negli organismi inferiori (Collis et al., 2006). Sarebbe importante capire il meccanismo molecolare del funzionamento di questa versione così semplificata

del sistema presente negli invertebrati che mancano pressoché completamente del complesso FA (l'unità costitutiva fondamentale per la monoubiquitinazione di FANCD2) ma che sfruttano le stesse proteine dei vertebrati per far funzionare il macchinario di riparazione a valle.

L'analisi delle proteine FA nel loro complesso quindi ha evidenziato che, sebbene alcune siano presenti solamente nelle specie superiori, altre quali FANCM, FANCL, FANCD2, FANCI, FANCO, FANCP e FANCD2 sono conservate in tutte le specie (Figura 4).

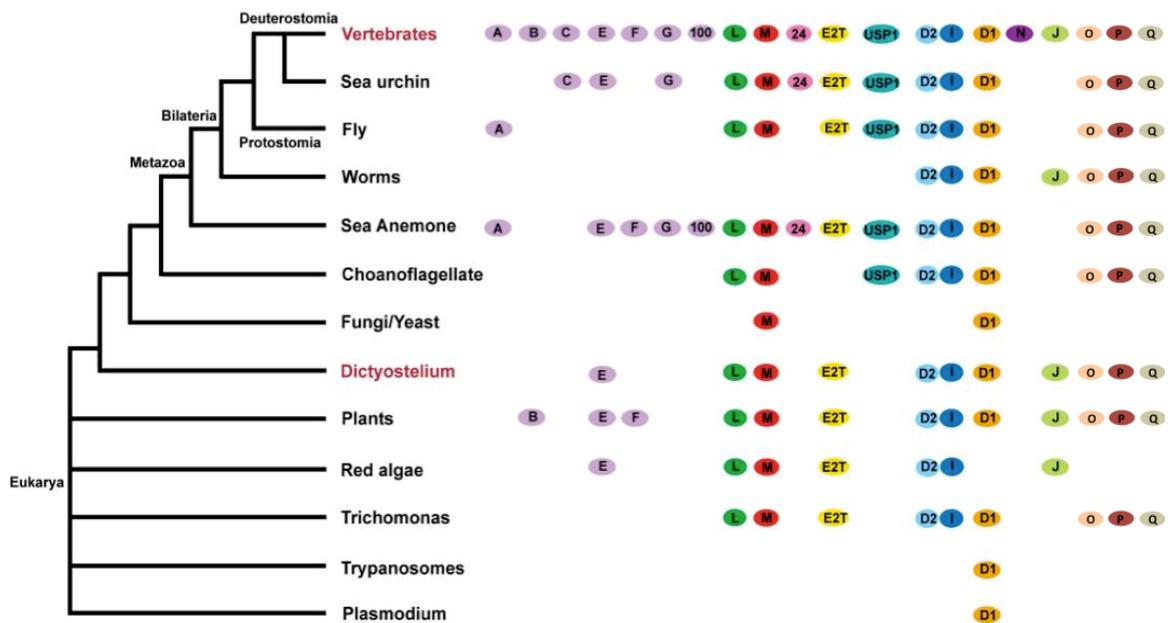


Figura 4. Conservazione evolutiva delle proteine FA negli eucarioti. Molte delle proteine FA costituiscono un complesso chiamato "core complex". Gli eucarioti semplici sembrano non possedere proteine ortologhe alla maggior parte dei componenti del complesso tuttavia, presentano le proteine a valle della monoubiquitinazione di FANCD2 (Adattato da Zhang et al., 2009).

La conservazione di molti componenti dall'uomo fino agli organismi unicellulari suggerisce il ruolo funzionale cruciale dei geni FA nel processo evolutivo (Patel and Joenje, 2007). Grazie ai numerosi studi effettuati, oggi è ampiamente accettato che le proteine FA siano implicate in un network di risposta al danno del DNA svolgendo un ruolo cruciale nel meccanismo della ricombinazione omologa in seguito a rotture del doppio filamento del DNA.

1.6.2 Classificazione funzionale delle proteine Fanconi

Da un punto di vista funzionale i sedici geni FA finora identificati sono classicamente suddivisi in tre sottogruppi (Figura 5):

- 1) Le proteine FANCA, -B, -C, -E, -F, -G, -L e -M, costituiscono il primo gruppo. Queste si assemblano in un complesso multiproteico chiamato “core complex” o più semplicemente complesso I stabilizzato da interazioni deboli tra differenti sub-complessi. La funzione di questo complesso, assieme a due “FA-interacting protein” chiamate FAAP24 e FAAP100, è il controllo della monoubiquitinazione di FANCD2 e FANCI, processo indispensabile per l’attivazione del sistema di riparazione del DNA (Moldovan and D’Andrea, 2009); mutazioni che colpiscono anche solo una subunità rendono le cellule incapaci di legare l’ubiquitina a FANCD2 e a FANCI in quanto viene compromessa la stabilità dell’intero complesso FA. E’ da notare che le uniche due subunità per le quali non sono mai state identificate mutazioni sono FAAP24 e FAAP100. Tuttavia il loro ruolo è da considerarsi comunque centrale poiché, il silenziamento di queste proteine, genera un fenotipo cellulare tipico della FA (Wang, 2007); FANCM e FAAP24 sembrano essere funzionalmente distinte dalle restanti proteine del complesso I in quanto non risultano costitutivamente associate ad esso bensì sono richieste per il riconoscimento dei siti della lesione e probabilmente guidano il complesso verso la sede del danno (Kim et al., 2008).
- 2) Il secondo gruppo è costituito da FANCD2 e FANCI che interagiscono tra di loro in un complesso dinamico (complesso II) che si muove, fuori e dentro la cromatina, in base alle modificazioni post-traduzionali cui va incontro. Quando attivato a seguito dell’ubiquitinazione, il complesso si localizza in corrispondenza dei foci nucleari indotti dal danno al DNA. Queste due proteine, simili per dimensioni e struttura, sono interdipendenti per la monoubiquitinazione. Infatti, in cellule prive di FANCD2, FANCI non è monoubiquitinata e viceversa.

- 3) Il terzo gruppo (complesso III) consiste delle proteine FANCD1/BRCA2, FANCN (o PALB2, partner and localizer of BRCA2), FANCI (o BRIP1, BRCA1 interacting protein), FANCO (o RAD51C), FANCP (o SLX4) e FANCG (o ERCC4, excision repair cross-complementing rodent repair deficiency group 4). Cellule difettive di uno solo di questi geni presentano livelli normali di FANCD2 monoubiquitinata indicando che le proteine corrispondenti funzionano a valle del complesso II o in una via parallela (Moldovan and D'Andrea, 2009). Le proteine FANCD1, FANCI, FANCO e FANCG sono strettamente legate alla riparazione delle rotture del doppio filamento di DNA mediante meccanismi di ricombinazione che sfruttano le sequenze omologhe presenti nel cromosoma integro per riparare le estremità danneggiate del DNA. La ricombinazione tra stampo e molecola difettosa permette la formazione di due molecole di DNA intatte e prive di errori. Per quanto riguarda FANCN invece, sembra sia importante per il reclutamento e l'accumulo di FANCD1 nella sede della riparazione; FANCP infine agisce come *scaffold* per diverse endonucleasi che agiscono in concerto al fine di risolvere eventuali strutture secondarie formatesi a seguito degli eventi di riparazione (Kim and D'Andrea, 2012).

Studi di interazione hanno messo in evidenza la formazione di diversi sub-complessi formati da proteine FA che si formano ciclicamente durante le diverse fasi del ciclo cellulare e permangono più o meno a lungo nel citoplasma. Per esempio, FANCA e FANCG formano un complesso molto stabile così come il complesso FANCB/FANCL/FAAP100 e FANCC/FANCE/FANCF (Hodson and Walden, 2012). Tuttavia, il ruolo funzionale di questi sub-complessi non è ancora stato chiarito.

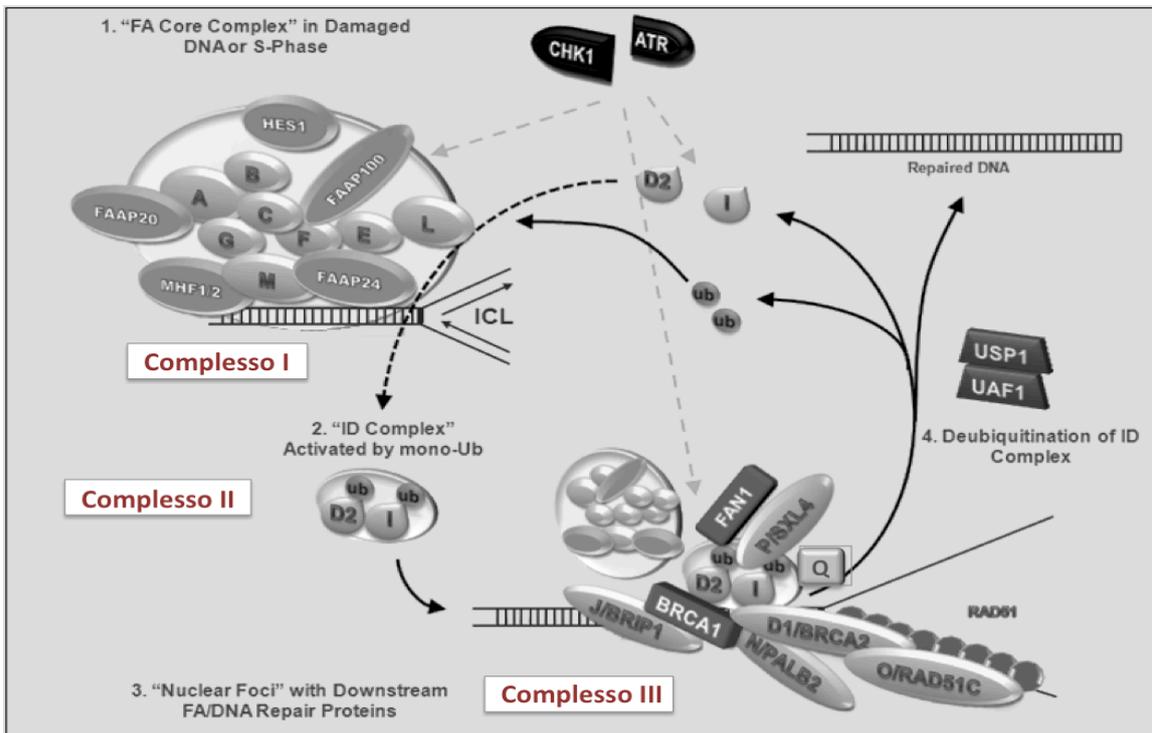


Figura 5. Rappresentazione schematica del pathway FA/BRCA. Un danno al DNA durante la fase S del ciclo cellulare blocca la forca replicativa e attiva i sensori del danno ATR e CHK1. L'attivazione di queste due proteine porta all'assemblaggio di 8 proteine FA nel complesso I e nelle successive ubiquitinazioni di FANCD2 e FANCI (complesso II). Il complesso così attivato si sposta nella sede del danno ovvero a livello dei foci nucleari dove interagisce, attivandole, con le proteine effettrici del riparazione del danno al DNA (Complesso III). Una volta che la riparazione è stata portata a termine, appositi enzimi deubiquitinanti disattivano il complesso II disattivandolo (Valeri et al., 2011).

1.6.3 Il pathway di riparazione del danno al DNA: i complessi I e II

I meccanismi attraverso i quali le proteine FA ed altri fattori esterni cooperano tra loro per recepire, segnalare e riparare il danno rimangono, purtroppo, uno degli aspetti più importanti del pathway ancora da chiarire.

L'attivazione del pathway FA-BRCA dipende, almeno in parte, da ATR (ataxia teleangiectasia and Rad3 related protein) e CHK1 (checkpoint chinase 1), due chinasi attivate a seguito di stress replicativi. Molte proteine FA, quali FANCM, FANCA, FANCE, e FANCG sono infatti fosforilate a seguito di danno al DNA. Tali proteine quindi, potrebbero agire trasducendo il segnale attraverso cui ATR segnala alla cellula la

presenza di problemi replicativi (Cohn and D'Andrea, 2008). A seguito degli eventi fosforilativi ed all'interazione tra le diverse proteine FA, conseguono modificazioni conformazionali che risultano nell'esposizione dei domini di localizzazione nucleare e alla conseguente entrata nel nucleo (Figura 6). Un elemento cruciale di attivazione è costituito inoltre da FANCM che, pur possedendo un dominio elicastico, funziona piuttosto come traslocatore ATP-dipendente, riconosce la forza replicativa bloccata, la stabilizza e richiama il core complex in sede (Collis et al., 2008).

A questo punto può avvenire la monoubiquitinazione di FANCD2 e FANCI, un processo chiave nel pathway FA. Si tratta di un evento di modificazione post-traduzionale che richiede l'instaurarsi di un legame covalente tra la glicina 76 dell'ubiquitina e il gruppo ϵ -amminico di una lisina presente sulla molecola bersaglio (Qiao et al., 2001). Si osserva questo meccanismo in numerosi processi cellulari che richiedono variazioni nella localizzazione subcellulare oppure nell'attività di alcuni componenti. Le otto proteine FA del gruppo I, FAAP24, FAAP100 e UBE2T (enzima ubiquitina-coniugato, E2), riunite in complesso, sono fondamentali per legare la molecola di ubiquitina a livello della Lys 561 e della Lys 523 rispettivamente di FANCD2 e FANCI (Jacquemont and Taniguchi, 2007). La stabilità del complesso è severamente compromessa quando una delle subunità non è prodotta o è difettosa; in entrambi i casi, l'assenza di una qualsiasi di queste componenti compromette la modificazione di FANCD2 e FANCI. La monoubiquitinazione di FANCI non sembra svolgere un ruolo diretto nel pathway FA, tuttavia agisce come una sorta di interruttore molecolare per permettere al pathway FA di attivarsi (de Winter and Joenje, 2009).

La reazione di ubiquitinazione è associata in modo particolare alla presenza di FANCL, una delle subunità del complesso, che rivestirebbe il ruolo di E3 in quanto l'unica ad avere un'attività ubiquitina ligasica grazie ai domini WD40 e PHD (plant homeodomain) che ne permettono l'interazione con l'enzima E2, UBE2T .

La monoubiquitinazione è un processo reversibile e regolato: l'enzima deubiquitinante specifico per questo pathway è la proteasi-1 ubiquitina specifica (USP1). Essa agirebbe come regolatore negativo del pathway FA rimuovendo l'ubiquitina da FANCD2 durante la fase G2 del ciclo cellulare (Alpi and Patel, 2009) (Figura 6).

In seguito alla monoubiquitinazione, FANCD2 e FANCI, sfruttando l'attività translocasica di FANCM, migrano nei foci nucleari in prossimità del sito della lesione al DNA dove

possono coordinare le attività di riparazione insieme alle proteine FA del complesso III nonché a numerosi altre proteine.

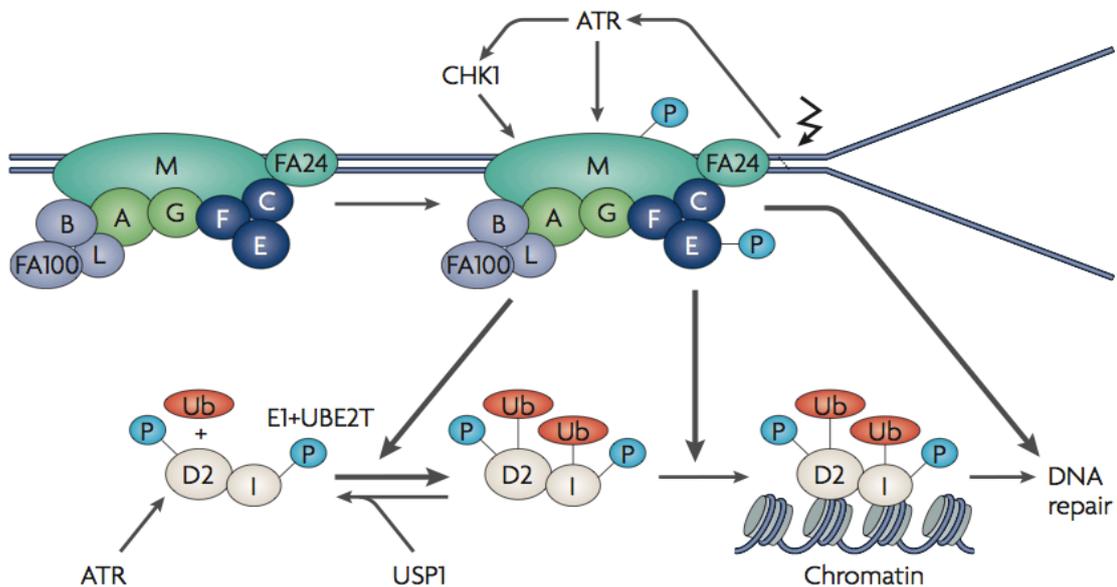


Figura 6. Modello per l'attivazione "a monte" del network FA/BRCA. Una forza replicativa bloccata attiva ATR e CHK1. Le due chinasi attivano il complesso FA e il complesso II tramite fosforilazione. Il complesso FA trasloca sul sito del danno al DNA grazie all'azione di FANCM e FAAP24 e qui, contribuisce a richiamare il complesso II monoubiquitinandolo. Un enzima deubiquitinante, USP1 rimuove l'ubiquitina da FANCD2 per spegnere il network al termine della riparazione del DNA (Wang, 2007).

Una caratteristica fondamentale del pathway FA è che la sua attività è regolata (dallo stato di ubiquitinazione di FANCD2) e inducibile solo nel caso di danno al DNA. E' di notevole importanza sottolineare come l'attività del pathway sia limitata alla fase S del ciclo cellulare. Può tuttavia essere indotta in risposta ad agenti che danneggiano il DNA quali radiazioni ionizzanti, DEB, MMC, idrossiurea, ecc. (Jacquemont and Taniguchi, 2007). Una regolazione di questo tipo potrebbe attuarsi agendo sulla presenza o sulla rimozione dell'ubiquitina da FANCD2, sull'assemblaggio a livello nucleare del complesso FA o, infine, sulla presenza o meno del complesso nel sito danneggiato (Patel and Joenje, 2007).

1.6.4 Il pathway di riparazione del danno al DNA: il complesso III

A seguito dell'ubiquitinazione, in prossimità dei foci nucleari, FANCD2-Ub agisce come sito di ancoraggio per alcune nucleasi quali FAN1 (FA-associated nuclease 1), FANCP/SLX4, SLX4-associated MUS81-EME1 e XPF-ERCC1. Il ruolo di queste nucleasi è di promuovere lo sblocco della forca replicativa mediante il taglio in prossimità del *cross-link* tra i due filamenti del DNA. Il risultato sarà una rottura del doppio filamento (double-strand break, DSB) superata mediante utilizzo di una specifica polimerasi a bassa fedeltà dal meccanismo TLS (translesion synthesis) (Wang, 2007). Quando la forca non è bloccata, la replicazione può proseguire, e il DSB sarà riparato successivamente mediante ricombinazione omologa e NER (nucleotide excision repair) (Kim and D'Andrea, 2012) (Figura 7). In particolare, FANCD2/ERCC1 è un endonucleasi di riparazione che entra in gioco a seguito dell'attivazione del NER coinvolta, se mutata, anche nello xeroderma pigmentoso.

Le altre proteine FA gruppo III ovvero FANCD1, FANCF, FANCG e FANCI svolgono un ruolo importante proprio nel processo di ricombinazione omologa. Queste, infatti, favoriscono la progressione tra i diversi intermedi della HR e facilitano l'accumulo di proteine chiave della riparazione a livello dei foci quali RAD51, BRCA1, BLM, TLS (translesion synthesis) polymerase, RPA e ATR.

La fase finale del pathway FA, ovvero l'attivazione della serie dei processi specifici di riparazione, prende avvio con la deubiquitinazione di FANCD2 ad opera di USP1. Questo evento probabilmente rimuove FANCD2 dalla sede del danno permettendo al complesso FANCD1/BRCA2 e FANCF di interagire con RAD51, la proteina chiave della ricombinazione omologa, attivandola (de Winter and Joenje, 2009).

