DEALING WITH THE MAIN CHALLENGES OF FANCONI ANEMIA MOLECULAR DIAGNOSIS AND THERAPY

ILARIA PERSICO

UNIVERSITÀ DEGLI STUDI DI TRIESTE

UAB Universitat Autònoma de Barcelona

UNIVERSITÀ DEGLI STUDI DI TRIESTE
-
UNIVERSITAT AUTÒNOMA DE BARCELONA

XXXV CYCLE OF THE RESEARCH DOCTORATE IN
SCIENZE DELLA RIPRODUZIONE E DELLO SVILUPPO
-
DOCTORADO EN GENÉTICA

DEPARTMENT OF MEDICAL SCIENCES
-
FACULTY OF BIOSCIENCES
DEPARTMENT OF GENETICS AND MICROBIOLOGY
GENOME INSTABILITY AND DNA REPAIR GROUP

DEALING WITH THE MAIN CHALLENGES OF FANCONI ANEMIA MOLECULAR DIAGNOSIS AND THERAPY

DOCTORAL DISSERTATION

ILARIA PERSICO

MARCH 2023

UNIVERSITÀ DEGLI STUDI DI TRIESTE

UAB Universitat Autònoma de Barcelona

UNIVERSITÀ DEGLI STUDI DI TRIESTE
-
UNIVERSITAT AUTÒNOMA DE BARCELONA

DEPARTMENT OF MEDICAL SCIENCES

-

FACULTY OF BIOSCIENCES

DEPARTMENT OF GENETICS AND MICROBIOLOGY

GENOME INSTABILITY AND DNA REPAIR GROUP

DEALING WITH THE MAIN CHALLENGES OF FANCONI ANEMIA MOLECULAR DIAGNOSIS AND THERAPY

DISSERTATION RESPECTFULLY SUBMITTED BY

ILARIA PERSICO

MARCH 2023

Within the framework of the Joint Supervision of Doctoral Thesis in the Doctoral Studies programs in SCIENZE DELLA RIPRODUZIONE E DELLO SVILUPPO and GENÉTICA between the Università degli Studi di Trieste and Universitat Autònoma de Barcelona, respectively, for the achievement of the Doctorate Degree with the Mention of International Doctorate.

Thesis Director and Tutor (Internal)	Thesis Director (Internal)	Thesis Director (External)	Author
Dr. Jordi Surrallés Calonge	Dr. Massimo Bogliolo	Dr. Adamo Pio D'Adamo	Ilaria Persico

Numquam ponenda est pluralitas sine necessitate

William of Ockham

ACKNOWLEDGMENTS

I would like to start by thanking my supervisors Prof. Adamo Pio D'Adamo, Prof. Jordi Surrallés, and Prof. Massimo Bogliolo, and Prof. Anna Savoia for their support, trust, and patience (especially patience) during this three-year-long journey called PhD.
I am also really thankful to Prof. Manuel Kaulich for having hosted me in his laboratory and, together with his amazing team, introduced me in the complicated, but extremely fascinating world of "CRISPR-Cas9 genome editing".
Thanks of course to all my laboratory colleagues for the joys and sorrows (and spritzes) shared together, and a special thanks to Dr. Roberta Bottega for having been my inspiring mentor since when I was just a little "Labrador's puppy".

I would like to thank the European Social Fund (ESF) - Operational Programme 2014/2020 of Friuli Venezia Giulia Region to have foster my education and research by granting me a PhD scholarship.

Thanks to my beloved, strong family.

Eventually, I would like to thank who embarked on this adventure with me three years ago (or much longer ago, indeed), who wanted to stay and continue to accompany me, who chose to join me, and also who decided to leave or had not option, but leave. And, of course, thanks to whom showed me that windows could also have a "vasistas" opening.

