UNIVERSITÀ
DI PAVIA

Dipartimento di Biologia e Biotecnologie “L. Spallanzani”

Laurea Magistrale in Biologia Sperimentale e Applicata

ALTERAZIONI TRASCRIZIONALI NELLA RISPOSTA AL DANNO AL DNA ASSOCIATA ALL’ANEMIA DI FANCONI

Relatore:
Prof.ssa Donata Orioli

Correlatore:
Dott.ssa Silvia Bione

Tesi Sperimentale di:
Alice Triunfo

INDICE

1. INTRODUZIONE 4
1.1 L'anemia di Fanconi: aspetti clinici e molecolari 4
1.2 La risposta al danno al DNA 5
1.3 Tipologie di danno al DNA: DSB e ICL 7
1.3.1 Le rotture del DNA a doppio filamento 7
1.3.2 I legami crociati interfilamento 8
1.4 Architettura molecolare e funzioni dei geni FA 9
1.4.1 Composizione genica e regolazione trascrizionale 9
1.4.2 Organizzazione funzionale del pathway FA/BRCA 17
1.4.3 Inversione e protezione delle forcelle replicative 20
1.5 Long non coding RNA e regolazione trascrizionale nella DDR 22

2. OBIETTIVI DELLA TESI 24

3. MATERIALI E METODI 25
3.1 Allestimento del sistema cellulare 25
3.1.1 Trattamento con MMC 25
3.1.2 Incorporazione e colorazione BrdU 26
3.1.3 Estrazione dell'RNA 27
3.1.4 Retrotrascrizione 28
3.1.5 Reazione di PCR e analisi del gel elettroforetico 29
3.1.6 Reazione di qPCR 30
3.2 Protocollo di analisi dei dati RNA-Seq 32
3.2.1 FastQC 32
3.2.2 RNA STAR 32
3.2.3 FeatureCounts 33
3.2.4 DESeq2 33
3.3 Analisi di arricchimento funzionale 34
3.4 Analisi dei lncRNA 35
3.5 Analisi integrata ChEA–ENCODE per la caratterizzazione dei promotori 35
3.6 Predizione dei ruoli biologici dei lncRNA tramite analisi KEGG 38

4. RISULTATI E DISCUSSIONE 39
4.1 Ottimizzazione di un sistema cellulare per l'induzione al danno al DNA 39
4.2 Analisi dei dati RNA-Seq 43
4.3 Analisi dei geni coinvolti nella via FA/BRCA 44
4.4 Analisi globale dell'espressione genica e funzionale 47

4.5 Analisi dei lncRNA differenzialmente espressi 53
4.6 Classificazione temporale dei pattern di espressione dei lncRNA e dei geni FA 55
4.7 Analisi dei fattori di trascrizione E2F1 ed E2F4 58
4.7.1 Incrocio tra siti di legame e potenziali promotori PCGs DEGs 61
4.7.2 Incrocio tra siti di legame e potenziali promotori di lncRNA DEGs 65

5. CONCLUSIONI 68

BIBLIOGRAFIA 71

1.INTRODUZIONE

1.1 L'anemia di Fanconi: aspetti clinici e molecolari

L'anemia di Fanconi (FA) è la sindrome da insufficienza midollare ereditaria più comune, è stata descritta per la prima volta nel 1927 dal Dr. Guido Fanconi, che osservò una famiglia di 3 fratelli che presentavano diverse anomalie fisiche e anemia perniciosa (Wiedemann HR, 1979).

La FA oggi è definita come una malattia genetica rara, caratterizzata da un modello di ereditarietà prevalentemente autosomico recessivo, che colpisce le proteine coinvolte nella riparazione del DNA e nella regolazione del ciclo cellulare. La fisiopatologia sottostante alla malattia è meglio compresa per quanto riguarda il sistema emopoietico. Infatti, le cellule staminali emopoietiche (HSC) con un *pathway* FA difettoso presentano livelli aumentati di danno al DNA con successiva morte cellulare data dall'attivazione del percorso p53, che avvia l'apoptosi.
La caratteristica della malattia è una progressiva deplezione della funzionalità del midollo osseo accompagnata da pancitopenia, malformazioni congenite ed un alto rischio di sviluppare tumori solidi ed ematologici (Malouf C. et al., 2022).

A causa dell'instabilità cromosomica negli individui affetti da questo disturbo, le comuni strategie chemioterapiche che inducono danni al DNA, risultano spesso inefficaci o controindicate. Le terapie si limitano al sollievo dei sintomi e l'aspettativa di vita media dei pazienti è compresa tra i 20 e i 30 anni, con pochi casi che superano i 40-50 anni (Mehta PA & Tolar J, 2018).

Una caratteristica cellulare tipica delle cellule derivate da pazienti con FA è che sono specificamente ipersensibili agli agenti antitumorali come mitomicina C, cisplatino e diepossibutano (D'Andrea e Grompe, 2003), che aumentano la rottura cromosomica, arrestano numerose cellule nella fase G2/M e inducono l'apoptosi. Clinicamente, l'elevato accumulo di prodotti di rottura cromosomica che si verifica durante il test di fragilità cromosomica con diepossibutano o mitomicina C, consente la diagnosi di FA (Auerbach, 2009). Questo test rimane il gold standard diagnostico, pur dovendo considerare che il mosaicismo somatico può attenuare la valutazione in alcuni individui (Mehta PA & Tolar J, 2018). Con l'avvento delle tecniche di sequenziamento di nuova generazione (NGS) oggi è possibile analizzare pannelli di geni o l'esoma clinico per identificare varianti nei geni associati alla FA. Per la valutazione dei portatori o per test prenatali, il sequenziamento mirato di uno specifico gene o di una mutazione nota rimane la strategia più indicata (Dufour C e Pierri F, 2022).

Dopo la diagnosi, viene monitorata la progressione dell'aplasia midollare per stabilire il momento più appropriato al trapianto di cellule staminali emopoietiche, che andranno a correggere le principali complicanze ematologiche. Questo trattamento, tuttavia, non è sufficiente poiché persistono complicanze extramidollari e rischio oncologico che impongono sorveglianza clinica continua.

Negli ultimi anni per superare i limiti del trapianto e la necessità di un donatore compatibile, sono stati raggiunti progressi clinici sulla terapia genica autologa delle cellule staminali emopoietiche mediante vettori lentivirali, che correggendo il gene mutato, aiutano ad ottenere una popolazione stabile e funzionale del midollo osseo, come dimostrato in uno studio europeo in cui il 62,5% (5 su 8) dei pazienti ha mostrato benefici clinici senza precondizionamento o genotossicità. (Río et al., 2024 - Malouf et al., 2022 -Molina-Estevez, et al., 2015).

1.2 La risposta al danno al DNA

Le cellule eucariotiche sono frequentemente esposte a stimoli dannosi esogeni ed endogeni che causano danni al DNA e minacciano la stabilità genomica. (Chen et al. 2024; Bauer et al. 2015). Una cascata di segnalazione chiamata risposta al danno del DNA (DDR) viene attivata per rilevare e riparare le lesioni genetiche o per indurre apoptosi, proteggendo in questo modo l'integrità del genoma. Il sistema di riparazione così coordina il riconoscimento delle lesioni, l'amplificazione del segnale di allarme, il controllo dei *checkpoint* e l'attivazione dei meccanismi di riparazione appropriati. A monte vi sono differenti sensori a seconda del tipo di lesione, gli enzimi attivati e reclutati dai sensori sono chinasi che fungono da trasduttori centrali: ATM (Ataxia-Telangiectasia Mutated), ad esempio, è responsabile della risposta alle rotture a doppio filamento del DNA (DSB) che presentano estremità libere e sono accessibili al macchinario di riparazione, mentre ATR (Ataxia Telangiectasia and Rad3-Related) si attiva in situazioni di stress replicativo che riguardano il DNA a singolo filamento (ssDNA) rivestito da RPA (Replication Protein A). Queste chinasi fosforilano numerosi target, tra cui mediatori come CHK1 (Checkpoint kinase 1) e CHK2 (Checkpoint kinase 2), che attivano i *checkpoint* del ciclo cellulare, un meccanismo essenziale per impedire la progressione del ciclo cellulare in presenza di danno al DNA.

Il *checkpoint* G1/S blocca l'ingresso in fase S per evitare la replicazione di DNA danneggiato, mentre quello in G2/M impedisce l'entrata in mitosi finché le lesioni non sono state riparate. L'attivazione di questo sistema di controllo consente alla cellula di arrestare temporaneamente il ciclo cellulare per attivare i meccanismi di riparazione appropriati. Per evitare un arresto

prolungato o un'uscita prematura dal *checkpoint*, questo sistema è sottoposto a una rigida regolazione (*Figura 1*). (Bartek & Lukas 2001).

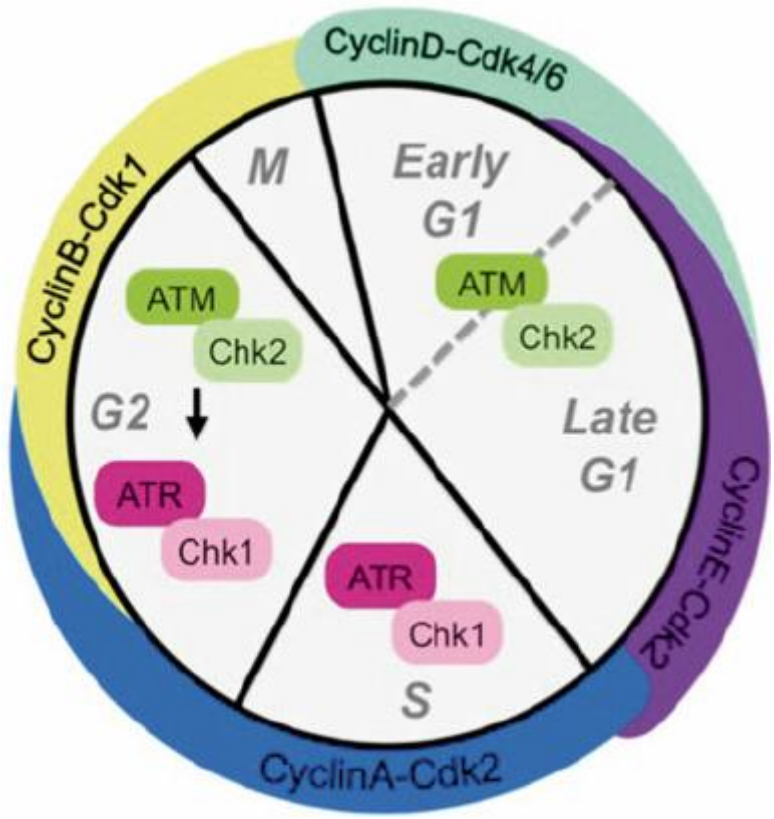


Figura 1 - Panoramica della segnalazione dei *checkpoint* del ciclo cellulare indotta dal danno al DNA dipendente da ATM e ATR (Jazayeri et al., 2005).

1.3 Tipologie di danno al DNA: DSB e ICL

1.3.1 Le rotture del DNA a doppio filamento

Le rotture del DNA a doppio filamento (DSB) rappresentano una delle forme più citotossiche di danno al DNA. Una singola DSB irrisolta è in grado di mettere a repentaglio la salute generale della cellula. La sua riparazione è mediata da due vie principali: la ricombinazione omologa (HR) e l'unione delle estremità non omologa (NHEJ).

NHEJ è una via attiva in tutto il ciclo cellulare ma è la via preferenziale in fase G1, quando non è presente un cromatide fratello per la ricombinazione omologa (Chapman et al., 2012). Si innesca con il rapido legame tra gli eterodimeri Ku70/Ku80 (X-ray repair cross-complementing 6 e 5) alle estremità del DSB, successivamente altri enzimi le sistemano se necessario e infine, una ligasi le salda assieme. È un meccanismo di riparazione rapido e non richiede un modello omologo, per questo motivo è più soggetto ad errori.

HR, invece, è una via ad alta fedeltà che richiede il cromatide fratello come stampo; pertanto, è preferenziale nella fase S e G2. Inizialmente viene rifinita la rottura: alcuni enzimi eliminano nucleotidi vicino al taglio creando un tratto di DNA a singolo filamento. Quel tratto viene protetto da RPA, che viene successivamente sostituito da RAD51 (radiation-sensitive 51); quest'ultimo forma un filamento che cerca la sequenza uguale sul cromatide fratello e permette di copiarla per ricostruire il tratto di DNA rotto. BRCA1 (Breast Cancer gene 1) favorisce l'avvio di questo processo, PALB2 (Partner And Localizer of BRCA2) collega BRCA1 a BRCA2 (Breast Cancer gene 2) con BRCA2 che aiuta a collocare e stabilizzare RAD51 sul DNA. Se uno di questi fattori manca, la HR non funziona bene e aumenta il rischio di errori e di instabilità del genoma. (Prakash et al., 2015; Foo & Xia 2022).

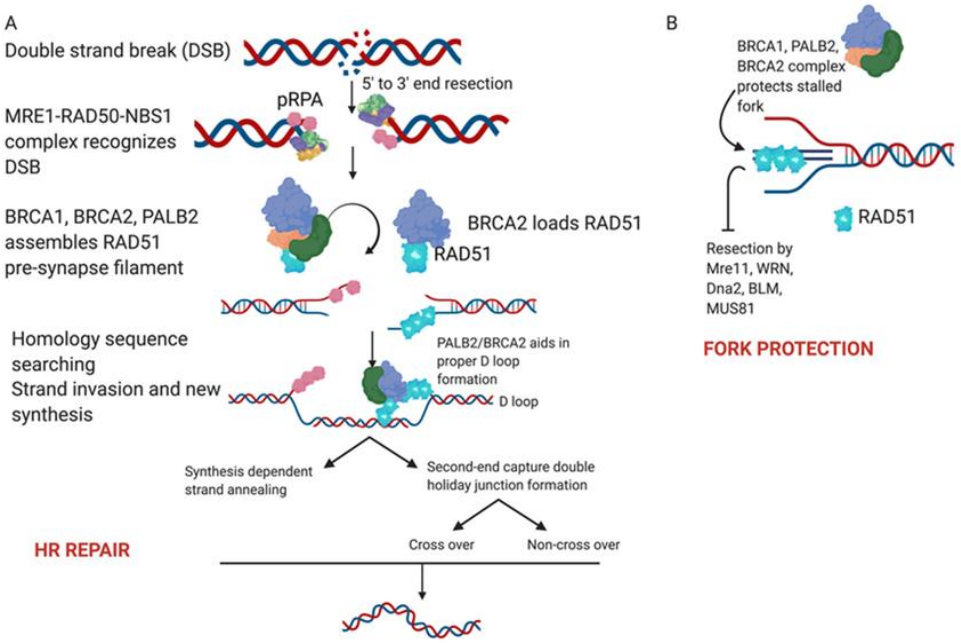


Figura 2 - Meccanismo della ricombinazione omologa nella riparazione delle rotture a doppio filamento. (Bouberhan et al., 2020)

1.3.2 I legami crociati interfilamento

La replicazione del genoma umano è un processo complesso che richiede l'attivazione e il mantenimento di forcelle replicative provenienti da migliaia di origini di replicazione durante la fase S; se queste incontrano ostacoli lungo il DNA, si bloccano e richiedono un'elaborazione rapida per impedirne il collasso, poichè questo porterebbe a generare un DSB che può contribuire ad eventi di ricombinazione spontanea (Bhattacharjee e Nandi, 2017).

Tra le principali cause di tali ostacoli vi sono i danni chimici al DNA, generati endogenamente o esogenamente da molecole e metaboliti che si legano covalentemente alla doppia elica. Tali danni si possono presentare sotto diverse forme, tra cui:

-molecole attaccate a singole basi (addotti di base)

-strutture ad anello (addotti ciclici)

-legami covalenti tra due basi della stessa elica (*crosslink* intrafilamento)

-legami covalenti tra le due eliche opposte del DNA (*interstrand crosslinks*, ICL)

-legami covalenti tra DNA e proteine (DNA–*protein crosslinks*)

Tra questi troviamo i legami crociati interfilamento (ICL), generati da reattivi capaci di legare due nucleotidi situati su filamenti opposti del DNA per formare un legame covalente tra le due eliche. I reattivi responsabili della formazione degli ICL si dividono in due categorie principali:

- Induttori endogeni: derivano dai normali processi fisiologici e comprendono metaboliti cellulari come i prodotti della perossidazione lipidica, del metabolismo degli amminoacidi o della demetilazione mediata dal citocromo P450, oltre ai radicali liberi. Tra le molecole più reattive vi sono le aldeidi (formaldeide, acetaldeide) e vari composti azotati generati dal metabolismo di nitrati e nitriti, come l'ossido nitrico (NO) e l'acido nitroso.
- Induttori esogeni: provengono dall'ambiente o da trattamenti farmacologici e includono aldeidi ambientali, mostarde azotate, agenti chemioterapici a base di platino (come il cisplatino), psoraleni, cloroetil-nitrosouree e composti batterici quali la mitomicina C (MMC), l'azinomicina B e la colibactina.

In questo contesto, un ruolo cruciale è svolto dai geni FA, che codificano le proteine deputate alla risoluzione di queste lesioni e facilitano il riavvio delle forcelle di replicazione bloccate; quindi, ogni mutazione in questi geni porta a difetti nel *pathway*, con conseguente compromissione della riparazione degli ICL. La persistenza di questi danni causa rotture e riarrangiamenti cromosomici, traducendosi in un'instabilità genomica elevata che risulta particolarmente dannosa per i tessuti ad alta proliferazione. (Fang et al., 2025 - Shukla et al., 2013).

1.4 Architettura molecolare e funzioni dei geni FA

1.4.1 Composizione genica e regolazione trascrizionale

Esistono almeno 22 geni FA (*FANCA, B, C, D1, D2, E, F, G, I, J, L, M, N, O, P, Q, R, S, T, U, V, W*). Come mostrato in Tabella 1, ciascuno contribuisce a una parte specifica del *pathway* FA/BRCA. Tutti quanti vengono trascritti dalla RNA polimerasi II. Invece delle TATA o CAAT box, molti geni FA hanno la regione a monte del sito di inizio della trascrizione ricca in GC, tipica dei geni *housekeeping.* Questo li rende particolarmente suscettibili ad eventi di metilazione epigenetica. La trascrizione non è costante: in condizioni di stress genotossico intervengono fattori che modulano la loro espressione per favorire l'arresto del ciclo e la riparazione.

Trattandosi di geni altamente espressi nei tessuti proliferativi come midollo osseo, linfociti e cellule epiteliali, un'inadeguata riparazione e l'accumulo di danni porta al crollo delle cellule staminali ematopoietiche e all'insufficienza midollare. Infatti, il midollo osseo di questi

pazienti mostra una proliferazione ridotta rispetto ai soggetti sani, con un numero maggiore di cellule bloccate nelle fasi G0 e G1 del ciclo cellulare e un aumento dei segnali di danno al DNA, evidenziato dalla presenza dell'istone γH2AX. Queste cellule sono anche incapaci di generare progenitori ematopoietici in coltura in vitro. (Ceccaldi et al. 2012).

Gene	Full Name	Chromosome Location	Function in DNA Damage Response
FANCA	FA complementation group A	16q24.3	Scaffold of FA core complex; essential for ICL repair
FANCB	FA complementation group B	Xq22.3	Core complex component; required for ubiquitination
FANCC	FA complementation group C	9q22.32	Core complex component; stabilizes FANCE
FANCD1	FA complementation group D1 (BRCA2)	13q13.1	Homologous recombination; interacts with RAD51
FANCD2	FA complementation group D2	3p25.3	Forms ID complex with FANCI; monoubiquitination for ICL repair
FANCE	FA complementation group E	6p21.31	Connects FANCC to FANCD2; stabilizes core complex
FANCF	FA complementation group F	11p14.3	Bridges FANCG and FANCA; stabilizes FA core
FANCG	FA complementation group G (XRCC9)	9p13.2	Core complex component; interacts with FANCF
FANCI	FA complementation group I	15q26.1	Partner of FANCD2 in ID complex; essential for FA pathway
FANCJ	FA complementation group J (BRIP1)	17q23.2	DNA helicase; interacts with BRCA1 for HR
FANCL	FA complementation group L	2p16.1	E3 ubiquitin ligase for FANCD2/FANCI monoubiquitination
FANCM	FA complementation group M	14q21.2	DNA translocase; recruits FA core to stalled forks
FANCN	FA complementation group N (PALB2)	16p12.2	BRCA2 partner; promotes homologous recombination
FANCO	FA complementation group O (RAD51C)	17q23.2	RAD51 paralog; involved in HR and ICL repair
FANCP	FA complementation group P (SLX4)	16p13.3	Scaffold for structure-specific nucleases in ICL repair
FANCQ	FA complementation group Q (ERCC4/XPF)	16p13.12	Endonuclease for ICL unhooking with ERCC1
FANCR	FA complementation group R (RAD51)	15q15.1	Central recombinase for homologous recombination
FANCS	FA complementation group S (BRCA1)	17q21.31	DNA repair and checkpoint signaling; HR
FANCT	FA complementation group T (UBE2T)	1q32.1	E2 ubiquitin-conjugating enzyme for FANCL activity
FANCU	FA complementation group U (XRCC2)	7q36.1	RAD51 paralog; promotes HR and fork stability
FANCV	FA complementation group V (MAD2L2/REV7)	1p36.22	Translesion synthesis; checkpoint control; ICL repair
FANCW	FA complementation group W (RFWD3)	16q23.1	E3 ubiquitin ligase; promotes HR and replication fork recovery

Tabella 1 - I 22 geni coinvolti nell'anemia di Fanconi (Fanconi Anemia Clinical Care Guidelines, 2020).

I 22 geni sono organizzati funzionalmente in tre livelli: un complesso centrale che attiva il segnale, un complesso ID2 che lo veicola direttamente sul DNA, e una serie di effettori deputati all'esecuzione della ricombinazione omologa, alle incisioni sugli ICL e al riavvio delle forcelle.

Il cuore del *pathway* è il complesso centrale FA, un assemblaggio di otto proteine: FANCA (Fanconi anemia complementation group A), FANCB (Fanconi anemia complementation group B), FANCC (Fanconi anemia complementation group C), FANCE (Fanconi anemia complementation group E), FANCF (Fanconi anemia complementation group F) , FANCG (Fanconi anemia complementation group G), FANCL (Fanconi anemia complementation group L) e FANCM (Fanconi anemia complementation group M) che ha un compito preciso: promuovere la monoubiquitinazione del complesso ID2, costituito da FANCD2 (Fanconi anemia complementation group D2) e FANCI (Fanconi anemia complementation group I). Questo singolo evento post-traduzionale consente a FANCD2 e FANCD1 (Fanconi anemia complementation group D1) di traslocare sulla cromatina, dove predispongono il reclutamento degli effettori a valle.

Questi effettori comprendono nucleasi e proteine regolatrici essenziali per le fasi di incisione e di ricombinazione omologa. Tra questi figurano FANCD1 (Fanconi anemia complementation group D1, storicamente noto come BRCA2) e FANCN (Fanconi anemia complementation group N, o PALB2), che formano un complesso molecolare che coordina il collocamento di RAD51 sul DNA; FANCJ (Fanconi anemia complementation group J, o BRIP1), una DNA elicasi che risolve conformazioni non canoniche del DNA; e FANCO (Fanconi anemia complementation group O, o RAD51C), FANCR (Fanconi anemia complementation group R, o RAD51) e FANCS (Fanconi anemia complementation group S, o BRCA1), componenti del complesso della ricombinazione. Il processo di taglio e incisione delle strutture con incroci inter-filamento è invece orchestrato dalle nucleasi FANCQ (Fanconi anemia complementation group Q, o ERCC4/XPF) e FANCU (Fanconi anemia complementation group U, o XRCC2). Infine, la via è regolata e completata da FANCT (Fanconi anemia complementation group T, o UBE2W), FANCV (Fanconi anemia complementation group V, o REV7) e FANCW (Fanconi anemia complementation group W, o RFWD3), che garantiscono la corretta risoluzione del segnale di riparazione.

Di seguito vengono analizzati nel dettaglio i singoli componenti del complesso, con le loro caratteristiche strutturali e il ruolo funzionale:

FANCA è un componente strutturale essenziale per l'integrità e funzione del complesso centrale. si configura come una proteina prevalentemente α-elicoidale, con un peso molecolare di circa 163 kDa. Questo si organizza in tre macro-regioni funzionali ed evolutivamente conservate:
Il dominio N-terminale è caratterizzato da un'estesa piattaforma di ripetizioni HEAT, questa regione è deputata al legame con FANCG, che stabilizza la conformazione aperta di FANCA e ne previene la degradazione prematura nel citoplasma. Il dominio centrale contiene il segnale di localizzazione nucleare (NLS). Questa regione, una volta che la proteina è stata attivata dallo stress replicativo, espone i residui necessari al trasporto nel nucleo. Il dominio C-terminale contiene un dominio di legame al DNA che mostra un'alta affinità per le giunzioni tra singolo

e doppio filamento (ssDNA-dsDNA). Inoltre, l'estremità C-terminale è la sede dell'interazione diretta con FANCF, agendo come un vero e proprio gancio molecolare che permette l'ancoraggio di FANCA al resto del complesso nucleare centrale. (Ianzano et al. 1997).

FANCB è l'unico gene della via FA a essere localizzato sul cromosoma X (X-linked). Questa caratteristica spiega perché tutti i casi documentati di pazienti con mutazioni in *FANCB* siano esclusivamente di sesso maschile. A differenza di *FANCA*, che opera come omodimero indipendente nel reclutamento iniziale, si associa in modo costitutivo con due proteine nel citoplasma, formando il cosiddetto sottocomplesso B-L-100 (costituito da FANCB, FANCL e FAAP100) in cui due unità di FANCB interagiscono centralmente fungendo da asse di simmetria per la dimerizzazione del sottocomplesso. Da un punto di vista strutturale, FANCB si configura come una proteina da impalcatura ricca di ripetizioni HEAT, priva di attività catalitica intrinseca.

FANCC pur essendo parte del complesso centrale nucleare, si localizza anche nel citoplasma e partecipa a circuiti di segnalazione come JAK/STAT e vie apoptotiche. Il gene può produrre diversi tipi di mRNA attraverso meccanismi di splicing alternativo che coinvolgono esoni non codificanti al 5' UTR e l'utilizzo di differenti segnali di poliadenilazione. In particolare, sono stati identificati due esoni alternativi nella regione iniziale del gene. La sua architettura terziaria è formata da alfa eliche strutturate in ripetizioni HEAT. Il dominio N-terminale è deputato all'interazione costitutiva citoplasmatica con FANCG e FANCF, fondamentale per stabilizzare conformazionalmente FANCC, proteggendolo dalla degradazione prematura prima del trasporto nel nucleo. Il dominio centrale contiene la regione di legame diretto per FANCE e il dominio C-terminale oltre a consolidare l'ancoraggio al complesso centrale nucleare, partecipa a vie di segnalazione citoplasmatiche extranucleari (risposta allo stress ossidativo e fattori anti-apoptotici) (Savoia et al. 1995).