L'ultimo elemento del pathway FA agente a valle della monoubiquitinazione è FANCF. Questa proteina risulta essere una ATPasi DNA-dipendente con funzione elicastica 5'-3', che lega direttamente i domini BRCT di BRCA1 mettendo in connessione il pathway FA con un'altra via di riparazione del DNA ancora poco chiara ma, che si suppone presenti delle analogie con il TLS (Wang, 2007). Inoltre, FANCF sembra essere in grado di rimuovere strutture di DNA che impediscono la riparazione della lesione (de Winter and Joenje, 2009) (Figura 7).

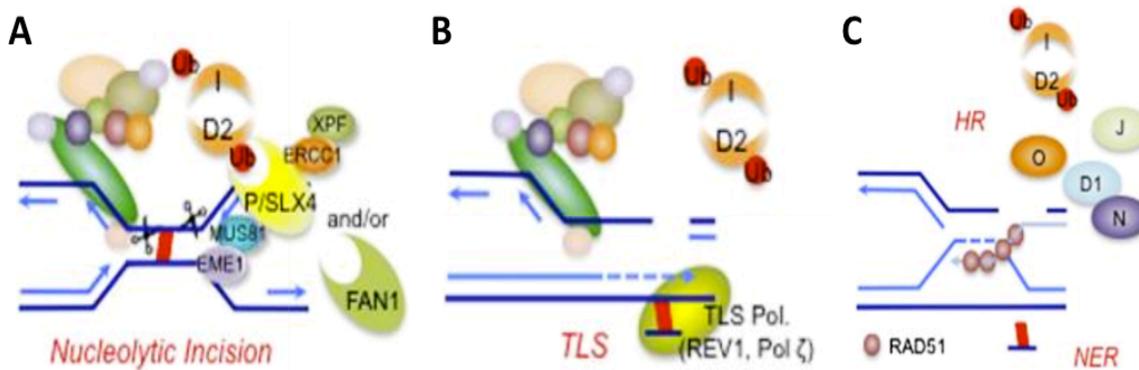


Figura 7. Modello per l'attivazione "a valle" nel network FA/BRCA. Le proteine FA interagiscono in un pathway comune per la risoluzione degli ICLs durante la fase S del ciclo cellulare. (A) FANCD2-Ub agisce da piattaforma per reclutare diverse nucleasi coinvolte nello sblocco della forza replicativa mediante tagli nucleotidici in prossimità dell'ICL; (B) Nel filamento stampo, i tagli operati dalle nucleasi originano un DSB mentre, a livello del filamento complementare, rimane legato il nucleotide crosslinkato (barra rossa). Tuttavia una polimerasi a bassa fedeltà quale la TLS, può superare l'ostacolo e procedere con la replicazione; (C) Il DSB può essere riparato mediante HR ad opera delle proteine FA del gruppo III e di altre proteine coinvolte nella riparazione del DNA tra cui RAD51 (Kim and D'Andrea, 2012).

Oggi è noto che il pathway FA costituisce solo una parte della complessa rete di meccanismi che controllano l'integrità del genoma. Ciò è dimostrato dalla scoperta di una vasta serie di interazioni tra le proteine FA ed altri componenti implicati nella riparazione del DNA.

Mutazioni in alcune di queste proteine (come ad esempio BLM, ATM, ATR) causano altre malattie della riparazione del DNA (Wang, 2007) come l'ataxia teleangiectasia e la sindrome di Bloom che condividono con la FA alcuni aspetti fenotipici e cellulari quali instabilità cromosomica, predisposizione alle neoplasie e ipersensibilità ad alcuni agenti come raggi ultravioletti, raggi X, sostanze alchilanti, nonché il coinvolgimento delle medesime pathway di riparazione. Nonostante ciò, presentano notevoli differenze che riflettono la complessità di questi processi e le diverse e distinte funzioni di ciascuno di essi. Tuttora sono in corso numerosi studi che confermeranno e chiariranno i ruoli delle varie proteine implicate allo scopo di delineare il network complessivo della riparazione del DNA.

1.6.5 Altre funzioni del pathway Fanconi

Oltre ad un ruolo diretto nella riparazione del danno al DNA, il pathway FA è sicuramente coinvolto anche in altre attività che rendono conto dei difetti delle cellule FA oltre l'instabilità cromosomica. Sebbene la monoubiquitinazione di FANCD2 da parte del complesso I sia la chiave del pathway, il controllo di questo processo potrebbe non essere l'unica funzione delle subunità del complesso. Infatti, un costrutto di FANCD2 ubiquitinato in posizione K561 riesce a complementare cellule DT40 knockout per FANCD2 (a suggerire che la posizione dell'ubiquitina potrebbe non essere importante) ma non quelle in cui il complesso I non è integro (Matsushita et al., 2005).

Il fenotipo cellulare delle cellule FA, oltre alla tipica instabilità cromosomica, è caratterizzato da un prolungato periodo di transizione del ciclo cellulare da G2 a M probabilmente dovuto alla mancata efficienza di riparazione del danno al DNA che non permette alla cellula di oltrepassare con successo il *check-point* dell'integrità genomica per poi proseguire alla mitosi (Tischkowitz and Hodgson, 2003).

Oltre alle conseguenze direttamente connesse ad un difetto di riparazione del DNA, l'aspetto fenotipico forse più caratteristico delle cellule FA è lo stato aumentato di stress ossidativo che è probabilmente imputabile a difetti del metabolismo mitocondriale (Pagano et al., 2012). Le cellule FA presentano un incremento della sensibilità all'ossigeno e la frequenza delle rotture cromosomiche è ridotta quando le cellule sono poste in ambiente ipossico. Molte sostanze chimiche che causano danno esogeno al DNA producono intermedi reattivi dell'ossigeno (ROS) ed è possibile che la sensibilità delle cellule FA a tali sostanze sia il risultato di un'incapacità di neutralizzare i ROS (Pagano et al., 2013). In accordo con questa ipotesi, numerosi studi legano la funzione dei geni FA al controllo dell'omeostasi ossido-riduttiva: (a) FANCA, FANCC e FANCG interagiscono con proteine direttamente implicate nel metabolismo mitocondriale (quali il citocromo P450) e nello stress ossidativo (Pagano et al., 2012); (b) FANCG si localizza anche a livello mitocondriale (Mukhopadhyay et al., 2006); (c) FANCI è un sensore del danno ossidativo (Suhasini et al., 2009); (e) alcuni antiossidanti, quali il resveratrolo, diminuiscono l'incidenza del cancro e i difetti dell'ematopoiesi nei topi *Fancd2*^{-/-} (Zhang et al., 2010). Tutte queste evidenze depongono a favore di un ruolo centrale delle

proteine FA nel controllo del metabolismo dell'ossigeno mediante meccanismi non ancora conosciuti.

Le cellule FA inoltre, presentano un'anomala regolazione dei meccanismi apoptotici sebbene non sia ancora chiara la natura del meccanismo alla base del difetto. Un incremento dell'apoptosi potrebbe essere correlato all'incapacità delle cellule FA di riparare i danni al DNA o alla mancata interazione delle proteine FA con proteine dell'apoptosi (Tischkowitz and Hodgson, 2003). Alcuni studi dimostrerebbero ad esempio che FANCC è implicata nella via di segnalazione apoptotica mediata da JAK/STAT (Fagerlie et al., 2004).

Nelle cellule FA è stata, inoltre, rilevata un'aumentata velocità di accorciamento dei telomeri e un contemporaneo aumento delle rotture telomeriche che potrebbero essere dovuti al fatto che le cellule ematopoietiche sono sottoposte ad un tasso di divisione ben più elevato delle cellule *wild type* per cercare di compensare lo stato di anemia (Hanson et al., 2001).

Nei pazienti, infine, è rilevabile una condizione di ematopoiesi inefficace e una riduzione dei livelli di alcuni fattori di crescita come IL-6, GM-CSF e IL-1. Questo concorre all'instaurarsi dei difetti emopoietici tipici dei pazienti FA in quanto, ne risulta un ambiente midollare alterato con conseguenti problemi nella regolazione della divisione e del differenziamento delle cellule ematopoietiche in risposta allo stress. Il tessuto emopoietico è uno dei più sensibili ai danni al DNA causati dalle radiazioni o dalla chemioterapia e questa è la ragione del fatto che malattie quali l'atassia teleangiectasia o lo xeroderma pigmentoso pur presentando problemi di instabilità genomica con aumento del rischio di sviluppare tumori, non hanno la stessa incidenza di insufficienza midollare vista nella FA. Sembra perciò possibile che le proteine FA giochino un ruolo critico nel mantenimento delle cellule staminali ematopoietiche: l'iniziale declino della loro funzionalità potrebbe originarsi a causa del difetto di mantenimento della stabilità telomerica e a causa dell'alto tasso di apoptosi delle cellule FA. In seguito, pressioni selettive potrebbero promuovere lo sviluppo di cloni cellulari mutanti (facilitato anche dall'instabilità genomica) che sfocerebbe in MDS e AML (Tischkowitz and Hodgson, 2003).

1.7 LA TERAPIA DELLA FA

La sopravvivenza dei pazienti FA raramente oltrepassa la terza o quarta decade di vita, sebbene sia molto variabile il decorso clinico e la severità del fenotipo. Anche la terapia quindi andrà adattata caso per caso in base all'età d'insorgenza dei sintomi e alla presenza o meno delle manifestazioni ematologiche.

Il trattamento con androgeni è stato molto utilizzato in passato allo scopo di incrementare l'attività residua del midollo osseo migliorando i livelli di emoglobina e la conta piastrinica. Oggi tuttavia questa terapia è pressoché abbandonata a causa degli effetti collaterali che includono mascolinizzazione e frequente insorgenza di adenomi epatici che spesso evolvono in epatocarcinomi.

Tuttora controversa rimane l'utilità della somministrazione di citochine e fattori di crescita emopoietici come il G-CSF (granulocyte-colony stimulating factor) o il GM-CSF (granulocyte-macrophage colony-stimulating factor) che potrebbero favorire una progressione verso una leucemia acuta (Tischkowitz and Hodgson, 2003).

L'unico trattamento risolutivo per le manifestazioni ematologiche è rappresentato dal trapianto allogenico di midollo osseo o di cellule staminali emopoietiche (SCT) di cordone ombelicale (Gluckman, 1990). Dato che le cellule FA sono più sensibili a diverse sostanze, il trapianto prevede un regime di condizionamento *ad hoc* per ridurre le complicanze indotte dalla tossicità molto elevata di alcune farmaci che sono comunemente usate in questa procedura. Allo stesso modo le terapie tumorali devono tener conto dei difetti nella riparazione del DNA; pertanto, i pazienti seguono un programma di prevenzione per la diagnosi precoce e la rimozione chirurgica delle cellule neoplastiche che il trapianto non può prevenire, così da evitare chemioterapia e radiazioni (Alter, 2007, Alter et al., 2003). Se il donatore è HLA compatibile la probabilità di sopravvivenza senza malattia, a tre anni dal trapianto, è di circa l'80-85%. Nel caso di trapianti a ridotta compatibilità la prognosi cala drasticamente passando al 20-40%. Una soluzione potrebbe essere la diagnosi genetica preimpianto (PGD), che consentirebbe di impiantare gli embrioni sani e HLA compatibili per utilizzare le cellule emopoietiche staminali del cordone ombelicale in un possibile trapianto nel fratello affetto (Kuliev et al., 2005, Steffann et al., 2005). Restano però aperti i dibattiti sull'eticità del procedimento che, attualmente, non può essere effettuato in Italia.

La FA è una malattia che si presterebbe ad applicazioni di terapia genica. Le cellule staminali ematopoietiche possono essere utilizzate in un trapianto autologo dopo una “cura” in vitro dopo introduzione del gene normale. E’ verosimile pensare che le cellule corrette possano facilmente prendere il sopravvento sulle cellule FA e ripopolando il midollo favorire il recupero delle funzioni ematopoietiche (Mankad et al., 2006). Ci sono stati e ci sono tuttora in atto trials di terapia genica per pazienti con mutazioni *FANCA* e *FANCC* mediante utilizzo di vettori lentivirali (MacMillan et al., 2011, Tolar et al., 2011). Questo approccio terapeutico è purtroppo ostacolato da numerosi fattori. Primo fra tutti, il fatto che i pazienti FA non possiedano cellule staminali in numero sufficiente per la trasduzione *ex vivo*; pertanto, è necessaria la messa a punto di nuove tecniche che permettano l’espansione delle poche cellule ottenute a seguito del prelievo di midollo nei pazienti. In secondo luogo, le cellule staminali dei pazienti sono trasdotte con scarsa efficienza anche mediante utilizzo di lentivirus. Inoltre, le cellule corrette *ex vivo* rappresentano un bersaglio da parte del sistema immunitario, in quanto esprimono una proteina “estranea”. Infine, la terapia genica pone sempre di fronte al rischio d’insorgenza di tumori nel caso di integrazione del vettore contenente il gene corretto in prossimità di un protooncogene (Kee and D’Andrea, 2012). Per questi motivi, fino ad ora, l’applicazione della terapia genica per correggere la FA non ha avuto successo. Aspetti che riguardano la messa a punto di nuovi vettori e una corretta manipolazione delle cellule progenitrici del sangue sono in continua fase di studio. Non è da escludere che informazioni più dettagliate sulla funzione delle proteine FA e la sperimentazione mediante utilizzo di animali *knockout* possano, in futuro, portare allo sviluppo di protocolli efficaci di terapia genica.

Solo recentemente, la possibilità di riprogrammare cellule somatiche mature in cellule staminali pluripotenti indotte (iPS), ha reso possibile pensare ad una strategia terapeutica alternativa per le malattie caratterizzate da difetti nel midollo osseo quali la FA (Raya et al., 2009). Il vantaggio di questa tecnologia è evidente: essendo le cellule staminali ematopoietiche inadeguate, l’ipotesi è quella di riprogrammare cellule somatiche non ematopoietiche per generare iPS aploidiche da correggere *in vitro* (Robinton and Daley, 2012).

2. SCOPO DELLA TESI

SCOPO DELLA TESI

L'Anemia di Fanconi (FA) è una malattia difficile da diagnosticare sia per l'eterogeneità delle manifestazioni cliniche e la sovrapposizione del quadro clinico con quello di altre patologie, sia per l'eterogeneità genetica dovuta alla presenza di mutazioni in almeno sedici diversi geni. Le prime difficoltà, quindi, si presentano al momento della diagnosi e in seguito alla scelta della strategia volta ad identificare le mutazioni.

Il sospetto clinico può essere confermato o meno dal test al DEB. È importante ricordare che non sono rari i casi in cui il test è di dubbia interpretazione per la presenza di un mosaicismo emopoietico dovuto alla reversione del fenotipo in cellule del sangue.

Sebbene la ricerca di mutazioni sia agevolata dal fatto che il 65% dei pazienti presenti alterazioni di *FANCA*, lo screening di questo gene è un processo complesso sia per il numero considerevole di esoni codificanti sia per l'ampio spettro di mutazioni finora identificate, tra cui anche grosse delezioni intrageniche (in circa il 20% degli alleli mutati) che richiedono, per l'individuazione, analisi specifiche, spesso non di routine. Inoltre, anche se non di frequente, una serie di pazienti presenta inevitabilmente alterazioni in geni mutati meno frequentemente rispetto a *FANCA*. A meno che non si disponga dell'analisi di complementazione, questi pazienti non sono generalmente oggetto di screening e rimangono "orfani" della diagnosi molecolare.

Questa tesi rientra in un progetto più ampio che ha lo scopo definire e continuamente migliorare la strategia per la diagnosi molecolare nella FA, cercando di sfruttare sia le conoscenze fornite dalla ricerca di base sia lo sviluppo di nuove tecnologie. Si tratta di un progetto in continua evoluzione che richiede la collaborazione di molti clinici che fanno parte del gruppo di studio delle aplasie dell'AIEOP (Associazione Italiana Ematologia Oncologia Pediatrica), in particolare il Dott. Carlo Dufour con i suoi collaboratori, il Dott. Enrico Cappelli e Dott.ssa Johana Svhan, per la raccolta dei dati clinici, l'allestimento di colture cellulari e l'analisi di complementazione.

In modo più specifico il lavoro di questa tesi ha permesso di integrare l'attività di ricerca di mutazioni svolta dalle colleghe Daniela De Rocco e Elena Nicchia mediante sequenziamento Sanger e di nuova generazione, con la caratterizzazione degli effetti delle mutazioni stesse a livello di mRNA e proteine. E' stato, inoltre, valutato se lo studio di espressione delle proteine FA potesse sostituire l'analisi di complementazione nella definizione del gene candidato. L'analisi di complementazione è un processo complesso che richiede competenze specifiche e tempi di indagine lunghi. Pertanto, non sempre è applicabile con il rischio di analizzare inutilmente geni non mutati e lasciare il paziente senza una diagnosi molecolare.

La conoscenza delle mutazioni è fondamentale per una corretta presa in carico dei pazienti FA non solo per la conferma della diagnosi clinica (che non sempre può limitarsi al solo test al DEB) ma anche per l'identificazione dei portatori tra i familiari e per eventuali diagnosi prenatali. Inoltre, mentre da un lato lo screening genetico fornisce una risposta adeguata e certa a livello diagnostico, dall'altro, conoscere le mutazioni arricchisce le conoscenze sulla malattia, consente di adeguare lo screening alla popolazione analizzata e migliorare, di conseguenza, la strategia diagnostica.

3. MATERIALI E METODI

MATERIALI E METODI

3.1 PAZIENTI, COLTURE CELLULARI, DNA E RNA

Sono stati inclusi nella tesi pazienti provenienti dai centri dell'AIEOP (Associazione Italiana Ematologia Oncologia Pediatrica) con diagnosi sospetta di FA sulla base del test di fragilità cromosomica (DEB test). Tutti i soggetti hanno fornito un consenso informato allo studio e i protocolli sono stati approvati dal comitato etico della struttura che ha arruolato i pazienti. Il materiale biologico dei pazienti analizzati (DNA e linee cellulari linfoblastoidi e di fibroblasti) proviene principalmente dall'Ospedale Pediatrico Gaslini di Genova.

Le colture cellulari linfoblastoidi (in sospensione) e di fibroblasti (in adesione) sono mantenute rispettivamente in terreno RPMI 1640 1X (L-Glutamine) (Invitrogen/Gibco) e DMEM (Sigma) con aggiunta di siero fetale bovino al 10% (Gibco), a 37°C in atmosfera al 5% di CO₂. Dopo aver estratto DNA, RNA o proteine, le linee cellulari sono state congelate in azoto liquido aggiungendo al terreno il 10% di DMSO.

Il DNA genomico dei diversi pazienti è stato estratto sia da linee cellulari linfoblastoidi che da fibroblasti. I pellet cellulari sono risospesi con 5 ml di soluzione di lisi (NaCl 10 mM, Tris- HCl 10 mM a pH 7.5, SDS 0,6%, EDTA 10mM a pH 8) a cui è aggiunta 1mg/ml di proteasi K (20 mg/ml) prima di incubare le cellule lisate a 50°C O/N. Il giorno seguente si aggiungono 5 ml di una soluzione di NaCl 75 mM in etanolo 100% freddo. La soluzione è invertita più volte per permettere la precipitazione del DNA. Si procede con una centrifugazione per 10' a 13000 rpm, al termine della quale si elimina il surnatante e si aggiungono 500 µl di EtOH 100%. In seguito, si procede con una nuova centrifugazione, si elimina il surnatante e si risospende in acqua.

L'RNA totale è stato estratto dalle cellule utilizzando il reagente TRIzol® (Invitrogen, Life Technologies) in accordo con le indicazioni del produttore.

Per valutare qualità e concentrazione di DNA e RNA estratti, è stata eseguita una lettura allo spettrofotometro NanoDrop-1000 (Celbio), mentre per valutare l'eventuale degradazione sono stati caricati 2µl di campione su gel d'agarosio 1%.

3.2 REAZIONE DI AMPLIFICAZIONE (PCR)

Le reazioni di PCR sono state condotte utilizzando la KAPA2G FastHotStart (KAPA BIOSYSTEMS). Ogni reazione di PCR è stata sottoposta al seguente ciclo termico:

- Attivazione della polimerasi: 95°C per 1'
- 30 cicli di amplificazione suddivisi in:
 1. denaturazione: 95°C per 10"
 2. appaiamento: T^m per 10"
 3. estensione: 72°C 1" per ampliconi < 1kb (4" per dimensioni maggiori)
- Estensione finale dell'amplificato: 72°C per 30".