ABBREVIATIONS

3Cs	Covalently-closed-circular synthesized
AF	Allele frequency
AIEOP	Italian Association of Pediatric Hematology and Oncology
ALDH2	Aldehyde dehydrogenase 2
ALDH9A1	Aldehyde dehydrogenase 9 family member A1
AML	Acute myeloid leukemia
ATM	Ataxia telangiectasia mutated
BLM	Bloom's syndrome
BM	Bone marrow
BMF	BMF Bone marrow failure
BSA	Bovine serum albumin
Cas9	Clustered regularly interspaced short palindromic repeats- associated protein 9
CB	Chromosomal breakage
cDNA	Complementary DNA
CFCs	Colony forming cells
CNV/s	Copy number variant/s
CO2	Carbonic anhydride
CRISPR	Clustered regularly interspaced short palindromic repeats

CRISPRa	CRISPR activation
CRISPRi	CRISPR interference
CRISPRko	CRISPR knockout
CtIP	Choline transporter-like protein
DAPI	4’-6’-diamidino-2-phenylindole
dbSNP	Single Nucleotide Polymorphism Database
DEB	Diepoxybutane
dH2O	Distilled hydrogen oxide or water
DIV/s	Deep intronic variant/s
DMEM	Dulbecco's Modified Eagle medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSB/s	Double-strand break/s
dsDNA	Double stranded DNA
EGFR	Epidermal growth factor receptor
EMA	European Medicines Agency
EtOH	Ethanol
FA	Fanconi anemia
FA-A	Fanconi anemia subtype A
FA/BRCA	Fanconi anemia/Breast cancer

FANCD2-Ub	FANCD2 mono-ubiquitination
FBS	Fetal bovine serum
FC	Fold change
FDA	Food and Drug Administration
FDR	False discovery rates
G2/M	G2/M phase cell cycle arrest
gDNA	Genomic DNA
GFP	Green fluorescent protein
gnomAD	Genome Aggregation Database
gRNA/s	Guide ribonucleic acid/s
GT	Gene therapy
GVHD	Graft-versus-host disease
GW	Genome wide
H2O	Hydrogen oxyde or water
HC	High content
HCS	High content screening
HGMD	Human Gene Mutation Database
HNSCCs	Head and neck squamous cell carcinomas
HR	Homologous recombination
HRR	Homologous recombination repair

HSC/s	Hematopoietic stem cell/s
HSCT	Hematopoietic stem cell transplantation
HSPC/s	Hematopoietic stem and progenitor cell/s
HTS	High throughput screening/s
HU	Hydroxyurea
ICL/s	Interstrand cross-link/s
ID	Interstitial deletion
Indels	Insertions and deletions
Kb	Kilobase
KO/s	Knockout/s
LAMTOR1/5	MAPK and MTOR activator 1/5
LB	Lysogeny broth
LCL/s	Lymphoblastoid cell line/s
LFC	Logarithmic fold change
LOF	Loss-of-function
MAF	Minor allele frequency
Maxi-preps	Maxi-preparations
MDS	Myelodysplastic syndrome
MF	Multiradial figure
Mini-preps	Mini-preparations

MLPA	Multiplex ligation-dependent probe amplification
MMC	Mitomycin C
MOI	Multiplicity of infection
mRNA	Messenger ribonucleic acid
NAD(P)H	Nicotinamide adenine dinucleotide phosphate
NESs	Nuclear export signal/s
NGS	Next generation sequencing
NHEJ	Non-homologous end joining
NLS	Nuclear localization signal
NNSplice	Splice Site Prediction by Neural Network
NQO1	NAD(P)H quinone dehydrogenase
NSLC	Non-small cell lung cancer
NT	Not-truncating
O/N	Overnight
OMIM	Online Mendelian Inheritance in Man
PAMs	Protospacer adjacent motifs
PARP	Poly (ADP-ribose) polymerase
PARPi	Poly (ADP-ribose) polymerase inhibitors
PB	Peripheral blood
PBS	Phosphate-buffered saline

PCR	Polymerase chain reaction
PD-1	Programmed cell death protein 1
PFA	Paraformaldehyde
PFC	Phenotypic fold change
PFCL	Primary fibroblast cell line
PFs	Primary fibroblasts
PHENOS	Skin pigmentation abnormalities, small head, small eyes, structural central nervous system abnormalities, otologic abnormalities and short stature
PI	Propidium iodide
Pol	Polymerase
Pol ı	Polymerase ı
qPCR	Quantitative or real time PCR
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RNAi	RNA interfering
ROS	Reactive oxygen species
RPMI	Roswell Park Memorial Institute
RT	Room temperature
RT-PCR	Reverse transcription PCR
SCCs	Squamous cell carcinomas