FANCE svolge il ruolo di proteina adattatrice all'interno della via poichè promuove la traslocazione nucleare di FANCC, tramite il dominio N-terminale. Il dominio centrale invece è deputato al legame diretto con FANCD2. Il dominio C-terminale partecipa all'ancoraggio strutturale con le altre componenti regolatorie del complesso centrale (in particolare FANCB e FANCL), consolidando la stabilità del macchinario proteico sul sito del danno al DNA.

FANCF, agisce come proteina da impalcatura, stabilizzando il complesso centrale. Il dominio N-terminale è deputato all'interazione diretta con il modulo formato da FANCC e FANCG. Il dominio centrale permette a FANCF di fare da ponte proteico, coordinando la distanza tra il modulo sensore (FANCA/G) e il modulo catalitico (B-L-100). Il dominio C-terminale è la regione che interagisce direttamente con l'estremità C-terminale di FANCA. Questa interazione agisce come un punto di ancoraggio finale, serrando l'intero complesso centrale in una conformazione chiusa e stabile. (Saberwal et al., 2009).

FANCG (noto anche come XRCC9) codifica per una proteina strutturale di impalcatura, caratterizzata dalla presenza di 7 motivi ripetuti di tetratricopeptide (TPR), costituiti da 34 aminoacidi ciascuno. In condizioni basali, questi domini mediano le interazioni tra le proteine necessarie per l'assemblaggio del complesso centrale. La regolazione funzionale avviene mediante modificazioni post-traduzionali coordinate sia con la progressione del ciclo cellulare che con la risposta genotossica: Durante la fase M, i residui di serina 383 (Ser383) e serina 387 (Ser387) vengono fosforilati, un evento mediato dalla chinasi dipendente da ciclina 1 (CDK1). Questa duplice fosforilazione determina il distacco di FANCG dalla cromatina al fine di disattivare temporaneamente la via di riparazione durante la divisione cellulare. Al contrario, in seguito a danno al DNA indotto da MMC, viene fosforilato il residuo N-terminale di serina 7 (Ser7). Questa specifica modificazione, guidata dalle chinasi di *checkpoint* del danno, è essenziale per indurre l'attivazione funzionale del complesso centrale e promuovere la successiva monoubiquitinatione del complesso FANCD2-FANCI sul sito della lesione.

FANCL rappresenta la subunità catalitica del complesso centrale, operando come una E3 ubiquitina ligasi. Mentre in vitro la proteina manifesta auto-ubiquitinazione, in vivo il suo ruolo è essenziale per la monoubiquitinazione di FANCD2. Matura tramite splicing alternativo, generando diverse isoforme. I domini ELF (ELF1 e ELF2): Posizionati all'N-terminale, sono domini strutturali che mediano l'interazione e l'ancoraggio stabile con le subunità FANCB e FAAP100. Insieme, formano il sub-complesso catalitico trimerico (B-L-100). Il dominio centrale di legame al substrato è la porzione che prende contatto con il modulo adattatore (FANCE-FANCC) per orientare nello spazio il complesso bersaglio. Il dominio RING finger al C-terminale è il vero e proprio sito catalitico. Questo dominio coordina due ioni zinco ed è deputato al reclutamento dell'enzima coniugante E2 legato all' ubiquitina, UBE2T (Ubiquitin-Conjugating Enzyme E2 T). Grazie a questa struttura, FANCL catalizza direttamente la monoubiquitinazione del dimero FANCD2-FANCI.

FANCM Presenta attività ATPasica indotta da ssDNA e dsDNA e attività di DNA traslocasi: ha un dominio N-terminale che agisce come una traslocasi del DNA, sfruttando l'idrolisi dell'ATP per muoversi lungo il filamento e promuovendo il modellamento delle forche di replicazione bloccate. La regione centrale di interazione contiene i siti di legame per il complesso centrale tramite FANCF e per le chinasi di *checkpoint* come ATR/ATRIP, fungendo da piattaforma di segnalazione per bloccare il ciclo cellulare in presenza di danno. Il dominio C-terminale è strutturalmente omologo alle nucleasi, ma è cataliticamente inattivo, ovvero mancano degli amminoacidi chiave per il taglio. Questo dominio si associa stabilmente al cofattore FAAP24. Il dimero FANCM-FAAP24 riconosce e si lega specificamente alle strutture del DNA aberranti strutture a Y, forche bloccate e *R-loops*. Sebbene le proprietà intrinseche di elicasi ed endonucleasi non siano ancora state dimostrate in modo diretto, forti evidenze filogenetiche ne suggeriscono la funzione: *FANCM* mostra infatti una stretta parentela con la proteina archeale Hef (Helicase-associated Endonuclease) deputata alla risoluzione delle

forche di replicazione, un'elevata omologia con le endonucleasi coinvolte nel pathway di riparazione per escissione di nucleotidi (NER) e con elicasi specializzate nel rimodellamento delle strutture replicative. Coerentemente con questo ruolo di sensore e coordinatore, FANCM va incontro a una rapida fosforilazione in risposta allo stress genotossico.

Il secondo gruppo è formato da *FANCI* e *FANCD2*, che agiscono come sensori diretti del danno. Vengono ubiquitinati e formano foci nucleari sui siti di ICL, rappresentando il nodo centrale che collega il complesso centrale agli effettori a valle.
FANCD2 è il fulcro operativo: nelle cellule normali esistono due isoforme, una non ubiquitinata (~155 kDa) e una monoubiquitinata (~162 kDa), dove la modifica cade sulla lisina 561. La regione C-terminale di FANCD2 contiene diversi residui amminoacidici acidi ed è indispensabile per il corretto reclutamento e la stabilizzazione del complesso FANCD2–FANCI sui siti di danno. FANCI è un fattore necessario per la sua monoubiquitinazione, pur non essendo una subunità del complesso centrale. Una prova di principio viene dalle cellule FA-I, che formano normalmente il complesso ma non monoubiquitinano FANCD2.

A valle si trovano gli effettori della ricombinazione omologa e delle incisioni sui legami crociati interfilamento. Tra questi troviamo FANCD1, il cui risvolto patologico è strettamente legato al dosaggio genico: i portatori monoallelici sono predisposti al carcinoma mammario e ovarico, mentre le mutazioni bialleliche causano il sottotipo FA-D1 dell'Anemia di Fanconi. La sua funzione principale consiste nella regolazione della riparazione per ricombinazione omologa, attuata attraverso il controllo della ricombinasi RAD51 o FANCR.
BRCA2 svolge ulteriori funzioni biologiche, tra cui la stabilizzazione delle forche di replicazione bloccate e la regolazione della citochinesi. Dal punto di vista strutturale, la proteina è caratterizzata da: 8 ripetizioni BRC, motivi amminoacidici deputati al legame con RAD51, che ne permettono la polimerizzazione su DNA a singolo filamento.
Una regione C-terminale organizzata in 5 domini, di cui uno elicoidale (HD), tre moduli di legame per oligonucleotidi/oligosaccaridi (OB1, OB2, OB3) e un dominio a torre (TD) inserito in OB2. Le pieghe OB mediano l'interazione con il ssDNA, mentre il dominio TD interagisce con il DNA a doppio filamento. Una regione di Interazione con DSS1 (Deleted in Split-Hand/Split-Foot Syndrome 1) che si associa ai domini HD, OB1 e OB2, stabilizzando la struttura di BRCA2; DSS1 costituisce inoltre un componente regolatorio del proteasoma 26S.

In parallelo, FANCJ, opera come un'ATPasi dipendente dal DNA e come una DNA elicasi con polarità 5'-3'. Dal punto di vista strutturale, la proteina presenta i sette motivi elicasici canonici affiancati da un'estensione C-terminale. FANCJ interagisce direttamente con il dominio BRCT di BRCA1, un legame strettamente dipendente dalla fosforilazione del residuo di serina 990 (Ser990). Pur rivestendo un ruolo di rilievo nel processo di ricombinazione omologa, l'attività di FANCJ non risulta necessaria per la monoubiquitinazione di FANCD2.

Attorno a BRCA1 e BRCA2 si organizza la proteina di impalcatura PALB2, mentre i paraloghi di RAD51, nello specifico RAD51C e XRCC2 contribuiscono attivamente alla formazione e

alla stabilizzazione dei filamenti di RAD51. Al fine di eseguire le incisioni enzimatiche necessarie all'apertura dei legami crociati inter-filamento, viene reclutato e attivato SLX4, il quale opera come vera e propria piattaforma nucleasica, agendo frequentemente in sinergia con il complesso XPF/ERCC4. Ulteriori modulatori molecolari, quali REV7/MAD2L2, sostengono i meccanismi di tolleranza al danno, garantendo la protezione e il successivo riavvio delle forche di replicazione bloccate; parallelamente, l'attività della E3 ubiquitina ligasi RFWD3, coordina e regola RAD51 durante le fasi di riparazione per ricombinazione omologa. (Taniguchi e D'Andrea, 2006).

Il gene *FANCA* rappresenta l'elemento più frequentemente alterato nell'Anemia di Fanconi, essendo responsabile del 60-70% dei casi totali della malattia. L'elevata incidenza di varianti patogenetiche a carico di questo locus è confermata dal registro pubblico ClinVar, che attualmente censisce 971 varianti patogene, un numero nettamente superiore rispetto a quello registrato per gli altri geni del *pathway*.

Dal punto di vista funzionale, la proteina costituisce un componente essenziale del complesso centrale, all'interno del quale stabilisce interazioni costitutive con FANCG e FANCF. Questo sottocomplesso si assembla inizialmente nel citoplasma per poi essere traslocato nel nucleo in risposta al danno al DNA. Sotto il profilo filogenetico e strutturale, FANCA è l'elemento più evolutivamente conservato e di maggiori dimensioni tra i geni della via di Fanconi. Mappato sul cromosoma 16, il locus si estende per circa 80 kilobasi (kb) e comprende 43 esoni, i quali codificano per una sequenza di 4,3 kb.

Per quanto riguarda le mutazioni: le ampie delezioni intrageniche rappresentano la classe di varianti patogenetiche più ricorrenti. L'insorgenza di tali macro-riarrangiamenti è legata all'elevata densità di ripetizioni Alu introniche; queste sequenze appartenenti alla famiglia dei retrotrasposoni SINE, agiscono da origini di eventi di ricombinazione cromosomica e crossing-over ineguale, determinando variazioni del numero di copie (CNV) la cui localizzazione coincide direttamente con i punti di rottura della delezione. Inoltre, gli elementi Alu possono alterare i normali meccanismi di splicing, inducendo il salto dell'esone con conseguente inclusione di esoni alternativi sia a monte, sia a valle.

Oltre a coordinare la riparazione dei legami crociati inter-filamento, FANCA protegge e stabilizza le forche di replicazione sottoposte a stress. Insieme a FANCM e al complesso BRCA1/BRCA2, la proteina previene la degradazione da parte delle esonucleasi del DNA neosintetizzato. Questa funzione protettiva si rivela essenziale anche in assenza di ICL, dunque ha un ruolo determinante nel preservare l'integrità strutturale delle forche in caso di stallo replicativo indotto da stress ossidativo o da deplezione del *pool* di deossiribonucleotidi (dNTP). (Schlacher et al., 2012).

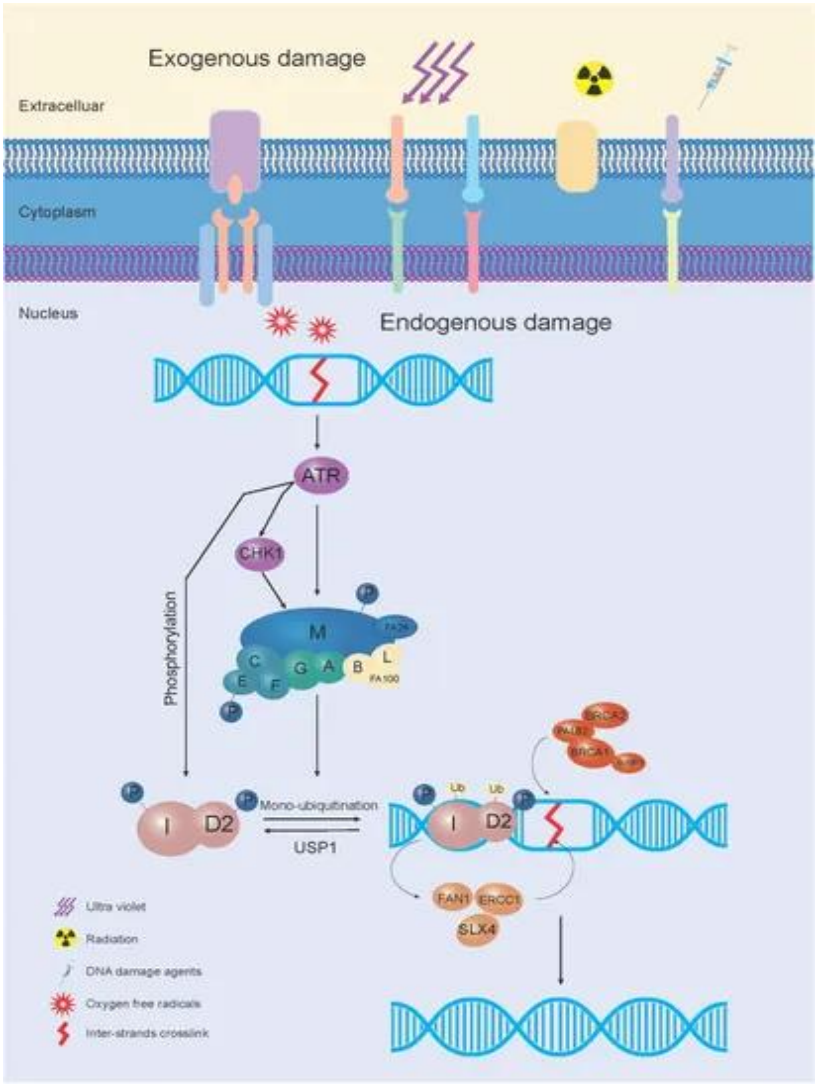


Figura 3 -Rappresentazione schematica del meccanismo del *pathway* FA utilizzato durante la riparazione del DNA (Fang et al., 2020)

1.4.2 Organizzazione funzionale del pathway FA/BRCA

Gli incroci interfilamento rappresentano ostacoli importanti per la replicazione perché legano covalentemente i due filamenti di DNA, impedendone la separazione e arrestando di fatto l'avanzamento della forcella. Quando una forcella in fase S imbatte in un ICL, sensori e rimodellatori contribuiscono a segnalare lo stallo del replisoma provocando la formazione di regioni di DNA a singolo filamento rapidamente rivestite da RPA. Questo rivestimento svolge una duplice funzione: da un lato protegge il filamento dall'azione esonucleasica, dall'altro recluta il complesso ATR-ATRIP e dei co-attivatori, avviando il *checkpoint* della replicazione e stabilizzando le forcelle per evitare collassi. Il reclutamento avviene tramite interazioni ATRIP–RPA–ssDNA, con contributi aggiuntivi mediati da altre proteine. (Ceccaldi, 2016; Tsui & Crismani, 2019).

In parallelo alla segnalazione dipendente da ATR, si attiva FANCM, un'ATPasi/elicasi specializzata nel riconoscere strutture di DNA anomale. FANCM si ancora fisicamente alla cromatina in corrispondenza della lesione, in un complesso con FAAP24 e con le proteine MHF1/MHF2 (FANCM-associated histone-fold protein 1 e 2) note anche come CENPS/CENPX, e interagisce con l'elicasi replicativa CMG (Cdc45-MCM2-7-GINS). In questo modo rimodella controllando la composizione del replisoma, blocca la forcella replicativa per evitare rotture a doppio filamento e ne facilita il successivo riavvio oltre l'ICL. Al contempo, attraverso collegamenti diretti, tra cui quello mediato da FANCF, promuove l'ingaggio del complesso centrale FA, portando operativamente la lesione dal livello di segnalazione al livello di esecuzione (Tsui & Crismani, 2019; Peake et al., 2022).

L'architettura del complesso centrale consente un coordinamento efficiente: il modulo BL100 (FANCB, FANCL, FAAP100) possiede l'attività catalitica, con FANCL che, come E3 ligasi dotata di dominio RING trasferisce una singola ubiquitina dai substrati degli enzimi E2, UBE2T/FANCT; il modulo CEF (FANCC, FANCE, FANCF) ha funzione di riconoscimento del substrato e collega il complesso centrale con FANCM, assicurando che i bersagli molecolari siano processati al momento giusto; il modulo AG20 (FANCA, FANCG, FAAP20) svolge un ruolo strutturale, stabilizzando il complesso e fungendo da ponte tra i moduli catalitici e di riconoscimento. L'azione concertata di questi sottocomplessi determina la monoubiquitinazione sequenziale di FANCD2 e poi di FANCI, evento cardinale della via di trasduzione del segnale (Peake et al., 2022).

La forma monoubiquitinata del dimero FANCD2–FANCI subisce un cambiamento conformazionale e si comporta come un morsetto scorrevole che si chiude sul DNA alla sede della lesione. Questo, stabilizza il sito di danno e funge da piattaforma di reclutamento per gli effettori a valle, in particolare per SLX4 (anche nota come FANCP), una proteina-*scaffold* che coordina l'azione di multiple endonucleasi e in risposta al reclutamento mediato da FANCD2/FANCI ubiquitinate, ingaggia l'eterodimero XPF–ERCC1 (FANCQ/ERCC4),

capace di incidere in prossimità dell'ICL, MUS81–EME1, che taglia strutture di giunzione che si formano durante il rimodellamento della forcella, e SLX1, deputata alla risoluzione di intermedi ramificati. La combinazione di queste attività consente il rilascio dell'ICL da uno dei due filamenti, generando tuttavia sul filamento opposto un adotto residuo legato a una singola base, che rappresenta un ingombro sterico e chimico non aggirabile dalle polimerasi replicative canoniche (Pol δ e Pol ε).

Per scongiurare un arresto prolungato, intervengono le polimerasi di sintesi di translesione a bassa fedeltà (TLS), polimerasi specializzate che replicano il DNA oltre la lesione, inserendo nucleotidi opposti all'ICL e allungando il filamento nascente. Le TLS polimerasi sono reclutate da PCNA (Proliferating Cell Nuclear Antigen) e dal complesso centrale FA ubiquinato. In modo coordinato, REV1 (Reversionless 1) inserisce un nucleotide opposto alla base modificata, mentre Pol ζ (composta dalle subunità REV3L/REV7 rispettivamente Subunità catalitica della DNA polimerasi zeta e Subunità regolatoria della DNA polimerasi zeta) estende la catena oltre il sito lesionato, ripristinando il substrato idoneo alla prosecuzione della sintesi convenzionale.

Il processo di inserzione è responsabile della maggior parte delle mutazioni osservate nei siti di riparazione ICL, ed è compensato a valle dalla riparazione ad alta fedeltà.

Le incisioni necessarie al rilascio, e in alcuni casi al collasso di forcelle, possono generare DSB che devono essere risolti senza introdurre ulteriore instabilità genomica. In fase S/G2 la via elettiva è la ricombinazione omologa (HR), che si apre con la resezione delle estremità del DSB per produrre ssDNA 3′; segue il caricamento di RAD51 sul ssDNA, evento dipendente dal complesso BRCA2–PALB2. L'estremità 3′ funge da primer per la sintesi di riparazione, una volta completata, gli intermedi ricombinativi vengono risolti, riportando la doppia elica a una struttura corretta. In questo tratto di percorso, l'elicasi FANCJ coopera con le proteine del mismatch repair MLH1 e PMS2 (mutL homolog 1 e Postmeiotic Segregation Increased 2) modulando sia la sintesi di translesione sia gli esiti della HR (Ceccaldi et al., 2016; Tsui & Crismani, 2019).

Completata la riparazione, deubiquitinasi quale USP1 (Ubiquitin-Specific Peptidase 1), rimuovono l'ubiquitina da FANCD2 e FANCI, permettendo l'apertura del morsetto e il suo distacco dal DNA. Questo rilascio è essenziale: libera fisicamente la sede di danno, normalizza la segnalazione locale e consente la ripresa coordinata della replicazione oltre la regione interessata (Ceccaldi et al., 2016; Peake et al., 2022; Che R, Zhang J et al., 2018).

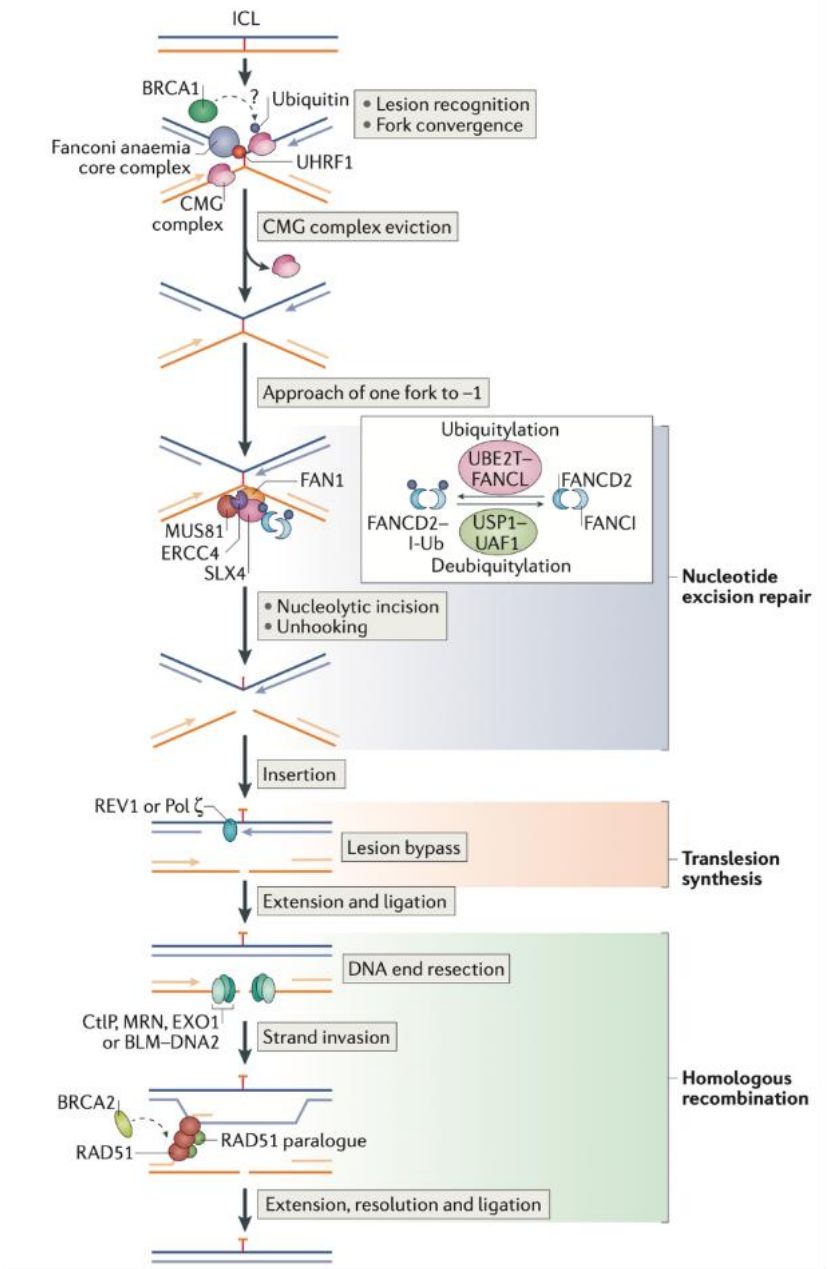


Figura 4 – Riparazione del percorso FA e ICLs (Ceccaldi et al., 2016)

1.4.3 Inversione e protezione delle forcelle replicative

Quando la replicazione rallenta — per carenza di dNTP indotta da idrossiurea oppure a causa di agenti che inducono legami crociati o da regioni genomicamente difficili come telomeri, centromeri o sequenze ripetute — si accumula DNA a singolo filamento. Questa condizione innesca l'inversione della forcella: un processo in cui i filamenti neosintetizzati si separano dai rispettivi stampi e si appaiano tra loro, convertendo la struttura replicativa a Y in una giunzione a quattro vie. In particolare, questo rimodellamento sequestra le estremità libere di DNA neosintetizzato, impedendone la degradazione incontrollata da parte delle nucleasi. Al contempo, il reclutamento di fattori di protezione come RAD51 e BRCA1/2 stabilizza l'intermedio, prevenendo il collasso della forcella in rotture del doppio filamento (DSB) e garantendo il tempo necessario per la risoluzione dell'ostacolo replicativo.
La transizione verso questa forma invertita è catalizzata da apposite traslocasi ed elicasi, mentre una classe distinta di elicasi si occupa di rimuovere la giunzione e riaprire la forcella quando le condizioni lo consentono.

La Figura 5 sintetizza il processo meccanico che porta alla regressione e successiva inversione della forcella replicativa:

A) L'accumulo di superavvolgimenti positivi genera uno stress topologico e una tensione meccanica a monte della forcella di replicazione. Quando tale stress supera una soglia critica, la forcella regredisce rilasciando la tensione accumulata
B) La regressione porta alla formazione della struttura intermedia completamente rilasciata.
C) Successivamente, l'introduzione di superavvolgimenti negativi (ad esempio ad opera della DNA girasi) promuove il ripristino della struttura nativa della forcella.

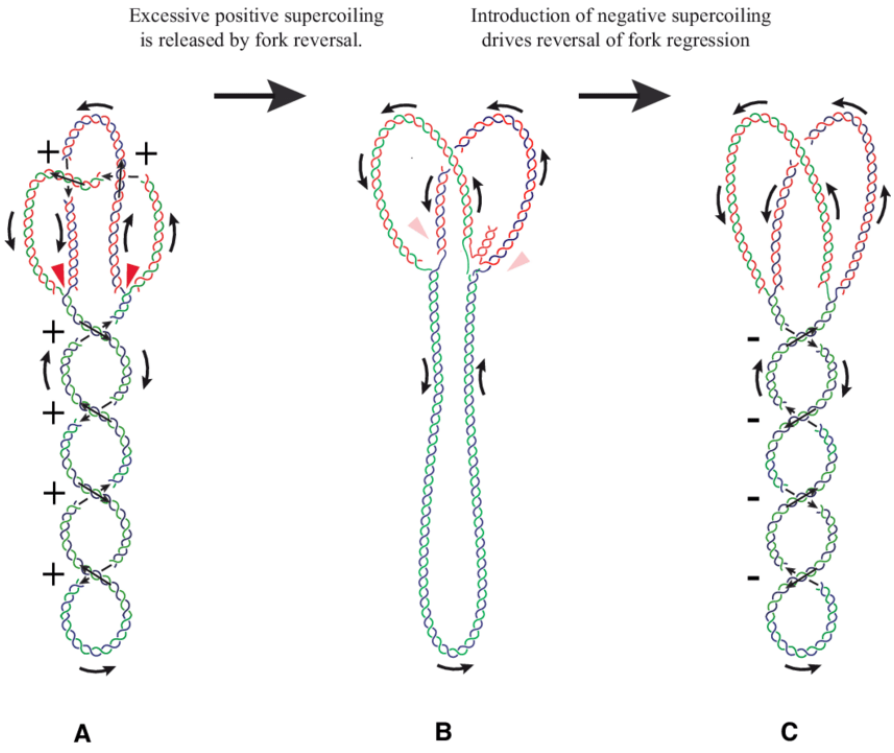


Figura 5 - Meccanismo di regressione e riapertura della forcella replicativa in risposta allo stress topologico (Schvartzman et al., 2019).