La presenza e la corretta lunghezza dei prodotti di amplificazione sono stati verificati attraverso elettroforesi su gel di agarosio (Seakem).

Lo screening del gene FANCA è stato condotto amplificando i 43 esoni del gene come riportato in Tabella 2.

ESONE	Tm	oligo senso 5'-3'	oligo antisenso 5'-3'	Lunghezza
1*	62 °C	1FN= ggaccaacatggccacgacg	1RN= cgaccggcggaggctctg	328
2*	62 °C	2FN=gtttcccagacctgcgcc	2RN=cgcccctcggtgttttc	361
3	58 °C	3FN= ggagcttgaatttttaaatgg	3RN= ttgctactacacacaccag	272
4+5	60 °C	4FN=gtataactctgggttcaggtg	5RN=gagaagcttggaattccc	701
6	62 °C	6FN2=gagaagcaagtgtggccaac	6RN=gcaatgcaatctagtctagtac	450
7	60 °C	7F_ART=tgggatttagttgaccttac	7R_ART=aagtgatggaacacttaa	221
8	62 °C	8F3=tgaagtggatggtctgtgcc	8RN=acacttggataaggacggct	241
9	60 °C	9F_NET=ttctctgtgtgatgcaggtatc	9R_NET=tgaccacagattcatgaggtat	151
10	60 °C	10F_NET=ttttgattaaggcctacagattg	10R_NET=cctcctctcacgcagttatcg	256
11	60 °C	11FT7=cacagtttgtgctgggattg	11RN=ctagaattcctggcatctcc	290
12	60 °C	12FN=caacttttgatctctgacttg	12RN=agacatcctgaacctctcg	150
13	60 °C	13FN=gtcacagctccatgtgttac	13 RN=tagggaaggcttactgag	400
14	60 °C	14FN=ggaatactgatcaccacg	14RN=aaagctgacagcaaggtgc	350
15	62 °C	15(2)FN =cactgcctaggtggccttc	15(2)RN=gctcagagcagatctgcagg	400
16+17	62 °C	16FN=agcactgtggatgttgaag	17RT7=gaggcaagaccagacatgag	500
18	60 °C	18FN=acaaggaagcacaggttgag	18 RN=catcagagcggagtctgcac	302
19	60 °C	19F_NET=gtgcacaagaagacttcataatg	19R_NET=agtccttgcttctacacaactg	243
20	60 °C	20F_NET=cttctctgtgttcagcatattc	20R_NET=agaagaacctggaagtagtcat	175
21	58 °C	21FN =caggctcatactgtacacag	21 RN=gagctcactcgggtggtg	304
22	62 °C	22FN=cgtctgccacacacagcagc	22RN=agccacagagctccaaccac	300
23	58 °C	23FN=gatgtgtgcagtcagcag	23 RN=ctaagcctctgcctaattgg	350
24	60 °C	24F_NET=ccttctgctgctcccgtcc	24R_NET=cagacttgcccagcaagag	306
25	60 °C	25F_NET=ccgctggtggttgattagctgt	25R_NET=tttcagggcactgaagacgaat	260
26	60 °C	26FN=ttcagtccctggaaaattc	26RN=atcaaacgagcatgtgtcact	313
27	60 °C	27FN=ccatccagttcggatgcact	27RN=agctgcgtaaacctgaaacg	286
28	62 °C	28FN=tgctcagccactcacgtg	28RN=atgatgcagggaaggaacg	350
29	62 °C	29F_NET=gacatggaggactgcgtatgag	29R_NET=gtggctgtgatgactggaactg	260
30	62 °C	30FN=cagctctcactagtgtgtg	30 RN=cctaagctcaattagatgggc	350
31	62 °C	31FN=aggagaactgggaacttcag	31RN=cacgcggcttaaatgaagtg	224
32	60 °C	32FN=cctgacccctgacgtctg	32RN=ctaggaccgtcatgagatgct	350
33	60 °C	33FN=ctgctctgagagtctgagc	33 RN=gctgctgttagccacag	350
34+35	60 °C	34FN=aggagcaccatccatggtg	35RN=agatggtaacaccctgatg	600
36	62 °C	36FN =ctgtagtggcctgtaggagca	36RN=ctcacacgagaggctgcc	350
37	60 °C	37 FN=gcacacctgattctgatgg	37 RN=gacaagcccagagaatagc	400
38	60 °C	38F2=tgtgaaatgaccagccag	38R2=tactcacacttccgcaaac	400
39	60 °C	39FN=tgccctcttctgtctgag	39 RN=caaggaacctcaaggagg	300
40+41	62 °C	40FN=cagggtacacgtcccacc	41RN=ccatagtctgatgctgt	650
42+43	58 °C	42FN=acagcatgcagactatgg	43RN=cactaaagcagtcaggagaga	600

Tabella2. Sequenze dei primer intronici e relative temperature di appaiamento utilizzati per l'amplificazione del gene *FANCA*. *esoni che richiedono l'aggiunta di DMSO alla mix a causa dell'elevata concentrazione di citosine e guanine nella sequenza.

Lo screening del gene *FANCG* è stato effettuato dividendo l'intero gene in 7 frammenti. I primer utilizzati per amplificare ogni frammento, le temperature di appaiamento e la lunghezza di ogni amplificato sono elencati in tabella 3.

F	ESONI	Tm	oligo senso 5'-3'	oligo antisenso 5'-3'	Lunghezza
1	1+2	60 °C	1F=ccttcctaagtccgcttctg	1R=cactccagtctcctctgtg	552
2	3+4	60 °C	2F=gggtgatctctgagttggg	2R=catgatgatgctactgcagc	757
3	5	60 °C	3F=gcctgtgtcctcaaggatc	3R=ggaggaccacttaaggtag	302
4	6+7+8	60 °C	4F=cattcagttccatgggcttc	4R=ctgccactgcctcattcatc	871
5	9+10+11	58 °C	5F=ccatctgtagctaagagac	5R=cacagtcctttctttaag	950
6	12	58 °C	6F=ggttcccttaatgtgtccc	6R=tgaacacctggatataatccc	451
7	13+14	62 °C	7F=acacagtggtatgcaggc	7R=gatggtgaagcagaaagccc	736

Tabella 3. Sequenze dei primer intronici e relative temperature di appaiamento utilizzati per l'amplificazione del gene *FANCG*.

Lo screening del gene *FANCB* è stato condotto amplificando i 9 esoni del gene come riportato in Tabella 4.

ESONE	Tm	oligo senso 5'-3'	oligo antisenso 5'-3'	Lunghezza
1 non-c	60 °C	1F= ctctgctttcacaactcc	1R= ggaatgacagaccaggtc	350
2 a	56 °C	2aF= ttagttacatgtgttag	2aR= gaagaatgctttgatgcc	600
2 b	56 °C	2bF= catgtgtgcaactgtgtg	2bR= gacaataagactccagaatg	800
3	53 °C	3F= ttacaaatgacaactacatg	3F= ccaccaatatttaaatgatag	400
4	56 °C	4F= gggaaactgttctgactgg	4F= aagcaccaagttgcttatac	300
5	58 °C	5F= actgttttgcctgagaag	5F= tcttgatccactgagcagc	400
6	56 °C	6F= cataagccctagccaaaagc	6F= agttgaaagatcttcatgac	300
7	54 °C	7F= ttttactgtgtgtatgac	7F= agttgaaagatcttctagac	500
8	58 °C	8F= ttgttgactgtactgctac	8F= cgctgttgaaccacacattg	500
9	58 °C	9F= gattactgcagtatattgtac	9F= ctactacagtaagcctcgg	600

Tabella 4. Sequenze dei primer intronici e relative temperature di appaiamento utilizzati per l'amplificazione del gene *FANCB*.

Lo screening del gene *FANCC* è stato condotto amplificando i 15 esoni del gene. I primer utilizzati per amplificare ogni frammento, le temperature di appaiamento e la lunghezza di ogni amplificato sono elencati in tabella 5.

ESONE	Tm	oligo senso 5'-3'	oligo antisenso 5'-3'	Lunghezza
1	59 °C	1F=agatggctccagcatagaag	1R=cggcacttcagtcaatacc	482
2	60 °C	2F=agaagcattcacgttccctc	2R=catcggcacttcagtcaatac	255
3	60 °C	3F=gaagcattcaggttcctt	3R=gttcataatgtaagcctctg	220
4	58 °C	4F=ctgtgaacttagtagtttgag	4R=caagataatgtaacagtgaag	351
5	60 °C	5F=gtaggcattgtacataaaagg	5R=acatctcttctggaggactg	372
6	58 °C	6F=actgatgtaacctgtttgc	6R=cataactgaacatccatttcc	240
7	59 °C	7F=gtgttcacgatgggttag	7R=gtctttccaacacaccacag	344
8	60 °C	8F=actgctgtgaccatgatgag	8R=tccaagcatctcctcaagg	330
9	58 °C	9F=cagtacagcagtggttcg	9R=cacaacattttctgattcatg	283
10	60 °C	10F=gagttgagatactgctgaag	10R=cttaaccttctgtggggcac	378
11	62 °C	11F=gttctgaccatgttagtactg	11R=tgggaagccaagcagactc	359
12	60 °C	12F=gtgaaccagaagtaagggc	12R=cagtaatgccttctctgtc	396
13	62 °C	13F=ggttgacttcagtgataag	13R=caggtgtcatggaagccaag	401
14	62 °C	14F=gactgcttggtgagcaagac	14R=cagctgtcaacctcttgac	550
15	62 °C	15F=ggatagggcttcttcaggg	15R=acaggtcatcacctgtctg	386

Tabella 5. Sequenze dei primers intronici e relative temperature di appaiamento utilizzati per l'amplificazione del gene *FANCC*.

Lo screening del gene *FANCD2* è stato condotto amplificando i 44 esoni del gene. Per brevità, si riportano in Tabella 6 le sequenze dei primer utilizzati per amplificare le regioni interessate da mutazioni nella nostra casistica, le temperature di appaiamento e la lunghezza di ogni amplificato.

ESONE	Tm	oligo senso 5'-3'	oligo antisenso 5'-3'	Lunghezza
7	52 °C	7F=caaagtgtttattccaggagc	7R=catcagggtactttgaacattc	342
12	55 °C	12F=tcatggtagagagactggactgtgc	12R=accctggagcaaatgacaacc	432
13	55 °C	13F=atggcaggaactccgatctt	13R=ccttattcagcatgcctgt	363

Tabella 6. Sequenze dei primer intronici e relative temperature di appaiamento utilizzati per l'amplificazione del gene *FANCD2*.

3.3 SEQUENZIAMENTO E ANALISI DELLE SEQUENZE

Dopo l'amplificazione, i frammenti sono stati purificati utilizzando 2µl di Exo/SAP e 5 µl di prodotto di PCR. La mix è portata a 35°C per 15' per attivare gli enzimi e poi a 85°C per 15' per inattivarli. I prodotti di PCR una volta purificati sono stati sottoposti a marcatura inserendoli in una mix costituita da: 7 µl di DNA, 0,6 µl di Big Dye, 0,5 µl di

primer For o Rev (10 μ M) e 3,9 μ l di H₂O per un totale di 12 μ l. I primer utilizzati sono specifici per ogni frammento da sequenziare. In seguito, sono state eseguite le reazioni di sequenza:

- 96°C per 1'
- 26 cicli costituiti da:
 1. 96°C per 10"
 2. T_m° primer per 5"
 3. 60°C per 2'

I prodotti della marcatura sono stati purificati utilizzando le colonnine MicroSpin G-50 Columns (GE Healthcare). La corsa elettroforetica è stata portata a termine in un sequenziatore automatico AB3130xl (Applied biosystems).

L'analisi delle sequenze ottenute dal sequenziamento diretto è stata condotta utilizzando il programma SeqMan del pacchetto Lasergene (DNASTAR). Per l'analisi degli SNP è stata usata come riferimento la banca dati dbSNP, gestita e fornita dall'NCBI che identifica e cataloga SNPs di genomi di vari organismi tra cui l'uomo (<http://www.ncbi.nlm.nih.gov/projects/SNP/>).

3.4 REAZIONE DI RETROTRASCRIZIONE E RT-PCR

Per la sintesi del cDNA a partire a 1 μ g di RNA totale è stato utilizzato il kit "Transcriptor First Strand cDNA Synthesis Kit" (Roche) in accordo con le indicazioni del produttore. Per la reazione di retrotrascrizione sono stati utilizzati sia oligo dT (2,5 pmol/ μ l) che Random esameri (60 pmol/ μ l) ed il seguente ciclo termico: 25°C per 10' > 50°C per 60' > 85°C per 5'.

Il cDNA ottenuto è stato amplificato come descritto precedentemente utilizzando coppie di primer specifiche disegnate sulla porzione codificante dei geni (Tabella 7).

GENE	MUTAZIONE	oligo senso 5'-3'	oligo antisenso 5'-3'	Tm (°C)
FANCA	c.709+1G>A	6F= GTGTGGCATCTTCACGTACA	10R= GATCTTGTGAGTGGAGGACT	60
	c.710-5T>C c.790C>T c.793-1G>C	7F= GACATGCATGCTGTGGGATC	10R= GCACCTCACGATCTTGTGAG	60
	c.1645C>T	14F= TTTGATGGCCCAGGCATTCG	22R= AGAGTTGGGTTCTGCCCTC	60
	c.1359+1G>C	10F= CACTCACAAGATCGTGAGG	16R= GTGTCTTGGCCAATGAGATG	60
	c.1776+7A>G	15F= CACACGAGGCTACCATGGC	20R= GAGACTCTATAAACGCCACA	60
	c.2778+1G>A	21F= GCTCTGCTGCTGAAGAGAAG	30R= AAATCCATCAGTGCCTTGACA	60
	c.3348+1G>A	31F= TGGCCGCACAGGAAATGAG	35R= CCAGTATGAAGTCGACCATC	60
	c.4011-1G>C	37F= CTAAGAAGCTGGACTGCGA	43R= TATTACACGGGAGCTGGGC	60
FANCG	c.924+1G>A	3GF= GGAGCTGACTGTCACCTGC	11GR= TTCCTCAGGTGTGGCCCG	60
	c.1144-1G>T	9GF= ACGCTGCAGAGCATTACTTG	11GR= TTCCTCAGGTGTGGCCCG	60
	c.1715G>A	12GF= GATGTAAGTCAGATGCGGC	14GR= AAGACTTTGGCAGAGATGTC	60

Tabella 7. Sequenze dei primer esonici e relative temperature di appaiamento utilizzati per l'amplificazione del cDNA dei geni *FANCA* e *FANCG* nelle reazioni di RT-PCR.

3.5 ANALISI BIOINFORMATICA DELLE MUTAZIONI MISSENSO E DEI SITI DI *SPLICING*

L'effetto delle mutazioni missenso è stato valutato utilizzando quattro programmi bioinformatici di predizione della patogenicità:

- PoliPhen-2 (<http://genetics.bwh.harvard.edu/pph2/>);
- Mutation Taster (<http://www.mutationtaster.org/>);
- Mutation Assessor (<http://mutationassessor.org/>);
- SIFT (<http://sift.jcvi.org>).

Considerando che ogni *software* utilizza differenti scale di predizione, gli *output* sono stati convertiti in un valore da 0 a 2. In particolare, è stato assegnato il valore 0, 1 e 2 alle predizioni di PolyPhen-2 corrispondenti a "benign", "possibly damaging", e "probably damaging"; è stato assegnato il valore 0 e 2 agli output "polymorphism" e "disease causing" di Mutation T@aster, rispettivamente; alle predizioni "neutral", "low", e "medium" di Mutation Assessor sono stati assegnati rispettivamente i valori di 0, 1 e 2; infine, per un *output* "tolerated" in SIFT è stato assegnato il valore di 0 e per "not tolerated" il valore di 2. Sommando tutti i valori di *output* ottenuti dai quattro diversi

software è stato calcolato punteggio di patogenicità per ogni mutazione (“pathogenicity score”).

L’analisi bioinformatica delle mutazioni che coinvolgono i siti di *splicing* è stata eseguita con l’ausilio di due *software*: Human Splicing Finder Version 2.4.1 (<http://www.umd.be/HSF/>) e Splice Site Prediction by Neural Network (NNSplice; http://www.fruitfly.org/seq_tools/splice.html).

3.6 IMMUNOBLOTTING

3.6.1 Estrazione delle proteine e determinazione della concentrazione

L’estrazione del lisato proteico totale e delle frazioni citoplasmatiche e nucleari è stata effettuata, rispettivamente mediante dei reagenti di estrazione proteica M-PER (Mammalian protein extraction reagent, Thermo Scientif-Pierce) e NE-PER (Nuclear and Cytoplasmic extraction reagent, Thermo Scientif-Pierce) con l’aggiunta di un cocktail di inibitori di proteasi (Sigma). L’analisi quantitativa degli estratti è stata effettuata mediante il kit “BCA Protein Assay” (Pierce®), utilizzando lo spettrofotometro Nanodrop ND-1000 (Celbio) per la determinazione delle concentrazioni.

3.6.2 Separazione delle proteine e trasferimento sul filtro di nitrocellulosa

La separazione delle proteine avviene sulla base peso molecolare in gel *precast* di poliaccrilammide denaturante. Sono stati utilizzati gel al 7,5% per l’immunoblotting di FANCD2 e a gradiente 4-20% per FANCA (Bio-Rad). Come marcatore per seguire la corsa elettroforetica è stato utilizzato il Precision Plus Protein Western C Standards (Bio-Rad). Prima di procedere alla separazione, gli estratti proteici sono stati diluiti in un buffer di caricamento costituito da 8,7% β -mercaptoetanol, 52,17% SDS, 0,217M TrisHCl pH 8, 0,026% blu di bromofenolo, 17,4% glicerolo e bolliti per 5 minuti a 95°C. In ogni pozzetto

del gel sono stati caricati 30 µg di lisato proteico ed è stato applicato un voltaggio non superiore a 250 V per il tempo necessario alla separazione delle proteine di interesse. Al termine, le proteine separate nel gel sono state trasferite su filtro di nitrocellulosa (Mini Nitrocellulose Transfer Packs, Bio-Rad) utilizzando il Trans-Blot® Turbo™ Transfer System (Bio-Rad) con programma preimpostato per il trasferimento di proteine ad alto peso molecolare.

3.6.3 Rilevazione delle proteine di interesse

I filtri di nitrocellulosa sono stati bloccati per 1 ora a temperatura ambiente in una soluzione di “blocking” contenente latte 5% e TBST (Tris-buffer salino 1X, Tween20 0,1%). Dopo tre lavaggi con TBST, i filtri sono stati incubati con l’anticorpo primario opportunamente diluito nella soluzione di *blocking*, per un’ora a temperatura ambiente. Come anticorpi primari sono stati utilizzati un anticorpo policlonale che riconosce la porzione C-terminale di FANCA e anticorpi monoclonali diretti contro ORC4 (1:500, ab68348, Abcam), HSP90 (1:4000, sc-13119, Santa Cruz), β-actina (1:4000, sc-4778, Santa Cruz) e la porzione N- terminale di FANCD2 (1:500, sc20022, Santa Cruz). Dopo tre lavaggi la membrana è stata incubata con l’anticorpo secondario appropriato, contro immunoglobuline di coniglio (primario policlonale) e di topo (primario monoclonale) coniugati con HRP (horseradish peroxidase) (Santa Cruz) diluiti 1:2000 in *blocking solution*, per 1 ora a temperatura ambiente.

Allontanato l’anticorpo secondario in eccesso, il filtro è stato sviluppato con il substrato chemiluminescente Super Signal® West Pico (Pierce), molto sensibile alla HRP. La luce emessa è stata infine rilevata dal ChemiDoc mediante il programma Quantity one (Bio-Rad).