shRNA	Short hairpin RNA
SI/s	Synthetic interaction/s
SL	Synthetic lethality
SNP/s	Single nucleotide polymorphism/s
SOC medium	Catabolite repression medium
SRB	Sulforhodamine B
SS	Splice site
SSB/s	Single-strand break/s
SV	Synthetic viability
t-NGS	Targeted NGS
TALEN	Transcription activator-like effector nuclease
TCA	Trichloroacetic acid
TKOv3	Toronto KnockOut v3
USA	United States of America
USP48	Ubiquitin specific peptidase 48
VACTERL-H	Vertebral abnormalities, anal atresia, cardiac abnormalities, tracheo-esophageal fistula, esophageal or duodenal atresia, renal abnormalities, upper limb abnormalities and hydrocephalus
VUS/s	Variant/s of unknown significance
WES	Whole exome sequencing
WGS	Whole genome sequencing

WT	Wild type
YF	Yellow fluorescence
YPF	Yellow fluorescent protein

TABLE OF CONTENTS

ABSTRACT

Fanconi anemia (FA) is a rare genetic DNA repair deficiency with vast genetic heterogeneity, multiple private mutations, and high mosaicism rate, still jeopardizing molecular diagnosis in case of variants with difficult characterization or unclear pathogenicity.
FA also shows bone marrow failure and enhanced susceptibility to hematologic and solid malignancies. Since current treatments only address hematopoietic defects, it is essential to find new systemic drug-based therapies for FA patients.

We sought to cope with these hurdles through:

i. Design of an optimized mutation screening strategy based on targeted next generation sequencing and complementary techniques for a faster and exhaustive FA molecular diagnosis.
We successfully characterized 14/14 patients from three different complementation groups, and reported a total of 23 genetically distinct alleles, including a founder one. We also developed a statistical analysis to infer copy-number variants (CNVs) from the NGS data, and, for all the 5 gross deletions predicted, we provided detection rates matching with those of CNVs gold standards. Moreover, we described the case of a patient with a large deletion and a known splicing variant compensated by a *de novo* insertion, elucidating a presumptive natural gene therapy phenomenon *in vivo*.

ii. Validation of cellular systems expressing three stable, but defective FANCA proteins (Arg951Gln, Thr1131Ala, Phe1263del) suitable for high content screening (HCS) to identify drug/s able to rescue mutants' phenotypes.
To generate the models, we transduced lentiviral vectors carrying the mutations into a *FANCA*-deficient line that expresses a fluorescent FANCD2 protein. These cells are unable to mono-ubiquitinate FANCD2 or promote its relocation to DNA repair foci, both hallmarks of FA pathway activity. Despite confirming mutant FANCA expression in all models, only the Phe1263del one exhibited a cellular phenotype compatible with the *FANCA*-null line; the other wild-type behaving mutants were instead discarded as HCS candidates. We will next monitor Phe1263del correction via fluorescent FANCD2 foci formation with the aim of finding a drug, possibly already approved by the Food and Drug Administration (FDA), for a personalized mutation-dependent therapy towards all FA clinical features.

iii. Performance of genome-wide (GW) clustered regularly interspaced short palindromic repeats (CRISPR) knockout (KO) screens to dissect novel synthetic interactions (SIs) with FA deficiency.
We transduced CRISPR-associated protein 9 (Cas9)-expressing *FANCA*-deficient and control lines with two GW KO libraries of guide RNAs (gRNAs), i.e., an in-house and the commercial Toronto KO version 3 (TKOv3) library. Over the screens we did not apply any biological challenge in order to analyze the end phenotypes deriving from the

proliferative competition between the cells subjected to gene KO. To keep gRNAs/phenotypes representation, screens were ended at ~10 cell doublings, and genomic DNA from all replicates was isolated and amplified with barcode-tagged primers for NGS. Bioinformatic and statical analyses have already been run for the in-house library screen, providing a preliminary roster of depleted (lethal SIs) and enriched (viable SIs) hits. Subsequently, we will contrast these results with those of the TKOv3 screen sequencing and validate the best hits to prove the SIs detected and, thus, provide novel druggable targets for FA treatment (viable SIs) and/or new indications for gene alterations sensitizing cancer cells to the inhibition of FA pathway or its interactors (lethal SIs).