La protezione delle forcelle è cruciale perché in assenza di adeguata tutela, l'azione incontrollata delle nucleasi può favorire DSB e generare delezioni e traslocazioni.

Il complesso FANCI–FANCD2 ubiquitinato agisce come un morsetto protettivo legandosi all'elicasi MCM (Minichromosome Maintenance Complex). Questa interazione ostacola l'accesso delle nucleasi e recluta i fattori necessari per il riavvio della sintesi.

Tuttavia, un blocco prolungato di FANCD2 sui siti di danno crea un ingombro sterico che impedisce il completamento della riparazione, aumentando il rischio di DSB. La sua de-ubiquitinazione mediata da USP1, è dunque fondamentale per il ripristino delle forcelle di replicazione, in particolare in condizioni di mancanza di BRCA1/2.

Da una prospettiva terapeutica, la protezione delle forcelle di replicazione e la riparazione tramite ricombinazione omologa, rappresentano fattori critici che contribuiscono alla chemioresistenza nei tumori deficienti per FANCD1 (BRCA2). In questi contesti, l'inibizione di FANCD2 o di altri componenti del complesso FA potrebbe rappresentare una strategia efficace per superare la resistenza ai trattamenti con inibitori di PARP (Poly ADP-ribose Polymerase) o cisplatino. Infatti, la sovraespressione di FANCD2 è stata correlata a un'aumentata resistenza al trattamento e ad una prognosi sfavorevole nei pazienti oncologici, suggerendo che la modulazione di questa via potrebbe avere importanti implicazioni cliniche.

La disfunzione di uno qualunque degli elementi del *pathway*: sensori, complesso centrale, enzimi di ubiquitinazione o substrati come FANCD2/FANCI, si traduce comunque in un'incapacità di proteggere le forcelle e di completare l'HR, con conseguente accumulo di DSB, esaurimento delle cellule staminali ematopoietiche e predisposizione a neoplasie. Le varianti genetiche riscontrate nei pazienti sono eterogenee con perdita completa della proteina o riduzione della sua funzione. Tra i geni della via, *FANCA* è quello più frequentemente mutato, la sua alterazione compromette la formazione e la stabilità del complesso centrale, impedendo la monoubiquitinazione di FANCD2/FANCI e bloccando l'attivazione del *pathway* (Kee & D'Andrea, 2012; Nalepa & Clapp, 2018; Lemonidis et al., 2023).

1.5 Long non coding RNA e regolazione trascrizionale nella DDR

Lo sviluppo delle tecnologie di sequenziamento a RNA (RNA-seq) e dei *microarray* ha permesso di rivelare l'elevata complessità del trascrittoma umano. Si stima che fino al 75–80% del genoma venga trascritto generando RNA non codificanti (ncRNA), mentre la percentuale di genoma umano che codifica per proteine è inferiore al 2%.

Le regioni non codificanti comprendono un'ampia varietà di elementi regolatori e funzionali; tra questi, una frazione rilevante è rappresentata dai geni per i lunghi RNA non codificanti (lncRNA). Con il sequenziamento di nuova generazione, ne sono stati identificati decine di migliaia di loci, dagli eucarioti unicellulari all'uomo, molti dei quali producono chiari effetti fenotipici in studi di perdita di funzione sia in vitro che in vivo. Tuttavia, il numero di lncRNA caratterizzati funzionalmente a livello molecolare o associati con certezza a patologie, si attesta tuttora nell'ordine delle centinaia, rappresentando una frazione marginale, ovvero ≤ 1-2% dei loci identificati. (Barbara Uszczynska-Ratajczak, 2018*).*

I lncRNA sono trascritti di RNA polimerasi II (RNAPII) privi di una *open reading frame* (ORF) e più lunghi di 200 nucleotidi. Questa dimensione permette di distinguerli dai piccoli RNA come i microRNA, gli RNA interagenti con piwi (piRNA), i piccoli RNA nucleolari (snoRNA) e i piccoli RNA interferenti (siRNA).

Gli RNA lunghi non codificanti condividono con gli RNA messaggeri (mRNA) diversi elementi fondamentali nella biogenesi, tra cui la trascrizione da parte dell'RNA polimerasi II, l'aggiunta del cappuccio in 5' e la poliadenilazione dell'estremità 3'. Tuttavia, rispetto agli mRNA sono espressi meno frequentemente, contengono meno esoni e presentano differenze nella fase di elongazione e terminazione della trascrizione: una quota significativa di lncRNA deriva da una polimerasi II che non segue i canonici segnali di regolazione e terminazione a causa di promotori più deboli ed enanchers meno potenti degli mRNA. Inoltre, molti lncRNA presentano modalità di splicing incompleto o splicing alternativo, portando il complesso di

sorveglianza nucleare a trattenerli e spesso degradarli, contribuendo alla loro predominante localizzazione nucleare. Nonostante ciò, una frazione minore si localizza e svolge funzioni nel citoplasma (Statello et al., 2021).

La crescente evidenza scientifica dimostra che i lncRNA sono molto più che sequenze non codificanti: essi costituiscono un livello aggiuntivo di regolazione genica con funzioni essenziali. Un caso rappresentativo è dato da *XIST* (X-inactive specific transcript), un lncRNA di circa 19 kb, coinvolto nella inattivazione del cromosoma X (XCI). In assenza di XCI, le femmine presenterebbero un'espressione doppia dei geni legati al cromosoma X rispetto ai maschi. *XIST*, viene trascritto dal cromosoma X destinato all'inattivazione e agisce in cis, rivestendolo e reclutando complessi proteici che rimodellano la cromatina, portando al silenziamento genico e alla formazione del corpo di Barr (Li et al., 2022).

Diversi studi hanno riportato un'alterazione dei livelli di espressione di singoli lncRNA in tumori di varia origine tissutale, e dati recenti indicano che i loro trascritti possono modulare l'attività genica in seguito a danno al DNA:

In un'analisi del trascrittoma di lncRNA in cellule di fibroblasti umani (GM0637) con DSB indotte dal farmaco radiomimetico neocarzinostatin (NCS), i ricercatori Wan et al. Nel 2013 hanno identificato un lncRNA denominato ANRIL (lungo RNA non codificante antisenso nel locus INK4) che era marcatamente sovraregolato nella risposta al danno al DNA.

ANRIL promuove la repressione del locus soppressore tumorale INK4/ARF (INK4B-ARF-INK4A). Questo locus si estende per circa 35 kb sul cromosoma umano 9p21 e comprende tre geni oncosoppressori strettamente collegati: *INK4a* (Cyclin Dependent Kinase Inhibitor 2A), *INK4b* (Cyclin Dependent Kinase Inhibitor 2B) e *ARF* (Alternative Reading Frame). Osservando una significativa induzione del gene nelle fasi tardive del danno (8-24 ore post-danno) è stato scoperto che *ANRIL* era sovraregolato nella fase avanzata del DDR. Mentre il silenziamento di ATM ha abolito l'induzione di *ANRIL*, suggerendo che fosse ATM-dipendente nel DDR. Analizzandone il promotore, hanno identificato un presunto sito di legame per E2F1(E2F Transcription Factor 1), successivamente test con saggi luciferasi hanno mostrato che l'attività del promotore aumentava durante il danno al DNA, ma veniva abolita dal silenziamento di E2F1. Analisi di immunoprecipitazione della cromatina (ChIP) hanno confermato il legame diretto di E2F1 al promotore di *ANRIL*, con arricchimento significativo dopo danno al DNA. Inoltre, i livelli di E2F1 aumentavano in maniera ATM-dipendente, dimostrando che ATM induce E2F1, che a sua volta attiva trascrizionalmente *ANRIL* nella DDR. Questo gene, quando viene sovraregolato interagisce con complessi repressivi per formare eterocromatina e reprimere l'espressione del locus *INK4B–ARF–INK4A*. Il suo silenziamento invece induce l'espressione di *INK4B*, *ARF* e *INK4A* in modo costantemente elevato durante la DDR. Poiché queste proteine funzionano come inibitori della chinasi ciclina-dipendente, contribuiscono all'arresto del ciclo cellulare nella risposta al danno

al DNA. Dunque, ANRIL ha un ruolo nel mediare l'efficienza della riparazione del DNA (Wan et al., 2013).

NEAT1 (nuclear enriched abundant transcript 1) invece, è un lungo RNA non codificante che funziona come componente strutturale fondamentale del *paraspeckle* nel nucleo. La sua espressione risulta frequentemente alterata nei tumori umani. Recentemente, si è scoperto che il gene può essere coinvolto nella carcinogenesi attraverso la regolazione del DDR. Il suo silenziamento ha sensibilizzato le cellule preneoplastiche alla morte cellulare indotta dal danno al DNA, con aumento del marcatore degli DSB, γ-H2AX. In cellule trattate con idrossiurea, l'inibizione di *NEAT1* ha ridotto la fosforilazione mediata da ATR della chinasi del *checkpoint* Chk1 e della proteina di replicazione RPA32 (Replication Protein A2), indicando che il lncRNA promuove la segnalazione di ATR e l'attivazione del *checkpoint* in risposta allo stress replicativo (Wu et al., 2015; Su et al., 2018).

2.Obiettivi della tesi

L'obiettivo principale di questo lavoro di tesi è quello di esaminare i profili trascrizionali dei geni della via dell'Anemia di Fanconi e dei lncRNA espressi in risposta ai danni al DNA indotti dalla Mitomicina C (MMC) nel modello cellulare HeLa.

Questo *pathway* protegge l'integrità del genoma e svolge un ruolo centrale nell'attivazione e nel coordinamento della risposta al danno al DNA (DDR), in particolare nella riparazione dei legami crociati inter-filamento (ICLs). Di conseguenza, investigare come l'espressione dei geni della via FA/BRCA venga regolata in condizioni di stress genotossico è importante per comprendere la patogenesi della malattia.

Mentre l'attivazione della via FA/BRCA a livello proteico è guidata da una cascata sequenziale di complessi funzionali che svolgono un ruolo nella riparazione del DNA, non è noto se la risposta trascrizionale sia un evento in cui tutti i trascritti vengono indotti simultaneamente oppure risponda ad una precisa programmazione temporale. Pertanto, l'analisi dell'espressione genica a diversi tempi post-trattamento risulta utile per comprendere se esista un controllo temporale a livello di RNA (sia codificante proteine che non codificante) in grado di riflettere la regolazione osservata a livello proteico.

L'approccio sperimentale volto a valutare la dinamica temporale dei profili trascrizionali ha previsto il trattamento con mitomicina C per indurre l'attivazione del DDR. I campioni sono stati raccolti ad intervalli temporali definiti e analizzati successivamente tramite RNA-seq, al fine di identificare i geni differenzialmente espressi (DEGs) e i relativi *pathway* associati. L'analisi a valle (*downstream)* si è inizialmente concentrata sulla via FA/BRCA per poi

estendersi all'identificazione dei lncRNA potenzialmente associati alla regolazione di questa via o alla risposta al danno al DNA. Sulla base dei risultati dell'analisi di espressione differenziale, i geni significativamente deregolati sono stati studiati per identificare quali risultavano sovra o sotto-regolati nei diversi intervalli di tempo.

Per approfondire il significato biologico di queste variazioni, è stata condotta un'analisi di arricchimento funzionale sui trascritti differenzialmente espressi. Questo approccio ha permesso di caratterizzare i processi biologici e i *pathway* KEGG associati a ciascun gruppo temporale, fornendo una panoramica dettagliata dei principali processi cellulari prevalentemente influenzati dal trattamento. Inoltre, per chiarire i meccanismi di regolazione trascrizionale alla base dei profili osservati, è stata condotta un'analisi integrata ChEA–ENCODE basata su dati di ChIP-seq, analizzando l'interazione dei fattori E2F1 (attivatore) ed E2F4 (repressore) con i promotori dei geni codificanti proteine e dei lncRNA.

Infine, per identificare i *pathway* biologici specificamente associati ai lncRNA regolati da E2F1 ed E2F4, è stata condotta un'analisi di associazione predittiva tramite il *database* lncHUB2, basata su matrici di co-espressione su larga scala, con l'obiettivo di proporre ipotesi funzionali sul loro ruolo nella risposta allo stress genotossico.

3.Materiali e metodi

3.1 Allestimento del sistema cellulare

Le cellule sono state seminate in piastre da 60 mm con 3 mL di terreno di coltura. I trattamenti sono stati inizialmente eseguiti ai timepoint di 1-2-4-6-8-16-24 ore. Successivamente, sulla base dei risultati preliminari sull'induzione del danno, sono state preparate otto piastre con una densità cellulare di 2×10^5 cellule per piastra, corrispondenti ai punti temporali 0, 2, 8, 12, 16, 20, 24 e 32 ore. Per un controllo a 24 ore è stata preparata una piastra con una densità di 1.2×10^5. In parallelo, sono state preparate altre quattro piastre per l'incorporazione di bromodesossiuridina (BrdU), corrispondenti ai punti temporali 0, 12, 16 e 20 ore. Per garantire la robustezza dei dati, sono stati condotti tre esperimenti indipendenti.

3.1.1 Trattamento con MMC

Quando le cellule hanno raggiunto circa il 60% di confluenza, è stata aggiunta Mitomicina C (MMC – Sigma-Aldrich M0503) a una concentrazione finale di 1 μM, preparata sciogliendo 2 mg in 4 mL di PBS 1X, in linea con la letteratura, perchè questa dose è efficace nell'indurre il danno al DNA senza risultare eccessivamente tossica. Il trattamento è stato mantenuto per la durata specifica di ciascun punto temporale.

Al termine del trattamento, il terreno è stato rimosso per aspirazione e le piastre sono state lavate con 3 mL di PBS 1X. Le cellule sono state poi separate con tripsina e risospese in 1.5 mL di PBS 1X. La sospensione è stata suddivisa in due aliquote uguali in provette separate: una destinata all'analisi western blot e una all'estrazione dell'RNA per l'analisi di RNA-seq. Le provette sono state centrifugate a 4000 rpm per 5 minuti a 4°C e il surnatante è stato rimosso con attenzione.

Dopo l'estrazione proteica con 75 µL di Laemmli 1X, le proteine sono state analizzate tramite Western Blot utilizzando gel di poliacrilammide all'8% e al 15% per valutare l'induzione del danno. Le analisi di western blot sono state impiegate per monitorare l'efficacia dell'induzione del danno al DNA, attraverso la valutazione di proteine coinvolte nella risposta al danno. In particolare, sono stati analizzati CHK1 e la sua forma fosforilata pCHK1 (S345), H2AX e la sua forma fosforilata γH2AX (S139), il marcatore Vimentina, oltre a FANCD2 e FANCA, principali componenti della via FA.

3.1.2 Incorporazione e colorazione BrdU

Per caratterizzare con precisione l'effetto del trattamento con MMC sulla progressione del ciclo cellulare, in concomitanza con il trattamento e 30 minuti prima della raccolta, sono stati aggiunti 3 µL di BrdU alle piastre per permetterne l'incorporazione. La BrdU è un analogo della timidina che viene integrato nel DNA durante la fase S. Al termine del trattamento, le cellule sono state raccolte e fissate in etanolo puro per la successiva analisi mediante citofluorimetria a flusso o FACS, utilizzata per rilevare la distribuzione delle popolazioni cellulari nelle diverse fasi del ciclo. Questo approccio permette di distinguere in modo quantitativo e temporale le cellule in fase G1, S e G2/M.

Per la rilevazione dell'incorporazione di BrdU, le cellule fissate sono state centrifugate a 2500 rpm per 10 minuti a 10°C e poi risospese in 1 mL di HCl 2N - 0.5% Triton X, con incubazione per 30 minuti a temperatura ambiente.

Dopo una seconda centrifugazione alle stesse condizioni, il pellet è stato risospeso in tampone Borax 0.1 M (pH 8.5) e incubato per 10 minuti a temperatura ambiente.
Le cellule sono state quindi incubate per 1 ora a temperatura ambiente con l'anticorpo primario anti-BrdU (mouse) diluito 1:100 in PBS 1X contenente Tween-0.5% e 1% di BSA. Dopo centrifugazione a 2500 rpm per 5 minuti a 10°C, il pellet è stato risospeso con l'anticorpo secondario anti-mouse 488 (goat) diluito 1:500 in PBS 1X con Tween-1.5% e 1% di BSA e incubato per un'ora a temperatura ambiente al buio. Le cellule sono state poi lavate con un'ulteriore centrifugazione e risospese in 500 µL di ioduro di propidio (PI) a 10 µg/mL in PBS 1X, per consentire l'analisi della distribuzione del DNA nelle diverse fasi del ciclo cellulare. Infine, i campioni sono stati analizzati mediante citofluorimetria a flusso (FACS).

3.1.3 Estrazione dell'RNA

L'estrazione dell'RNA è stata eseguita utilizzando il RNeasy Mini Kit (QIAGEN, cat. n. 74106). Le cellule sono state lisate aggiungendo la soluzione RLT fornita nel kit, integrata con 1% di β-mercaptoetanolo. Per la rimozione del DNA genomico, il lisato cellulare omogeneizzato è stato trasferito in una colonna di eliminazione del DNA (gDNA eliminator) inserita in una provetta da 2 mL. Il campione è stato centrifugato per 30 secondi a ≥8000 × g (≈13.000 rpm). Dopo la centrifugazione, la colonna è stata eliminata e il surnatante recuperato. A quest'ultimo è stato aggiunto un volume equivalente di etanolo al 70%, miscelando accuratamente mediante pipettaggio. La soluzione è stata quindi caricata sulle colonne di purificazione fornite dal kit. Poiché ciascuna colonna può contenere fino a 700 µL di soluzione, il passaggio è stato ripetuto nel caso di volumi superiori.

Il protocollo di lavaggio è stato eseguito con centrifugazione per 15 secondi a 10.000 rpm, seguita dall'eliminazione del surnatante; una successiva aggiunta di 700 µL di tampone RW1, seguita da incubazione a temperatura ambiente per 1 minuto, quindi centrifugazione per 15 secondi a 10.000 rpm e rimozione del surnatante. Aggiunta di 500 µL di tampone RPE, incubazione a temperatura ambiente per 1 minuto, centrifugazione per 15 secondi a 10.000 rpm e rimozione del surnatante. Aggiunta di ulteriori 500 µL di tampone RPE, incubazione a temperatura ambiente per 1 minuto, seguita da centrifugazione per 2 minuti a 10.000 rpm.

Completata questa fase, la provetta di raccolta è stata sostituita con una nuova fornita dal kit e si è effettuata una centrifugazione aggiuntiva di 1 minuto a 13.000 rpm per rimuovere eventuali residui di etanolo, che potrebbero interferire con le analisi successive.

Infine, la colonna è stata trasferita in una provetta sterile da 1,5 mL e l'RNA è stato separato con 30 µL di acqua priva di RNasi. Dopo un'incubazione a temperatura ambiente per 5 minuti, l'RNA legato alla membrana è stato recuperato mediante centrifugazione per 1 minuto a 10.000 rpm.

Per eliminare eventuali tracce residue di DNA genomico, i campioni di RNA estratto (~30 µL) sono stati trattati con il kit TURBO DNase (Ambion™ AM1907, Thermo Fisher Scientific). A ciascun campione sono stati aggiunti 3,5 µL di tampone 10× TURBO DNase e 2 µL di DNasi, con incubazione a 37°C per 20 minuti.

Per rimuovere l'enzima, è stato aggiunto un volume pari a 0,2× (7 µL) di DNase Inactivation Reagent, una miscela di anticorpi specifici per la DNasi. La soluzione è stata mescolata manualmente ogni 30 secondi per 2 minuti, quindi centrifugata per 1 minuto a 13.000 rpm. Il surnatante che ne è risultato (~35 µL) è stato trasferito con cura in una provetta sterile da 1,5 mL, evitando il contatto con il pellet di resina. In caso di contaminazione da resina, il campione è stato nuovamente mescolato e sottoposto a una seconda centrifugazione prima del trasferimento finale.

La quantificazione e la purezza dell'RNA sono state valutate mediante spettrofotometria Nanodrop, misurando l'assorbimento a 260 nm per determinare la concentrazione dell'RNA estratto. La purezza è stata stimata tramite i rapporti A260/A280 e A260/A230.

3.1.4 Retrotrascrizione

Per confermare ulteriormente a livello molecolare l'attivazione delle risposte cellulari al danno, dopo aver notato l'induzione di danno al DNA, successivamente l'RNA è stato retrotrascritto in cDNA La retrotrascrizione è stata eseguita utilizzando il kit SuperScript® IV First-Strand Synthesis System (Invitrogen™, codice: 18090050). In questo processo, gli mRNA, che rappresentano circa il 5% dell'RNA totale, sono stati retrotrascritti impiegando oligo(dT)$_{(20)}$ come primer. Per facilitare la retrotrascrizione degli mRNA con code poli-A parzialmente degradate, sono stati aggiunti degli esameri casuali. In una provetta da 0,5 mL sono stati aggiunti i seguenti componenti:

- Oligo(dT)$_{(20)}$ (50 µM): 0,5 µL

- Esameri casuali (50 ng/µL): 0,5 µL

- Mix di dNTPs (10 mM ciascuno): 1 µL

- RNA totale

- H_2O fino a un volume finale di 11 µL

Per il controllo negativo (RT–), non è stato aggiunto RNA. Le provette sono state quindi incubate a 65°C per 5 minuti.

Successivamente, è stato aggiunto un mix contenente:

-Tampone 5× SSIV: 4 µL

-DTT (100 mM): 1 µL

-Inibitore delle ribonucleasi (40 U/µL): 1 µL

Infine, è stato aggiunto 1 µL di "SuperScript® IV RT" (200 U/µL) e la retrotrascrizione è stata eseguita secondo il seguente programma:

-10' a 23°C per l'appaiamento dei primer

-10' a 50°C per l'attività della trascrittasi inversa

-10' a 80°C per la denaturazione dell'enzima

Al termine della reazione, è stato aggiunto 1 µL di RNasi H (2 U/µL) e le provette sono state incubate a 37°C per 20 minuti per degradare l'RNA ibridato al cDNA.

3.1.5 Reazione di PCR e analisi del gel elettroforetico

Per verificare l'assenza di contaminazione da DNA genomico, è stata eseguita una reazione di PCR, seguita da analisi mediante elettroforesi su gel su GAPDH.
Per le reazioni di PCR è stato utilizzato 1/20 del cDNA precedentemente ottenuto tramite retrotrascrizione (pari a 1 µL). Il volume finale della reazione è stato di 50 µL, e come enzima è stata impiegata la GoTaq Flexi DNA Polymerase (Promega, codice: M8291).
I primer utilizzati per la reazione sono riportati di seguito (Figura 6):

Primer forward: 5' – TCAAGAAGGTGGTGAAGCAGG – 3'

Primer reverse: 5' – ACCAGGAAATGAGCTTGACAAA – 3'

Le condizioni della PCR erano le seguenti:

-2' a 95°C

-30'' a 90°C

-30'' a 58°C

-30'' a 72°C

-5' a 72°C

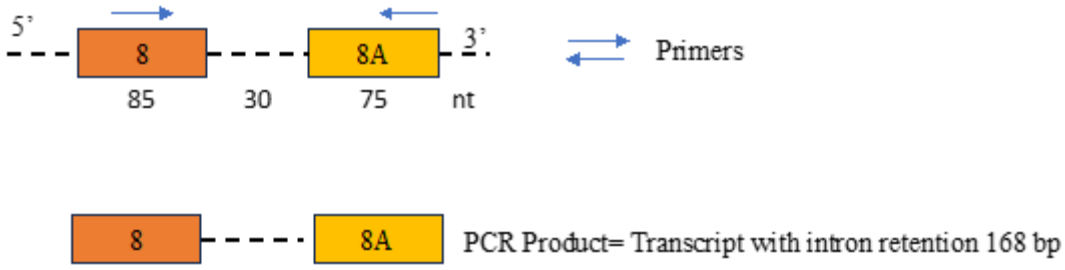


Figura 6 – Localizzazione dei siti di legame dei primer su GAPDH

Poiché l'introne tra i due esoni sui quali sono mappati i primer è di piccole dimensioni, in caso di residui di DNA genomico tale regione verrebbe amplificata anche in un controllo negativo. Per verificare ciò e assicurarsi dell'assenza di contaminazione genomica, è fondamentale includere il controllo negativo (RT–), ossia un campione non sottoposto a retrotrascrizione, e confermare che non siano presenti bande nel gel di elettroforesi.

Se nel controllo negativo comparisse una banda alla stessa altezza del prodotto di amplificazione genomica (168 bp), ciò indicherebbe la presenza di contaminazione da DNA genomico.
I risultati dell'analisi elettroforetica per il gene GAPDH hanno confermato l'assenza di contaminazione genomica, poiché non sono state osservate bande nei controlli negativi (Figura 7).

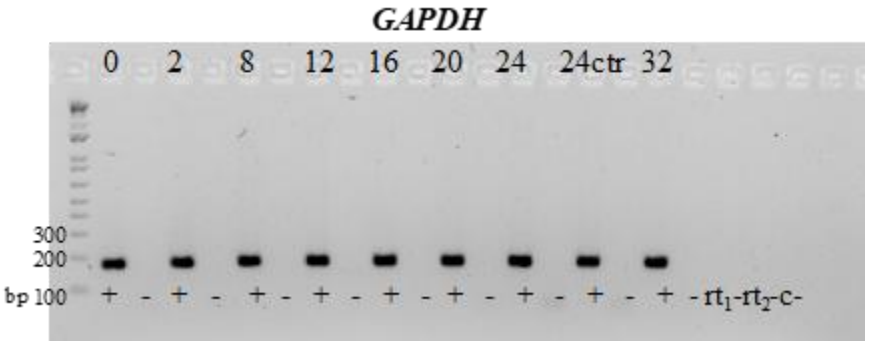


Figura 7 – risultati dell'elettroforesi su gel di GAPDH

3.1.6 Reazione di qPCR

Successivamente, è stata effettuata una quantificazione dell'espressione genica mediante PCR quantitativa in tempo reale (qPCR) su geni coinvolti nella via dell'Anemia di Fanconi, confrontando i livelli relativi di trascrizione (2-ΔCt) nei diversi tempi di trattamento. Per ciascun gene di interesse sono stati progettati dei primer specifici, che permettono di amplificare una regione precisa del gene. Per la reazione è stato utilizzato il kit SYBR Green qPCR.