3.7 MULTIPLEX LIGATION-DEPENDENT PROBE AMPLIFICATION (MLPA)

Per l’analisi è stato utilizzato il kit SALSA MLPA P031/P032-A2 FANCA sviluppato dalla MRC-Holland. Per ogni reazione sono stati utilizzati 100 ng di DNA. In accordo con

il protocollo del kit, dopo un'iniziale denaturazione di 5' a 98°C, il DNA è stato sottoposto ad un rigoroso ciclo termico. Sono stati aggiunti 1,5 µL di SALSA probemix e 1,5 µL di buffer MLPA con termociclatore a 25°C a cui è seguito 1 minuto a 54°C. Poi, i campioni sono stati lasciati per 16 ore a 60°C per permettere l'ibridizzazione delle sonde. La ligazione si ottiene abbassando la temperatura a 54°C e aggiungendo 32 µL di ligase-65 mix. Seguono un'incubazione per 15' a 54°C e il riscaldamento per 5' a 98°C. L'ultima fase è rappresentata dalla PCR, che si ottiene in termociclatore a 60°C aggiungendo a 10 µl dei prodotti di ligazione, 4 µl di SALSA PCR buffer e 26 µl di acqua in un volume finale di 30 µl. La reazione prevede 30 cicli suddivisi in: denaturazione del DNA a 95°C per 30", appaiamento del primer universale a 60°C per 10" ed estensione a 60°C per 30". Infine, è stato eseguito il completamento dell'estensione a 72°C per 20'. I campioni sono stati sottoposti ad una corsa di elettroforesi capillare con il sequenziatore AB3130xl e, i dati generati sono stati importati nel software GeneMapper v.4 per l'analisi dei risultati e l'analisi statistica.

4. RISULTATI

RISULTATI

4.1 STRATEGIA DIAGNOSTICA PER LA DIAGNOSI MOLECOLARE DELL'ANEMIA DI FANCONI

Diverse possono essere le strategie che permettono di formulare una diagnosi molecolare della FA. Come accennato negli obiettivi, la diagnosi è ostacolata da numerosi fattori che devono necessariamente essere presi in considerazione nel proporre e validare un algoritmo di screening efficace.

Il nostro percorso, maturato dall'esperienza degli ultimi anni, prevede l'arruolamento di pazienti con diagnosi sospetta sulla base delle evidenze cliniche (Figura 8).

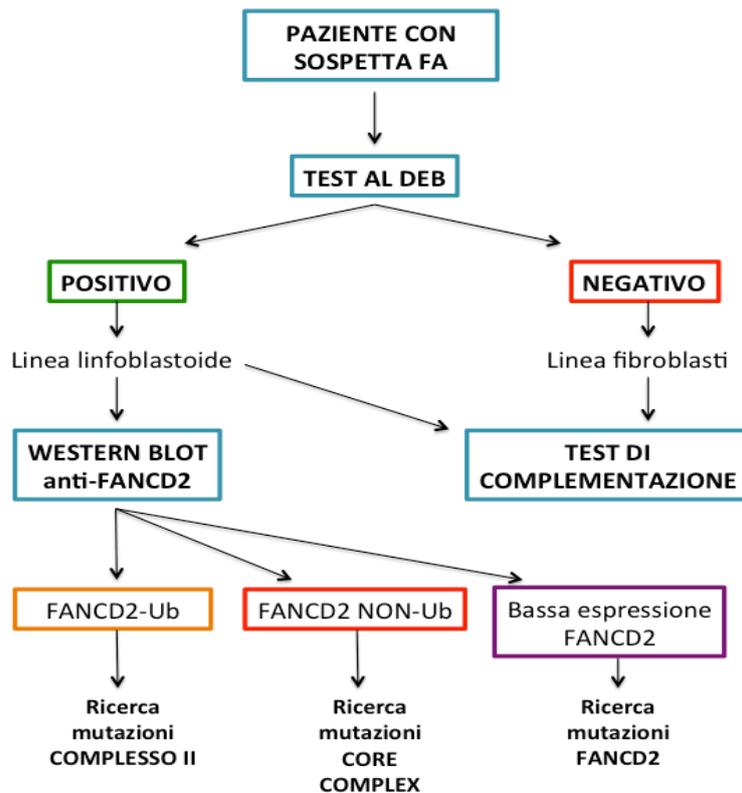


Figura 8. Rappresentazione schematica della strategia di diagnosi da noi utilizzata. Le cellule del paziente sospetto FA sono sottoposte al test al DEB per confermare la diagnosi. Se il test è negativo si

ipotizza la presenza di mosaicismo e si procede con il test di complementazione a partire da fibroblasti primari. In caso di test positivo invece su linea cellulare linfoblastoide si eseguono la complementazione e l'analisi delle proteine mediante Western blot per determinare il gene coinvolto. Ottenuta l'indicazione in merito al gene mutato, si procede con lo screening del gene per la ricerca delle mutazioni con ausilio di eventuale MLPA per la ricerca delle delezioni intrageniche che interessano il gene FANCA.

In questo caso, i linfociti ottenuti da un prelievo di sangue periferico vengono, prima di tutto, sottoposti al test al DEB, per verificarne la fragilità cromosomica. Se il test è positivo, si procede con l'allestimento di una linea cellulare linfoblastoide per ottenere il materiale biologico utile alla ricerca del gene candidato mediante analisi di complementazione e delle proteine. Al contrario, se il risultato è incerto, si sospetta che il paziente sia un mosaico e si procede all'allestimento di fibroblasti primari con i quali condurre le stesse indagini proposte condotte su linfociti (test al DEB) e linfoblasti (analisi di complementazione e delle proteine).

Il test di complementazione, condotto in collaborazione con il Dr. Dufour (IRCCS G. Gaslini – Genova), prevede l'utilizzo di vettori retrovirali in ciascuno dei quali è stato inserito il cDNA *wild type* dei diversi geni FA. Dopo l'infezione e l'esposizione a diverse concentrazioni di DEB o MMC, si valuta la percentuale di sopravvivenza delle cellule. Una curva di sopravvivenza analoga a quella delle cellule controllo indica che il cDNA esogeno inserito corrisponde al gene mutato nella linea cellulare del paziente in quanto riesce a "complementare", permettendo così alle cellule di acquisire un fenotipo normale. Se invece la curva di sopravvivenza non cambia rispetto alle cellule FA *wild type* significa che il cDNA introdotto "non complementa" e che il gene alterato nella linea cellulare è diverso da quello trasdotto (Figura 9). Di routine, è eseguita l'analisi di complementazione utilizzando, in successione, i vettori corrispondenti ai geni FANCA, FANCG e FANCC ovvero quelli più frequentemente mutati nei pazienti FA.

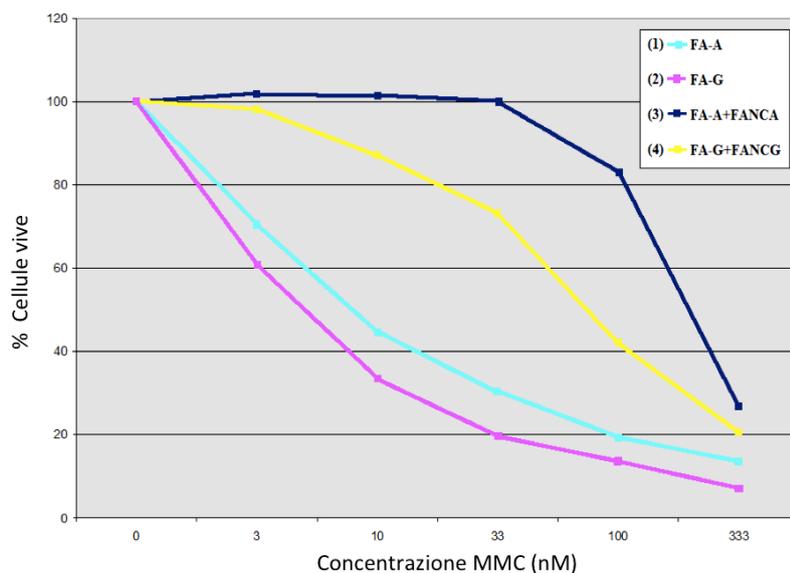


Figura 9. Curve di sopravvivenza alla MMC delle linee linfoblastoidi dei pazienti FA42 e FA98. (1) Linea cellulare di FA42 in cui sono state identificate due mutazioni di *FANCA*; (2) Linea cellulare del paziente FA98 in cui sono state identificate due mutazioni di *FANCG*; (3) Linea cellulare del paziente FA42 complementata con il gene *FANCA*; (4) Linea cellulare del paziente FA98 complementata con il gene *FANCG*; In ascissa: concentrazione crescente di mitomicina C; In ordinata: tasso di sopravvivenza delle cellule in coltura misurata mediante citofluorimetria.

L'analisi di complementazione è complessa: i tempi di indagini sono relativamente lunghi e richiede una formazione adeguata sull'utilizzo di vettori retrovirali. Per questo motivo, abbiamo valutato se l'analisi delle proteine potesse rappresentare un approccio alternativo alla complementazione. Si presuppone, infatti, che in presenza di mutazioni le proteine codificate, oltre ad essere non funzionali, siano rapidamente degradate e quindi instabili. Inoltre, l'analisi di FANCD2 consente, valutandone lo stato di ubiquitinazione, di determinare se il gene candidato si trova a monte o a valle di questo elemento chiave del pathway FA. Infatti, se FANCD2 non è monoubiquitinata (FANCD2-nonUb), possiamo supporre che sia mutato uno dei geni del complesso I (*FANCA*;-*B*;-*C*;-*E*;-*F*;-*G*;-*L*;-*M*), se invece lo è (FANCD2-Ub), l'alterazione è da ricercare nei geni *FANCI*, *BRCA2*, *FANCI*, *SLX4*, *RAD51C* o *ERCC4* che agiscono a valle nel pathway.

A titolo di esempio, sono illustrati quattro diversi casi di immunoblotting eseguito su estratti proteici di linee linfoblastoidi (Figura 10). Nei controlli (WT) sono presenti le bande che corrispondono a FANCA e ad entrambe le forme, monoubiquitinata e non monoubiquitinata, di FANCD2. Nella linea FA85, invece, l'assenza della proteina FANCA e

la presenza della sola forma non monoubiquitinata di FANCD2 è indice di un'alterazione a livello del gene *FANCA*. Nel paziente FA86, l'immunoblotting evidenzia l'espressione di FANCA ma non quella di FANCD2-Ub, indicando che una delle subunità (esclusa FANCA almeno teoricamente; vedi sotto) del complesso I è alterata. Una situazione diversa si presenta per il paziente FA91 in cui, essendo espressa correttamente FANCA ma essendo FANCD2 a livelli molto bassi (Kalb et al., 2007), si ha l'indicazione di mutazioni in *FANCD2*. In FA2 si osserva una situazione paragonabile ai campioni di controllo. Mentre il risultato del test al DEB era indubbiamente positivo nei primi tre casi, in FA2 il test non era conclusivo, nonostante un quadro clinico compatibile con quello FA. Sospettando, pertanto, una condizione a mosaico di reversione del fenotipo è stata fondamentale l'analisi dei fibroblasti che ha permesso di assegnare questo paziente al gruppo FA-A.

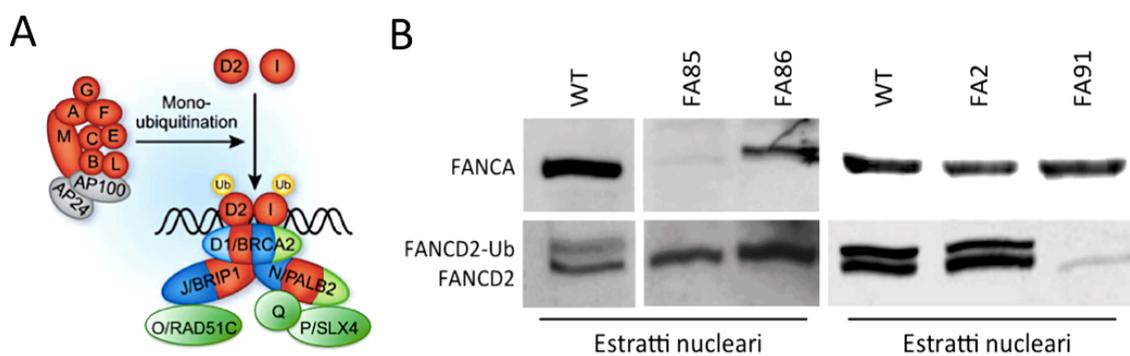


Figura 10. Immunoblotting per l'identificazione del gene candidato. (A) Rappresentazione schematica del pathway FA in cui l'integrità del complesso I è indispensabile per la monoubiquitinazione di FANCD2 e l'attivazione della proteina a valle del pathway FA (Levy-Lahad, 2010). (B) Esempi di western blot eseguiti su estratti proteici ottenuti da linee cellulari linfoblastoidi di due controlli (WT) e quattro individui affetti. I pazienti FA85 e FA86 hanno mutazioni nei geni rispettivamente *FANCA* e *FANCB*, subunità del complesso I; FA91 invece è mutato in *FANCD2*. Il paziente FA2 pur manifestando un pattern proteico simile ai controlli, è stato ulteriormente indagato come un potenziale caso di mosaicismo, permettendo di identificare mutazioni in *FANCA*.

Come accennato nell'introduzione, i pazienti cosiddetti "mosaico" rappresentano circa il 20% dei casi; presentano un mosaicismo associato a reversione del fenotipo cellulare, a livello emopoietico che può influire, in base al grado, anche sensibilmente sull'interpretazione del test al DEB (Waisfisz et al., 1999). La corretta diagnosi di questi pazienti riveste un ruolo importante non solo in termini di consulenza genetica e diagnosi prenatale, ma anche per quanto riguarda le strategie terapeutiche.

All'identificazione del gene candidato segue lo screening di mutazioni che, nel caso di *FANCA*, prevede anche un'analisi con MLPA (multiplex ligation-dependent probe amplification). Infatti, il 20% degli alleli *FANCA* sono caratterizzati da ampie delezioni del gene (Castella et al., 2011) che, se in emizigosi, sono impossibili da riconoscere mediante amplificazione del DNA sequenziamento.

Se non ci sono indicazioni circa il gene candidato, la strategia diagnostica prevede di procedere direttamente con l'analisi di *FANCA*, il gene maggiormente mutato, cui seguirà lo screening di *FANCG* e *FANCC*. Lo screening di questi tre geni consente di identificare le mutazioni nell'85% dei pazienti FA.

4.2 ANALISI MOLECOLARE DELLA CASISTICA ITALIANA

Nel corso degli ultimi tre anni, l'applicazione della strategia diagnostica descritta nel paragrafo precedente (Figura 8) alle famiglie FA arruolate nell'ambito dell'AIEOP (Associazione Italiana di Ematologia e Oncologia Pediatrica), ha consentito l'identificazione delle mutazioni in un considerevole numero di pazienti che, aggiungendosi a quelli diagnosticati in precedenza, ha portato a 100 il numero delle famiglie FA con diagnosi molecolare (De Rocco et al., 2014).

Dall'analisi è emerso che la maggior parte dei probandi (n=85) hanno mutazioni in *FANCA*. Nelle altre 15 famiglie sono state invece identificate alterazioni di *FANCG* (n= 9), *FANCC* (n= 3), *FANCD2* (n= 2), e *FANCB* (n= 1) (Tabella 9; Figura 11, A). È interessante notare l'elevata percentuale di eterozigoti composti sul totale dei pazienti a suggerire che, forse, la FA possa essere più frequente nella popolazione di quanto riportato dalla letteratura (Figura 11, B). Di tutte le varianti identificate, 108 risultano essere potenzialmente patogenetiche: 55 sono già presenti nel database delle mutazioni FA, consultabile all'indirizzo <http://www.rockefeller.edu/fanconi/>, e 53 sono nuove varianti non riportate nelle banche dati degli SNP.

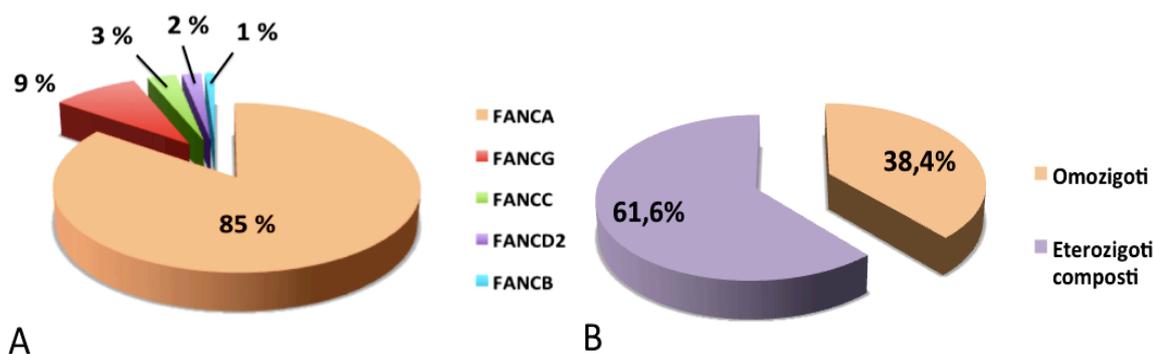


Figura 11. (A) Distribuzione delle famiglie FA arruolate in Italia (n=100) tra i diversi gruppi di complementazione. (B) Percentuale dei pazienti omozigoti ed eterozigoti composti sul totale dei pazienti FA identificati (n=99). Dal totale è stato escluso il paziente appartenente al gruppo di complementazione FA-B, emizigote per una mutazione nel gene *FANCB*.

4.2.1 Mutazioni del gene FANCA

Considerando l'insieme dei pazienti appartenenti al gruppo di complementazione FA-A, 27 sono omozigoti e 58 sono eterozigoti composti per 91 diverse mutazioni identificate in 162 dei 170 alleli *FANCA* mutati attesi. In 8 famiglie, non è stato individuato il secondo allele mutato (Figura 12). Si riscontrano tutte le categorie di mutazioni: ampie delezioni intrageniche (n=24), piccole delezioni/inserzioni *in frame* (n=11) o *frameshift* (n=26), mutazioni missenso (n=37) mutazioni nonsense (n=37) e mutazioni di *splicing* (n=27) (Figura 12) (De Rocco et al., 2014). A sostegno del fatto che la maggior parte delle mutazioni sono private, 45 delle 91 mutazioni identificate non sono mai state descritte prima.

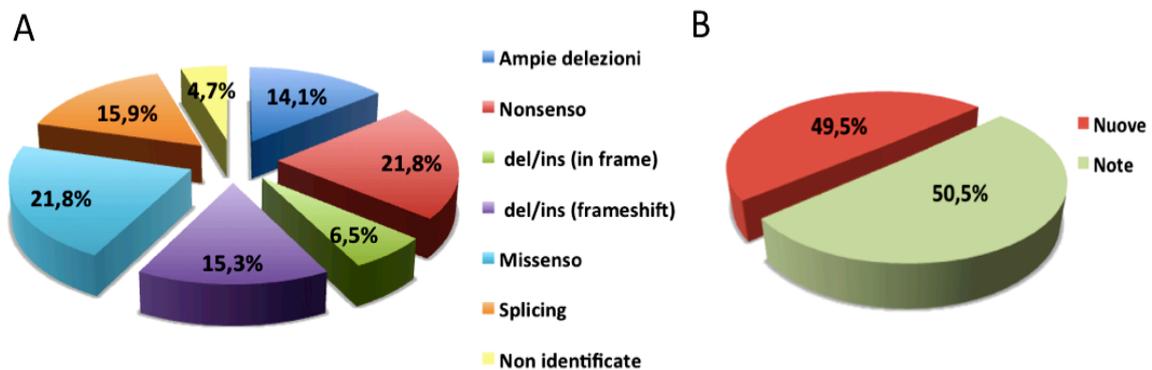


Figura 12. (A) Distribuzione dei diversi tipi di mutazioni identificate nei 170 alleli *FANCA*. In 8 alleli non è stata identificata alcuna mutazione; (B) Percentuale di mutazioni nuove e note sul totale delle diverse mutazioni identificate nel gene *FANCA* (n=91). L'elenco delle mutazioni FA è disponibile all'indirizzo <http://www.rockefeller.edu/fanconi/>.

4.2.2 Caratterizzazione delle mutazioni

L'identificazione di una variante genica spesso non è sufficiente ad attribuirne la patogenicità. Per questo motivo, in alcuni casi, come per delezioni intrageniche, le alterazioni di *splicing* e le sostituzioni aminoacidiche, sono stati condotti ulteriori studi volti alla caratterizzazione di tali mutazioni.