The achievement of the goals of this thesis project will provide a comprehensive strategy for the current challenges of FA, proceeding from the bench to the bedside and back.

Sensitive data (i.e., denomination of the candidate genes identified by CRISPR screening) have been partially omitted from the thesis content for publication and/or patent purposes.

RIASSUNTO

L'anemia di Fanconi (*Fanconi anemia*, FA) è una sindrome genetica rara causata da un difetto nella riparazione del danno al DNA con ampia eterogeneità genetica, numerose mutazioni private ed elevato tasso di mosaicismo. Tali fattori compromettono tuttora la formulazione di una diagnosi molecolare in caso di varianti di difficile caratterizzazione o significato patogenetico incerto. La FA è caratterizzata da aplasia midollare e un'aumentata suscettibilità allo sviluppo di tumori ematologici e solidi. Dato che i trattamenti attualmente disponibili si rivolgono ai soli difetti ematopoietici, è essenziale identificare nuove terapie farmacologiche sistemiche per il trattamento dei pazienti FA.

Abbiamo cercato di far fronte a tali criticità mediante:

i. Progettazione di una strategia di screening mutazionale ottimizzata basata sul sequenziamento mirato di nuova generazione (*targeted next generation sequencing*, t-NGS) e su tecniche complementari per una diagnosi molecolare di FA più rapida ed esaustiva.
   Abbiamo caratterizzato con successo 14/14 pazienti appartenenti a tre gruppi di complementazione differenti e riportato un totale di 23 alleli geneticamente distinti, tra cui uno fondatore. Abbiamo anche sviluppato un'analisi statistica per inferire le varianti del numero di copie (*copy-number variants*, CNVs) dai dati di NGS e, per tutte le 5 ampie delezioni predette, abbiamo fornito tassi di rilevazione sovrapponibili a quelli delle tecniche di riferimento per le CNVs. Ulteriormente, abbiamo descritto il caso di un paziente con un'ampia delezione e una mutazione di *splicing* nota compensata da un'inserzione *de novo*, rivelando un probabile fenomeno di terapia genica naturale *in vivo*.

ii. Validazione di sistemi cellulari basati sull'espressione di tre proteine FANCA mutanti, ma stabili (Arg951Gln, Thr1131Ala, Phe1263del), per uno *screening* ad alto contenuto di farmaci (*high content screening*, HCS) al fine di trovare molecole capaci di correggerne il fenotipo aberrante.
   Per generare i modelli, abbiamo trasdotto vettori lentivirali con le mutazioni in una linea deficitaria di *FANCA* e che esprime una proteina FANCD2 fluorescente. Tali cellule non sono in grado di mono-ubiquitinare FANCD2, né di promuoverne la traslocazione a foci di riparazione del DNA, entrambi elementi caratteristici dell'attività della via FA di riparazione del DNA. Nonostante la conferma dell'espressione della proteina FANCA mutante in tutti i modelli, solo quello del mutante Phe1263del ha mostrato un fenotipo cellulare compatibile con la linea deficitaria di *FANCA*; gli altri che mostravano un fenotipo normale sono stati invece esclusi come candidati per il HCS. Successivamente, monitoreremo la correzione del mutante Phe1263del tramite la formazione di foci fluorescenti di FANCD2 con l'obiettivo di individuare una molecola, possibilmente già approvata dalla *Food and Drug Administration* (FDA), per una terapia mutazione-dipendente personalizzata verso ogni caratteristica clinica della FA.