La qPCR permette di misurare l'accumulo dei prodotti di amplificazione durante ciascun ciclo termico mediante il tracciamento della fluorescenza. L'impiego del fluoroforo intercalante SYBR Green consente la quantificazione del cDNA target proporzionalmente all'aumento della fluorescenza emessa a seguito del legame specifico con il DNA a doppio filamento. I primer impiegati sono i seguenti:

FANCA_L
Primer forward (F1): 5′-TGCCAGCCTTCAGAGACAGAGG-3′ 22 nt 66,4°C
Primer reverse (R1): 5′-GTGGAAGAACTGCTCGCATCTG-3′ 22 nt 64,5°C
Amplicone: 133 bp

FANCA_all
Primer forward (F2): 5′-CAGGACCTGAATGCCCTTTTGC-3′ 22 nt 67,1 °C
Primer reverse (R2): 5′-AGCATAGGCCTCAGAACTGTCACAG-3′ 25 nt 65,2 °C

Amplicone: 99 bp

FANCD2
Primer forward (F1): 5′-GGCAGCAAGACATCCTCCTCTG-3′ 22 nt 65,7 °C
Primer reverse (R1): 5′-GACCTCAATGTCCAGCTCTCGG-3′ 22 nt 65,6 °C
Amplicone: 171 bp

FANCG
Primer forward (F1): 5′-GGGACCTGGCCTTGTTACTAGAGAC-3′ 25 nt 64,6 °C
Primer reverse (R1): 5′-GCATCTAATTCCTCAGCTGGGGG-3′ 23 nt 67,1 °C
Amplicone: 109 bp

PALB2
Primer forward (F1): 5′-AGTGGGACCCTTTCTGATCAACAAG-3′ 25 nt 65,8 °C
Primer reverse (R1): 5′-GCATCCCTTGGACCTCAGCAAAAG-3′ 24 nt 69,0 °C
Amplicone: 127 bp

Per l'analisi dell'espressione genica relativa, è stato calcolato il valore di $2^{-}\Delta Ct$, dove Ct (cycle threshold) è il ciclo in cui la fluorescenza supera una soglia e indica la quantità di cDNA presente per ciascun gene e ΔCt rappresenta la differenza tra il valore di Ct del gene di interesse (ad esempio FANCA) e quello del gene di controllo (GAPDH), che viene utilizzato come riferimento perché la sua espressione è stabile.

Questo metodo consente di quantificare l'espressione relativa del gene di interesse rispetto al controllo. Un valore di $2^{-}\Delta Ct > 1$ indica che il gene di interesse è sovraespresso rispetto al controllo nel campione analizzato, mentre un valore < 1 indica una sottoregolazione.

In seguito alle analisi del western blot, del ciclo cellulare e dell'espressione genica sono stati selezionati i campioni ritenuti più rappresentativi. Sono stati scelti i tempi 0h, 16h, 20h e 24h per descrivere le diverse fasi della risposta al danno. Il campione a 0h rappresenta la condizione di controllo, mentre abbiamo selezionato i tempi 16h, 20h e 24h perchè nelle analisi mostravano progressivamente le fasi di attivazione della risposta cellulare. Dato che questi *timepoint* sono stati esplicativi per lo sviluppo della risposta biologica nel tempo, li abbiamo selezionati per la successiva analisi di sequenziamento.

3.2. Protocollo di analisi dei dati RNA-Seq

3.2.1. FastQC

La qualità delle sequenze grezze è stata valutata utilizzando FastQC (v 0.12.1), un software di controllo qualità per file in formato FASTQ, in grado di individuare eventuali problematiche nelle letture di sequenziamento. L'analisi è stata eseguita su tutti i campioni con il seguente comando:

```
fastqc *.fq.gz -o FASTQC *fq.gz
```

FastQC genera grafici che mostrano la distribuzione della qualità lungo le letture, tramite un calcolo del Phred score, un indice che misura la qualità delle basi chiamate durante il sequenziamento. Il programma analizza inoltre il contenuto in basi GC, utile per individuare eventuali bias sistematici, e rileva le sequenze sovrarappresentate, che possono indicare la presenza di contaminanti o adattatori di sequenziamento residui.
I report generati sono stati esaminati per verificare che tutte le letture soddisfacessero i criteri di qualità richiesti prima di procedere con le analisi successive.

3.2.2. RNA STAR

L'allineamento delle letture è stato eseguito utilizzando RNA STAR (Galaxy Version 2.7.11a galaxy1), uno strumento progettato per l'allineamento di dati RNA-Seq.
RNA STAR gestisce letture che si allineano a regioni non contigue del genoma di riferimento, una caratteristica tipica dei dati di RNA-Seq, che riflette la struttura intronica degli mRNA e le giunzioni di splicing. Queste letture vengono mappate identificando le giunzioni di splicing, che rappresentano i punti di connessione tra esoni e introni. L'analisi è stata condotta su letture *paired-end*, utilizzando come genoma di riferimento GRCh38 o hg38. Per fornire informazioni strutturali sui geni è stato utilizzato un file di annotazione in formato GTF proveniente da GENCODE v.46, che contiene informazioni dettagliate sulla localizzazione e sulla struttura dei geni, comprese giunzioni di *splicing*, esoni, introni e varianti di *splicing*. L'allineamento è stato effettuato utilizzando un indice pre-costruito del genoma di riferimento, fornendo il file GTF citato prima, per migliorare l'accuratezza nella mappatura delle giunzioni di *splicing*. È stata impiegato un range di 30 nucleotidi attorno ai siti di giunzione annotati per ottimizzare il rilevamento delle giunzioni. Al termine dell'allineamento sono stati generati sia i file BAM contenenti le letture allineate, sia i conteggi genici, successivamente utilizzati per l'analisi dell'espressione genica.

3.2.3. FeatureCounts

La quantificazione dell'espressione genica è stata eseguita utilizzando il pacchetto featureCounts, parte della *suite* Subread (versione 2.20.0, Bioconductor 3.20). La quantificazione è stata effettuata sui file BAM generati in precedenza, utilizzando l'annotazione genomica GENCODE v46.

Per i dati *paired-end* sono stati applicati parametri restrittivi, richiedendo che entrambi gli estremi di ciascuna lettura fossero correttamente mappati (requireBothEndsMapped = TRUE), mentre il conteggio dei frammenti chimerici è stato disabilitato (countChimericFragments = FALSE).
L'unità di conteggio è stata impostata a livello di esone (GTF.featureType = "exon"), utilizzando l'attributo "gene_id" per aggregare i conteggi a livello di gene (GTF.attrType = "gene_id").

3.2.4. DESeq2

L'analisi di espressione differenziale è stata condotta utilizzando il pacchetto DESeq2 (versione 1.46.0, Bioconductor), un metodo basato sulla distribuzione binomiale negativa. La matrice dei conteggi grezzi ottenuta da featureCounts è stata importata senza applicare filtri preliminari, al fine di massimizzare la sensibilità dell'analisi. I dati dei campioni sono stati strutturati raggruppando le repliche biologiche in base ai tempi di trattamento (0h, 16h, 20h e 24h dopo esposizione a MMC).

L'oggetto DESeqDataSet è stato creato utilizzando la formula ~ *condition*, dove "condition" rappresenta i diversi punti temporali sperimentali.

L'analisi è stata eseguita mediante la funzione DESeq(), che: stima i fattori di normalizzazione per la correzione delle dimensioni dei campioni, calcola la dispersione specifica per gene e applica una regressione binomiale negativa per stimare i fold-change e testare la significatività statistica.

Per ciascun confronto tra condizioni (0h vs 16h, 0h vs 20h, 0h vs 24h) è stata condotta un'analisi differenziale utilizzando i seguenti parametri validati:

- Filtraggio indipendente (independentFiltering = TRUE) per aumentare la potenza statistica;

- Soglia Alpha = 0.1 per il filtraggio indipendente;

- Correzione per test multipli mediante il metodo Benjamini-Hochberg (FDR);

- Rimozione degli outlier basata sul test di Cook (cooksCutoff = TRUE).

È stato successivamente applicato un filtro basato sull'espressione media (baseline ≥ 10) per escludere dall'analisi i geni con livelli di espressione estremamente bassi, riducendo così il rumore di fondo. Il p-value misura la probabilità che le differenze d'espressione osservate tra le condizioni siano dovute unicamente al caso (ipotesi nulla). Nell'analisi di RNA-seq, poiché vengono condotti test statistici simultanei su decine di migliaia di geni, si applica la correzione per test multipli tramite il metodo di Benjamini-Hochberg, per ottenere l'adjusted p-value (o False Discovery Rate, FDR). Questo parametro riduce il rischio di falsi positivi, garantendo che i geni identificati come differenzialmente espressi con un adjusted p-value inferiore a 0,05 abbiano una probabilità trascurabile di essere scoperti per errore.

I geni sono stati quindi considerati differenzialmente espressi quando presentavano un p-value corretto (adjusted p-value) minore di 0.05. Per ciascun confronto sono stati calcolati i seguenti parametri: numero totale di geni analizzati, numero di geni con p-value corretto significativo (< 0.05, < 0.01, < 0.005) e numero di geni sovraregolati (up-regulated) e sottoregolati (down-regulated) ai vari livelli di significatività.

3.3. Analisi di arricchimento funzionale

L'analisi funzionale è stata condotta sui geni differenzialmente espressi (DE) comuni ai tre punti temporali sperimentali (16h, 20h e 24h dopo trattamento con MMC), senza distinzione tra geni sovraregolati e sottoregolati. Questo approccio è stato adottato per individuare i processi biologici modulati in modo coerente durante l'intera risposta al danno genotossico indotto da MMC.

Per l'analisi dell'arricchimento funzionale è stata utilizzata la piattaforma DAVID (Database for Annotation, Visualization and Integrated Discovery). I geni differenzialmente espressi comuni con adjusted p-value < 0.05, sono stati annotati rispetto alla categoria Gene Ontology – GO_BP_FAT (Biological Process), utilizzando l'intero genoma umano come riferimento. La scelta di focalizzarsi esclusivamente sui processi biologici è stata motivata dall'interesse principale nelle vie di risposta cellulare attivate in seguito al trattamento con MMC. I termini GO sono stati considerati significativamente arricchiti quando presentavano un p-value corretto secondo Benjamini < 0.05 (FDR). Per identificare *pattern* biologici coerenti, i termini GO arricchiti sono stati raggruppati in *cluster* funzionali. Il punteggio di arricchimento (Enrichment Score, ES) di ciascun *cluster* è stato calcolato come la media geometrica negativa dei p-value dei termini GO inclusi, convertita in scala logaritmica ($-\log_{10}$). (Huang da W, et al., 2009) I primi dieci *cluster* sono stati selezionati in base al valore di ES e rinominati da C1 a C10, assegnando a ciascuno un nome descrittivo (es. "C1: Regulation of Gene Expression").

Per valutare il coinvolgimento specifico dei geni del *pathway* FA/BRCA nei processi biologici identificati, è stata condotta un'analisi di sovrarappresentazione (over-representation analysis)

mediante test esatto di Fisher. Questo test permette di verificare se la proporzione di geni FA/BRCA presenti in ciascun *cluster* sia significativamente superiore a quella attesa per caso.

3.4 Analisi dei lncRNA

L'analisi dei lncRNA è stata condotta selezionando i trascritti annotati come lncRNA e risultati differenzialmente espressi (p-adjusted < 0.05) nei confronti temporali effettuati. Le liste dei lncRNA sono state confrontate mediante il software Venny (Venn Diagram Generator), al fine di generare diagrammi di Venn e classificare i trascritti in base alla loro presenza o assenza nei differenti intervalli di tempo. Sulla base dell'andamento qualitativo dei livelli di espressione nei quattro tempi sperimentali, i lncRNA significativi sono stati assegnati a sette gruppi (A–H). La classificazione è stata effettuata definendo per ciascun trascritto una relazione ordinale tra i valori di espressione. La distribuzione dei lncRNA sovraregolati e sottoregolati nei diversi intervalli temporali è stata inoltre confrontata con quella dei geni codificanti per proteine. Per la visualizzazione e interpretazione dei risultati, sono state generate heatmap di raggruppamento gerarchico dei geni DE e grafici di distribuzione dei geni FA/BRCA nei *cluster* funzionali, utilizzando i pacchetti R ggplot2 (v3.5.1) e pheatmap (v1.0.12).

3.5 Analisi integrata ChEA–ENCODE per la caratterizzazione dei promotori

Per identificare i potenziali fattori di trascrizione (TF) coinvolti nella regolazione dei geni appartenenti al *pathway* della Fanconi Anemia (FA) e al restante trascrittoma differenzialmente espresso, è stato utilizzato lo strumento ChEA (ChIP-X Enrichment Analysis). Questo strumento bioinformatico consente di effettuare un'analisi di arricchimento basata su dati sperimentali di ChIP-seq, confrontando la lista di geni fornita con database di interazioni tra TF e bersagli molecolari. I geni del complesso centrale FA sono stati dunque inseriti nel portale ChEA, che ha restituito una lista ordinata di fattori di trascrizione potenzialmente associati, calcolata mediante test statistici di arricchimento. Tra i risultati ottenuti, il fattore E2F1 è risultato il più significativo ed è stato selezionato assieme ad E2F4 noto per interagire con complessi di repressione complementari. I dataset sperimentali relativi ai siti di legame dei fattori di trascrizione E2F1 ed E2F4 sono stati scaricati dal portale ENCODE, dal dataset dell'esperimento ChIP-seq, con analisi condotte da Peggy Farnham del laboratorio della Keck school of Medicine of USC, in cellule HeLa-S3. Il file dei siti di legame per E2F4 è: accession ENCSR000EVL (formato narrowPeak) e per E2F1: accession ENCFF399JUZ (formato narrowPeak). I file sono stati convertiti in formato BED per utilizzarli come set di picchi per l'intersezione genomica. Al fine di confermare sperimentalmente il legame al promotore dei geni, sono state definite artificialmente le regioni del promotore di ciascun trascritto da

utilizzare nell'analisi. A partire dalla versione GENCODE v46, sono state estratte le coordinate dei siti di inizio della trascrizione (TSS) per tutti i trascritti annotati, le regioni promotoriali dei geni codificanti e non codificanti sono state costruite estendendo 3000 bp *upstream* e 3000 bp *downstream* dal TSS annotato. Le regioni così ottenute sono state salvate in formato BED per utilizzarle come input nelle analisi successive. L'intersezione tra i promotori definiti e i siti di legame dei fattori di trascrizione è stata effettuata con lo strumento Bedtools Intersect (versione 2.31.0), questo strumento consente di confrontare insiemi di coordinate genomiche e individuare le sovrapposizioni tra due file in formato BED. In questa analisi, l'intersezione è stata considerata positiva quando almeno 1 bp del promotore si sovrapponeva a un picco ChIP-seq del fattore selezionato. Un esempio di comando utilizzato:

```
bedtools intersect -a lncRNA_promoters.bed -b E2F4_peaks.bed -u
> lncRNA_with_E2F4.bed.
```

L'opzione -u consente di riportare solo le regioni del primo file (-a) che presentano almeno una sovrapposizione con il secondo (-b), evitando duplicati. Analoghe procedure sono state applicate per E2F1 e per l'intersezione tra i due dataset, al fine di identificare i trascritti i cui promotori risultano occupati da uno o entrambi i fattori di trascrizione.

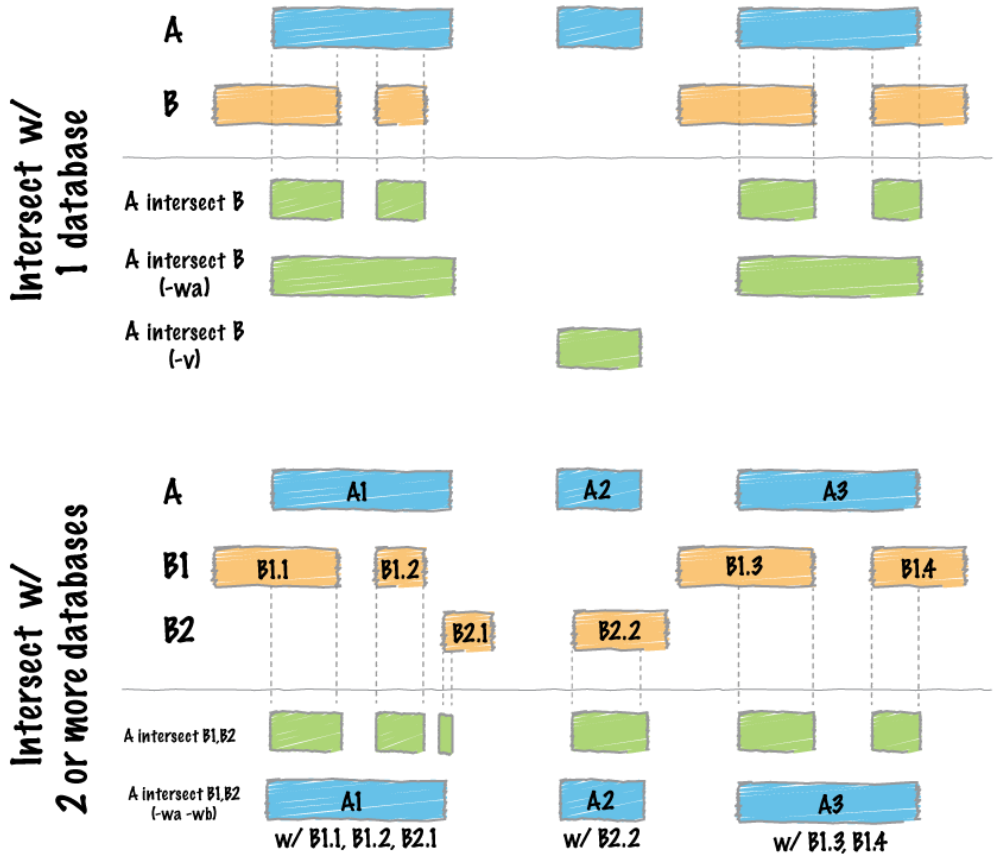


Figura 8 – funzionamento dello strumento bedtools intersect

Per verificare in modo indipendente la correttezza dei risultati ottenuti dall'intersezione, ho integrato l'analisi bioinformatica con una conferma visiva mediante l'utilizzo del Genome Browser di ENCODE e dello UCSC Genome Browser. Questi strumenti permettono di osservare direttamente la collocazione genomica dei segnali ChIP-seq e la struttura dei geni, offrendo una validazione qualitativa della presenza dei siti di legame all'interno dei promotori dei trascritti analizzati.

Il Genome Browser di ENCODE è stato utilizzato per visualizzare i profili ChIP-seq di E2F1 ed E2F4 generati in cellule HeLa-S3. In questo contesto, i picchi rappresentano regioni di arricchimento del fattore di trascrizione identificate mediante algoritmi di *peak calling* (ad esempio MACS2). Ogni picco corrisponde a una finestra genomica in cui la densità di frammenti ChIP-seq supera significativamente il *background*, indicando un potenziale sito di legame. La traccia di intensità associata permette di valutare la forza del segnale, mentre i dati forniti (signalValue, p-value, q-value) aggiungono informazioni sulla qualità statistica del picco. La visualizzazione ha consentito di confermare che i picchi ricadessero effettivamente

entro la regione ±3 kb dal TSS, escludendo false sovrapposizioni dovute alla sola elaborazione dei file BED.

Parallelamente, lo UCSC Genome Browser è stato impiegato per verificare la struttura genica e l'esatta posizione dei siti d'inizio della trascrizione, utilizzando l'annotazione GENCODE v46. Questo passaggio è risultato essenziale in presenza di geni con più isoforme, poiché la definizione accurata del TSS influisce direttamente sulla delimitazione della regione del promotore. L'utilizzo del browser ha permesso di confermare la corrispondenza tra le coordinate utilizzate nelle analisi e le annotazioni del trascritto differenzialmente espresso.

3.6 Predizione dei ruoli biologici dei lncRNA tramite analisi KEGG

L'associazione tra lncRNA e *pathway* KEGG fornita successivamente da lncHUB2 si basa su un'analisi di co-espressione. In particolare, per ogni lncRNA viene calcolata la correlazione di Pearson tra il suo livello di espressione e quello dei geni codificanti proteine in un ampio dataset di campioni. Successivamente, i geni più correlati vengono confrontati con insiemi di geni appartenenti a *pathway* annotati nel database KEGG. Per ciascun *pathway*, si calcola la correlazione media tra il lncRNA e i geni del set, e si valuta la significatività statistica di questa associazione mediante un test (right-tailed p-value). I *pathway* vengono quindi ordinati in base alla significatività, fornendo una lista di processi biologici potenzialmente associati al lncRNA. Questa procedura non implica una funzione confermata, ma rappresenta una predizione funzionale indiretta: se il lncRNA mostra un profilo di espressione simile a quello dei geni di un determinato *pathway*, si ipotizza che possa avere un ruolo regolatorio in quel processo. L'uso dei termini KEGG da lncHUB2, quindi, consente di formulare ipotesi biologiche sul possibile coinvolgimento del lncRNA in specifici *pathway*, basandosi su evidenze di correlazione statistica.

4. Risultati e Discussione

4.1. Ottimizzazione di un sistema cellulare per l'induzione al danno al DNA

Per caratterizzare la risposta al danno al DNA (DDR) e l'attivazione della via dell'Anemia di Fanconi (FA), le cellule HeLa sono state trattate con mitomicina C 1 μM e analizzate a diversi intervalli temporali.

L'efficacia del trattamento è stata monitorata valutando le alterazioni proteiche di componenti chiave della risposta cellulare: l'attivazione dei *checkpoint* di danno (pCHK1/CHK1 e γH2AX/H2AX) e l'ingaggio della via FA attraverso la monoubiquitinazione di FANCD2 e l'espressione del complesso centrale (FANCA). L'analisi è stata condotta parallelamente alla quantificazione della Vimentina.

CHK1 è una chinasi centrale nei *checkpoint* del ciclo cellulare e la sua forma fosforilata (pCHK1) rappresenta un marcatore diretto dell'attivazione della risposta al danno al DNA. Quando CHK1 è fosforilata, è attivata e svolge il suo ruolo nel bloccare la progressione del ciclo cellulare. L'aumento dei livelli di pCHK1 indica che la cellula ha riconosciuto il danno e ha attivato meccanismi per rallentare la progressione del ciclo cellulare e consentire la riparazione.

H2AX invece è un istone che in seguito a danni come le rotture a doppio filamento (DSB), viene fosforilato nella forma γH2AX. Questa modifica si accumula nei siti di danno.

La Vimentina è una proteina usata come controllo interno di carico per entrambi i gel. Poiché la sua intensità rimane costante in tutti gli intervalli di tempo, dimostra che in ogni corsia è stata caricata la stessa quantità di proteina totale. Questo garantisce che le variazioni viste per le altre proteine siano veri effetti biologici del trattamento con MMC e non errori tecnici.

Per quanto riguarda la via FA, FANCD2 è una proteina fondamentale che, in risposta al danno, subisce monoubiquitinazione. All'analisi mediante western blot, FANCD2 appare come un doppietto: la banda superiore corrisponde alla forma monoubiquitinata, mentre quella inferiore rappresenta la forma non modificata. La comparsa del doppietto è quindi un marcatore dell'attivazione della via. Anche FANCA, è stata analizzata per una valutazione più completa dell'attivazione di questa via. La sua presenza indica che il complesso FA è espresso e permette di escludere che eventuali difetti osservati nella monoubiquitinazione di FANCD2 siano dovuti a problemi a monte, come la perdita di espressione di FANCA.

Come mostrato in Figura 9, già dopo 8 ore di esposizione a MMC si osserva un'evidente attivazione della risposta al danno, evidenziata dall'aumento della fosforilazione di CHK1 sul

residuo Ser345 (pCHK1), indicatore dell'attivazione della via ATR e del blocco del ciclo cellulare. Contemporaneamente, a 8 ore si rileva la comparsa della banda superiore del doppietto di FANCD2, corrispondente alla sua forma monoubiquitinata, marcatore dell'attivazione funzionale della via FA necessario per la riparazione dei legami crociati (ICL). L'espressione costante di *FANCA* a tutti gli intervalli temporali, indica che il complesso FA è espresso e permette di escludere che eventuali difetti osservati nella monoubiquitinazione di FANCD2 siano dovuti a problemi a monte, come la perdita di espressione di *FANCA*.

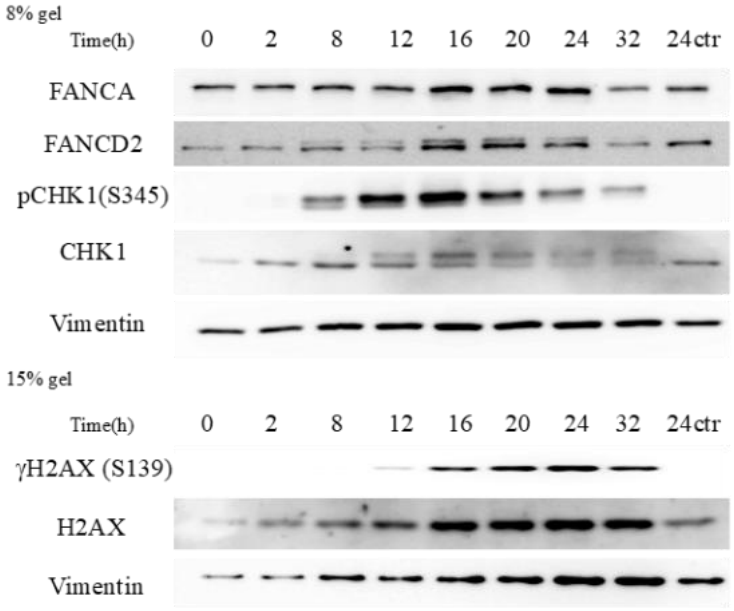


Figura 9 – risultati dell'analisi del western blot

A partire da 12 ore di trattamento, si osserva inoltre un progressivo incremento del segnale di γH2AX, a conferma dell'accumulo di rotture del doppio filamento e della risposta al danno sostenuto da MMC.

Infine, i livelli costanti e uniformi della Vimentina in tutte le corsie confermano il corretto bilanciamento del carico proteico, attestando che le variazioni cinetiche osservate rappresentano una specifica risposta biologica allo stress genotossico indotto dal trattamento.