Ampie delezioni intrageniche. L'assenza di prodotti specifici di amplificazione o l'analisi MLPA, ci ha permesso di identificare 16 diverse delezioni intrageniche di *FANCA*, tra cui 5 riarrangiamenti mai descritti precedentemente. Oltre alle delezioni che interessano l'intero gene (Centra et al., 1998), all'interno della nostra casistica abbiamo rilevato delezioni che variano in ampiezza da un singolo esone (esoni 4 o 15) fino alla rimozione di 26 esoni (Tabella 9). A titolo di esempio si riporta il caso del paziente FA12, in cui la mancata amplificazione dell'esone 15 aveva fatto sospettare la presenza di una delezione in omozigosi (Figura 13). Non sono rari i casi in cui si osserva il cosiddetto "allelic drop-out". In questi casi, la presenza di rari SNP nelle regioni di appaiamento dei primer impedisce la reazione di amplificazione di uno specifico prodotto. Per confermare la delezione della regione contenente l'esone 15 quindi, abbiamo amplificato con i primer intronici 14F e 17R il DNA gnomico. Mentre nel controllo, come atteso, si ottiene un prodotto di 9 kb, in FA12 la banda è di circa 7 kb, a sostegno della presenza di una delezione di circa 2 kb (Figura 13, A). Entrambi i prodotti sono evidenti nel fratello (Fra, Figura 13, A) del probando che risulta pertanto un portatore sano della delezione. Sebbene i punti di rottura non siano ancora stati caratterizzati, è probabile che siano avvenuti in corrispondenza delle numerose ripetizioni *Alu* presenti nelle porzioni introniche del gene *FANCA* (Morgan et al., 1999). Disponendo della linea cellulare FA12, la delezione dell'esone 15 è stata, poi, confermata a livello dell'RNA. Mediante RT-PCR in FA12 si produce una banda di 431 bp, in cui l'analisi di sequenza (Figura 13, C e D) dimostra l'assenza dell'esone 15. La stessa delezione è stata poi caratterizzata allo stesso modo nelle famiglie FA7 e FA40.

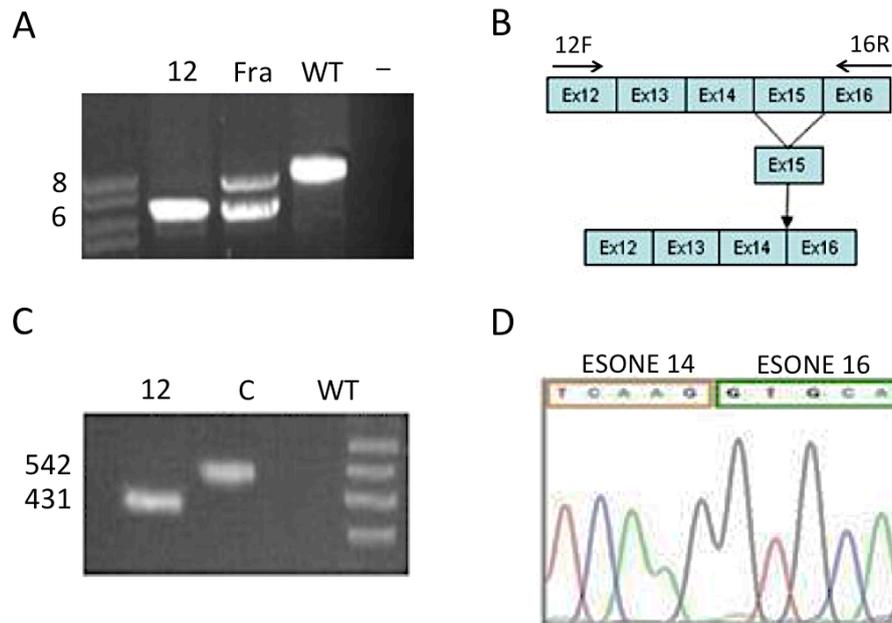


Figura 13. Identificazione di una delezione intragenica nella famiglia FA12. (A) Identificazione a livello genomico della delezione dell'esone 15. La PCR è stata condotta su gDNA estratto da linee cellulari linfoblastoidi utilizzando primer intronici 14F e 17R. Il probando (12) presenta una banda di circa 7kb, presente nel fratello in eterozigosi (Fra), e assente nel controllo *wild type* (WT); (B) Schema dell'organizzazione degli esoni nel cDNA e localizzazione dei primers utilizzati per visualizzare la delezione dell'esone 15; (C) RT-PCR utilizzando RNA estratto dalla linea cellulare che rivela una banda di 430 bp nel paziente e un'altra di 542 bp in un controllo *wild type* (C WT); (D) Elettroferogramma della sequenza diretta della banda di 431 estratta e purificata dal gel in 12, che evidenzia l'assenza dell'esone 15.

Mutazioni di splicing. Tutte le varianti introniche (n=22) sono state considerate come possibili cause di alterazioni a livello dello splicing. Delle 22 varianti identificate, 12 sono state caratterizzate funzionalmente considerando il loro effetto sulla maturazione dell'RNA. Oltre a 4 mutazioni descritte in lavori precedenti (c.283+3A>C; c.893+1G>T; c.1567-20A>G, and c.2504+2T>C) (Savino et al., 2003), altre 5 alterazioni (c.710-5T>C, c.1359+1G>C, c.1776+7A>G, c.2778+1G>A, and c.3348+1G>A) sono state associate rispettivamente con lo *skipping* degli esoni 8, 14, 19, 28 e 33 (Tabella 9). In presenza della mutazione c.793-1G>C, l'RT-PCR ha invece evidenziato la presenza di due messaggeri distinti: una banda debole associata alla perdita dell'esone 9 ed una più marcata associata ad un secondo messaggero contenente gli ultimi 57 nucleotidi dell'introne 8 a seguito di riconoscimento di un sito criptico di splicing intronico. Una

simile condizione si osserva per la mutazione c.4011-1G>C con riconoscimento di un sito criptico all'interno dell'introne 41.

Condizione diversa è quella della mutazione c.709+1G>A identificata nel paziente FA13 (Figura 14). Mediante RT-PCR abbiamo verificato che la distruzione del sito di splicing naturale favorisce l'attivazione di un sito criptico nell'introne 7. Questo evento causa l'inclusione nell'RNA maturo dei primi 30 nucleotidi dell'introne 7 che non alterando il codice di lettura consente alla proteina di esprimersi stabilmente. Come discusso di seguito (vedi analisi proteine) si pone l'interrogativo se l'espressione della proteina FANCA, anche se mutata, possa svolgere qualche attività residua.

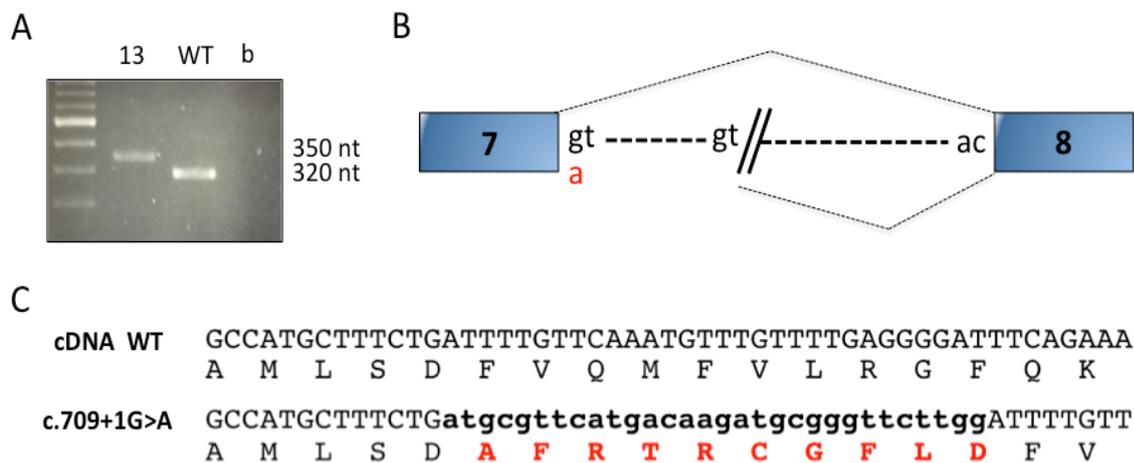


Figura 14. Caratterizzazione molecolare della mutazione di splicing c.709+1G>A identificata nel paziente FA13. (A) RT-PCT effettuata utilizzando RNA estratto dalla linea cellulare linfoblastoide del paziente (13) e di un controllo sano (WT). E' possibile rilevare la presenza di un'unica banda di circa 350 nucleotidi in FA13 (l'altro allele è deleto) anziché la banda corrispondente al corretto processamento del messaggero; (B) Rappresentazione schematica degli eventi di splicing (linee sottili tratteggiate) associati al corretto processamento del messaggero (sopra) e a quello mutato (sotto) che, a seguito della distruzione del "gt" naturale, prevede il riconoscimento di un sito criptico di splicing interno all'introne 7; (C) Sequenza del cDNA *wild type* e mutato verificata mediante sequenziamento Sanger. In grassetto è indicata la porzione di 30 nucleotidi dell'introne 7 inclusa nell'mRNA del paziente FA13 che ripristina la corretta cornice di lettura; in rosso invece, sono indicati i 10 aminoacidi inclusi nella proteina mutante stabile.

Non solo le alterazioni dei siti di splicing ma anche le mutazioni nonsense possono interferire con il corretto processo di maturazione dell'mRNA sfuggendo alla

degradazione tramite NMD (Nonsense-Mediated Decay) (Kervestin and Jacobson, 2012). Tale effetto si è difatti osservato in presenza di due mutazioni, la c.790C>T (p.Asp237Glyfs*4) e la c.1645C>T (p.Pro543Hisfs*26), che si associano alla rimozione rispettivamente degli esoni 8 e 18. In questi due casi come in quelli finora descritti, lo splicing alternativo produce un effetto di *frameshift* sul codice di lettura (Tabella 9) tuttavia, il ripristino della corretta cornice di lettura a seguito di eventi di splicing alternativo associati, come nel caso della c.709+1G>A alla produzione di una proteina stabile, hanno messo in evidenza la necessità di verificare a livello di RNA anche l'effetto delle mutazioni nonsense.

Mutazioni missense e piccole delezioni in-frame. In *FANCA* sono state identificate 23 varianti dal significato patogenetico incerto: 2 piccole delezioni *in frame* e 21 varianti missenso putative (Tabella 8). Entrambe le delezioni (p.Glu1239_Arg1243del ep.Phe1263del), come pure 9 sostituzioni amminoacidiche sono presenti nella banca dati delle mutazioni FA (<http://www.rockefeller.edu/fanconi/>), e non in quella degli SNP, sostenendo così il loro ruolo nella malattia. Per altre 5 varianti il significato patogenetico è invece dubbio: sono presenti nella banca dati FA ma anche in quella degli SNP. Tra queste, abbiamo considerato c.4198C>T (rs149851163) e c.3971C>T (rs182657062) come patogenetiche ("P") in quanto la loro frequenza nella popolazione FA è significativamente più alta rispetto a quella della popolazione di controllo, come riportato sia in 1000 Genomes (<http://www.1000genomes.org>) che in Exome Sequencing Project (<http://evs.gs.washington.edu/EVS/>). Al contrario, abbiamo considerato non patogenetici ("NP") gli SNPs rs139235751 (c.1874G>C), rs17233141 (c.2574C>G), rs17233497 (c.3263C>T) per la loro relativa frequenza anche nella popolazione di controllo.

Delle 21 varianti missenso identificate nella nostra coorte, 7 sono nuove alterazioni identificate in uno o due alleli *FANCA*, il cui significato biologico rimane incerto (VUS, "variant of uncertain significance") (Tabella 8).

Il ruolo patogenetico di tutte le 21 sostituzioni amminoacidiche è stato valutato anche mediante utilizzo di 4 programmi di predizione bioinformatica. Gli output dei diversi programmi sono stati riassunti in un unico valore compreso tra 0 e 8, come descritto in

Materiali e Metodi. Sebbene il risultato delle predizioni bioinformatiche vada considerato con cautela, 10 delle 11 varianti classificate come “P”, patogenetiche, hanno ricevuto uno score ≥ 4 . E’ interessante notare come per quasi tutte le varianti l’assegnazione di patogenicità su base genetica e bioinformatica coincida. Allo stesso modo, le varianti considerate in precedenza come “NP” o VUS hanno ricevuto un valore < 4 . E’ importante notare che, nonostante il loro incerto significato, queste varianti colpiscono il locus *FANCA* in pazienti che presentano sull’altro allele alterazioni certamente patogenetiche, rassicurandoci almeno sulla corretta attribuzione del gruppo di complementazione (Tabella 8 e Tabella 9).

NM_000135.2 (SNP)	Proteina (MAF as %) ^a	No. di alleli FA (nostra casistica)	No. di alleli FA (database mutazioni FA)	Patogenicità su base genetica (patient ID) ^e	Score di patogenicità in base ai tools bioinformatici
		MAF (%) tra gli alleli FA ^b			
c.1450G>C	p.Glu484Gln	1	0	nd	8
c.2851C>T	p.Arg951Trp	1	14	P	8
c.3490C>T	p.Pro1164Ser	4	1	P	8
c.4198C>T (rs149851163)	p.Arg1400Cys (1000G: 0.05) (ESP: 0.02)	1	4	P	8
		0.25 ($p=0.04$)			
c.1874G>C (rs139235751)	p.Cys625Ser (1000G: 0.14) (ESP: 0.23)	1	1 ^c	NP (FA56)	7
		0.10			
c.2852G>A	p.Arg951Gln	5	8	P	7
c.3239G>A	p.Arg1080Gln	1	1	P	7
c.3971C>T (rs182657062)	p.Pro1324Leu (1000G: 0.05) (ESP: 0.00)	4	5	P	7 ^f
		0.46 ($p=0.003$)			
c.2051T>C	p.Leu684Pro	1	3	P	6 ^g
c.2290 C>T	p.Arg764Trp	1	1	P	6
c.3164G>A	p.Arg1055Gln	1	1	P	6
c.352G>C	p.Ala118Pro	1	0	nd	5 ^h
c.4075G>T	p.Asp1359Tyr	1	1	P	5
c.3692A>C	p.His1231Pro	1	0	nd	4
c.3798G>A	p.Met1266Ile	1	0	NP (FA79)	3
c.457C>G	p.Gln153Glu	1	0	NP (FA65)	2
c.2770G>T	p.Asp924Tyr	2	0	NP (FA35)	2
c.2574C>G (rs17233141)	p.Ser858Arg (1000G: 0.50) (ESP: 0.53)	1	5 ^d	NP	1
		0.31			
c.2738A>C	p.His913Pro	7	4	P	1 ^h
c.3263C>T (rs17233497)	p.Ser1088Phe (1000G: 3.34) (ESP: 5.92)	1	3	NP (FA45)	1
		0.20			
	p.His1343Gln	1	0	nd	0

Tabella 8. Predizione di patogenicità in 21 sostituzioni aminoacidiche del gene FANCA. ^aMAF, Minor Allele Frequency, sulla base di quanto riportato nei 2.184 alleli del 1000 Genome database (<http://www.1000genomes.org/>) ed i 12,996 alleli dell'Exome Sequencing Project (ESP, <http://evs.gs.washington.edu/EVS/>); ^bFrequenza stimata sulla base dei 1960 alleli FA dei 980 pazienti FA-A di cui 85 della nostra casistica e 895 descritti nel database delle mutazioni FA

(<http://www.rockefeller.edu/fanconi/>). Sono riportati solo i *p-values* FA vs controlli ≤ 0.05 ; ^cErroneamente riportata come c.1873G>C (p.Cys625Ser) in Castella et al., 2011; ^dIdentificata in altri lavori (Tamary et al., 2000, Wijker et al., 1999); ^ePazienti con altre due mutazioni patogenetiche; ^fPresenti studi funzionali (Adachi et al., 2002, Ameziane et al., 2008); ^gMutazione de novo identificata nel paziente FA48; ^hMutazione persa nella linea linfoblastoide revertita (De Rocco et al., 2014).

4.2.3 Effetto del fondatore delle mutazioni in FANCA

Come accennato la maggior parte delle mutazioni *FANCA* sono private, identificate solo in solo uno (n=61) o due (n=17) dei 162 alleli *FANCA*. Tuttavia, cinque sono relativamente frequenti, rappresentando insieme il 26% (42/162) degli alleli *FANCA* caratterizzati nella nostra popolazione (Tabella 9, Figura 13). Anche se non conosciamo l'aplotipo, ipotizziamo che tre mutazioni siano il risultato di un effetto del fondatore in Campania (c.790C>T e c.2840C>G) e in Sicilia (c.2738A>C). Le altre due mutazioni (c.3558dup e c.3788_3790del) essendo relativamente frequenti nella popolazione FA (<http://www.rockefeller.edu/fanconi/>) sono molto probabilmente il risultato di eventi mutazionali indipendenti.

GENE	MUTAZIONE	ALLELI ITALIANI IDENTIFICATI
FANCA	c.790C>T (skipping esone 8)	10/162
	c.2840C>G (p.Ser947X)	8/162
	c.2738A>C (p.His913Pro)	7/162
	c.3558dupG (frameshift)	7/162
	c.3788_3790 delTCT p.Phe1263del	7/162
FANCG	c.1715G>A (p.Trp572*)	4/10



Figura 15. Mutazioni ricorrenti tra gli alleli FA della casistica italiana. Le mutazioni c.790C>T e c.2840C>G sono ricorrenti in Campania mentre le mutazioni c.2738A>C e c.1715G>A (gene *FANCG*) sono frequenti rispettivamente in Sicilia e Sardegna.

4.2.4 Mutazioni in geni non-FANCA

Le famiglie appartenenti ai gruppi di complementazione diversi da FA-A sono 15 (Figura 11). Per quanto riguarda i pazienti FA-G, 8 sono omozigoti e 1 eterozigote composto per 9 diverse alterazioni del gene *FANCG* (Tabella 9).

Anche in questo gene si ritrovano tutte le categorie di mutazioni (3 mutazioni *frameshift* causate da piccole delezioni o duplicazioni, tre nonsense, due mutazioni di splicing e una delezione *in frame*). E' di particolare interesse la mutazione c.1715G>A (p.Trp572*) associata ad un prodotto di RT-PCR aberrante a causa di uno splicing alternativo. Questa mutazione si ritrova in omozigosi in due pazienti apparentemente non imparentati di origine sarda a suggerire un quarto effetto fondatore in Italia (Figura 15).

Dei tre pazienti FA-C, due sono omozigoti per mutazioni nonsense o *frameshift* note, mentre il terzo è un eterozigote composto (p.Gln13* e p.Lys231del). Ancora più rari sono i pazienti FA-D2. Ne abbiamo caratterizzati solo due: un omozigote e l'altro eterozigote composto per mutazioni missenso e nonsense (Borriello et al., 2007, Kalb et al., 2007). Infine, una sola famiglia è mutata in *FANCB*: l'effetto patogenetico della mutazione p.Phe118Ser, mai identificata in precedenza, è supportato dai programmi bioinformatici di predizione (Tabella 9) (De Rocco et al., 2014).

Di tutti i pazienti giunti alla nostra attenzione per la diagnosi molecolare, solo in 3 casi non è stato possibile identificare né le mutazioni né il gene FA responsabile, per cui ipotizziamo che possa essere mutato un nuovo gene FA. Non possiamo, tuttavia, escludere che le loro mutazioni siano sfuggite all'analisi nonostante tutti i 16 geni FA siano stati analizzati con un sistema di sequenziamento di nuova generazione (Ion Torrent PGM™, Life Technologies).