iii. Realizzazione di *clustered regularly interspaced short palindromic repeats (CRISPR) knockout* (*KO*) *screening* su scala genomica per identificare nuove interazioni sintetiche (*synthetic interactions*, SIs) con i geni FA.
Abbiamo trasdotto una linea cellulare deficitaria di *FANCA* e una di controllo, che esprimevano entrambe *CRISPR-associated protein 9* (Cas9), con due librerie *KO* di RNA guida (*guide RNA*, gRNAs), una generata in laboratorio ("libreria interna") e la commerciale *Toronto KO version 3* (TKOv3). Durante gli *screen* non abbiamo applicato alcuno stimolo biologico esterno poiché interessati ad analizzare i fenotipi derivanti dalla competizione proliferativa tra le cellule soggette a soppressione genica. Per mantenere la rappresentatività gRNA/fenotipi, abbiamo concluso gli *screen* attorno alle 10 divisioni cellulari, estratto il DNA genomico e lo abbiamo amplificato con *primers* contenti sequenze "codici a barre" per NGS. Abbiamo già condotto l'analisi bioinformatica e statistica dello *screen* con la "libreria interna", stilando una lista preliminare di geni deleti (SIs letali) e arricchiti (SIs vitali). In seguito, compareremo tali risultati con quelli di sequenziamento dello *screen* con la libreria TKOv3 e valideremo i migliori candidati per comprovare le SIs identificate e, pertanto, individuare nuovi bersagli farmacologici per il trattamento della FA (SIs vitali) e/o indicazioni per alterazioni genetiche in grado di sensibilizzare le cellule tumorali all'inibizione della via FA o dei suoi interattori (SIs letali).

Il conseguimento degli obiettivi di questo progetto di tesi fornirà una strategia comprensiva per le attuali sfide poste dalla FA, procedendo dal laboratorio al letto del paziente e viceversa.

Dati sensibili (i.e., denominazione dei geni candidati individuati mediante *CRISPR screening*) sono stati parzialmente omessi dal contenuto della tesi a scopo di pubblicazione e/o brevetto.

RESUMEN

La anemia de Fanconi (*Fanconi anemia*, FA) es un trastorno genético raro de la reparación del ADN con amplia heterogeneidad genética, múltiples mutaciones privadas y alta tasa de mosaicismo. Todo esto hace que el diagnóstico molecular se vea perjudicado, sobre todo en el caso de variantes de difícil caracterización o de patogenicidad dudosa.
Los pacientes con FA también presentan insuficiencia de la medula ósea y mayor susceptibilidad a tumores sólidos y neoplasias hematológicas. Dado que los tratamientos actuales solo abordan los defectos hematopoyéticos, es indispensable encontrar nuevas terapias farmacológicas sistémicas para los pacientes FA.

Intentamos hacer frente a estos obstáculos a través de:

i. Diseño de una estrategia optimizada de cribado mutacional basada en secuenciación dirigida de nueva generación (*targeted next generation sequencing*, t-NGS) y en técnicas complementarias para un diagnóstico molecular de FA más rápido y exhaustivo.
   Logramos caracterizar 14/14 sujetos pertenecientes a tres grupos de complementación diferentes, y reportamos un total de 23 alelos genéticamente distintos, incluido uno fundador. También desarrollamos un análisis estadístico para inferir variantes de numero de copia (*copy-number variants*, CNVs) a partir de los datos de NGS y, para las 5 deleciones pronosticadas, proporcionamos tasas de detección coincidentes con las de las técnicas de referencia para el estudio de CNVs. Además, describimos el caso de un paciente con una gran deleción y una variante de *splicing* conocida compensada por una inserción *de novo*, resultando en un potencial efecto de terapia génica natural *in vivo*.

ii. Validación de sistemas celulares de tres proteínas FANCA mutadas pero estables (Arg951Gln, Thr1131Ala, Phe1263del), aptos para un cribado de alto contenido de fármacos (*high content screening*, HCS), con el fin de identificar molécula/s capaces de rescatar sus fenotipos.
   Para generar los modelos, transducimos vectores lentivirales con las mutaciones en una línea deficiente para *FANCA* que expresa una proteína FANCD2 fluorescente. Estas células no son capaces de mono-ubiquitinar FANCD2 o promover su reubicación en focos de reparación del ADN, características propias de la actividad de la ruta FA. A pesar de confirmar la expresión de la proteína FANCA mutante en todos los modelos, solo el del mutante Phe1263del mostró un fenotipo celular compatible con la línea deficiente para *FANCA*; por el contrario, los otros dos modelos con comportamiento de tipo salvaje se descartaron como candidatos para el HCS. Posteriormente, monitorearemos la corrección de Phe1263del a través de la formación de focos FANCD2 fluorescentes con el objetivo de encontrar un fármaco, posiblemente ya aprobado por *la Food and Drug Administration* (FDA), para una terapia mutación-dependiente personalizada para todos los rasgos clínicos de FA.