Per caratterizzare con precisione l'effetto del trattamento con MMC sulla progressione del ciclo cellulare, le cellule sono state sottoposte a incorporazione di bromodeossiuridina (BrdU). Successivamente, sono state analizzate mediante citofluorimetria a flusso o FACS, per rilevare la distribuzione delle popolazioni cellulari nelle diverse fasi del ciclo.

Nel campione non trattato a 0h, la distribuzione delle cellule risulta equilibrata tra le fasi del ciclo: il 39,32% delle cellule si trova in G1, il 44,28% in S e il 14,42% in G2/M. Questa distribuzione è tipica di una popolazione proliferante in condizioni fisiologiche. Dopo 12 ore di esposizione a MMC, si può osservare un cambiamento marcato: la percentuale di cellule in fase S aumenta ad arrivare al 77,28%, mentre quelle in G1 e G2/M scendono rispettivamente al 12,40% e al 9,17%. Questo dato indica che il trattamento con MMC provoca un arresto della progressione del ciclo cellulare nella fase S a causa dell'accumulo di danni al DNA che impediscono il proseguimento verso le fasi successive. L'analisi più dettagliata della fase S mostra inoltre l'accumulo nella porzione iniziale della fase S chiamata *early* S (e-S), data dalla MMC che blocca le cellule precocemente per l'attivazione dei checkpoint di sorveglianza che rilevano le anomalie date dagli ICL. (Figura 10)

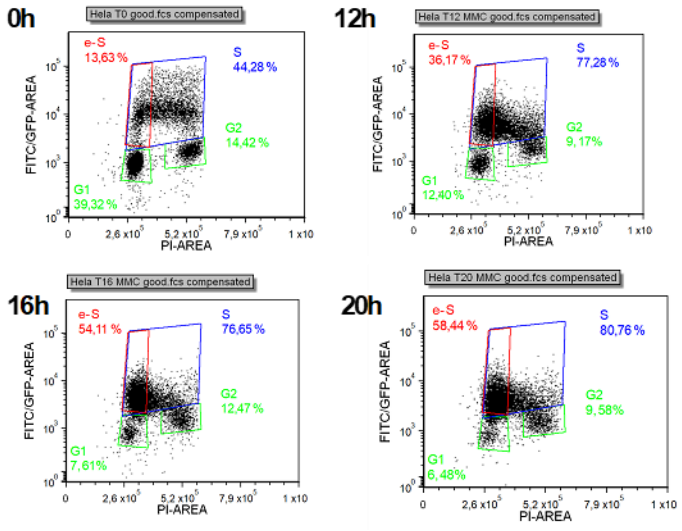


Figura 10 – incorporazione BrdU

Per valutare se la risposta al danno al DNA indotto da MMC comporti anche una regolazione a livello trascrizionale dei geni della via dell'Anemia di Fanconi, è stata eseguita un'analisi di qPCR confrontando i livelli relativi di trascrizione (2-ΔCt) nei diversi tempi di trattamento.

Un valore di 2^-ΔCt maggiore di 1 indica che il gene di interesse è sovraespresso rispetto al controllo e i risultati ottenuti nei tre replicati hanno evidenziato un progressivo incremento dell'espressione dei geni target, in particolare FANCA, FANCD2, FANCG e PALB2.

Come mostrato in Figura 11, i livelli relativi di trascrizione del gene nei primi punti temporali a 2 h, 8 h e 12 h si mantengono pressoché inalterati rispetto al livello basale di controllo a 0 h. Analogamente, il controllo non trattato a 24 h mostra valori sovrapponibili a quelli iniziali,

confermando che le variazioni trascrizionali osservate sono un effetto del trattamento genotossico e non del tempo di coltura in incubatore.

A partire da 16 ore dal trattamento si evidenzia un progressivo e costante incremento della trascrizione di *FANCA*, che raggiunge l'apice tra le 24 e le 32 ore, registrando un aumento di oltre 5 volte rispetto al controllo basale. Tale induzione trascrizionale risulta statisticamente significativa a 24 h e 32 h rispetto allo 0 h (p-adj < 0,01), riflettendo un'attivazione coordinata e prolungata della via FA/BRCA in risposta all'accumulo di lesioni al DNA.

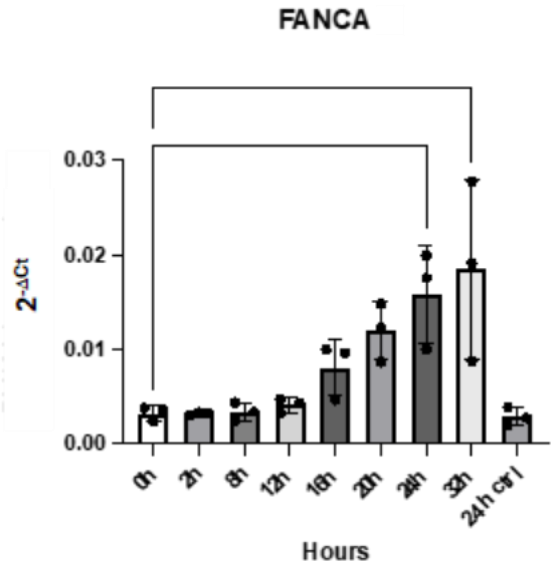


Figura 11 - risultato della qPCR di FANCA normalizzato con GAPDH (** padj<0.01)

Nel loro insieme, l'analisi proteica mediante western blot, lo studio del ciclo cellulare e la quantificazione dell'espressione genica tramite qPCR hanno consentito di definire l'andamento temporale della risposta cellulare allo stress genotossico.

Sulla base di questi risultati, sono stati scelti i tempi 0h, 16h, 20h e 24h, considerati i più significativi per descrivere le diverse fasi della risposta al danno, dalla situazione iniziale di assenza di danno fino all'attivazione della via FA/BRCA.

4.2. Analisi dei dati RNA-Seq

Sono stati inviati a sequenziamento 3 replicati biologici (rispettivamente denominati A, B, C) dei campioni a 0, 16, 20 e 24h (denominati rispettivamente 1, 2, 3, 4). Una volta ricevuti i dati grezzi, è stato effettuato il controllo di qualità tramite FastQC. L'analisi FastQC ha indicato *reads* di alta qualità con un Phred Score elevato, mostrando una precisione superiore al 99,9% nella lettura delle basi. Inoltre, la percentuale di GC nei campioni sequenziati era superiore al 50%, confermando la qualità e la robustezza del sequenziamento (Tabella 2).

Sample	Raw reads	Q30 (%)	GC (%)	Uniquely mapped reads number	Uniquely mapped reads %
A1	185048884	89.31	55.80	68003252	73.50%
A2	168944636	90.28	53.51	63179820	74.79%
A3	172344354	90.47	53.29	67659088	78.52%
A4	190261778	89.83	56.10	68660676	72.17%
B1	162601528	89.32	54.77	58799941	72.32%
B2	199393824	91.19	52.15	76495395	76.73%
B3	222355130	91.32	52.69	87473245	78.68%
B4	200095760	91.50	52.24	78331824	78.29%
C1	178098384	90.35	53.55	66487510	74.66%
C2	166081192	90.90	52.39	64495604	77.67%
C3	175979010	91.72	51.26	70350792	79.95%
C4	175045496	92.05	51.15	70270123	80.29%

Tabella 2 - Risultati FastQC e RNA STAR summary

Dopo il controllo qualità, le *reads* sono state allineate al genoma umano (hg38) utilizzando RNA STAR (GTF GENCODE Release 46). Tutti i campioni hanno avuto una percentuale di *reads* univocamente mappate superiore al 70%, indicando che la maggior parte di queste sono state correttamente allineate, con contaminazione o errori minimi. Il conteggio delle letture mappate su ciascun gene è stato effettuato tramite featureCounts, ottenendo una matrice per 70.611 geni in 12 campioni. Questo file è stato usato come input per l'analisi di espressione differenziale. L'analisi di espressione differenziale è stata condotta con DESeq2, confrontando i diversi intervalli di tempo rispetto al controllo.

4.3. Analisi dei geni coinvolti nella via FA/BRCA

La caratterizzazione dei profili d'espressione ha preso in esame il pannello dei geni appartenenti alla via FA/BRCA, identificati consultando i *database* KEGG e Reactome. La selezione è stata effettuata esplorando il *pathway* Fanconi Anemia e ottenendo una lista di 58 geni. Oltre al complesso centrale della FA, questi geni appartengono al *checkpoint* replicativo e stabilità delle forcelle, ricombinazione omologa (HR) ad alta fedeltà, nucleasi/risoluzione di strutture del DNA, risoluzione degli intermedi HR, Mismatch Repair (MMR), sintesi translesione (TLS), ubiquitinazione e regolazione trascrizionale.

Tra i 58 geni analizzati, un numero significativo ha mostrato una regolazione differenziale già a 16 ore dal trattamento, con 33 geni significativamente regolati (padj < 0,05), di cui 32 sovraregolati e solo 1 sottoregolato. Questo andamento si conferma ai *timepoint* successivi, con 39 geni differenzialmente espressi a 24 ore, suggerendo una progressiva attivazione della via di riparazione del DNA (Tabella 3).

gene_name	padj_16h	log2FC_16h	padj_20h	log2FC_20h	padj_24h	log2FC_24h
ATR	1,46E-01	2,24E-01	7,85E-01	5,05E-02	3,48E-01	1,45E-01
ATRIP	2,70E-02	2,71E-01	4,25E-07	5,36E-01	3,76E-11	6,82E-01
BLM	4,60E-13	9,94E-01	5,32E-16	1,09E+00	1,18E-15	1,08E+00
BRCA1	5,21E-10	9,05E-01	3,40E-15	1,11E+00	4,96E-15	1,10E+00
BRCA2	5,20E-11	1,11E+00	1,03E-16	1,37E+00	8,89E-21	1,53E+00
BRIP1	5,62E-23	1,99E+00	4,57E-33	2,38E+00	5,68E-39	2,58E+00
CENPS	6,74E-01	1,68E-01	7,47E-01	1,21E-01	3,90E-01	2,76E-01
CENPX	6,57E-01	-1,57E-01	1,99E-01	-3,55E-01	2,30E-01	-3,28E-01
DCLRE1A	1,77E-02	4,63E-01	4,91E-02	3,82E-01	2,16E-02	4,30E-01
DCLRE1B	1,47E-05	5,70E-01	1,96E-09	7,50E-01	7,17E-11	8,05E-01
EME1	1,59E-02	4,15E-01	1,66E-03	5,01E-01	6,02E-05	6,16E-01
EME2	6,98E-01	-1,58E-01	7,37E-01	-1,27E-01	6,42E-01	-1,64E-01
ERCC1	1,46E-01	-2,91E-01	3,10E-01	-2,05E-01	9,34E-01	-2,12E-02
ERCC4	3,74E-03	3,53E-01	4,92E-04	4,02E-01	7,57E-06	4,98E-01
FAAP100	8,82E-01	4,27E-02	2,61E-01	2,28E-01	2,70E-02	3,97E-01
FAAP20	3,47E-01	-2,74E-01	6,85E-01	-1,25E-01	8,18E-01	-7,23E-02
FAAP24	1,24E-02	4,70E-01	3,22E-06	7,78E-01	5,53E-08	8,89E-01
FAN1	6,83E-02	2,72E-01	4,23E-01	1,31E-01	4,32E-02	2,81E-01
FANCA	3,09E-09	8,21E-01	1,70E-17	1,14E+00	2,09E-25	1,37E+00
FANCB	4,88E-15	1,06E+00	3,22E-14	1,01E+00	3,43E-16	1,08E+00
FANCC	7,52E-01	-5,41E-02	3,78E-01	-1,19E-01	8,20E-01	-3,48E-02
FANCD2	6,08E-07	5,37E-01	3,46E-09	6,16E-01	1,52E-10	6,60E-01
FANCE	9,91E-01	-3,23E-03	6,76E-01	-8,12E-02	9,39E-01	-1,66E-02
FANCF	4,26E-01	-3,95E-01	3,54E-03	-1,14E+00	2,84E-03	-1,16E+00
FANCG	1,64E-16	9,45E-01	1,98E-31	1,30E+00	2,31E-45	1,56E+00
FANCI	2,10E-11	8,70E-01	1,24E-15	1,01E+00	1,64E-23	1,25E+00
FANCL	9,51E-02	4,21E-01	4,00E-01	2,27E-01	3,34E-01	2,50E-01
FANCM	9,22E-04	4,83E-01	5,69E-05	5,57E-01	7,83E-08	7,18E-01
HES1	4,32E-03	1,08E+00	7,74E-05	1,39E+00	1,07E-03	1,17E+00
MLH1	6,71E-02	2,24E-01	4,52E-02	2,31E-01	3,70E-03	3,13E-01
MUS81	2,25E-03	4,20E-01	1,11E-09	7,54E-01	1,22E-11	8,28E-01
PALB2	1,62E-06	6,19E-01	6,31E-13	8,82E-01	5,89E-14	9,14E-01
PMS2	6,19E-04	-3,83E-01	5,06E-08	-5,66E-01	1,10E-05	-4,62E-01
POLH	9,39E-01	1,53E-02	2,29E-01	1,57E-01	8,32E-07	5,22E-01
POLI	6,57E-01	-1,26E-01	9,34E-01	2,42E-02	5,97E-01	1,29E-01
POLK	1,26E-08	8,10E-01	3,39E-12	9,61E-01	2,53E-15	1,08E+00
POLN	5,74E-01	-4,16E-01	6,84E-01	-2,88E-01	1,57E-01	-8,39E-01
RAD51	2,10E-15	9,52E-01	2,33E-17	1,00E+00	2,49E-23	1,16E+00
RAD51C	7,94E-01	5,86E-02	5,87E-01	1,00E-01	3,17E-02	3,14E-01
REV1	8,48E-01	5,36E-02	7,64E-01	7,11E-02	9,78E-01	-7,13E-03
REV3L	2,14E-04	7,82E-01	6,47E-09	1,15E+00	1,73E-10	1,24E+00
RMI1	2,72E-13	1,10E+00	7,69E-18	1,26E+00	2,01E-23	1,45E+00
RMI2	1,75E-02	7,35E-01	2,17E-03	8,81E-01	2,06E-03	8,78E-01
RPA1	6,16E-01	-1,08E-01	2,94E-01	-1,85E-01	7,03E-02	-2,88E-01
RPA2	6,81E-16	9,14E-01	6,82E-26	1,16E+00	2,99E-39	1,43E+00
RPA3	3,01E-02	4,22E-01	2,35E-02	4,22E-01	7,77E-08	8,93E-01
RPA4	7,09E-04	1,31E+00	1,04E-05	1,59E+00	1,10E-06	1,73E+00
RPS27A	4,39E-01	-1,98E-01	6,97E-01	-1,03E-01	8,96E-01	3,62E-02
SLX4	4,09E-01	-2,24E-01	8,77E-01	4,71E-02	8,16E-01	6,63E-02
TELO2	1,59E-02	4,55E-01	1,19E-03	5,66E-01	4,74E-07	8,32E-01
TOP3A	4,55E-31	1,25E+00	7,93E-54	1,64E+00	2,29E-71	1,88E+00
TOP3B	6,41E-01	-2,74E-01	4,45E-01	-3,83E-01	6,53E-01	-2,34E-01
UBA52	8,36E-02	-5,74E-01	5,97E-01	-2,03E-01	7,20E-02	-5,60E-01
UBB	3,53E-01	-2,91E-01	8,16E-01	-8,10E-02	2,12E-01	-3,41E-01
UBC	6,19E-01	-2,08E-01	8,55E-01	7,73E-02	3,10E-01	-3,42E-01
UBE2T	5,20E-18	1,03E+00	5,65E-39	1,50E+00	3,76E-56	1,80E+00
USP1	4,58E-06	7,20E-01	6,29E-07	7,62E-01	1,42E-08	8,52E-01
WDR48	6,17E-03	5,07E-01	2,96E-04	6,24E-01	1,30E-03	5,59E-01

Tabella 3 - Geni via FA/BRCA, DEGS evidenziati.

L'analisi dettagliata dell'espressione dei geni appartenenti al complesso centrale FA rivela una regolazione trascrizionale eterogenea, sebbene a livello proteico si osservi una regolazione generalmente coordinata (Tabella 3). Nel complesso AG20, *FANCA* e *FANCG* mostrano un'espressione crescente nel tempo, mentre *FAAP20* non cambia. Nel complesso ID2, *FANCI* aumenta progressivamente mentre *FANCD2* cresce da 0h ma si stabilizza ai *timepoint* successivi. Nel complesso CEF, invece, *FANCC* e *FANCE* non variano, mentre *FANCF* mostra una riduzione. Infine, nel complesso BL100, *FANCB* aumenta da 0h e poi resta stabile, mentre *FANCL* non varia. *FAAP100* mostra espressione relativamente costante ai primi intervalli di tempo, ma una lieve differenza significativa tra 0h e 24h.

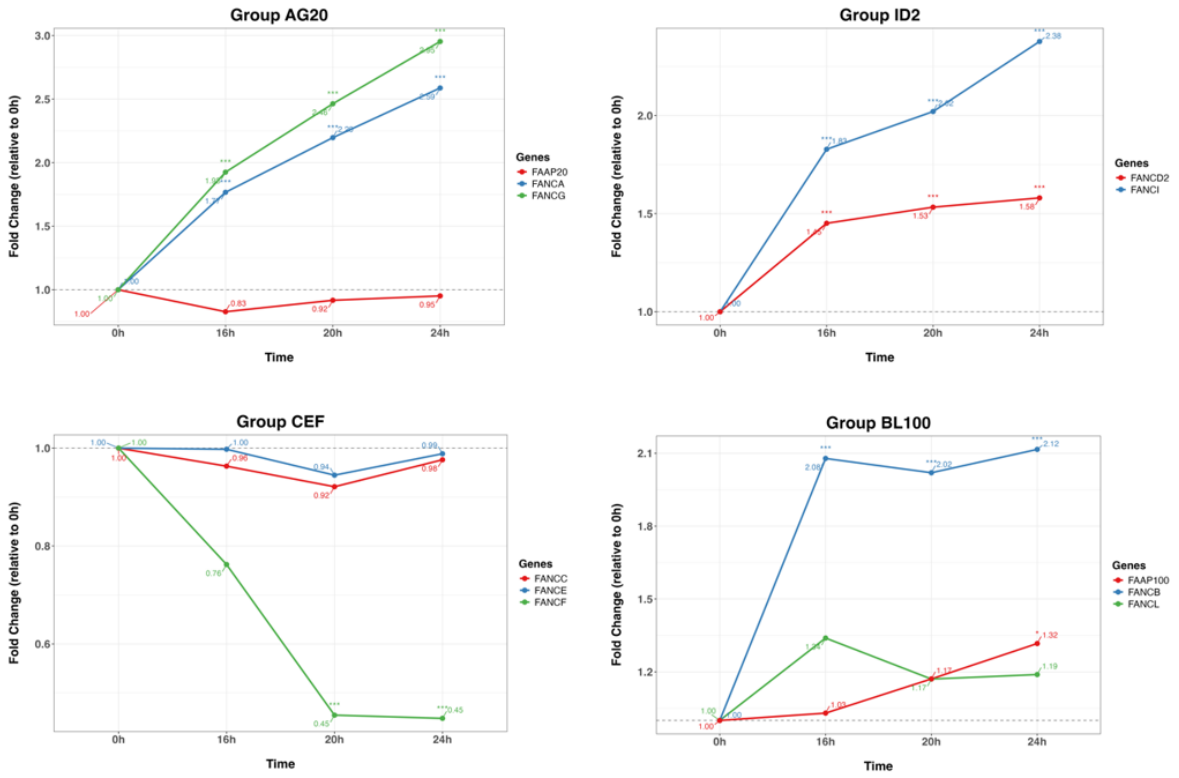


Figura 12 - Schema di espressione temporale dei geni del complesso centrale della via FA/BRCA. I valori sono normalizzati rispetto al punto temporale 0h (fold change = 1, indicato dalla linea tratteggiata) e rappresentano la media di tre repliche biologiche. La significatività statistica rispetto al punto temporale 0h è indicata da asterischi (*p < 0,05, **p < 0,01, ***p < 0,005).

In conclusione, questa eterogeneità nella regolazione trascrizionale, osservata all'interno dei singoli complessi, suggerisce che, se da un lato le analisi funzionali al FACS e i marcatori del western blot confermano la nota attivazione coordinata del complesso proteico FA in risposta al danno, le analisi di espressione genica via RNA-seq indicano che la cellula non risponde allo stress genotossico mediante un'induzione trascrizionale massiva di tutti i componenti della via, ma applica una modulazione gene-specifica.

4.4. Analisi globale dell'espressione genica e funzionale

Per avere una visione più ampia dell'effetto del trattamento, sono stati analizzati tutti i geni. A 16 ore, 4.505 geni risultano differenzialmente espressi in maniera significativa (padj < 0,05), con prevalenza di sovraregolati (2.833) rispetto ai sottoregolati (1.672). A 20 ore, il numero di geni sale a 6.899 con prevalenza di sovraregolati (4.124) rispetto ai sottoregolati (2.775), mentre a 24 ore la risposta trascrizionale raggiunge il picco, con 7.740 geni DE significativi, di cui 4.470 sovraregolati e 3.270 sottoregolati.

Per identificare con maggiore rigore i principali geni codificanti per proteine coinvolti nella risposta e ridurre la presenza di possibili falsi positivi derivanti dall'elevato numero di trascritti analizzati, è stata applicata una soglia di significatività più stringente (padj < 0,005). Su questo dataset selezionato è stata quindi condotta un'analisi di sovrapposizione temporale mediante diagramma di Venn (Figura 13), per identificare i geni differenzialmente espressi in comune tra i tre confronti temporali. In totale abbiamo trovato 5.540 geni a padj < 0,005.

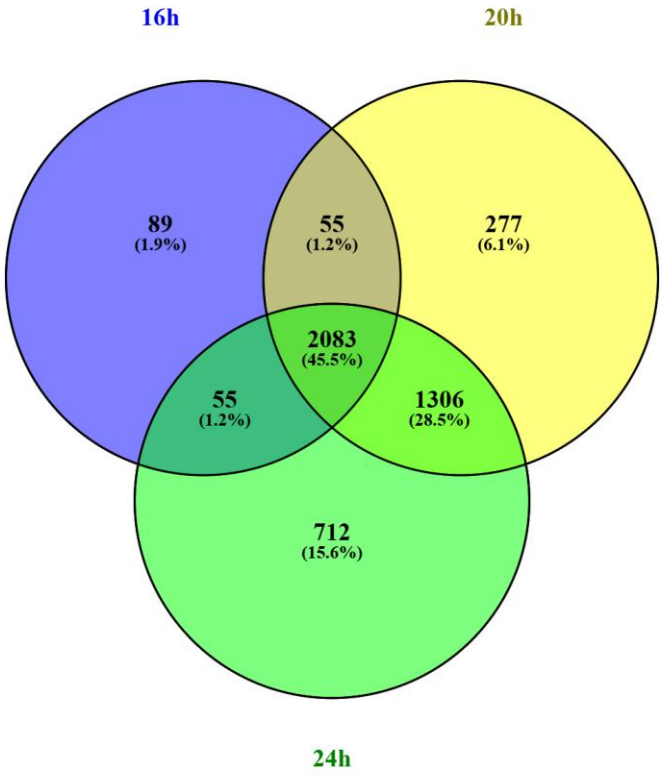


Figura 13 - distribuzione dei 5.540 geni a padj < 0,005 negli intervalli temporali a 16h,20h,24h

Questa selezione ha permesso di identificare un set ristretto di geni condivisi potenzialmente coinvolti in processi biologici rilevanti. Per approfondire il significato funzionale di questi

geni, è stata condotta un'analisi di arricchimento funzionale tramite DAVID (Database for Annotation, Visualization and Integrated Discovery). Dall'analisi sono stati selezionati dieci principali *cluster* contenenti i processi biologici più rilevanti (Tabella 4 e Figura 14). I valori di "Average Fold Enrichment" e "Enrichment Score" rappresentano rispettivamente l'abbondanza media di espressione di ciascun gruppo di geni rispetto al controllo e il grado di arricchimento di quei gruppi tra i geni maggiormente differenziati. Questi valori danno una misura dell'importanza e della rilevanza dei diversi processi nelle condizioni sperimentali analizzate.

Cluster_ID	Cluster_Name	N_Terms	Total_Unique_Genes	Min_PValue	Min_FDR	Avg_FoldEnrichment	Enrichment_Score
C1	Regulation of gene expression	27	1329	2,77E-85	2,48E-81	1,5	28,4
C2	DNA Unwinding	3	45	1,76E-12	3,15E-10	3,1	11,1
C3	DDR	31	715	2,89E-35	3,24E-32	2,7	9,85
C4	Cell cycle	58	449	6,57E-29	5,90E-26	2,3	8,59
C5	DNA replication	21	121	6,18E-24	3,26E-21	4,8	7,52
C6	Protein modification	23	654	4,34E-15	8,30E-13	1,5	4,93
C7	Structural organization	57	1061	2,05E-18	5,11E-16	2,1	4,85
C8	Centrosome cycle	10	94	7,40E-08	7,39E-06	2,8	4,36
C9	Nucleic acid metabolism	21	584	5,96E-20	2,06E-17	2,0	4,31
C10	Cell Death	15	358	2,10E-06	1,56E-04	1,7	3,82

Tabella 4 - Tabella riassuntiva dei 10 cluster principali.

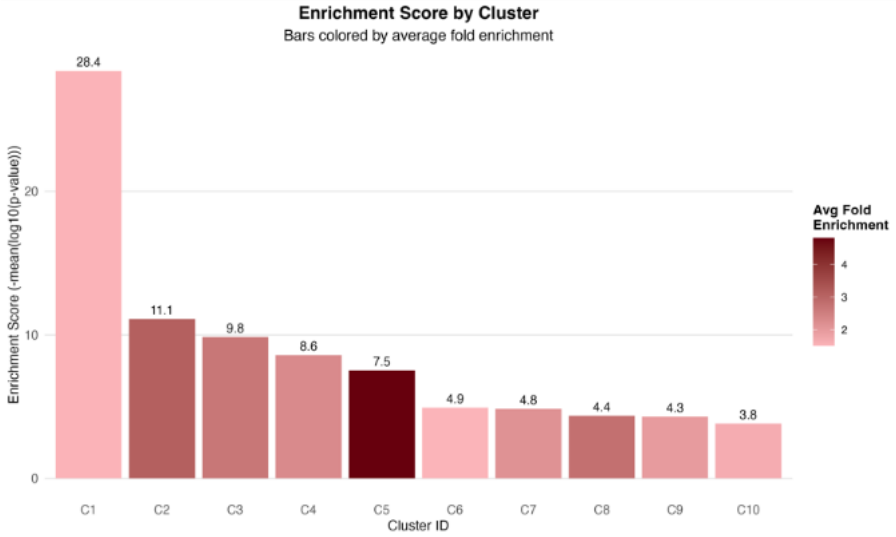


Figura 14 - Rappresentazione grafica dei 10 *cluster* principali.