Esone/ Introne	FANCA cDNA (NM_000135.2)	Number. of alleles	FANCA RNA and/or protein (NP_000126.2)	Mutation classification	References
i-1_i6	c.-42- ?_522+?del	2	nd	large deletion (removal of initiation codon)	FA database
i-1_i43	c.-42- ?_5481+?del	2	no protein	large deletion (entire gene)	Savino et al, 2003; Centra et al, 1999
1	c.2T>A	1	p.?	initiation codon	Novel
1	c.11C>A	1	p.Ser4*	nonsense	FA database
1	c.50dup	1	p.Arg18Profs*19	small duplication (frameshift)	Novel
1	c.66G>A	1	p.Trp22*	nonsense	FA database
2	c.123_124del	1	p.Lys42Ilefs*21	small deletion (frameshift)	Novel
i2	c.190-1G>T	1	nd	splicing (potential)	Savino et al, 2003
i2_i3	c.190- ?_283+?del	1	p.Val64Alafs*43	large deletion (frameshift)	FA database
i3	c.283+3A>C	2	r.190_283del (e3 skipping) p.Val64Alafs*43	splicing (frameshift)	Savino et al, 2003
i3_i20	c.284- ?_1826+?del	4	p.Gly95Gluufs*31	large deletion (frameshift)	Novel
i3_i4	c.284- ?_426+?del	1	p.Gly95Gluufs*38	large deletion (frameshift)	Novel
4	c.352G>C	1	p.Ala118Pro	missense	Novel
5	c.457C>T	1	p.Gln153*	nonsense	Novel
5	c.457C>G	1	p.Gln153Glu	missense	Novel
i5_i30	c.523- ?_2981+?del	1	p.Ser175Leufs*5	large deletion (frameshift)	Novel
i5_i31	c.523- ?_3066+?del	2	p.Ser175_Gln1022del	large deletion (in frame)	Novel
6	c.548G>A	1	p.Trp183*	nonsense	Novel
6	c.549G>A	1	p.Trp183*	nonsense	Novel
i6	c.596+1G>T	1	nd	splicing (potential)	Novel
i7	c.709+1G>A	1	nd	splicing (potential)	FA database
i7	c.710-5T>C	2	r.710_792del (e8 skipping) p.Asp237Glyfs*4	splicing (frameshift)	FA database
7	c.619G>T	1	p.Gly207*	nonsense	Novel
8	c.790C>T	10	r.710_792del (e8 skipping) p.Asp237Glyfs*4	(putative nonsense) splicing (frameshift)	FA database
i8	c.793-1G>C	2	r.793_826del (e9 skipping) p.Val1265Leufs*20 r.[793-1g>c; 792_793ins793-57_793- 1] (cryptic site in i8) p.Val265Lysfs*34	splicing (frameshift)	Novel
i9	c.826+3del	3	r.[709+1a>g; 709_710ins709+1_709+3 0] (cryptic site in intron 7) p.250_251insGlyAlaPheMet ThrArgCysGlyPheLeu	splicing (inframe)	Savino et al, 2003
i10	c.893+1G>T	1	r.[893+1g>u; 893_894ins893+1_893+2 15] (cryptic site in i10) p.Phe879Valfs*19	splicing (frameshift)	Savino et al, 2003
i10	c.893+5G>A	1	nd	splicing (potential)	Novel

i10_i15	c.894- ?_1359+?del	1	p.Phe299Profs*72	large deletion (frameshift)	FA database
i10_i33	c.894- ?_3348+?del	1	p.Trp298*	large deletion (nonsense)	FA database
11	c.987-990del	2	p.His330Alafs*4	small deletion (frameshift)	FA database
i11	c.1006+1G>T	1	nd	splicing (potential)	Novel
13	c.1115_1118del	2	p.Val372Alafs*42	small deletion (frameshift)	FA database
13	c.1126C>T	1	p.Gln376*	nonsense	Novel
i14	c.1359+1G>C	4	r.1226_1359del (e14 skipping) p.Asp409Glyfs*31	splicing (frameshift)	Savino et al, 2003
i14_i15	c.1360- ?_1470+?del	4	r.1360_1470del p.Ala454_Gln490del	large deletion (in frame)	FA database
i14_i20	c.1360- ?_1826+?del	1	p.Ala454Serfs*3	large deletion (frameshift)	FA database
i14_i28	c.1360- ?_2778+?del	1	p.Ala454_His926del	large deletion (in frame)	Novel
15	c.1450G>C	1	p.Glu484Gln	missense	Novel
i15_i17	c.1471- ?_1626+?del	1	p.Val491_Glu542del	large deletion (in frame)	FA database
i16	c.1567-20A>G	2	r.[1567-20a>g; 1566_1567ins1566+1_15 67-1] (i16 retention) p.Val523Alafs*49	splicing (frameshift)	Savino et al, 2003
i17_i21	c.1627- ?_1900+?del	1	p.Pro543Metfs*6	large deletion (frameshift)	Savino et al, 2003
18	c.1645C>T	1	r.1645c>u + r.1627_1715del (e18 skipping) Gln549* + p.Pro543Hisfs*26	nonsense + splicing (frameshift)	FA database
i19	c.1776+7A>G	1	r.1716_1776del (e19 skipping) p.Ile573Serfs*12	splicing (frameshift)	Novel
i19	c.1777- 7_1779del	1	nd	splicing (potential)	FA database
i20_i28	c.1827- ?_2778+?del	1	p.Arg609Serfs*2	large deletion (frameshift)	FA database
i20_i29	c.1827- ?_2852+?del	2	r.1827_2852del p.Ala610_Arg951del	large deletion (in frame)	FA database
21	c.1874G>C	1	p.Cys625Ser	missense	Novel
21	c.1850_1859del	1	p.Leu617Profs*20	small deletion (frameshift)	Novel
22	c.2005C>T	2	p.Gln669*	nonsense	Savino et al, 2003
23	c.2051T>C	1	p.Leu684Pro	missense	FA database
25	c.2314C>T	2	p.Gln772*	nonsense	Savino et al, 2003
25	c.2290 C>T	1	p.Arg764Trp	missense	FA database
i26	c.2504+2T>C	1	r.[2504+2u>c; 2504_2505ins2504+1_25 04+129] (cryptic site in i26) p.Phe*	splicing (nonsense)	Savino et al, 2003
27	c.2535_2536del	1	p.Cys846Glnfs*20	small deletion (frameshift)	FA database
27	c.2574C>G	1	p.Ser858Arg	missense	Savino et al, 2003
28	c.2638C>T	1	p.Arg880*	nonsense	FA database
28	c.2738A>C	7	p.His913Pro	missense	FA database
28	c.2770G>T	2	p.Asp924Tyr	missense	Novel

i28	c.2778+1G>A	1	r.2602_2778del (e28 skipping) p.Phe868_His926del	splicing (inframe)	FA database
i28	c.2778+83C>G	1	nd	splicing (potential)	Savino et al, 2003
29	c.2840C>G	8	p.Ser947*	nonsense	Savino et al, 2003
29	c.2852G>A	5	p.Arg951Gln	missense	FA database
29	c.2812_2830dup p	2	p.Asp944Glyfs*5	small duplication (frameshift)	Savino et al, 2003
29	c.2851C>T	1	p.Arg951Trp	missense	FA database
30	c.2853dup	2	p.Gln952Alafs*10	small duplication (frameshift)	Novel
30	c.2932C>T	1	p.Gln978*	nonsense	Novel
32	c.3109_3137del insG	1	p.Pro1037Alafs*14	small indel (frameshift)	Novel
32	c.3164G>A	1	p.Arg1055Gln	missense	FA database
32	c.3239G>A	1	p.Arg1080Gln	missense	FA database
33	c.3263C>T	1	p.Ser1088Phe	missense	FA database
i33	c.3348+1G>A	1	r.3240_3348del (e33 skipping) p.Ile1081Glyfs*10	splicing (frameshift)	FA database
34	c.3367G>T	1	p.Gly1123*	nonsense	Novel
35	c.3490C>T	4	p.Pro1164Ser	missense	FA database
36	c.3558dup	7	p.Arg1187Glyfs*28	small duplication (frameshift)	Savino et al, 2003
37	c.3638_3639del	3	p.Pro1213Argfs*64	small deletion (frameshift)	Savino et al, 2003
37	c.3660del	2	p.Asn1221Thrfs*26	small deletion (frameshift)	Novel
37	c.3715_3729del	2	p.Glu1239_Arg1243del (GluGluAsnIleArg)	small deletion (in frame)	FA database
37	c.3692A>C	1	p.His1231Pro	missense	Novel
37	c.3761_3762dup p	1	p.Glu1255Argfs*12	small duplication (frameshift)	Savino et al, 2003
38	c.3788_3790del	10	p.Phe1263del	small deletion (in frame)	FA database
38	c.3798G>A	1	p.Met1266Ile	missense	Novel
i38	c.3829-9G>A	1	nd	splicing (potential)	Novel
40	c.3971C>T	4	p.Pro1324Leu	missense	FA database
i40	c.4011-1G>C	1	r.4011_4033del (cryptic site recognition in e41) p.Ser1337Argfs*80	splicing (frameshift)	Novel
41	c.4029T>G	1	p.His1343Gln	missense	Novel
41	c.4075G>T	1	p.Asp1359Tyr	missense	Savino et al, 2003
42	c.4198C>T	1	p.Arg1400Cys	missense	FA database
42	c.4258G>T	1	p.Glu1420*	nonsense	Novel
i42	c.4261- 19_4261-12del	1	nd	splicing (potential)	FA database
i42	c.4261-2A>C	1	nd	splicing (potential)	Novel
Exon/Int ron	FANCB cDNA (NM_152633.2)	Number of alleles	FANCB protein (NP_001018123.1)	Mutation classification	References
3	c.353T>C	1	p.Phe118Ser	missense	Novel
Exon/Int ron	FANCC cDNA (NM_000136.2)	Number of alleles	FANCC protein (NP_000127.2)	Mutation classification	References
2	c.37C>T	1	p.Gln13*	nonsense	FA database
2	c.67del	2	p.Asp231Ilefs*23	small deletion (frameshift)	FA database

8	c.692_694del	1	p.Lys231del	small deletion (in frame)	Novel
15	c.1642C>T	2	p.Arg548*	nonsense	FA database
Exon/Intron	FANCD2 cDNA (NM_001018115.1)	No. of alleles	FANCD2 protein (NP_001018125.1)	Mutation classification	References
7	c.458T>C	2	p.Leu153Ser	missense	Borriello et al, 2007
12	c.904C>T	1	p.Arg302Trp	missense	FA database
13	c.1092G>A	1	p.Trp364*	nonsense	FA database
Exon/Intron	FANCG cDNA (NM_004629.1)	Number of alleles	FANCG RNA and/or protein (NP_004620.1)	Mutation classification	References
4	c.336del	2	p.Arg113Glyfs*39	small deletion (frameshift)	Novel
i7	c.924+1G>A	1	r.778_924del (e7 skipping) p.Gly260_Glu308del	splicing (inframe)	Novel
i9	c.1144-1G>T	2	r.1144_1168del (cryptic site in e10) p.Phe382Glyfs*13	splicing (frameshift)	Novel
10	c.1182_1192del insC	2	p.Glu395Trpfs*5	small indel (frameshift)	FA database
10	c.1199_1204del	2	p.Ala400_Ala401del	small deletion (in frame)	Novel
10	c.1223_1226dup	1	p.Ala410Glnfs*10	small duplication (frameshift)	Novel
10	c.1344C>A	2	p.Cys448*	nonsense	Novel
13	c.1715G>A	4	r.1715g>a + r.1637_1760del (e13 skipping) p.Trp572* + p.Asn547Leufs*6	nonsense + splicing (frameshift)	FA database
14	c.1788T>A	2	p.Tyr596*	nonsense	Novel

Tabella 9. Mutazioni identificate in 100 famiglie FA nella popolazione italiana (De Rocco et al., 2014).

4.3 ANALISI DELL'ESPRESSIONE PROTEICA PER LA DIAGNOSI DELL'ANEMIA DI FANCONI

Nonostante la strategia descritta nei paragrafi precedenti (Figura 8) sia stata applicata con successo alla maggior parte dei pazienti FA della nostra casistica, si tratta comunque di un processo complesso che fornisce risultati a volte dubbi senza garantire il successo della diagnosi molecolare. Come accennato precedentemente, un grosso "scoglio" nell'algoritmo proposto in Figura 8 è rappresentato dal test di complementazione.

Per questo motivo, abbiamo valutato la possibilità di sostituirlo con l'analisi dell'espressione delle proteine FA. La strategia illustrata in Figura 8 prevede già l'analisi di FANCD2 per stabilire se la mutazione è da ricercarsi nel complesso I, II o III del pathway. Pertanto, il passo successivo è stato definire i livelli di espressione di FANCA, il cui gene è mutato nella maggior parte dei casi.

4.3.1 Espressione della proteina FANCA e correlazione con il gruppo di complementazione

Per quanto riguarda lo studio FANCA, l'analisi è stata condotta su 44 linee cellulari linfoblastoidi in cui era stata confermata, almeno in una prima fase, la sensibilità al DEB o alla MMC. In 12 casi non abbiamo rilevato alcuna espressione di FANCA mentre negli altri 32 la banda corrispondente alla proteina risultava visibile anche se con livelli molto variabili di espressione rispetto alla linea *wild-type*. Per quanto riguarda il primo gruppo (senza FANCA), in tutte linee cellulari erano presenti mutazioni di FANCA. Del secondo gruppo (FANCA espressa), 8 appartenevano al gruppo di complementazione FA-G (Figura 16). In questi casi l'espressione di FANCA è di circa il 10%-20% rispetto ai controlli. Dalla letteratura è noto, infatti, che FANCG interagisce con FANCA e che mutazioni nel gene FANCG non solo impediscono a FANCA di migrare nel nucleo ma addirittura ne riducono l'emivita a livello citoplasmatico (Garcia-Higuera et al., Blood, 2000). In accordo con questi dati, esprimendo il gene FANCG nella linea FA92

si ripristina l'espressione di FANCA come determinato in collaborazione con il Dott. Doufur (IRCCS G. Gaslini, Genova) (Figura 16, FA93+G).

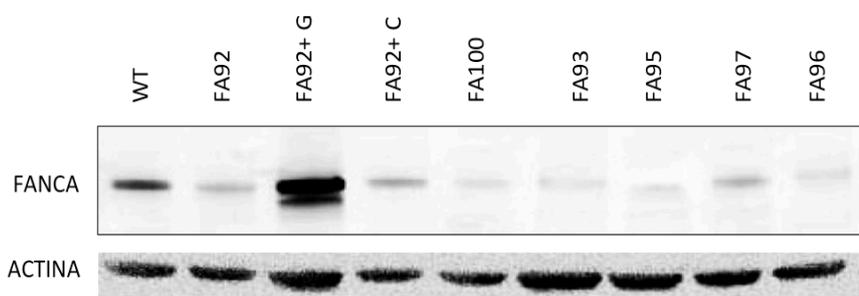


Figura 16. Immunoblotting eseguito su estratti proteici ottenuti da linee cellulari linfoblastoidi di un controllo sano (WT) e da sei affetti appartenenti al gruppo di complementazione FA-G (FA92, FA100, FA93, FA95, FA97, FA96). Nei pazienti con mutazioni nel gene *FANCG* i livelli di espressione di sono molto bassi rispetto al controllo *wild type*. La linea cellulare del paziente FA92 è stata corretta con il cDNA rispettivamente di *FANCC* (FA92+C) e di *FANCG* (FA92+G) per confermare la correzione del fenotipo dopo inserimento del gene corretto.

Per quanto riguarda le altre 24 linee che esprimono FANCA, 20 appartengono al gruppo di complementazione FA-A. In questi casi i livelli però variavano tra il 20 e il 100% rispetto ai controlli. Infine, 4 linee linfoblastoidi hanno mutazioni nei geni *FANCC* (n=2), *FANCB* (n=1) e *FANCD2* (n=1). In accordo con la letteratura, fatta eccezione per *FANCG*, l'espressione di FANCA non è alterata da mutazioni nelle subunità del complesso I o in *FANCD2*/*FANCI*. Purtroppo, la carenza di anticorpi specifici contro queste subunità impedisce un'analisi estesa a tutti componenti del pathway.

LINEE CELLULARI n=44	FANCA non espressa (0-5%)	FA-A n=12
	bassa espressione FANCA (5-20%)	FA-G n=8
	alta espressione FANCA (>20%)	FA-C n=2
		FA-B n=1
		FA-D2 n=1
	FA-A n=20	

Tabella 10. Espressione della proteina FANCA nelle linee cellulari linfoblastoidi dei pazienti FA appartenenti ai diversi gruppi di complementazione.

4.3.2 Gruppo di complementazione FA-A: correlazione tra genotipo ed espressione di FANCA

La maggior parte (n=32) delle linee linfoblastoidi analizzate presenta mutazioni in *FANCA*. Dato che in queste linee il livello di espressione di *FANCA* varia da 0% al 100% rispetto ai controlli, abbiamo correlato la quantità di *FANCA* espressa con le mutazioni responsabili della malattia. Delle 12 in cui non abbiamo rilevato *FANCA*, 11 hanno alterazioni cosiddette “deleterie”, quali mutazioni nonsense, *frameshift* o ampie delezioni, in entrambi gli alleli. In un caso (FA60), un eterozigote composto per una delezione *frameshift* e una potenziale mutazione di splicing, è presente una debole banda corrispondente a circa il 10% di espressione proteica. In questo caso, ipotizziamo che la seconda mutazione determini la formazione di più prodotti di *splicing* alternativi, uno dei quali associato con una piccola delezione/inserzione in frame in grado da generare una proteina *FANCA* stabile.

In altri 20 casi, invece, l'espressione di *FANCA* varia dal 50% al 100%. Poiché in letteratura è riportato che le mutazioni missenso sono associate a stabilità di *FANCA* (Castella et al., 2011), abbiamo quantificato l'espressione di *FANCA* in duplicato o triplicato rispetto ad un controllo citoplasmatico (β -actina). Tutti i pazienti (n=19) sono portatori di mutazioni missenso o delezioni *in frame* in omozigosi o in eterozigosi. La quantificazione del livello di espressione rivela una correlazione tra il numero di questi alleli e la quantità di *FANCA* (Figura 17, B). Se presenti due alleli (n=8), la proteina è espressa a livelli paragonabili a quelli delle cellule *wild type*. Se invece è presente un solo allele (10 pazienti) rileviamo solo il 50% di proteina.

Le uniche due eccezioni che abbiamo osservato sono rappresentate dalle linee linfoblastoidi FA29 e FA78. Contrariamente a quanto atteso per la presenza di una duplicazione *frameshift* e un'ampia delezione, in FA29 la proteina *FANCA* è espressa al 50% rispetto al controllo. L'analisi genetica della linea cellulare mediante MLPA ha rivelato una situazione complessa in cui almeno il 20% delle cellule sembrano aver perso la delezione (ripristinando un allele *wild type*) evento che può essere la conseguenza di una conversione genica. Anche FA78, con il 100% di proteina espressa, presenta un livello superiore di *FANCA* rispetto a quanto atteso sulla base delle mutazioni identificate (una mutazione missenso e la delezione dell'esone 4). Tuttavia, anche in

questo caso, un evento di conversione genica verificata mediante MLPA rende conto del ripristino della funzionalità del secondo allele a livello della linea cellulare.