iii. Realización de cribados *clustered regularly interspaced short palindromic repeats (CRISPR) knockout* (*KO*) en todo el genoma para diseccionar nuevas interacciones sintéticas (*synthetic interactions*, SIs) con los genes FA.
Transducimos una línea FANCA deficiente y su control, ambos caracterizados por la expresión de *CRISPR-associated protein 9* (Cas9), con dos librerías *KO* de ARN guía (*guide RNA*, gRNA), una librería generada en laboratorio ("librería interna") y una comercial denominada *Toronto KO version 3* (TKOv3). Durante los cribados no introdujimos ningún estimulo externo, con el fin de poder analizar los fenotipos resultantes de la competencia proliferativa entre las células sometidas a supresión genética. Para mantener la representación de gRNAs/fenotipos los cribados fueron detenidos después de 10 duplicaciones celulares, tras lo cual aislamos el ADN genómico de todas las réplicas y lo amplificamos con cebadores con código de barras para NGS. A continuación, se llevó a cabo el análisis bioinformático y estadístico del cribado de la "librería interna", redactando una lista preliminar de genes con baja representación (SI letales) y enriquecidos (SI viables). A continuación, compararemos estos resultados con los de secuenciación del cribado con la biblioteca TKOv3 y validaremos los mejores candidatos para probar las SIs identificadas y, por lo tanto, proporcionar nuevas dianas farmacológicas para la FA (SI viables) y/o indicaciones para alteraciones genéticas capaces de sensibilizar las células tumorales a la inhibición de la ruta FA o de sus interactores (SI letales).

El logro de los objetivos de este proyecto de tesis proporcionará una estrategia integral para los actuales desafíos de la FA, desde el laboratorio hasta la cama del paciente y vuelta.

Datos confidenciales (i.e., la denominación de los genes candidatos identificados mediante el cribado con tecnología *CRISPR*) han sido parcialmente omitidos del contenido de la tesis para fines de publicación y/o patente.

1. INTRODUCTION

Fanconi anemia (FA) is a rare genome instability syndrome due to mono-, biallelic or x-linked variants in 23 DNA repair genes [1]. First defined in 1927 by the Swiss pediatrician Guido Fanconi [2], the condition affects 1/160,000 newborns and displays an estimated carrier frequency of 1/200. FA has been described in all population groups, even though it mainly occurs in specific ethnic backgrounds due to founder effect [3].

1.1 CLINICAL FEATURES

FA exhibits a wide phenotypic heterogeneity, a trait which often undermines the reliability of diagnosis based on the sole clinical signs. Among the diverse features, the hematological abnormalities are regarded as FA major life-threatening conditions within the first two decades of life. Consistently, previous to the recent implementation of optimized therapies, FA was classified as a typically pediatric disorder [4]. Most patients manifest progressive pancytopenia evolving to bone marrow failure (BMF) in 80% cases before age 10, and develop hematological malignancies, as acute myeloid leukemia (AML) and myelodysplastic syndrome (MDS), with a cumulative incidence of up to 25% during their life [5], [6], [4]. In the early adulthood, FA individuals become more prone to solid tumors, mainly to head and neck squamous cell carcinomas (HNSCCs), reaching a cumulative incidence of 50% beyond the fourth decade of life [4], [7]. Similarly esophagus and anogenital tracts are often affected by solid tumors [7], [8] and specific genetic subtypes have been associated to augmented breast and ovarian cancer predisposition [9]. Whether compared to age-matched healthy subjects, FA patients show a 500-fold greater risk to suffer from AML [7], and over 700-fold increased hazard from HNSCCs, paired to a very aggressive course with 2-year survival rates <50% [9]; FA is indeed reported as the commonest inherited BMF syndrome with the highest cancer susceptibility [6].
The disorder is further related to congenital multi-organ malformations and other recurrent features, grouped into the VACTERL-H (Vertebral abnormalities, Anal atresia, Cardiac abnormalities, Tracheo-esophageal fistula, Esophageal or duodenal atresia, Renal abnormalities, upper Limb abnormalities and Hydrocephalus) and PHENOS (skin Pigmentation abnormalities, small Head, small Eyes, structural central Nervous system abnormalities, Otologic abnormalities and Short stature) acronyms, respectively, and frequently coupled in patients [11], [12].