In particolare, il gruppo "Regolazione dell'espressione genica" (C1) emerge come il più rilevante, con il valore FDR p più significativo (2,48E-81) e lo score di arricchimento più alto (28,4), comprendente 1329 geni distribuiti in 27 GO terms. Il suo ruolo centrale suggerisce una

rilevante regolazione trascrizionale nei processi cellulari analizzati. Anche il gruppo "DDR" (C3), coinvolto nella risposta al danno al DNA, mostra forte significatività (p-value FDR 3,24E-32) con 715 geni e 31 GO terms, evidenziando l'attivazione dei meccanismi di riparazione genomica. Questo si collega ad altri gruppi rilevanti come "Ciclo cellulare" (C4, p-value FDR 5,90E-26, 449 geni) e "Replicazione del DNA" (C5, p-value FDR 3,26E-21), che sottolineano il coinvolgimento di processi essenziali per l'integrità genomica e la proliferazione cellulare. Il gruppo "DNA Unwinding" (C2), pur contenendo solo 3 GO terms e 45 geni, è caratterizzato da una notevole specificità funzionale (arricchimento 3.1) e significatività statistica (p-value FDR 3,15E-10), suggerendo il ruolo chiave dello svolgimento del DNA nei processi analizzati. Altri gruppi importanti includono "Organizzazione strutturale" (C7, p-value FDR 5,11E-16), "Metabolismo degli acidi nucleici" (C9, p-value FDR 2,06E-17) e "Modificazione proteica" (C6, p-value FDR 8,30E-13), che rappresentano processi cellulari di base ampiamente coinvolti nelle condizioni studiate. Il gruppo "Morte cellulare" (C10, p-value FDR 1,56E-04) mostra un impatto più moderato ma comunque significativo.

L'analisi di similarità tra *cluster*, basata sull'indice di Jaccard, ha rivelato la presenza di una rete di interazioni funzionali tra i diversi gruppi di processi biologici. L'indice di Jaccard, che quantifica la sovrapposizione genica tra due gruppi (valori da 0 a 1), ha evidenziato i seguenti collegamenti chiave (Figura 15):

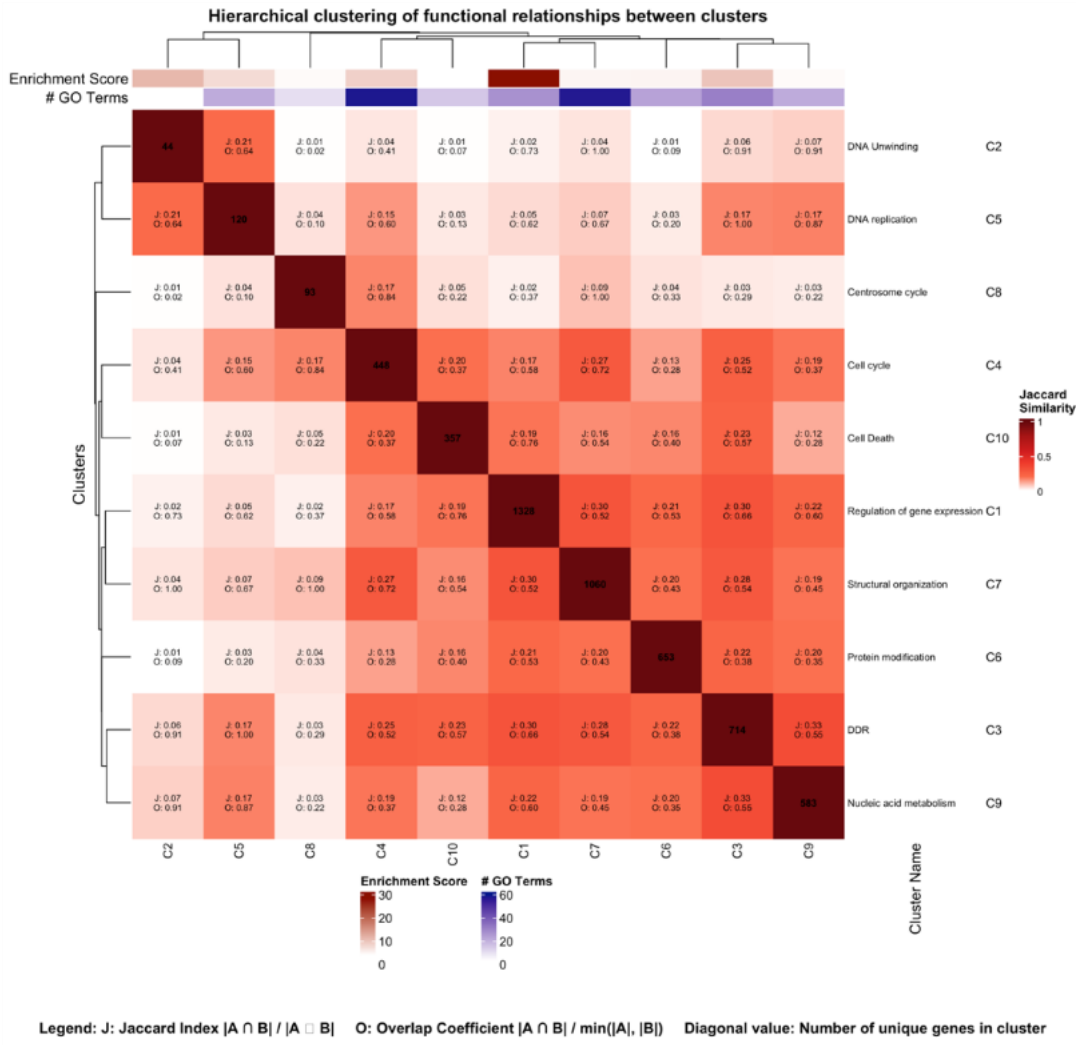


Figura 15 - Heatmap che mostra le relazioni funzionali tra cluster.

Elevata interconnessione tra gruppi regolatori: I gruppi C1 (regolazione dell'espressione genica), C3 (DDR - risposta al danno) e C7 (organizzazione strutturale) mostrano il maggior grado di sovrapposizione (Jaccard tra 0,28-0,30), suggerendo una stretta coordinazione tra regolazione genica, risposta al danno e organizzazione cellulare.

Relazione tra risposta al danno e metabolismo degli acidi nucleici: I cluster C3 (DDR) e C9 (metabolismo acidi nucleici) mostrano elevata similarità (Jaccard = 0,33), evidenziando il legame tra meccanismi di riparo e metabolismo degli acidi nucleici.

Connessione tra ciclo cellulare e organizzazione strutturale: I gruppi C4 (ciclo) e C7 (organizzazione strutturale) hanno sovrapposizione rilevante (Jaccard = 0,27), riflettendo il legame tra ciclo cellulare e i cambiamenti nell'organizzazione della cellula. In conclusione,

l’analisi funzionale e di sovrapposizione evidenzia che la risposta alla MMC non si limita all'attivazione dei classici geni di riparazione del DNA. Come atteso dalla letteratura (Deans & West, 2011), la cellula mette in atto una risposta integrata e coordinata: prima di tutto modifica la struttura della cromatina e lo svolgimento del DNA (C2 e C7) per rendere il danno accessibile, riorganizza il metabolismo degli acidi nucleici (C9) per sostenere la riparazione e, nel frattempo, mantiene sotto controllo l'innesco della morte cellulare (C10). L'insieme di questi gruppi riflette quindi una precisa strategia biologica in cui la riprogrammazione dell'espressione genica lavora in parallelo con la riparazione del danno per garantire la sopravvivenza della cellula.

Successivamente, è stata esaminata la distribuzione dei geni della via FA/BRCA nei gruppi funzionali identificati. Dei 58 geni analizzati, 25 sono stati trovati nei gruppi, pari al 43,1% del totale.
L’analisi ha rivelato una marcata concentrazione di questi geni nei gruppi C3 (DDR - risposta al danno) e C9 (metabolismo acidi nucleici), che contengono ciascuno 24 geni FA/BRCA. Questo dato è particolarmente rilevante poiché indica che oltre il 41% dell’intera via FA/BRCA è associata ai processi di riparo e metabolismo degli acidi nucleici, sottolineando il ruolo centrale di questi meccanismi nella stabilità genomica. Anche il gruppo C7 (organizzazione strutturale) mostra arricchimento significativo, includendo 17 geni FA/BRCA (29,3%) (Figura 16). Ciò suggerisce che la regolazione della struttura della cromatina e dell’organizzazione cellulare può essere fondamentale per il funzionamento della via FA/BRCA, facilitando probabilmente l’accesso dei complessi di riparo ai siti danneggiati. Nel complesso, questa distribuzione evidenzia il coinvolgimento predominante della via FA/BRCA nei processi di risposta al danno, metabolismo degli acidi nucleici e organizzazione strutturale, rafforzando l’idea di una coordinazione funzionale tra questi processi nella protezione dell’integrità genomica.

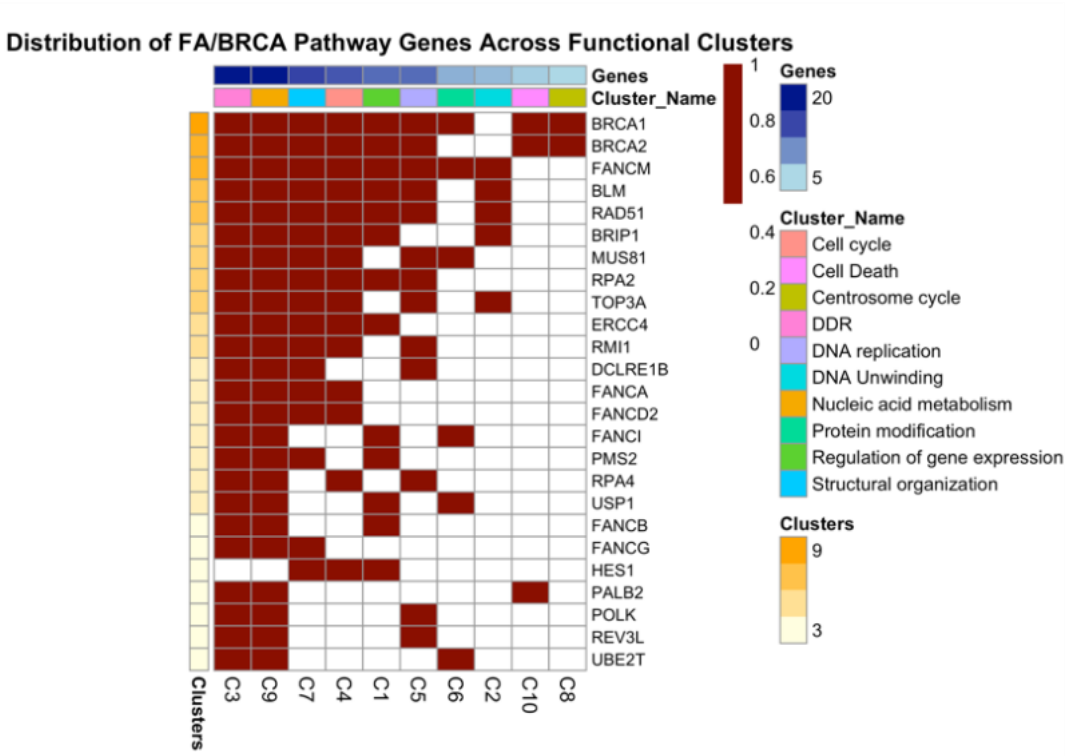


Figura 16 - Heatmap della distribuzione dei geni FA/BRCA nei cluster funzionali individuati.

L'analisi di arricchimento funzionale ha mostrato che il gruppo C2 (DNA Unfolding) possiede il valore di fold enrichment più elevato (37,89 con pvalue FDR=3,37E-07), seguito dal gruppo C5 (DNA Replication) con un arricchimento di 36,12 (pvalue FDR=8,63E-17). Anche il gruppo C9 mostra un arricchimento notevole (13,73; pvalue FDR=9,81E-21). Questi valori molto significativi sottolineano quanto i geni FA/BRCA siano fondamentali nei processi di svolgimento e replicazione del DNA, nonché nel metabolismo dei nucleotidi. Questa specificità funzionale evidenzia come tali geni rappresentino componenti essenziali nei meccanismi di mantenimento dell'integrità genomica, suggerendo che la loro disfunzione possa compromettere gravemente l'efficienza di tali processi (Tabella 5).

Cluster_Name	N_FA_BRCA_Genes	N_Total_Genes	Fold_Enrichment	P_value	FDR
C2: DNA Unwinding	5	44	37,89	2,02E-07	3,37E-07
C5: DNA replication	13	120	36,12	2,59E-17	8,63E-17
C9: Nucleic acid metabolism	24	583	13,73	9,81E-22	9,81E-21
C3: DDR	24	714	11,21	1,11E-19	5,53E-19
C4: Cell cycle	15	448	11,16	2,82E-12	7,04E-12
C8: Centrosome cycle	2	93	7,17	3,18E-02	3,53E-02
C7: Structural organization	17	1060	5,35	7,49E-09	1,50E-08
C1: Regulation of gene expression	13	1328	3,26	1,22E-04	1,75E-04
C6: Protein modification	6	653	3,06	1,32E-02	1,65E-02
C10: Cell Death	3	357	2,80	9,18E-02	9,18E-02

Tabella 5 - Risultati dell'analisi di arricchimento dei geni della via FA/BRCA.

4.5 Analisi dei lncRNA differenzialmente espressi

Nel contesto dell'anemia di Fanconi, è noto che i geni codificanti per proteine coinvolti nella via FA/BRCA vengono attivati in risposta a danno genotossico, mentre il ruolo dei lncRNA nella regolazione di questa risposta rimane ancora poco caratterizzato. Per questo motivo è stato analizzato il loro profilo trascrizionale in risposta al danno indotto da MMC, mettendolo a confronto con quello dei geni codificanti proteine, al fine di inquadrare possibili meccanismi regolatori coinvolti nel processo della risposta cellulare.

A partire dai dati di RNA-seq precedentemente elaborati, è stata effettuata un'analisi mirata sui lncRNA differenzialmente espressi per comprendere la loro dinamica trascrizionale nella risposta al danno indotto. Sono stati selezionati i trascritti con padj < 0.05, considerati significativamente deregolati, e successivamente confrontati tra diversi intervalli temporali (16h, 20h, 24h rispetto a 0h). A 16 ore, 550 lncRNA risultano differenzialmente espressi in maniera significativa, di cui sovraregolati 259 e sottoregolati 291. A 20 ore, il numero di lncRNA deregolati sale a 1.082 con 470 sovraregolati e 612 sottoregolati, mentre a 24 ore la risposta trascrizionale raggiunge il picco, con 1.344 lncRNA differenzialmente espressi, di cui 521 sovraregolati e 823 sottoregolati.

Per identificare profili comuni di regolazione, i lncRNA differenzialmente espressi nei vari confronti sono stati sovrapposti utilizzando lo strumento Venny, generando un diagramma di Venn. Questo approccio ha permesso di classificare i trascritti deregolati tra coppie di intervalli temporali (16h vs 20h, 16h vs 24h, 20h vs 24h), comuni a tutti i confronti ed esclusivi di ciascun intervallo temporale.

Questa classificazione ha evidenziato quali trascritti sono coinvolti in maniera persistente e quali mostrano una regolazione specifica in determinati momenti della risposta al danno. In totale, sono stati identificati 1.554 lncRNA espressi in almeno una condizione, di cui 476 presenti in tutti gli intervalli temporali, come illustrato nella Figura 17.

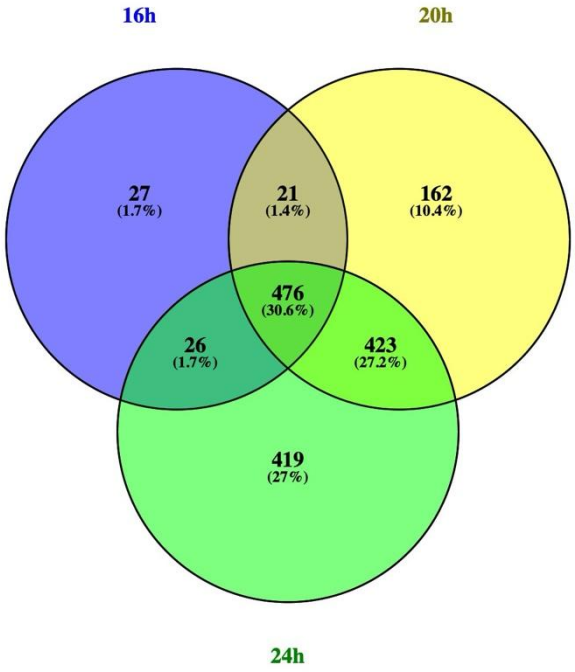


Figura 17 – Diagramma di Venn che mostra i 1.554 lncRNA espressi nelle diverse condizioni temporali analizzate

Dopo aver identificato i lncRNA differenzialmente espressi, questi sono stati confrontati con i geni codificanti proteine (PCGs) anch'essi differenzialmente espressi, per valutare se nella risposta trascrizionale i due gruppi mostrassero un profilo analogo o divergente nel tempo. La tabella in figura riporta, per ciascun intervallo temporale (0h vs 16h, 20h, 24h), il numero di geni sovraregolati, sottoregolati e il totale dei geni differenzialmente espressi (Figura 18).

	0h vs 16h			0h vs 20h			0h vs 24h		
	UP	**DOWN**	**TOT**	**UP**	**DOWN**	**TOT**	**UP**	**DOWN**	**TOT**
PCGs	2833 (63%)	1672 (37%)	4505	4124 (60%)	2775 (40%)	6899	4470 (58%)	3270 (42%)	7740
lncRNAs	259 (47%)	291 (53%)	550	470 (43%)	612 (57%)	1082	521 (39%)	823 (61%)	1344

Figura 18 - PCGs: la maggior parte è sovraregolata (es. 63% a 16h, 58% a 24h).
lncRNA: distribuzione più equilibrata, con tendenza alla sottoregolazione (es. 61% a 24h).

A differenza dei geni codificanti per proteine — in cui prevale una risposta prevalentemente induttiva — i lncRNA mostrano principalmente una risposta di sottoregolazione, che diventa dominante nella fase tardiva (24h). Questo andamento riflette la duplice funzione dei lncRNA come regolatori epigenetici e trascrizionali: da un lato, lo spegnimento di una vasta classe di questi geni consente di arrestare i processi cellulari ordinari e rimuovere l'inibizione su geni chiave della risposta al danno; dall'altro, l'induzione di un numero più ristretto di lncRNA guida la regolazione dei meccanismi di riparazione.

4.6 Classificazione temporale dei pattern di espressione dei lncRNA e dei geni FA

Al fine di verificare se i lncRNA DEGs potessero condividere meccanismi di co-regolazione con i 58 geni della via FA identificati come DEGs, è stata effettuata una classificazione basata sui profili di espressione. I 1.554 lncRNA significativamente differenziati (padj < 0,05) sono stati suddivisi in sette gruppi, dalla A alla H, per facilitare l'interpretazione dei diversi profili di espressione. (Figura 19)

Ogni gruppo riflette una specifica dinamica di risposta:

Gruppo A (incremento lineare nel tempo, 0h < 16h < 20h < 24h): 56/1.554

Gruppo B (incremento a 16h e stabilizzazione, 0h < 16h = 20h = 24h): 420/1.554

Gruppo C (incremento tardivo, 0h = 16h < 20h ≤ 24h): 423/1.554

Gruppo E (regolazione solo a 24h): 419/1.554

Gruppo F (regolazione solo a 20h): 162/1.554

Gruppo G (regolazione solo a 16h): 27/1.554

Gruppo H (andamento variabile): 47/1.554

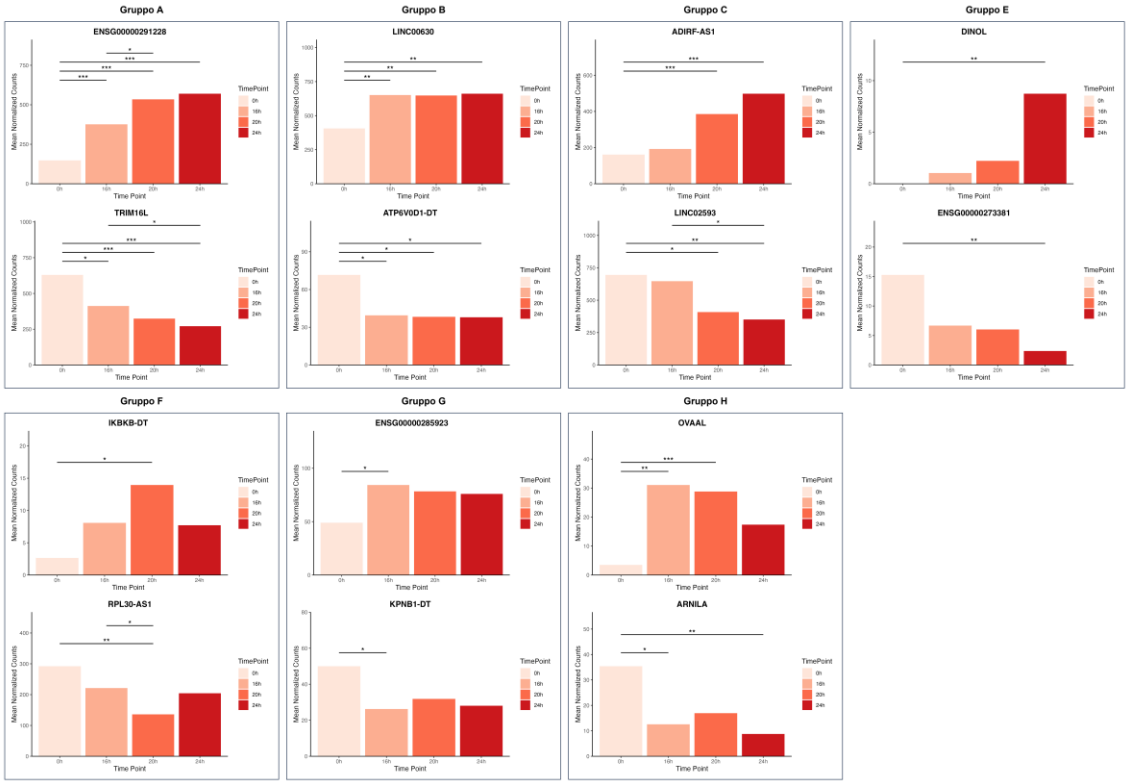


Figura 19 - i lncRNA sono stati suddivisi in gruppi in base ai loro pattern temporali di espressione: alcuni mostrano un aumento progressivo (Gruppo A), altri una variazione precoce (Gruppo B) o tardiva (Gruppo C), mentre altri ancora presentano una regolazione specifica a singoli time-point (Gruppi E, F e G) o andamento variabile (Gruppo H)

Escludendo i 47 trascritti ad andamento variabile, la distribuzione dei lncRNA tra i gruppi con dinamica definita rivela come la risposta sia quasi equamente ripartita tra tre profili dominanti: il Gruppo C (risposta tardiva, 28,1%), il Gruppo B (attivazione precoce e persistente, 27,9%) e il Gruppo E (regolazione tardiva a 24h, 27,8%). I restanti gruppi a risposta specifica nell'intervallo di tempo o progressiva, rappresentano una frazione più contenuta, guidata dal Gruppo F (10,8%), seguito dal Gruppo A (3,7%) e dal Gruppo G (1,7%).

Successivamente l'analisi focalizzata sui 58 geni della via FA ha evidenziato che la maggior parte di essi ha mostrato un incremento progressivo nel tempo, coerente con il loro ruolo nella risposta al danno cumulativo al DNA (Figura 20).

La distribuzione dei 58 geni della via FA nei vari gruppi risulta così suddivisa:

Gruppo A (incremento lineare): 11/58

Gruppo B (incremento a 16h e stabilizzazione): 16/58

Gruppo C (incremento tardivo): 6/58

Gruppo D (decrescita dopo 16h): 2/58

Gruppo NV (nessuna variazione significativa): 23/58

Questa distribuzione ha mostrato un incremento progressivo nel tempo. Nelle variazioni significative, il profilo dominante è rappresentato dal Gruppo B (incremento precoce e stabilizzazione), seguito dal Gruppo A (incremento lineare e continuo). I restanti geni mostrano una frazione più contenuta guidata dal Gruppo C (risposta tardiva) e, infine, dal Gruppo D (decrescita dopo 16 ore) come si vede nella Figura 20.

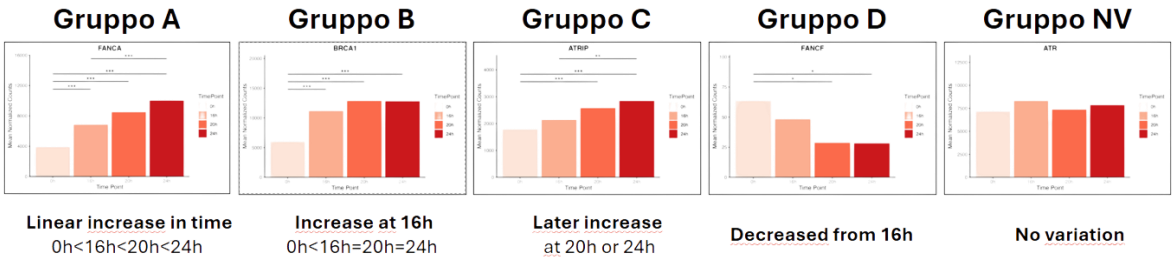


Figura 20 - Andamento temporale dell'espressione dei geni nei cinque gruppi (A, B, C, D, NV), con evidenza delle diverse dinamiche di variazione ai time point 0h, 16h, 20h e 24h.

In conclusione, l'analisi comparativa evidenzia una diversa dinamica temporale tra la risposta dei lncRNA e quella dei geni del *pathway* dell'Anemia di Fanconi.

In questi ultimi, la risposta appare omogenea e orientata all'attivazione sostenuta: oltre il 57% dei geni della via FA mostra, infatti, un incremento progressivo e continuo nel tempo. Al contrario, i lncRNA mostrano una risposta distribuita su più fasi temporali. Questa differente distribuzione suggerisce che, mentre i PCG garantiscono la presenza continua delle proteine di

riparazione del danno, i lncRNA intervengono con funzioni regolatorie nel tempo, accompagnando i diversi passaggi della risposta cellulare al danno.