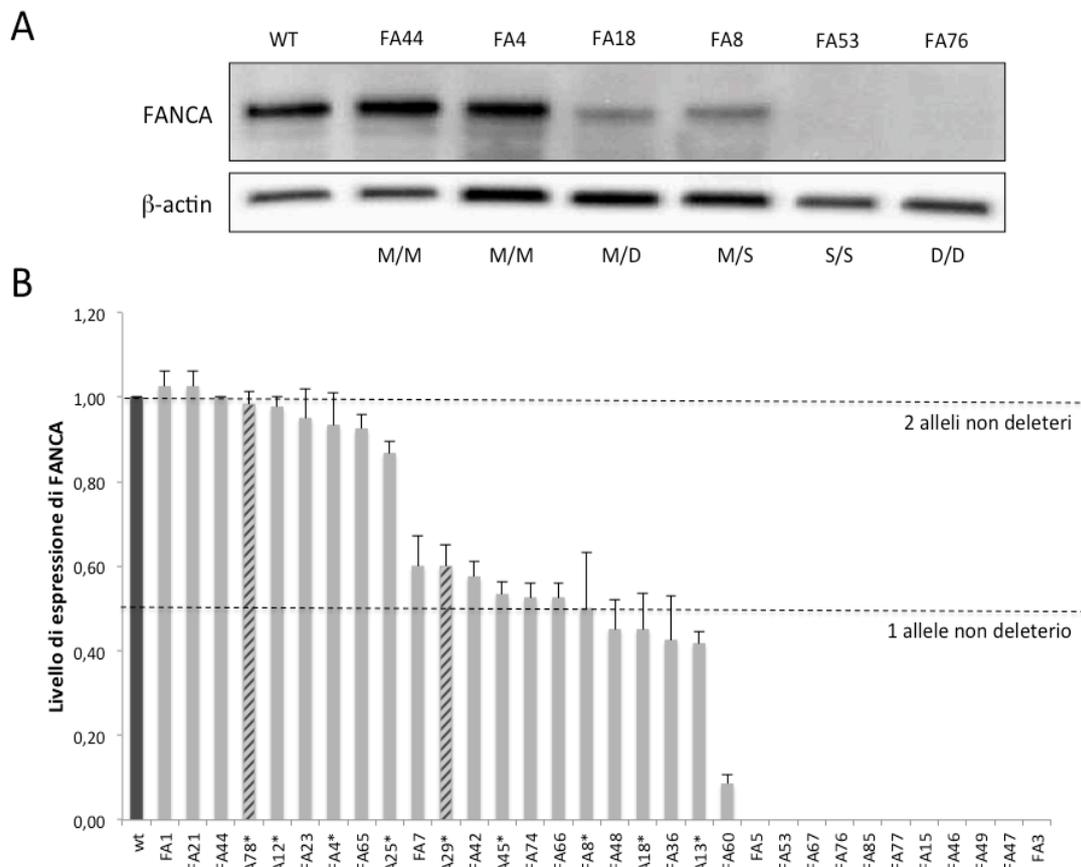


Figura 17. Espressione di FANCA in linee cellulari FA-A. (A) Western blot rappresentativo del livello di espressione di FANCA in LCLs derivati da pazienti portatori di diverse combinazioni di mutazioni del gene *FANCA*. WT, wild-type; M, mutazione missense; N, mutazione nonsense; S, mutazione di splicing (*frameshift*); D, ampia delezione (*frameshift*). (B) Quantificazione dei livelli di espressione di FANCA in 32 linee cellulari di pazienti appartenenti al gruppo si complementazione FA-A. Il livello di proteina è stato calcolato mediante densitometria in rapporto all'espressione della stessa proteina nella linea WT e normalizzato (β -actina). In ordinata è riportata la media di due esperimenti indipendenti; *media di tre esperimenti indipendenti; Le barre grigliate si riferiscono ai due pazienti per i quali la linea cellulare è revertita in coltura ripristinando un allele wild type.

4.3.3 Espressione della proteina FANCA nei pazienti con ampie delezioni in frame

Il dato più sorprendente è emerso dall'analisi dei pazienti con delezione intrageniche di *FANCA* che, nonostante la rimozione di uno o più esoni, consentono di generare molecole non solo di mRNA ma anche di proteina stabili. Rispettano questa condizione tutti i casi (n=6) di linee cellulari con delezioni *in frame* in uno o entrambi gli alleli in cui osserviamo rispettivamente il 50% e il 100% di FANCA (Figura 18, A). In particolare, i due eterozigoti composti (FA45 e FA66) per la stessa delezione degli esoni 21_29 (p.Ala610_Arg951del) esprimono una proteina di 120 kDa circa, corrispondente alla rimozione dei 342 aminoacidi codificati dagli esoni dal 21 al 29. In entrambi i casi non osserviamo la banda corrispondente alla proteina di peso molecolare *wild type* per la presenza delle mutazioni c.1645C>T (p.Pro543Hisfs*26) e c.1776+7A>G (p.Ile573Serfs*12) rispettivamente, nel secondo allele. Allo stesso modo, la delezione dell'esone 15, in omozigosi nel paziente FA12 e in eterozigosi in FA7 porta alla produzione di una proteina stabile evidenziabile come banda di peso molecolare inferiore rispetto a quella *wild type* (Figura 18, A). Nel paziente FA65 (c.3490C>T; c.1471-?_1626+?del), invece, è possibile distinguere due bande di FANCA, una di 163 kD tradotta dall'allele colpito dalla mutazione missenso e l'altra di circa 157 kD, relativa all'allele con delezione degli esoni 16 e 17.

Infine, includiamo in questo gruppo anche la mutazione di splicing (c.709+1G>A) identificata in FA13. In questo caso era stata identificata, oltre a questa mutazione, una nonsense (c.457C>T, p.Gln153*). L'osservazione di una banda con peso molecolare paragonabile a quello della proteina *wild type* ci ha spinto ad indagare, le conseguenze della c.709+1G>A a livello dell'mRNA. Come riportato precedentemente (Figura 14) questa mutazione determina il riconoscimento di un sito criptico donatore a valle del "gt" naturale con conseguente inserzione di 30 nucleotidi intronici ma senza alterare il codice di lettura. Di conseguenza, si ha la produzione di una proteina FANCA di 10 aminoacidi più lunga della *wild type* (r.[709+1g>a;709_710ins709+1-709+30]), differenza troppo piccola per essere apprezzata in western blot (Figura 18, A).

4.3.4 Localizzazione nucleare e citoplasmatica di FANCA mutata

L'osservazione della produzione di molecole di FANCA stabilmente espresse seppur mutate apre una serie di interrogativi circa la potenziale attività residua di queste proteine e la conseguente associazione con un fenotipo più lieve nei pazienti che le esprimono (Adachi et al., 2002, Castella et al., 2011).

Il pathway FA è implicato nella riparazione del DNA e, per questo motivo, i suoi elementi hanno una localizzazione nucleare. Ipotizzare una funzionalità residua implicherebbe almeno la possibilità, anche parziale, delle molecole mutate di FANCA di migrare nel nucleo. Sebbene gli esperimenti di localizzazione nucleare e citoplasmatica di FANCA siano ancora in corso, i dati preliminari condotti su alcune linee cellulari non confermano l'ipotesi. In Figura 18, C è riportato, a titolo di esempio, l'analisi condotta in FA45. L'immunoblotting dei lisati citoplasmatici e nucleari mostra che, mentre nel controllo *wild type* FANCA è rilevabile sia nel nucleo sia nel citoplasma, nella linea del paziente, dove è presente la delezione in frame degli esoni 21_29 e una mutazione nonsense, è visibile la banda di 130 kD corrispondente alla proteina deleta nella frazione citoplasmatica ma non nel nucleo. In accordo con i dati di letteratura (Castella et al., 2011), è molto probabile che tutte le alterazioni associate a stabilità della proteina che abbiamo identificato impediscano a FANCA di migrare nel nucleo e regolare i processi per il controllo dell'integrità del genoma. Non possiamo tuttavia escludere, vista la loro localizzazione non solo nucleare ma anche citoplasmatica, che le proteine FA possano svolgere funzioni accessorie non ancora caratterizzate e forse in grado di modulare il fenotipo nei pazienti se espresse anche nella forma mutata.

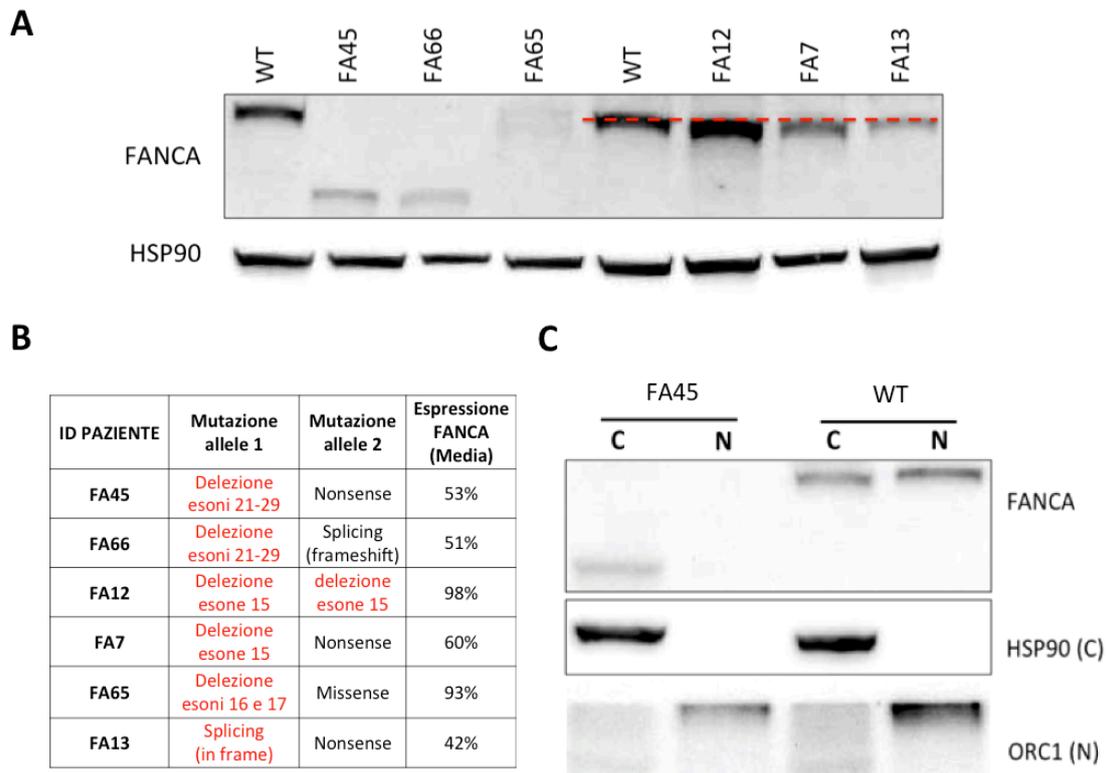


Figura 18. (A) Western blot relativo all'espressione della proteina FANCA in linee linfoblastoidi derivati da pazienti portatori di ampie delezioni in frame del gene *FANCA* (FA45, FA66, FA12, FA7 e FA65) e di una inserzione in frame (FA13). WT, *wild type*. HSP90, controllo di caricamento. (B) Linee cellulari che esprimono una proteina FANCA di dimensioni alterata - a causa di delezioni o inserzioni in frame - ma stabile. Il livello di proteina è stato calcolato mediante densitometria in rapporto all'espressione della stessa proteina nella linea WT e normalizzato (HSP90). (C) Western blot relativo all'espressione della proteina FANCA nella linea ICLs del paziente FA45 eseguito da preparazioni proteiche con separazione di nuclei e citoplasma. C, frazione citoplasmatica; N, frazione nucleare; HSP90, heat shock protein 90, marcatore di controllo citoplasmatico; ORC1, origin recognition complex 1, marcatore di controllo nucleare.

5. DISCUSSIONE

DISCUSSIONE

Nella FA, così come in molte altre malattie caratterizzate da eterogeneità genetica, la diagnosi molecolare è un processo complesso e non sempre conclusivo. Un primo ostacolo è rappresentato dai 16 diversi geni implicati nella malattia. In molti casi, l'analisi di complementazione fornisce un valido supporto all'identificazione del gene che causa causativo prima di procedere all'analisi molecolare. Anche nella nostra casistica, l'utilizzo di vettori retrovirali che esprimono i geni per *FANCA*, *FANCG* e *FANCC* ci hanno permesso, a partire dalla linea cellulare linfoblastoide o di fibroblasti, di assegnare il corretto gruppo di complementazione a 41 probandi.

L'analisi di complementazione è però un processo complesso che richiede competenze non sempre disponibili. Per questo motivo abbiamo valutato se l'analisi dell'espressione delle proteine FA, in particolare quella di *FANCA*, potesse rappresentare un approccio alternativo alla ricerca del gene candidato. Per questo motivo abbiamo analizzato 44 linee cellulari di pazienti appartenenti ai gruppi di complementazione FA-A, FA-G, FA-D2, FA-C e FA-B. Nonostante abbiamo osservato che la presenza/assenza della banda relativa alla proteina non sia associata alla presenza/assenza di mutazioni nel gene *FANCA*, lo studio dell'espressione di proteina corrispondente può, fornire utili informazioni per lo screening di mutazioni. Infatti, se la proteina non è espressa è indice che il paziente ha mutazioni in *FANCA*. Potrebbe trattarsi di ampie delezioni intrageniche *in frame* quando *FANCA* è rilevata con un peso molecolare alterato. Più problematica è la situazione in cui *FANCA* è espressa. Livelli molto bassi dovrebbero indurci a ipotizzare mutazioni in *FANCG*. Un'espressione quantificata al 50% di quella dei controlli, suggerisce di ricercare le mutazioni, tra cui una missenso, nel gene *FANCA*. Di ancora più difficile interpretazione è il caso in cui quando *FANCA* è espressa a livelli paragonabili a quelli delle cellule *wild type* perchè non possiamo stabilire qual è il gene mutato. Pertanto se da un lato questa analisi è da preferire per la maggior rapidità rispetto all'analisi di complementazione, dall'altro mostra una serie di limiti legati alla

stabilità delle proteine mutate a alla mancanza di anticorpi che riconoscono tutte le proteine FA in modo specifico.

Sebbene dallo studio dell'espressione di FANCA per l'identificazione del gene candidato siano emersi alcuni limiti, lo stesso studio pone l'interrogativo circa l'esistenza di mutazioni missenso, se non addirittura ampie delezioni *in frame*, che essendo stabilmente espresse potrebbero conservare qualche funzionalità residua associata ad un fenotipo "mild" analogamente a quanto avviene nella distrofia muscolare di Duchenne. In questa patologia (X-linked), si riconoscono infatti due forme, la forma classica, più grave, associata ad una mutazione nonsense o *frameshift* nel gene della distrofina ed una forma "mild", la sindrome di Becker, che si verifica quando, a seguito di una delezione intragenica, la proteina risulta danneggiata nelle parti non terminali e presenta un peso molecolare minore (Tinsley et al., 1993).

Se questo fosse il caso anche nella FA, si potrebbero definire importanti correlazioni genotipo/fenotipo. A questo proposito, mentre Faivre et al. rileva uno sviluppo precoce dell'anemia nei pazienti FA-A che non esprimono proteina rispetto a quelli che esprimono una forma alterata di FANCA (Faivre et al., 2000), Castella et al. non confermano il dato. Indipendentemente dall'espressione o meno di FANCA mutata, non osservano differenze nella suscettibilità alla fragilità cromosomica, spontanea o indotta, nell'età in cui compaiono i disturbi ematologici né in tipo e gravità delle malformazioni che affliggono i pazienti (Castella et al., 2011).

La nostra casistica, così come quella di Castella et al. (2011), non è sufficientemente ampia per permetterci di formulare qualche correlazione tra espressione di FANCA, anche se mutata, e un fenotipo clinico più lieve. Un prerequisito fondamentale per il raggiungimento di questo obiettivo potrebbe essere l'istituzione di un database internazionale che raccolga tutti i dati clinici e proteici dei pazienti al fine di stabilire se alcune mutazioni possano manifestare effetti meno deleteri di altre.

Gli studi sulla funzionalità residua dei mutanti proteici possono essere utili anche per identificare domini proteici importanti per la funzione cellulare. Oltre al dominio di localizzazione nucleare, ai residui sottoposti a fosforilazione e a regioni deputate all'interazione con FANCG (Garcia-Higuera et al., 2000, Lightfoot et al., 1999), non sono stati caratterizzati altri domini in FANCA. I risultati di questo studio, così come quello di

Castella et al. (2011), dimostrano che tutte le sostituzioni aminoacidiche analizzate, indipendentemente dalla loro posizione, portano ad una conseguenza funzionale comune: una proteina FANCA alterata è espressa a livelli *wild type* nonostante l'incapacità della stessa di migrare nel nucleo impedendo al pathway FA di funzionare. E' ormai noto che le proteine FA rivestono funzioni accessorie, sebbene poco conosciute, anche a livello citoplasmatico. Ne è un esempio il forte stato di stress ossidativo riportato nei pazienti e l'osservazione di alterazioni metaboliche e mitocondriali dovute all'ambiente pro-ossidante tipico di queste cellule (Pagano et al., 2012). Inoltre, è noto che l'interattore diretto di FANCA, FANCG, migra nel mitocondrio dove interagisce con componenti chiave della catena respiratoria (Mukhopadhyay et al., 2006). E' quindi lecito ipotizzare che l'eventuale funzionalità residua delle proteine FANCA alterate sia legata a questo aspetto che merita, pertanto, approfondimenti futuri.

Nei casi in cui non siano disponibili informazioni sul gene candidato, l'algoritmo diagnostico in FA si basa su una strategia che prevede, in una prima battuta, l'analisi diretta di *FANCA*, il gene mutato nel 65% degli individui affetti (<http://www.rockefeller.edu/fanconi/>). In accordo con la letteratura, anche se con una frequenza relativamente più elevata, la maggioranza (85%) delle famiglie nella coorte italiana, infatti, presentano mutazioni in questo gene (De Rocco et al., 2014). Si tratta principalmente di mutazioni private, anche se ci sono evidenze di effetti fondatore in Campania, Sicilia e Sardegna, informazione che può essere sfruttata se si conosce l'origine geografica delle famiglie.

In generale, lo screening di mutazione di *FANCA* è complicato non solo dalle grandi dimensioni del gene, dalla mancanza di *hot-spot* mutazionali e dalla presenza di ampie delezioni intrageniche, spesso in eterozigosi, che richiedono tecniche più sofisticate, quali l'MLPA, per la loro caratterizzazione. Per queste ragioni, alcune mutazioni potrebbero sfuggire all'analisi, come infatti si è verificato nel 6% degli alleli *FANCA* dei nostri pazienti.

Ogni qualvolta *FANCA* non risulta alterato, si procede con la ricerca di mutazioni negli altri geni FA, iniziando con quelli più frequentemente mutati in una specifica popolazione. I dati ottenuti dall'analisi della nostra coorte, indica che il 9% delle famiglie presentano mutazioni nel gene *FANCG*. Solo in pochi casi sono state identificate

mutazioni nei geni *FANCC*, *FANCD2*, e *FANCB*, solo grazie alle informazioni ottenute dall'analisi di complementazione, dall'analisi dell'espressione proteica oppure dalle tecniche di sequenziamento di nuova generazione (NGS) (Borriello et al., 2007, De Rocco et al., 2014).

La presenza di mosaicismo emopoietico è un altro aspetto che può rendere il processo diagnostico nella FA ancora più complesso. Quasi il 25% delle linee cellulari linfoblastoidi convertono il fenotipo (instabilità cromosomica) da alterato a *wild type*, come si evidenzia dalla perdita della sensibilità al DEB o alla MMC (Lo Ten Foe et al., 1996). Ciò ostacola il corretto inquadramento del paziente, interferendo con il test al DEB e l'analisi di complementazione e/o proteica. Nella nostra casistica, il fenomeno della reversione del fenotipo nelle linee cellulari linfoblastoidi si osserva con una frequenza inferiore rispetto a quella riportata in letteratura (De Rocco et al., 2014). Ciò è molto probabilmente da attribuire ai brevi periodi in cui manteniamo in coltura le cellule, rallentando la selezione di quelle in cui si è ripristinato il fenotipo *wild type*. E' interessante ricordare i meccanismi implicati in questo processo: sono stati descritti fenomeni di *crossing over* intragenico, conversione genica, scomparsa di una mutazione germinale ("*back mutation*") e mutazioni compensatorie *in cis*, tutti processi che portano alla ricostituzione di un allele *wild type*. Anche se la reversione del fenotipo cellulare può avere conseguenze importanti in termini di prognosi, non sono mai stati condotti studi prospettici per valutare l'evoluzione della malattia in questi pazienti (Gross et al., 2002).

Da quanto detto appare chiara la complessità della diagnosi molecolare della FA, che si traduce inevitabilmente in un incremento dei tempi di attesa per il paziente. Nel nostro laboratorio abbiamo affrontato il problema valutando non solo l'applicabilità dell'analisi dell'espressione proteica ma anche l'utilizzo delle tecnologie di NGS. Utilizzando il sistema Ion Torrent PGM™ (Life Technologies), siamo riusciti a caratterizzare già alcuni pazienti, tra cui uno appartenente al gruppo di complementazione FA-L. E' da sottolineare che, essendo stati descritti solo due casi (Ali et al., 2009, Meetei et al., 2003), *FANCL* è, tra i geni FA, quello meno frequentemente mutato e nella pratica diagnostica di *routine* pazienti con alterazioni in questo gene non sono identificati se non dopo l'analisi dei numerosi esoni codificanti dei geni FA, condizione improponibile a causa dei tempi e dei costi elevati dell'indagine.