4.7 Analisi dei fattori di trascrizione E2F1 ed E2F4 nella regolazione dei lncRNA e dei geni FA

Per approfondire ulteriormente i meccanismi di regolazione trascrizionale, sono stati esaminati i potenziali fattori di trascrizione (TF) coinvolti nel controllo dei geni appartenenti al *pathway* dell'Anemia di Fanconi e successivamente del trascrittoma differenzialmente espresso. A questo scopo, è stato utilizzato lo strumento ChEA (ChIP-X Enrichment Analysis), che possiede un database di interazioni TF–target, selezionate tramite dati sperimentali di ChIP-seq.

L'inserimento dei 58 geni del percorso FA nel portale ChEA ha restituito una lista ordinata di fattori di trascrizione potenzialmente associati. Tra questi, E2F1 è risultato il più significativamente arricchito. È stato quindi selezionato assieme a E2F4, noto per agire in modo complementare come repressore trascrizionale attraverso il reclutamento di complessi di modulazione della cromatina.

La scelta di focalizzarsi su E2F1 ed E2F4 è stata ulteriormente supportata dal loro ruolo documentato in letteratura nella risposta al danno al DNA e nella regolazione del ciclo cellulare: E2F1 funge da attivatore trascrizionale durante la fase S e in condizioni di danno al DNA, mentre E2F4 opera prevalentemente in G0/G1 contribuendo al silenziamento genico. La regolazione trascrizionale dei geni FA è strettamente dipendente dalla progressione del ciclo cellulare, ciò assicura che i livelli di proteine del complesso siano massimi in fase S, quando il rischio di danno al DNA è più elevato (Spardy et al., 2007). Inoltre, in uno studio sui fibroblasti normali, è stato evidenziato come la sovraespressione di E2F1 inducesse l'aumento dei livelli di mRNA di FANCC, mentre in linee cellulari HeLa i geni FANCA, FANCC e FANCJ risultano direttamente sovraregolati da E2F1 in saggi reporter. Al contrario, altri geni come FANCD2 non mostrano la medesima risposta (Tategu et al., 2007).

Il fattore E2F4 invece opera all'interno del complesso repressivo DREAM reclutandosi sui motivi regolatori dei promotori dei geni FA (tra cui FANCD2, FANCI, FANCC) per mediarne il silenziamento in condizioni di quiescenza o di arresto del ciclo cellulare via p53/p21 (Jaber et al., 2016).

Alla luce del ruolo biologico dei fattori E2F1 ed E2F4 emerso in letteratura, è stata condotta un'analisi in silico per verificare se potessero contribuire alla regolazione trascrizionale dei

lncRNA e dei geni codificanti proteine differenzialmente espressi nel nostro modello di danno indotto da MMC.

A tal fine, è stata utilizzata l'annotazione genica basata sul riferimento genomico GRCh38.p14 estratta da GENCODE, versione v46 e sono state analizzate le regioni del promotore definite nell'intervallo di ±3000 bp rispetto al sito di inizio della trascrizione (TSS) per valutare l'eventuale sovrapposizione con i picchi di arricchimento ottenuti da esperimenti di ChIP-seq per E2F1 ed E2F4, che sono stati scaricati dal portale ENCODE. i siti di legame dei fattori di trascrizione e le regioni definite attorno ai TSS sono stati confrontati mediante bedtools intersect, che ha restituito un file contenente le regioni dei trascritti che presentavano promotori occupati da almeno uno dei due fattori di trascrizione, insieme all'annotazione completa del picco stesso. In particolare, la validazione visiva condotta tramite i Genome Browser (UCSC ed ENCODE) ha evidenziato un chiaro arricchimento del segnale ChIP-seq in corrispondenza delle regioni regolative dei geni analizzati. Come illustrato per il gene FANCA (Figura 21), è stato osservato un marcato picco di legame situato subito a monte del TSS. La posizione e l'intensità del segnale risultano coerenti con l'occupazione diretta del promotore da parte del fattore di trascrizione, confermando che i geni analizzati presentano elementi regolativi accessibili al controllo da parte di E2F1 ed E2F4.

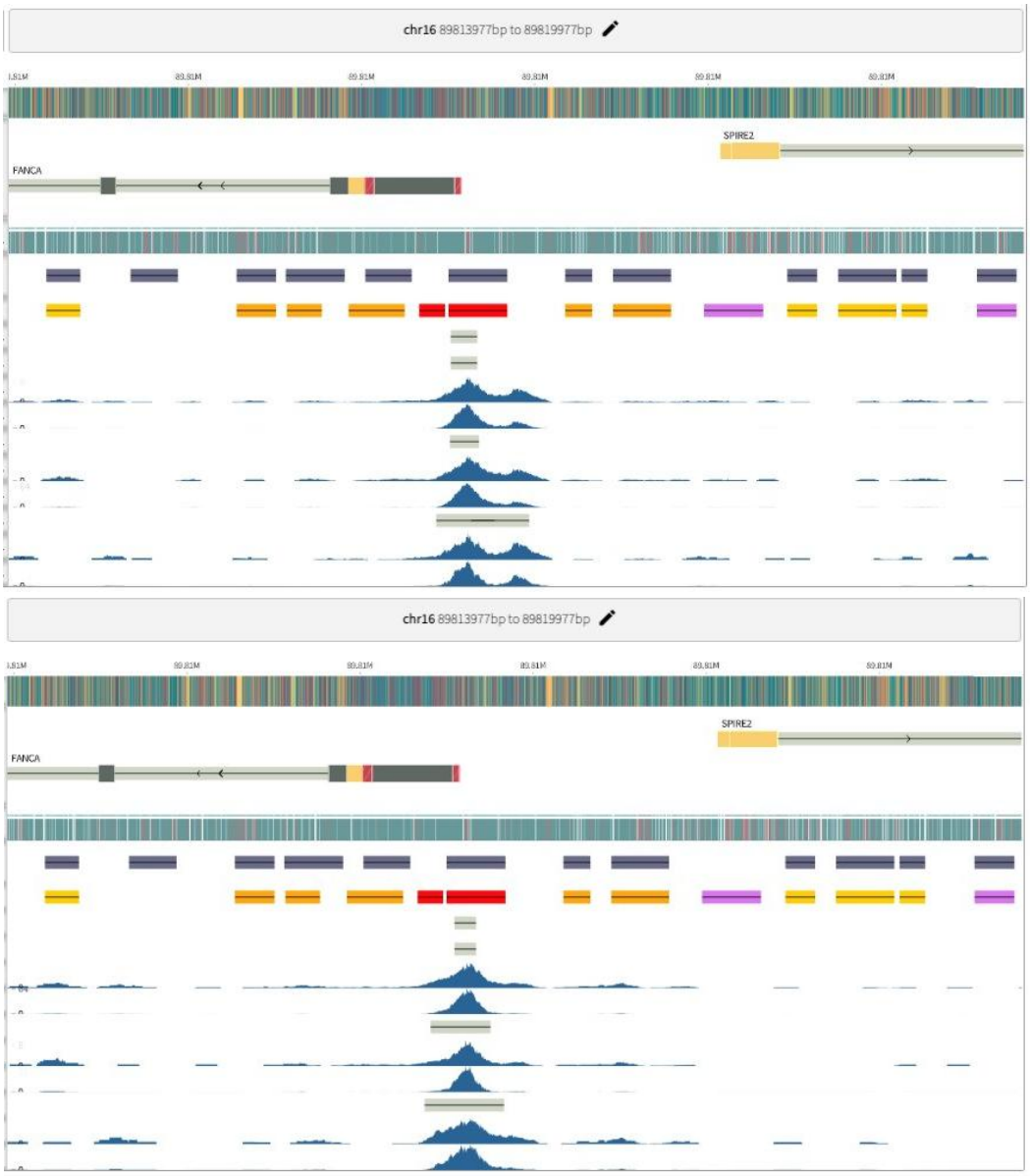


Figura 21 - intersezione tra promotore e fattore di trascrizione rispettivamente di E2F1 (alto) ed E2F4 (basso) per il gene FANCA

Nel browser ENCODE in figura, ciascun picco compare come un segmento orizzontale posizionato lungo il genoma di riferimento, rappresentato come una regione che possiede una traccia di intensità che riflette il livello di arricchimento del fattore in quella particolare coordinata genomica. L'altezza del segnale è proporzionale al numero di letture ChIP-seq allineate in quella regione e costituisce un'indicazione della forza relativa del legame: valori più elevati corrispondono tipicamente a un legame più robusto o più frequente.

4.7.1 Incrocio tra siti di legame e potenziali promotori PCGs DEGs

L'analisi dell'intersezione tra i siti di legame di E2F1 ed E2F4 e le regioni definite del promotore dei geni codificanti differenzialmente espressi ha messo in evidenza un coinvolgimento esteso di entrambi i fattori di trascrizione. 1890 presentano un picco di E2F1 nel loro promotore, mentre 1272 geni mostrano un picco riconducibile a E2F4. L'ampiezza di questi gruppi indica che l'attivazione del programma trascrizionale di risposta al danno al DNA include un forte controllo trascrizionale mediato dai fattori E2F1 ed E2F4, coerentemente con il loro ruolo nella transizione G1/S e nella risposta allo stress replicativo.

Per identificare i geni condivisi è stato utilizzato un diagramma di Venn, che ha consentito di visualizzare le sovrapposizioni tra gruppi di geni associati ai vari fattori di trascrizione e valutare quali fossero esclusivi di un gruppo o comuni a più gruppi. Da questa analisi è emerso che:

-1014 presentano un picco ChIP-seq esclusivo per E2F1 nel loro promotore
-396 presentano un picco ChIP-seq esclusivo per E2F4 nel loro promotore
-876 mostrano overlap nei loro promotori con entrambi i fattori (E2F1 + E2F4)

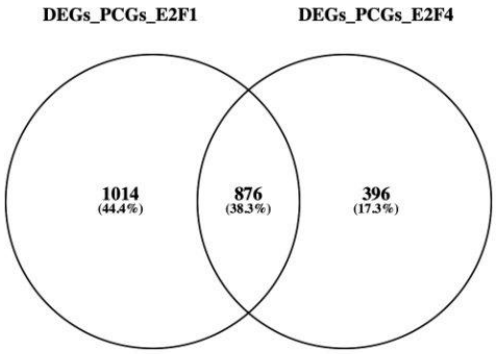


Figura 22 - Il diagramma di Venn illustra la ripartizione dei PCGs differenzialmente espressi in base alla presenza di E2F1 o E2F4 nel promotore

Per analizzare le funzioni biologiche associate ai geni sopra citati, è stata effettuata un'analisi di arricchimento Gene Ontology (GO) utilizzando lo strumento DAVID, per verificare se la regolazione dei fattori di trascrizione fosse effettivamente collegata ai processi attesi nel modello della FA.

La tabella risultante dall'analisi (Tabella 6), ordinata secondo il valore di FDR/Benjamini, mostra un arricchimento significativo per un insieme di processi altamente coerenti con la fisiopatologia della FA e con i ruoli canonici dei fattori di trascrizione E2F1 ed E2F4. Tra i termini GO maggiormente arricchiti emergono infatti:

-Risposta al danno al DNA (DDR) con FDR $\approx 9.9\times10^{-52}$ è risultato il processo biologico più rappresentato all'interno dei geni target di E2F1-E2F4, indicando una forte convergenza tra regolazione trascrizionale e attivazione dei meccanismi di riparo del DNA.

-Riparazione del DNA (incluse categorie specifiche come Riparazione delle rotture a doppio filamento). Questi processi sono strettamente associati alla ricombinazione omologa e alla stabilità genomica, aspetti centrali nella fisiologia cellulare della FA.

-Ciclo cellulare, regolazione del ciclo cellulare e progressione G1/S. Questi termini riflettono direttamente il ruolo canonico di questi fattori di trascrizione nel controllo dell'ingresso in fase S e dei checkpoint attivi in risposta al danno genomico.

-Processi metabolici degli acidi nucleici Indicano un coinvolgimento più ampio nella regolazione della trascrizione, replicazione e gestione dell'integrità genomica.

Term	Genes	Count	P-Value	Fold Enrichment	FDR
DNA damage response	11.30%	259	1.27e-55	2.75	9.92e-52
cellular response to stress	17.75%	407	1.42e-53	2.11	5.58e-50
DNA metabolic process	10.34%	237	3.94e-48	2.67	1.03e-44
cell cycle process	11.30%	259	4.55e-47	2.50	8.92e-44
regulation of cell cycle	12.69%	291	5.11e-46	2.32	8.02e-43
cell cycle	11.82%	271	6.03e-45	2.38	7.88e-42
DNA repair	7.85%	180	7.59e-42	2.90	8.50e-39
regulation of nucleobase-containing compound metabolic process	30.79%	706	1.69e-38	1.52	1.66e-35
regulation of cell cycle process	9.33%	214	2.13e-37	2.46	1.85e-34
mitotic cell cycle	7.85%	180	9.34e-36	2.64	7.32e-33

Tabella 6– Tabella dei termini GO significativamente arricchiti nei target E2F1/E2F4.

Un elemento particolarmente rilevante è la forte coerenza dei processi arricchiti in questa analisi rispetto a quelli emersi nell'analisi globale sui geni con padj < 0,005. Questi risultati indicano in modo chiaro che i geni associati al legame di E2F1 ed E2F4 convergono sui processi biologici maggiormente coinvolti nella risposta FA indotta da danno al DNA. In particolare, i *target* dei due fattori risultano significativamente arricchiti in *pathway* chiave quali la risposta al danno al DNA e la regolazione del ciclo cellulare. L'analisi specifica di E2F1 ed E2F4 mostra come una parte rilevante di quei processi visti nell'analisi globale (Tabella 4 – Figura 14) possa essere direttamente controllata da questi fattori di trascrizione, indicando una possibile modulazione trascrizionale integrata tra danno al DNA e *checkpoint* del ciclo cellulare.

È stato inoltre realizzato un diagramma di Venn più specifico, sui 58 geni coinvolti nella via FA/BRCA, con l'obiettivo di analizzare come questi geni si distribuiscono rispetto ai siti di legame dei fattori di trascrizione E2F1 ed E2F4 (Figura 23). In totale sono stati trovati 36 geni appartenenti al *pathway*, di questi:

- nr. 6 si trovano esclusivamente su E2F1, nel gruppo troviamo: *REV1*, *SLX4*, *FAN1*, *HES1*, *UBB, UBC*. REV1 svolge un ruolo cruciale nella sintesi translesione durante la fase S, un contesto biologico in cui E2F1 agisce come attivatore genico (Denechaud et al., 2017). SLX4 è una proteina di supporto per l'assemblaggio di nucleasi essenziale per la risoluzione dei crosslink del DNA nella fase di replicazione, ambito in cui E2F1 modula l'espressione di molti geni coinvolti nella risposta allo stress replicativo (Fouad & Hauton, 2021). HES1 collabora con E2F1 nella regolazione di alcuni target, come CTNND2 in cellule di carcinoma prostatico; in presenza di E2F1 funge da co-regolatore (Lu et al., 2010). Inoltre, dato che E2F1 è regolato tramite ubiquitinazione, UBB e UBC possono svolgere funzioni indirette nel modulare la sua stabilità.
- nr. 9 sono stati trovati esclusivamente su E2F4: *EME1*, *FAAP24*, *FANCD2*, *FANCG*, *FANCM*, *RAD51C*, *RPA2*, *POLI*, *POLN*. Tra questi FANCD2 ha un'espressione regolata in coordinamento con la via Rb/E2F4, coinvolgendo E2F4 (Hoskins et al., 2008). FAAP24 Interagisce con il complesso centrale FA, la sua espressione è modulata in stato di arresto cellulare, suggerendo un coinvolgimento del complesso E2F4/p130 (Ciccia et al., 2007). RAD51C come membro della famiglia RAD51, viene regolato da E2F4 nei contesti di stress replicativo e quiescenza (Orhan et al., 2021).
- nr. 23 sono in comune in entrambi i fattori di trascrizione: *ATR*, *BLM*, *BRCA2*, *BRIP1*, *CENPS*, *DCLRE1A*, *DCLRE1B*, *RAD51*, *REV3L*, *RMI1*, *RPA1*, *TOP3A*, *FANCA*, *FANCB*, *FANCC*, *FANCE*, *FANCI*, *FANCL*, *MLH1*, *UBE2T*, *USP1*, *PALB2*, *POLH*. Durante la fase G1/S, E2F1 promuove l'espressione dei geni *target*, mentre nella fase quiescente (G0), in G1 o a seguito dell'attivazione dei checkpoint di controllo, E2F4 reprime la trascrizione reclutando le proteine p130 e p107. (Hoskins et al., 2008). Per esempio, tra questi geni *RAD51* è direttamente represso in condizioni di ipossia da complessi E2F4/p130, mentre E2F1 lo attiva in fase S o in riparazione del danno (Bindra & Glazer, 2007).

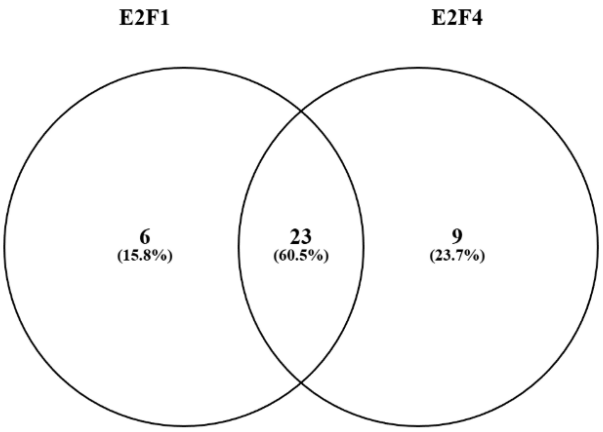


Figura 23 *– Diagramma di Venn dei 58 geni FA/BRCA confrontati con i target di E2F1 ed E2F4.*

4.7.2 Incrocio tra siti di legame e potenziali promotori di lncRNA DEGs

L'analisi sui geni differenzialmente espressi è stata successivamente estesa anche ai lncRNA, seguendo la stessa metodologia applicata ai geni codificanti per proteine, mettendo in luce un numero rilevante di lncRNA che potrebbero essere regolati da E2F1 ed E2F4. In totale, 280 lncRNA presentano siti di legame per E2F1 nel promotore, mentre 178 lncRNA mostrano siti di binding per E2F4. Per chiarire la modalità di controllo trascrizionale di tipo esclusivo o combinato, anche in questo caso è stato realizzato un diagramma ed è emerso che:

-142 lncRNA hanno un picco ChIP-seq esclusivo per E2F1 nel promotore

-40 lncRNA hanno un picco ChIP-seq esclusivo per E2F4 nel promotore

-138 lncRNA mostrano invece la presenza di entrambi i fattori (E2F1 + E2F4) nei loro promotori

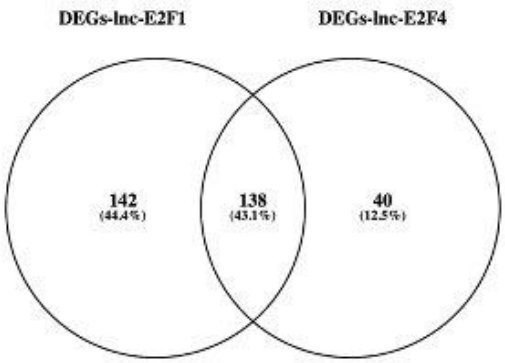


Figura 24 – Diagramma di Venn dei lncRNA confrontati con i target di E2F1 ed E2F4.

A partire dai 138 lncRNA inizialmente identificati come potenziali target dei fattori di trascrizione E2F1 ed E2F4, è stato possibile recuperare annotazioni funzionali predittive sul database lncHUB2. Questo database permette di associare tramite correlazione di Pearson, attuata tra lncRNA e geni codificanti proteine, una lista di potenziali *pathway* KEGG per ciascun lncRNA, fornendo metriche come il p-value e la Mean Pearson Correlation (MPC) per valutare la forza dell'associazione.

Il gruppo iniziale è stato ristretto a causa dell'incompleta caratterizzazione di molti trascritti non codificanti all'interno dei *repository* pubblici e sono state ottenute informazioni per 88 lncRNA. Esaminando i *pathway* KEGG associati a ciascun lncRNA, è emerso che 31 lncRNA su 88 (pari al 35,2%) presentano un'associazione con la via dell'Anemia di Fanconi. Alcuni lo riportano come *pathway* principale, mentre per altri compare in altre posizioni significative. Come illustrato nella Figura 25, il grafico riporta i primi 15 *pathway* KEGG più frequenti dell'intero dataset, classificati in base al numero di volte in cui ciascun *pathway* è stato identificato tra gli 88 lncRNA, unitamente ai rispettivi valori medi di MPC e -Log10(p-value).

Per valutare la rilevanza di questo arricchimento è fondamentale considerare sia la Mean Pearson Correlation (MPC) sia la significatività statistica espressa come -Log10(p-value):

L'MPC è un parametro che misura la forza della correlazione lineare media tra l'espressione di ciascun lncRNA e i geni codificanti proteine appartenenti ad uno specifico *pathway*. Dei valori elevati indicano una correlazione nell'espressione più forte e biologicamente rilevante. Nel nostro dataset, la via dell'Anemia di Fanconi registra l'MPC medio più alto in assoluto (0,028),

dimostrando che l'associazione riscontrata non è un artefatto statistico, ma riflette una co-regolazione trascrizionale coordinata tra i lncRNA analizzati e i geni del *pathway* FA.

Il valore di questo risultato emerge chiaramente dal confronto diretto con la categoria maggiormente rappresentata in termini di frequenza assoluta, "Ribosome biogenesis in eukaryotes" (n = 32). Sebbene il processo di biogenesi ribosomiale coinvolga un numero di lncRNA superiore di una singola unità rispetto alla via della FA (n = 31), esso registra un valore di MPC medio contenuto (0,011) e una significatività statistica inferiore con -Log10(p-value) = 2,10. Al contrario, la via dell'Anemia di Fanconi combina l'alta frequenza con un MPC medio quasi triplicato (0,028) e un'elevata significatività statistica con -Log10(p-value) = 2,30. In sintesi, si configura come il segnale biologicamente più consistente dell'intero dataset.

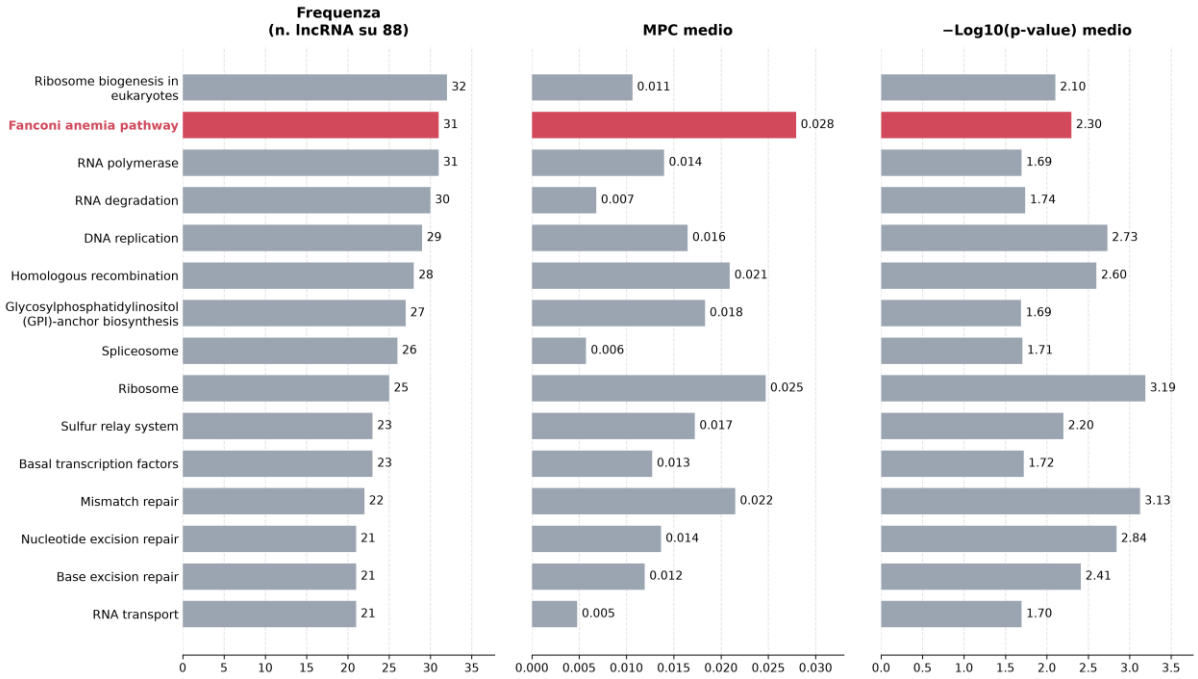


Figura 25 - Risultati dell'analisi di co-espressione (lncHUB2): i grafici mostrano per ogni via: la frequenza (lncRNA associati/88), la correlazione media di Pearson (MPC) e la significatività statistica con -Log10(p-value) . In rosso è evidenziata la via dell'Anemia di Fanconi.

Nel grafico si nota un chiaro arricchimento anche per altre vie direttamente coinvolte nella riparazione del DNA e nella stabilità del genoma.

In particolare, sebbene il termine "Ribosome" registri il secondo valore di correlazione più alto (n = 25, MPC = 0,025, -Log10(p-value) = 3,19), processi specifici di riparazione come il “Mismatch Repair” (n = 22, MPC = 0,022, -Log10(p-value) = 3,13) e la Ricombinazione

Omologa (Homologous recombination, n = 28, MPC = 0,021, -Log10(p-value) = 2,60) mostrano anch'essi valori di MPC molto elevati, posizionandosi subito dopo. A questi si aggiungono la Replicazione del DNA (DNA replication, n = 29, MPC = 0,016, -Log10(p-value) = 2,73) e gli altri meccanismi di riparazione per escissione, come il "Nucleotide Excision Repair" (NER, n = 21, MPC = 0,014, -Log10(p-value) = 2,84) e il "Base Excision Repair" (BER, n = 21, MPC = 0,012, -Log10(p-value) = 2,41).

La comparsa contemporanea di questi *pathway* ha un senso biologico preciso: la via dell'Anemia di Fanconi lavora infatti a stretto contatto con la ricombinazione omologa e con la replicazione del DNA per riparare i danni complessi (come gli ICLs) durante la fase di divisione cellulare. Trovare alti livelli di correlazione tra questi *pathway* e i lncRNA analizzati suggerisce che i fattori E2F1 ed E2F4 regolino in modo coordinato un intero gruppo di geni dedicati alla riparazione del DNA.

5.Conclusioni

Il presente lavoro di tesi ha permesso di caratterizzare la dinamica temporale della risposta cellulare allo stress genotossico indotto da Mitomicina C (MMC), combinando analisi molecolari in vitro e in silico per chiarire la regolazione trascrizionale della via dell'Anemia di Fanconi (FA/BRCA), delle vie funzionali correlate e dei relativi lncRNA regolatori.

L'analisi condotta sulle cellule HeLa ha evidenziato una distinzione temporale tra l'attivazione della cascata proteica e la risposta trascrizionale. Mentre l'attivazione del *checkpoint* del danno (pCHK1), l'accumulo di γH2AX e la monoubiquitinazione di FANCD2 si innescano tempestivamente a partire da 8 ore dal trattamento parallelamente all'arresto del ciclo cellulare in fase S, la risposta trascrizionale dei geni della via FA/BRCA mostra un andamento più tardivo ma prolungato. L'analisi qPCR ha dimostrato un'induzione significativa a partire dalle 16 ore, raggiungendo un picco tra le 24 e le 32 ore, permettendo di selezionare 4 punti temporali chiave (0h, 16h, 20h e 24h) destinati all'analisi di RNA-seq.

L'analisi di RNA-seq sui 12 campioni complessivi (3 replicati biologici per i 4 intervalli temporali) ha confermato l'elevata qualità del dataset generato (FastQC Phred Score con accuratezza > 99,9%, percentuale in GC > 50%, allineamento RNA STAR su hg38 > 70%). L'analisi differenziale condotta con DESeq2 su 58 geni chiave della via FA/BRCA su KEGG e Reactome ha rivelato un'attivazione progressiva. Di particolare rilevanza è l'eterogeneità trascrizionale osservata tra i singoli componenti dei complessi proteici FA (AG20, ID2, CEF, BL100): mentre alcuni fattori (FANCA, FANCG, FANCB, FANCI) aumentano progressivamente nel tempo, altri rimangono stabili o mostrano diminuzione come FANCF.

Ciò dimostra che la cellula applica una modulazione genica mirata e non un'induzione massiva e indifferenziata dell'intera via.

L'estensione dell'analisi all'intero genoma ha evidenziato una risposta coordinata nel tempo, culminante a 24 ore con 7.440 geni significativamente deregolati. Applicando una soglia di significatività stringente (p-adj < 0,005) sui geni codificanti proteine, l'analisi mediante diagramma di Venn ha identificato un totale di 5.540 geni, in comune tra gli intervalli di tempo o appartenenti al singolo *timepoint*. L'analisi di arricchimento funzionale condotta con DAVID su questo dataset ha identificato 10 cluster biologici chiave: accanto alla centralità della Regolazione dell'espressione genica (C1, FDR p = $2{,}48 \times 10^{-81}$) e del DDR (C3, FDR p = $3{,}24 \times 10^{-32}$), il gruppo DNA Unwinding (C2) ha mostrato una specificità funzionale (Fold Enrichment = 37,89 per i geni FA/BRCA). L'analisi di sovrapposizione funzionale tra i gruppi di geni (Indice di Jaccard) ha confermato che la risposta all'MMC costituisce un programma biologico in cui la riorganizzazione della cromatina (C2 e C7) e del metabolismo nucleico (C9) lavora in parallelo con il controllo del ciclo cellulare (C4) e della sopravvivenza/morte cellulare (C10).

L'analisi mirata sui trascritti non codificanti ha identificato 1.554 lncRNA differenzialmente espressi (p-adj < 0,05), di cui 476 stabili in tutti i *timepoint.* La comparazione con i geni codificanti per proteine (PCGs) ha rivelato un comportamento trascrizionale divergente: mentre nei PCGs prevale una risposta di sovraregolazione, nei lncRNA si osserva una risposta prevalentemente di sottoregolazione, che diventa dominante a 24h (823 sottoregolati vs 521 sovraregolati). La suddivisione dei trascritti in gruppi con diverse dinamiche temporali (dalla A alla H) ha ulteriormente rafforzato questa distinzione:

Nei geni del pathway FA, la risposta è fortemente omogenea e orientata all'attivazione sostenuta (oltre il 57% del totale e circa il 77% dei DEGs significativi appartiene ai gruppi ad aumento precoce o lineare, come i Gruppi A e B), garantendo un accumulo prolungato delle proteine della riparazione.

Nei lncRNA DEGs, escludendo la quota ad andamento variabile (Gruppo H, 3,0%), circa l'84% si distribuisce in modo quasi equivalente tra attivazione precoce e persistente (Gruppo B, 27,9%) e risposte ad attivazione o regolazione tardiva (Gruppo C, 28,1%; Gruppo E, 27,8%), affiancate da trascritti regolati in timepoint specifici (Gruppo F, 10,8%, Gruppo A 3,7%, Gruppo G 1,7%).

Tale differenza dimostra che, mentre i geni codificanti proteine assicurano la presenza continua dei macchinari proteici di riparazione, i lncRNA agiscono come regolatori temporali specifici che modulano i diversi passaggi della risposta cellulare al danno.

L'integrazione con i dati di ChIP-seq (ENCODE) e l'intersezione sulle regioni del promotore (±3000 bp dal TSS, annotazione GENCODE v46) mediante bedtools intersect ha dimostrato un coinvolgimento diretto di E2F1 ed E2F4. 1.890 geni codificanti proteine DEGs presentano

un picco di E2F1 e 1.272 un picco di E2F4 (1.014 esclusivi E2F1, 396 esclusivi E2F4 e 876 condivisi). L'analisi GO su DAVID ha confermato la convergenza di questi target sui processi chiave della FA: la Risposta al danno al DNA (DDR) è risultata la categoria più arricchita (FDR $p \approx 9{,}9 \times 10^{-52}$), insieme alla Riparazione del DNA / DSBs, alla Progressione G1/S e al Metabolismo degli acidi nucleici. Nello specifico del pathway FA/BRCA, 36 geni su 58 (62,1%) sono direttamente occupati da questi fattori, riflettendo il ruolo di E2F1 nell'attivazione della sintesi translesione (TLS), della risoluzione dei legami crociati interfilamento e del controllo proteico in fase S.

L'estensione dell'analisi ChIP-seq ai lncRNA DEGs ha identificato 280 trascritti con sito di legame per E2F1 e 178 per E2F4 (142 esclusivi E2F1, 40 esclusivi E2F4 e 138 condivisi con entrambi i fattori nei promotori). L'analisi delle annotazioni funzionali predittive condotta tramite lncHUB2 su 88 lncRNA caratterizzati ha rivelato che la via dell'Anemia di Fanconi rappresenta il principale bersaglio biologico regolato da questa rete di trascritti non codificanti (n = 31, pari al 35,2% dei lncRNA analizzati). La via FA registra l'MPC medio più alto dell'intero dataset con 0,028, e -Log10(p-value) = 2,30, configurandosi come il segnale biologico più consistente, superando processi a elevata frequenza assoluta come Ribosome biogenesis con n = 32, MPC = 0,011 e -Log10(p-value) = 2,10. La contemporanea presenza di correlazioni elevate con altri pathway coinvolti nella stabilità genomica – quali Homologous recombination (n = 28, MPC = 0,021), Mismatch repair (n = 22, MPC = 0,022), DNA replication (n = 29, MPC = 0,016)), NER (n = 21, MPC = 0,014) e BER (n = 21, MPC = 0,012) – evidenzia la presenza di un programma di regolazione trascrizionale coordinato dai lncRNA, volto alla riparazione dei danni strutturali al DNA.

In conclusione, questo studio dimostra che la risposta cellulare allo stress genotossico indotto da MMC, si articola secondo un programma trascrizionale nel tempo, nel quale l'attivazione dei geni codificanti proteine è accompagnata dall'azione dei lncRNA, caratterizzati da andamenti di espressione differenti. La dimostrazione che i fattori E2F1 ed E2F4 regolano direttamente sia i geni della via dell'Anemia di Fanconi sia i lncRNA coinvolti nella stabilità genomica pone anche le basi per future validazioni funzionali in vitro, volte anche a chiarire il ruolo molecolare specifico dei lunghi trascritti non codificanti nella patogenesi dell'Anemia di Fanconi.

Bibliografia

Auerbach AD. Fanconi anemia and its diagnosis. Mutat Res. 2009 Jul 31;668(1–2):4–10. doi: 10.1016/j.mrfmmm.2009.01.013. PMID: 19622403; PMCID: PMC2742943.

Bartek J, Lukas J. Mammalian G1- and S-phase checkpoints in response to DNA damage. Curr Opin Cell Biol. 2001 Dec;13(6):738–747. doi: 10.1016/S0955-0674(00)00280-5. PMID: 11698191

Bauer NC, Corbett AH, Doetsch PW. The current state of eukaryotic DNA base damage and repair. Nucleic Acids Res. 2015 Dec 2;43(21):10083–10101. doi: 10.1093/nar/gkv1136. PMID: 26519467 PMCID: PMC4666366

Bhattacharjee S, Nandi S. DNA damage response and cancer therapeutics through the lens of the Fanconi Anemia DNA repair pathway. Cell Commun Signal. 2017;15(1):41. doi: 10.1186/s12964-017-0195-9. PMID: 29017571. PMCID: PMC5635482.

Bindra RS, Glazer PM. Repression of RAD51 gene expression by E2F4/p130 complexes in hypoxia. Oncogene. 2007;26:2048–2057. doi:10.1038/sj.onc.1210001. PMID:17334390

Ceccaldi R, Parmar K, Mouly E, Delord M, Kim JM, Regairaz M, Pla M, Vasquez N, Zhang QS, Pondarre C, Peffault de Latour R, Gluckman E, Cavazzana-Calvo M, Leblanc T, Larghero J, Grompe M, Socié G, D'Andrea AD, Soulier J. Bone Marrow Failure in Fanconi Anemia Is Triggered by an Exacerbated p53/p21 DNA Damage Response that Impairs Hematopoietic Stem and Progenitor Cells. Cell Stem Cell. 2012;11(1):36–49. doi: 10.1016/j.stem.2012.05.013. PMID: 22683204. PMCID: PMC3392433.

Ceccaldi R, Sarangi P, D'Andrea AD. The Fanconi anaemia pathway: new players and new functions. Nat Rev Mol Cell Biol. 2016;17(6):337–349. doi: 10.1038/nrm.2016.48. PMID: 27145721.

Chapman JR, Taylor MRG, Boulton SJ. Playing the end game: DNA double-strand break repair pathway choice. Mol Cell. 2012;47(4):497–510. doi: 10.1016/j.molcel.2012.07.029. PMID: 22920291.

Che R, Zhang J, Nepal M, Han B, Fei P. Multifaceted Fanconi Anemia Signaling. Trends Genet. 2018 Mar;34(3):171–183. doi: 10.1016/j.tig.2017.11.006. PMID: 29254745 PMCID: PMC5858900

Chen J, Potlapalli R, Quan H, Chen L, Xie Y, Pouriyeh S, Sakib N, Liu L, Xie Y. Exploring DNA Damage and Repair Mechanisms: A Review with Computational Insights. BioTech. 2024;13(1):3. doi: 10.3390/biotech13010003. PMID: 38247733 PMCID: PMC10801582

Ciccia A, Ling C, Coulthard R, Yan Z, Xue Y, Meetei AR, Laghmani EH, Joenje H, McDonald N, de Winter JP, Wang W, West SC. Identification of FAAP24, a Fanconi anemia core complex protein that interacts with FANCM. Mol Cell. 2007;25:331–343. PMID:17289582 DOI: 10.1016/j.molcel.2007.01.003

D'Andrea AD, Grompe M. The Fanconi anaemia/BRCA pathway. Nat Rev Cancer. 2003 Jan;3(1):23–34. doi: 10.1038/nrc970. PMID: 12509764.

Deans AJ, West SC. DNA interstrand crosslink repair and cancer. Nat Rev Cancer. 2011;11(7):467–480. doi: 10.1038/nrc3088. PMID: 21701511 PMCID: PMC3560328

Denechaud PD, Fajas L, Giralt A. E2F1, a Novel Regulator of Metabolism. Front Endocrinol. 2017;8:311. doi:10.3389/fendo.2017.00311. PMID: 29176962 PMCID: PMC5686046

Dufour C, Pierri F. Modern management of Fanconi anemia. Hematology Am Soc Hematol Educ Program. 2022 Dec 9;2022(1):649–657. doi: 10.1182/hematology.2022000393. PMID: 36485157; PMCID: PMC9821189.

Fang C, Zhu Z, Cao J, Huang J, Xu Y. Comprehensive review on Fanconi anemia: insights into DNA interstrand cross-links, repair pathways, and associated tumors. Orphanet J Rare Dis. 2025;20:389. doi: 10.1186/s13023-025-03896-w. PMID: 40739565 PMCID: PMC12312369

Foo TK, Xia B. BRCA1-dependent and independent recruitment of PALB2-BRCA2-RAD51 in the DNA damage response and cancer. Cancer Res. 2022 Jul 12;can.22.1535. doi: 10.1158/0008-5472.CAN-22-1535. PMCID: PMC9481714 PMID: 35819255

Fouad S, Hauton D, D'Angiolella V. E2F1: Cause and Consequence of DNA Replication Stress. Front Mol Biosci. 2021;7:599332. doi:10.3389/fmolb.2020.599332. PMID:33665206.

Hoskins EE, Gunawardena RW, Habash KB, Wise-Draper TM, Jansen M, Knudsen ES, Wells SI. Coordinate regulation of Fanconi anemia gene expression occurs through the Rb/E2F pathway. Oncogene. 2008;27(35):4798–4808. doi:10.1038/onc.2008.121. PMID:18438432. PMCID:PMC2756820.

Ianzano L, D'Apolito M, Centra M, Savino M, Levran O, Auerbach AD, Cleton-Jansen AM, Doggett NA, Pronk JC, Tipping AJ, Gibson RA, Mathew CG, Whitmore SA, Apostolou S, Callen DF, Zelante L, Savoia A. The genomic organization of the Fanconi anemia group A (FAA) gene. Genomics. 1997;41(3):309–314. doi: 10.1006/geno.1997.4675.

Jaber S, Toufektchan E, Lejour V, Bardot B, Toledo F. p53 downregulates the Fanconi anaemia DNA repair pathway. Nat Commun. 2016 Apr 1;7:11091. doi: 10.1038/ncomms11091. PMCID: PMC4821997.

Kee, Y., & D'Andrea, A. D. (2012). Molecular pathogenesis and clinical management of Fanconi anemia. Journal of Clinical Investigation, 122(11), 3799–3806. PMID: 23114602 PMCID: PMC3484428 DOI: 10.1172/JCI58321

Lemonidis, K., Rennie, M. L., Arkinson, C., Chaugule, V. K., Clarke, M., Streetley, J., & Walden, H. (2023). Structural and biochemical basis of interdependent FANCI–FANCD2 ubiquitination. The EMBO Journal, 42(3), e111898. PMID: 36385258 PMCID: PMC9890228 DOI: 10.15252/embj.2022111898

Li J, Ming Z, Yang L, Wang T, Liu G, Ma Q. Long noncoding RNA XIST: Mechanisms for X chromosome inactivation, roles in sex-biased diseases, and therapeutic opportunities. Genes Dis. 2022 Apr 29;9(6):1478-1492. doi: 10.1016/j.gendis.2022.04.007. PMID: 36157489; PMCID: PMC9485286.

Lu J-P, Zhang J, Kim K, Case TC, Matusik RJ, Chen Y-H, Wolfe M, Nopparat J, Lu Q. Human homolog of Drosophila Hairy and enhancer of split 1, Hes1, negatively regulates δ-catenin (CTNND2) expression in cooperation with E2F1 in prostate cancer. Mol Cancer. 2010;9:304. doi:10.1186/1476-4598-9-304 PMID: 21106062 PMCID: PMC3009707

Malouf C, Loughran SJ, Wilkinson AC, Shimamura A, Río P. Translational research for bone marrow failure patients. Exp Hematol. 2022 Jan;105:18–21. doi: 10.1016/j.exphem.2021.11.004. PMID: 34801643; PMCID: PMC9165438

Mehta PA, Tolar J. Fanconi anemia. GeneReviews®. 2002 Feb 14 [aggiornato 2018 Mar 8]. Seattle (WA): University of Washington, Seattle. PMID: 20301575. Bookshelf ID: NBK1401

Molina-Estevez FJ, Nowrouzi A, Lozano ML, Galy A, Charrier S, von Kalle C, Guenechea G, Bueren JA, Schmidt M. Lentiviral-Mediated Gene Therapy in Fanconi Anemia-A Mice Reveals Long-Term Engraftment and Continuous Turnover of Corrected HSCs. Curr Gene Ther. 2015;15(6):550–562. doi: 10.2174/1566523215666150929110903. PMID: 26415575

Nalepa, G., & Clapp, D. W. (2018). Fanconi anaemia and cancer: an intricate relationship. Nature Reviews Cancer, 18(3), 168–185. PMID: 29376519 DOI: 10.1038/nrc.2017.116

Orhan E, Velazquez C, Tabet I, Sardet C, Theillet C. Regulation of RAD51 at the Transcriptional and Functional Levels: What Prospects for Cancer Therapy? Cancers (Basel). 2021;13(12):2930. doi:10.3390/cancers13122930.

Peake JD, Noguchi E. Fanconi anemia: current insights regarding epidemiology, cancer, and DNA repair. Hum Genet. 2022;141(7):1083–1102. doi: 10.1007/s00439-022-02462-9. PMID: 35596788

Prakash R, Zhang Y, Feng W, Jasin M. Homologous recombination and human health: The roles of BRCA1, BRCA2, and associated proteins. Cold Spring Harb Perspect Biol.

2015;7(4):a016600. doi: 10.1101/cshperspect.a016600. PMID: 25833843. PMCID: PMC4382744.

Río P, Zubicaray J, Navarro S, Gálvez E, Sánchez-Domínguez R, Nicoletti E, et al. Haematopoietic gene therapy of non-conditioned patients with Fanconi anaemia-A: results from open-label phase 1/2 (FANCOLEN-1) and long-term clinical trials. Lancet. 2024;404(10471):2584–2592. doi: 10.1016/S0140-6736(24)01880-4. PMID: 39642902.

Saberwal G, Horvath E, Hu L, Zhu C, Hjort E, Eklund EA. The Interferon Consensus Sequence Binding Protein (ICSBP/IRF8) Activates Transcription of the FANCF Gene during Myeloid Differentiation. J Biol Chem. 2009;284(48):33242–33254. doi: 10.1074/jbc.M109.010231. PMID: 19801548 PMCID: PMC2785167

Savoia A, Centra M, Ianzano L, de Cillis GP, Zelante L, Buchwald M. Characterization of the 5' region of the Fanconi anaemia group C (FACC) gene. Hum Mol Genet. 1995 Aug;4(8):1321–1326. doi: 10.1093/hmg/4.8.1321. PMID: 7581369.

Schlacher, K., Wu, H., & Jasin, M. (2012). A distinct replication fork protection pathway connects Fanconi anemia tumor suppressors to RAD51–BRCA1/2. Cancer Cell, 22(1), 106–116. PMID: 22789542 PMCID: PMC3954744 DOI: 10.1016/j.ccr.2012.05.015

Shukla P, Solanki A, Ghosh K, Vundinti BR. DNA interstrand cross-link repair: understanding role of Fanconi anemia pathway and therapeutic implications. Eur J Haematol. 2013;91(5):381–393. doi: 10.1111/ejh.12169. PMID: 23859405

Spardy N, Duensing A, Charles D, Haines N, Nakahara T, Lambert PF, Duensing S. The human papillomavirus type 16 E7 oncoprotein activates the Fanconi anemia (FA) pathway and causes accelerated chromosomal instability in FA cells. J Virol. 2007;81:13265–13270. doi:10.1128/JVI.01121-07. PMID:17898070 PMCID: PMC2169120

Statello, L., Guo, C.-J., Chen, L.-L., & Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. Nature Reviews Molecular Cell Biology, 22(2), 96–118. DOI: 10.1038/s41580-020-00315-9 PMID: 33353982 PMCID: PMC7754182

Su, M., Wang, H., Wang, W., Wang, Y., Ouyang, L., Pan, C., Xia, L., Cao, D., & Liao, Q. (2018). LncRNAs in DNA damage response and repair in cancer cells. Acta Biochimica et Biophysica Sinica, 50(5), 433–439. DOI: 10.1093/abbs/gmy022 PMID: 29579111

Taniguchi T, D'Andrea AD. Molecular pathogenesis of Fanconi anemia: recent progress. Nat Rev Mol Cell Biol. 2006;7(9):701–713. doi: 10.1038/nrm1989. PMID: 16493006

Tategu M, Arauchi T, Tanaka R, Nakagawa H, Yoshida K. Systems Biology-Based Identification of Crosstalk between E2F Transcription Factors and the Fanconi Anemia Pathway. Gene Regul Syst Bio. 2007;1:1–8. PMID: 19936073; PMCID: PMC2759144.

Tsui V, Crismani W. The Fanconi anemia pathway and fertility. Trends Genet. 2019;35(3):199–214. doi: 10.1016/j.tig.2018.12.003. PMID: 30683429

Uszczynska-Ratajczak, B., Lagarde, J., Frankish, A., Guigó, R., & Johnson, R. (2018). Towards a complete map of the human long non-coding RNA transcriptome. Nature Reviews Genetics, 19(9), 535–548. DOI: 10.1038/s41576-018-0017-y PMID: 29795125 PMCID: PMC6451964

Wan, G., Mathur, R., Hu, X., Liu, Y., Zhang, X., Peng, G., & Lu, X. (2013). Long non-coding RNA ANRIL (CDKN2B-AS) is induced by the ATM-E2F1 signaling pathway. Cellular Signalling, 25(5), 1086–1095. DOI: 10.1016/j.cellsig.2013.02.006 PMID: 23416462 PMCID: PMC3675781

Wiedemann H-R. Guido Fanconi (1892–1979) in memoriam. Eur J Pediatr. 1979 Nov;132(3):131–132. doi: 10.1007/BF00442427. PMID: 389632.

Wu, Y., Yang, L., Zhao, J., Li, C., Nie, J., Liu, F., Zhuo, C., et al. (2015). Nuclear-enriched abundant transcript 1 as a diagnostic and prognostic biomarker in colorectal cancer. Molecular Cancer, 14, 191. DOI: 10.1186/s12943-015-0461-4 PMID: 26627110 PMCID: PMC4681121

TOOLS

BEDTools: Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics, 26(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

ChEA: Lachmann, A., Xu, H., Krishnan, J., Berger, S. I., Mazloom, A. R., & Ma'ayan, A. (2010). ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. Bioinformatics, 26(19), 2438–2444. https://doi.org/10.1093/bioinformatics/btq466

DAVID: Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols, 4(1), 44–57. https://doi.org/10.1038/nprot.2008.211

DESeq2: Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15(12), 550. https://doi.org/10.1186/s13059-014-0550-8

ENCODE: Davis, C. A., Hitz, B. C., Sloan, C. A., Chan, E. T., Davidson, J. M., Gabdank, I., Hilton, J. A., Jain, K., Baymuradov, U. K., Narayanan, A. K., Onate, K. C., Graham, K., Miyasato, S. R., Dunham, I., Kundaje, A., Cherry, J. M. (2018). The Encyclopedia of DNA Elements (ENCODE): data portal update. Nucleic Acids Research, 46(D1), D794–D801. https://doi.org/10.1093/nar/gkx1081

Ensembl / GENCODE: Frankish, A., Diekhans, M., Jungreis, I., Lagarde, J., Loveland, J. E., et al. (2023). GENCODE 2023. Nucleic Acids Research, 51(D1), D1014–D1022. https://doi.org/10.1093/nar/gkac1071

FastQC: Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

FeatureCounts: Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics, 30(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Ggplot2: Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. https://doi.org/10.1007/978-3-319-24277-4

KEGG: Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., & Ishiguro-Watanabe, M. (2023). KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Research, 51(D1), D587–D592. https://doi.org/10.1093/nar/gkac963

lncHUB2: Marino, G. B., Wojciechowicz, M. L., Clarke, D. J. B., Kuleshov, M. V., Xie, Z., Jeon, M., Lachmann, A., & Ma'ayan, A. (2023). lncHUB2: aggregated and inferred knowledge about human and mouse lncRNAs. Database (Oxford), 2023, baad009. https://doi.org/10.1093/database/baad009 PMID: 36869839 PMCID: PMC9985331

Pheatmap: Kolde, R. (2019). pheatmap: Pretty Heatmaps. R package version 1.0.12. https://CRAN.R-project.org/package=pheatmap

Reactome: Gillespie, M., Jassal, B., Stephan, R., Milacic, M., Rothfels, K., et al. (2022). The Reactome pathway knowledgebase 2022. Nucleic Acids Research, 50(D1), D687–D692. https://doi.org/10.1093/nar/gkab1028

STAR: Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

UCSC Genome Browser: Navarro Gonzalez, J., Zweig, A. S., Speir, M. L., Schmelter, D., Rosenbloom, K. R., et al. (2021). The UCSC Genome Browser database: 2021 update. Nucleic Acids Research, 49(D1), D1046–D1057. https://doi.org/10.1093/nar/gkaa1070

Venny: Oliveros, J. C. (2007–2015). Venny. An interactive tool for comparing lists with Venn's diagrams. https://bioinfogp.cnb.csic.es/tools/venny/

FONTI DELLE FIGURE E DELLE TABELLE

Figura 1: Jazayeri A, Falck J, Lukas C, Bartek J, Smith GCM, Lukas J, Jackson SP. ATM- and cell cycle–dependent regulation of ATR in response to DNA double-strand breaks. Nat Cell Biol. 2005 Jul; 7(7):643–650. doi: 10.1038/ncb1270.

Figura 2: Bouberhan S, Philp L, Hill S, Al-Alem LF, Rueda B. Exploiting the prevalence of homologous recombination deficiencies in high-grade serous ovarian cancer. Cancers (Basel). 2020 May 11; 12(5):1206. doi: 10.3390/cancers12051206. PMID: 32403357; PMCID: PMC7281458.

Tabella 1: Sroka I, Frohnmayer L, Van Ravenhorst S, Wirkkula L (eds). Fanconi Anemia Clinical Care Guidelines. Fifth Edition. Fanconi Cancer Foundation; 2020.

Figura 3: Fang CB, Wu HT, Zhang ML, Liu J, Zhang GJ. Fanconi Anemia Pathway: Mechanisms of Breast Cancer Predisposition Development and Potential Therapeutic Targets. Front Cell Dev Biol. 2020 Apr 2; 8:160. doi: 10.3389/fcell.2020.00160. PMID: 32300589; PMCID: PMC7142266.

Figura 4: Ceccaldi R, Sarangi P, D'Andrea AD. The Fanconi anaemia pathway: new players and new functions. Nat Rev Mol Cell Biol. 2016;17(6):337–349. doi: 10.1038/nrm.2016.48. PMID: 27145721.

Figura 5: Schvartzman JB, Hernández P, Krimer DB, Dorier J, Stasiak A. Closing the DNA replication cycle: from simple circular molecules to supercoiled and knotted DNA catenanes. Nucleic Acids Res. 2019 Aug 22; 47(14):7182–7198. doi: 10.1093/nar/gkz586. PMID: 31276584; PMCID: PMC6698734.