Anche se utilizzo di tecniche di NGS risolveranno una serie di problematiche dovute all'eterogeneità genetica e alla presenza di molte mutazioni private, un altro limite del processo di diagnosi molecolare della FA, come pure di molte altre malattie, è rappresentato dall'individuazione di un numero relativamente elevato di sostituzioni aminoacidiche. Rappresentano, per esempio, nella nostra coorte il 22% di tutte le varianti identificate nel gene FANCA. Mentre è generalmente indiscusso l'effetto deleterio delle ampie delezioni e delle mutazioni nonsense o *frameshift*, il ruolo patogenetico delle alterazioni *missense*, al contrario, non è immediatamente evidente. Di solito, sono considerate patogenetiche le varianti non incluse nel database SNP. Tuttavia, fintanto che i progetti di sequenziamento su larga scala, come ad esempio il 1.000 Genomes (<http://www.1000genomes.org/>) o l'Exome Sequencing Project (<http://evs.gs.washington.edu/EVS/>) sono in corso, accadrà che anche le varianti patogenetiche più rare saranno annotate nei database di SNP.

La patogenicità delle varianti missenso dovrebbe essere confermata da studi funzionali che prevedono indagini complesse non sempre disponibili nei laboratori che si occupano di diagnosi molecolare. In alternativa, si può ricorrere all'utilizzo di strumenti bioinformatici. Sono stati sviluppati numerosi *software* in grado di predire l'effetto delle sostituzioni aminoacidiche sulla funzione, prendendo in considerazione le caratteristiche chimico-fisiche delle proteine stesse, nonché la conservazione dei residui aminoacidici e comparando gli effetti delle mutazioni sulla proteina mutante. I *software* sono di facile utilizzo e forniscono un risultato in tempi molto rapidi, anche se possono fornire predizioni non corrette (Hicks et al., 2011).

Pertanto, utilizzando sia i dati genetici che quelli bioinformatici integrando più strumenti di predizione, abbiamo cercato di classificare le sostituzioni aminoacidiche identificate nella nostra casistica in base alla loro patogenicità predetta. È interessante notare che, per la maggior parte delle varianti i due approcci, genetico e informatico, concordano nel definire la patogenicità o meno di una determinata variante.

Gli studi funzionali con l'analisi di complementazione rappresentano, tuttavia, l'unico strumento nelle condizioni in cui l'identificazione di alleli missenso non può garantire la corretta assegnazione di un gruppo di complementazione. Un'attribuzione non corretta potrebbe avere importanti conseguenze in diagnosi prenatale e nell'identificazione dei portatori asintomatici tra i membri della famiglia.

Sebbene le varianti con incerto significato (VUS) continueranno a rappresentare un problema, l'applicazione di sistemi per il sequenziamento di nuova generazione (NGS) avrà un forte impatto nei casi in cui, come la FA, i geni spiegano la malattia nella quasi totalità delle famiglie. Approcci di NGS sono stati precedentemente utilizzati con successo nella FA (Ameziane et al., 2012, Chandrasekharappa et al., 2013) e, in accordo con questi lavori, la piattaforma Ion Torrent PGM™ da noi utilizzata è particolarmente promettente. Siamo pertanto fiduciosi di poter applicare questa procedura nello screening molecolare della FA senza alcuna indagine pre-screening di completazione o espressione proteica.

6. CONCLUSIONI

CONCLUSIONI

Lo scopo di questa tesi è stato quello di definire la miglior strategia per la diagnosi molecolare nella FA valutando la possibilità di sfruttare l'analisi proteica per l'identificazione del gene candidato. Sebbene tale approccio abbia dimostrato dei limiti, lo studio dell'espressione proteica merita di essere preso in considerazione, sia perché può fornire informazioni utili allo screening di mutazioni, sia per la prospettiva di identificare mutazioni che potrebbero associarsi ad un fenotipo del paziente meno grave. A questo proposito, l'istituzione di un database internazionale per la raccolta dei dati clinici, molecolari e di espressione proteica rappresenterà un requisito fondamentale per accertare eventuali correlazioni genotipo-fenotipo. Sarà, inoltre, opportuno avviare studi per determinare se alcune mutazioni, tra quelle associate alla sintesi di proteine stabili, svolgano o meno un'attività residua da ricercare sia nell'ambito che al di fuori del pathway FA (per esempio a livello mitocondriale).

Il lavoro di questa tesi ha permesso, inoltre, di contribuire alla caratterizzazione e agli effetti delle varianti identificate per definirne la patogenicità. Questo aspetto rimane di primaria importanza perché, sebbene le tecniche di sequenziamento di nuova generazione (NGS) renderanno sempre più rapida ed efficace la ricerca delle mutazioni nella FA, l'identificazione di una variante non sarà sempre sufficiente a chiarirne il ruolo nella malattia.

7. BIBLIOGRAFIA

BIBLIOGRAFIA

- Adachi, D., Oda, T., Yagasaki, H., Nakasato, K., Taniguchi, T., et al. 2002. Heterogeneous activation of the Fanconi anemia pathway by patient- derived FANCA mutants. *Hum Mol Genet*, 11, 3125-34.
- Ali, A. M., Kirby, M., Jansen, M., Lach, F. P., Schulte, J., et al. 2009. Identification and characterization of mutations in FANCL gene: a second case of Fanconi anemia belonging to FA-L complementation group. *Hum Mutat*, 30, E761-70.
- Alpi, A. F. & Patel, K. J. 2009. Monoubiquitylation in the Fanconi anemia DNA damage response pathway. *DNA Repair (Amst)*, 8, 430-5.
- Alter, B. P. 2007. Diagnosis, genetics, and management of inherited bone marrow failure syndromes. *Hematology Am Soc Hematol Educ Program*, 29-39.
- Alter, B. P., Greene, M. H., Velazquez, I. & Rosenberg, P. S. 2003. Cancer in Fanconi anemia. *Blood*, 101, 2072.
- Ameziane, N., Errami, A., Leveille, F., Fontaine, C., De Vries, Y., et al. 2008. Genetic subtyping of Fanconi anemia by comprehensive mutation screening. *Hum Mutat*, 29, 159-66.
- Ameziane, N., Sie, D., Dentro, S., Ariyurek, Y., Kerkhoven, L., et al. 2012. Diagnosis of fanconi anemia: mutation analysis by next-generation sequencing. *Anemia*, 2012, 132856.
- Auerbach, A. D. 2009. Fanconi anemia and its diagnosis. *Mutat Res*, 668, 4-10.
- Borriello, A., Locasciulli, A., Bianco, A. M., Criscuolo, M., Conti, V., et al. 2007. A novel Leu153Ser mutation of the Fanconi anemia FANCD2 gene is associated with severe chemotherapy toxicity in a pediatric T-cell acute lymphoblastic leukemia. *Leukemia*, 21, 72-78.
- Castella, M., Pujol, R., Callen, E., Trujillo, J. P., Casado, J. A., et al. 2011. Origin, functional role, and clinical impact of Fanconi anemia FANCA mutations. *Blood*, 117, 3759-69.
- Centra, M., Memeo, E., D'apolito, M., Savino, M., Ianzano, L., et al. 1998. Fine exon-intron structure of the Fanconi anemia group A (FAA) gene and characterization of two genomic deletions. *Genomics*, 51, 463-7.
- Chandrasekharappa, S. C., Lach, F. P., Kimble, D. C., Kamat, A., Teer, J. K., et al. 2013. Massively parallel sequencing, aCGH, and RNA-Seq technologies provide a comprehensive molecular diagnosis of Fanconi anemia. *Blood*, 121, e138-48.
- Cohn, M. A. & D'andrea, A. D. 2008. Chromatin recruitment of DNA repair proteins: lessons from the fanconi anemia and double-strand break repair pathways. *Mol Cell*, 32, 306-12.
- Collins, N. B., Wilson, J. B., Bush, T., Thomashevski, A., Roberts, K. J., et al. 2009. ATR-dependent phosphorylation of FANCA on serine 1449 after DNA damage is important for FA pathway function. *Blood*, 113, 2181-90.

- Collis, S. J., Barber, L. J., Ward, J. D., Martin, J. S. & Boulton, S. J. 2006. C. elegans FANCD2 responds to replication stress and functions in interstrand cross-link repair. *DNA Repair (Amst)*, 5, 1398-406.
- Collis, S. J., Ciccica, A., Deans, A. J., Horejsi, Z., Martin, J. S., et al. 2008. FANCM and FAAP24 function in ATR-mediated checkpoint signaling independently of the Fanconi anemia core complex. *Mol Cell*, 32, 313-24.
- De Rocco, D., Bottega, R., Cappelli, E., Cavani, S., Criscuolo, M., et al. 2014. Molecular analysis of Fanconi anemia: the experience of the Bone Marrow Failure Study Group of the Italian Association of Pediatric Onco-Hematology. *Haematologica*, In Press.
- De Winter, J. P. & Joenje, H. 2009. The genetic and molecular basis of Fanconi anemia. *Mutat Res*, 668, 11-9.
- Fagerlie, S. R., Koretsky, T., Torok-Storb, B. & Bagby, G. C. 2004. Impaired type I IFN-induced Jak/STAT signaling in FA-C cells and abnormal CD4+ Th cell subsets in Fancg^{-/-} mice. *J Immunol*, 173, 3863-70.
- Faivre, L., Guardiola, P., Lewis, C., Dokal, I., Ebell, W., et al. 2000. Association of complementation group and mutation type with clinical outcome in fanconi anemia. European Fanconi Anemia Research Group. *Blood*, 96, 4064-70.
- Ferrer, M., Rodriguez, J. A., Spierings, E. A., De Winter, J. P., Giaccone, G., et al. 2005. Identification of multiple nuclear export sequences in Fanconi anemia group A protein that contribute to CRM1-dependent nuclear export. *Hum Mol Genet*, 14, 1271-81.
- Garcia-Higuera, I., Kuang, Y., Denham, J. & D'andrea, A. D. 2000. The fanconi anemia proteins FANCA and FANCG stabilize each other and promote the nuclear accumulation of the Fanconi anemia complex. *Blood*, 96, 3224-30.
- Gluckman, E. 1990. Radiosensitivity in Fanconi anemia: application to the conditioning for bone marrow transplantation. *Radiother Oncol*, 18 Suppl 1, 88-93.
- Gross, M., Hanenberg, H., Lobitz, S., Friedl, R., Herterich, S., et al. 2002. Reverse mosaicism in Fanconi anemia: natural gene therapy via molecular self-correction. *Cytogenet Genome Res*, 98, 126-35.
- Hanson, H., Mathew, C. G., Docherty, Z. & Mackie Ogilvie, C. 2001. Telomere shortening in Fanconi anaemia demonstrated by a direct FISH approach. *Cytogenet Cell Genet*, 93, 203-6.
- Hicks, S., Wheeler, D. A., Plon, S. E. & Kimmel, M. 2011. Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. *Hum Mutat*, 32, 661-8.
- Hirschhorn, R. 2003. In vivo reversion to normal of inherited mutations in humans. *J Med Genet*, 40, 721-8.
- Hodson, C. & Walden, H. 2012. Towards a molecular understanding of the fanconi anemia core complex. *Anemia*, 2012, 926787.
- Jacquemont, C. & Taniguchi, T. 2007. The Fanconi anemia pathway and ubiquitin. *BMC Biochem*, 8 Suppl 1, S10.
- Joenje, H., Lo Ten Foe, J. R., Oostra, A. B., Van Berkel, C. G., Rooimans, M. A., et al. 1995. Classification of Fanconi anemia patients by complementation analysis: evidence for a fifth genetic subtype. *Blood*, 86, 2156-60.
- Kalb, R., Neveling, K., Hoehn, H., Schneider, H., Linka, Y., et al. 2007. Hypomorphic mutations in the gene encoding a key Fanconi anemia protein, FANCD2, sustain a

- significant group of FA-D2 patients with severe phenotype. *Am J Hum Genet*, 80, 895-910.
- Kee, Y. & D'andrea, A. D. 2012. Molecular pathogenesis and clinical management of Fanconi anemia. *J Clin Invest*, 122, 3799-806.
- Kervestin, S. & Jacobson, A. 2012. NMD: a multifaceted response to premature translational termination. *Nat Rev Mol Cell Biol*, 13, 700-12.
- Kim, H. & D'andrea, A. D. 2012. Regulation of DNA cross-link repair by the Fanconi anemia/BRCA pathway. *Genes Dev*, 26, 1393-408.
- Kim, J. M., Kee, Y., Gurtan, A. & D'andrea, A. D. 2008. Cell cycle-dependent chromatin loading of the Fanconi anemia core complex by FANCM/FAAP24. *Blood*, 111, 5215-22.
- Kuliev, A., Rechitsky, S., Tur-Kaspa, I. & Verlinsky, Y. 2005. Preimplantation genetics: Improving access to stem cell therapy. *Ann N Y Acad Sci*, 1054, 223-7.
- Langevin, F., Crossan, G. P., Rosado, I. V., Arends, M. J. & Patel, K. J. 2011. Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice. *Nature*, 475, 53-8.
- Levy-Lahad, E. 2010. Fanconi anemia and breast cancer susceptibility meet again. *Nat Genet*, 42, 368-9.
- Lightfoot, J., Alon, N., Bosnoyan-Collins, L. & Buchwald, M. 1999. Characterization of regions functional in the nuclear localization of the Fanconi anemia group A protein. *Hum Mol Genet*, 8, 1007-15.
- Lo Ten Foe, J. R., Roomans, M. A., Bosnoyan-Collins, L., Alon, N., Wijker, M., et al. 1996. Expression cloning of a cDNA for the major Fanconi anaemia gene, FAA. *Nat Genet*, 14, 320-3.
- Macmillan, M. L., Hughes, M. R., Agarwal, S. & Daley, G. Q. 2011. Cellular therapy for fanconi anemia: the past, present, and future. *Biol Blood Marrow Transplant*, 17, S109-14.
- Mankad, A., Taniguchi, T., Cox, B., Akkari, Y., Rathbun, R. K., et al. 2006. Natural gene therapy in monozygotic twins with Fanconi anemia. *Blood*, 107, 3084-90.
- Mathew, C. G. 2006. Fanconi anaemia genes and susceptibility to cancer. *Oncogene*, 25, 5875-84.
- Matsushita, N., Kitao, H., Ishiai, M., Nagashima, N., Hirano, S., et al. 2005. A FancD2-monoubiquitin fusion reveals hidden functions of Fanconi anemia core complex in DNA repair. *Mol Cell*, 19, 841-7.
- Meetej, A. R., De Winter, J. P., Medhurst, A. L., Wallisch, M., Waisfisz, Q., et al. 2003. A novel ubiquitin ligase is deficient in Fanconi anemia. *Nat Genet*, 35, 165-70.
- Moldovan, G. L. & D'andrea, A. D. 2009. How the fanconi anemia pathway guards the genome. *Annu Rev Genet*, 43, 223-49.
- Morgan, N. V., Tipping, A. J., Joenje, H. & Mathew, C. G. 1999. High frequency of large intragenic deletions in the Fanconi anemia group A gene. *Am J Hum Genet*, 65, 1330-41.
- Mukhopadhyay, S. S., Leung, K. S., Hicks, M. J., Hastings, P. J., Youssoufian, H., et al. 2006. Defective mitochondrial peroxiredoxin-3 results in sensitivity to oxidative stress in Fanconi anemia. *J Cell Biol*, 175, 225-35.
- Pagano, G., Talamanca, A. A., Castello, G., D'ischia, M., Pallardo, F. V., et al. 2013. From clinical description, to in vitro and animal studies, and backward to patients:

- oxidative stress and mitochondrial dysfunction in Fanconi anemia. *Free Radic Biol Med*, 58, 118-25.
- Pagano, G., Talamanca, A. A., Castello, G., Pallardo, F. V., Zatterale, A., et al. 2012. Oxidative stress in Fanconi anaemia: from cells and molecules towards prospects in clinical management. *Biol Chem*, 393, 11-21.
- Patel, K. J. & Joenje, H. 2007. Fanconi anemia and DNA replication repair. *DNA Repair (Amst)*, 6, 885-90.
- Qiao, F., Moss, A. & Kupfer, G. M. 2001. Fanconi anemia proteins localize to chromatin and the nuclear matrix in a DNA damage- and cell cycle-regulated manner. *J Biol Chem*, 276, 23391-6.
- Raya, A., Rodriguez-Piza, I., Guenechea, G., Vassena, R., Navarro, S., et al. 2009. Disease-corrected haematopoietic progenitors from Fanconi anaemia induced pluripotent stem cells. *Nature*, 460, 53-9.
- Robinton, D. A. & Daley, G. Q. 2012. The promise of induced pluripotent stem cells in research and therapy. *Nature*, 481, 295-305.
- Savino, M., Borriello, A., D'apolito, M., Criscuolo, M., Del Vecchio, M., et al. 2003. Spectrum of FANCA mutations in Italian Fanconi anemia patients: identification of six novel alleles and phenotypic characterization of the S858R variant. *Hum Mutat*, 22, 338-9.
- Savino, M., Ianzano, L., Strippoli, P., Ramenghi, U., Arslanian, A., et al. 1997. Mutations of the Fanconi anemia group A gene (FAA) in Italian patients. *Am J Hum Genet*, 61, 1246-53.
- Steffann, J., Frydman, N., Burlet, P., Gigarel, N., Feyereisen, E., et al. 2005. [Extending preimplantation genetic diagnosis to HLA typing: the Paris experience]. *Gynecol Obstet Fertil*, 33, 824-7.
- Suhasini, A. N., Sommers, J. A., Mason, A. C., Voloshin, O. N., Camerini-Otero, R. D., et al. 2009. FANCI helicase uniquely senses oxidative base damage in either strand of duplex DNA and is stimulated by replication protein A to unwind the damaged DNA substrate in a strand-specific manner. *J Biol Chem*, 284, 18458-70.
- Tamary, H., Bar-Yam, R., Shalmon, L., Rachavi, G., Krostichevsky, M., et al. 2000. Fanconi anaemia group A (FANCA) mutations in Israeli non-Ashkenazi Jewish patients. *Br J Haematol*, 111, 338-43.
- Taniguchi, T. & D'andrea, A. D. 2006. Molecular pathogenesis of Fanconi anemia: recent progress. *Blood*, 107, 4223-33.
- Tinsley, J. M., Blake, D. J., Pearce, M., Knight, A. E., Kendrick-Jones, J., et al. 1993. Dystrophin and related proteins. *Curr Opin Genet Dev*, 3, 484-90.
- Tischkowitz, M. D. & Hodgson, S. V. 2003. Fanconi anaemia. *J Med Genet*, 40, 1-10.
- Tolar, J., Adair, J. E., Antoniou, M., Bartholomae, C. C., Becker, P. S., et al. 2011. Stem cell gene therapy for fanconi anemia: report from the 1st international Fanconi anemia gene therapy working group meeting. *Mol Ther*, 19, 1193-8.
- Valeri, A., Martinez, S., Casado, J. A. & Bueren, J. A. 2011. Fanconi anaemia: from a monogenic disease to sporadic cancer. *Clin Transl Oncol*, 13, 215-21.
- Waisfisz, Q., Morgan, N. V., Savino, M., De Winter, J. P., Van Berkel, C. G., et al. 1999. Spontaneous functional correction of homozygous fanconi anaemia alleles reveals novel mechanistic basis for reverse mosaicism. *Nat Genet*, 22, 379-83.
- Wang, W. 2007. Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins. *Nat Rev Genet*, 8, 735-48.

- Wijker, M., Morgan, N. V., Herterich, S., Van Berkel, C. G., Tipping, A. J., et al. 1999. Heterogeneous spectrum of mutations in the Fanconi anaemia group A gene. *Eur J Hum Genet*, 7, 52-9.
- Zhang, Q. S., Marquez-Loza, L., Eaton, L., Duncan, A. W., Goldman, D. C., et al. 2010. *Fancd2*^{-/-} mice have hematopoietic defects that can be partially corrected by resveratrol. *Blood*, 116, 5140-8.