The plethora of clinical manifestations and their expressivity, ranging from no evidence to fatal features, complicate the establishment of a clear genotype-phenotype correlation [13], as well as increase the phenotypical overlap of FA with other conditions (i.e. Seckel syndrome, Nijmegen breakage syndrome, Diamond-Blackfan anemia, Shwachman-Diamond syndrome, thrombocytopenia absent radius syndrome, dyskeratosis congenita) [14].

1.2 CELLULAR PHENOTYPE AND DIAGNOSTIC TESTS

FA cells show a distinctive hypersensitivity to DNA interstrand cross-links (ICLs)-inducing agents (e.g., mitomycin C (MMC), diepoxybutane (DEB), hydroxyurea (HU) and cisplatin), leading to increased chromosome fragility (Figure 1), G2/M cell cycle prolongation or arrest due to genome integrity checkpoint failure, and reduced survival rates in comparison with normal cells [15], [16]. Augmented chromosomal breakage and multiradial figures caused by DNA cross-linkers provide unique FA cellular markers, which have been exploited for the development of first-line diagnostic tests (i.e., DEB test) [17], [18] (Figure 1). DEB is the main ICL-inducing drug used for this purpose by virtue of its high stability and specificity for FA probands' detection [19]. Despite the wide variability of clastogens sensitivity among patients, these assays have greatly contributed to overcome the difficulties of diagnosis based on the sole clinical manifestations [14].

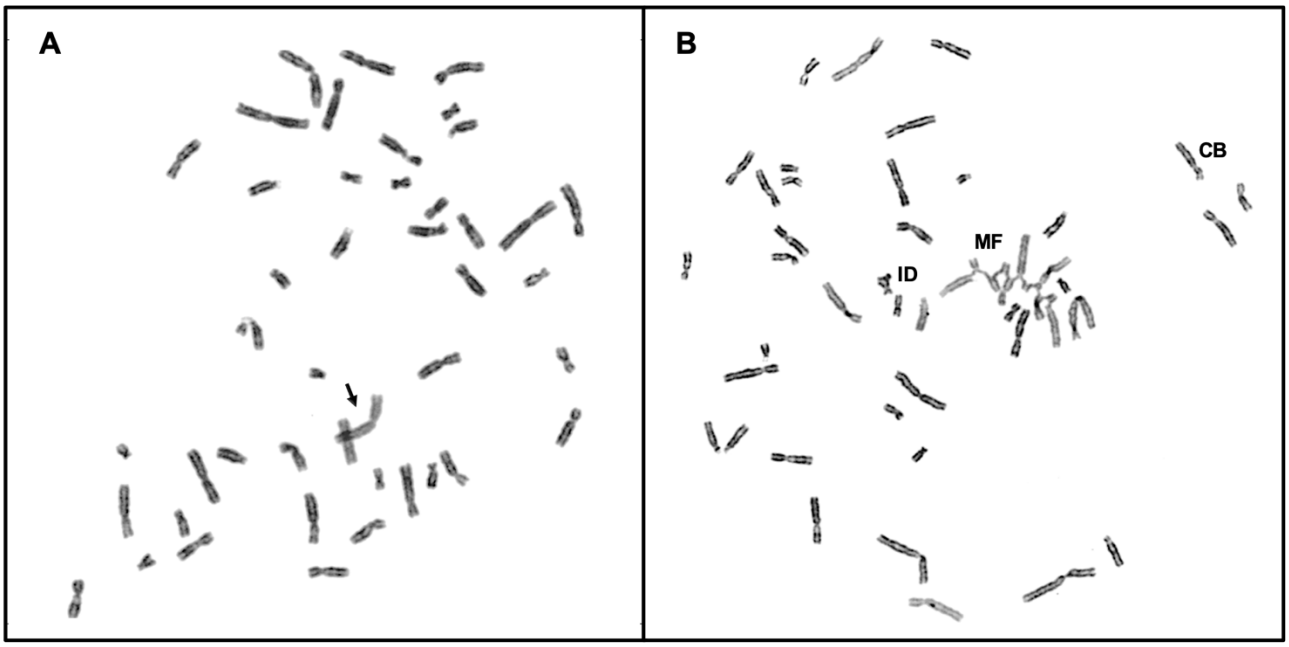


Figure 1. Comparison of metaphase spreads of FA lymphocytes with no (A) and 0.1 µg/ml DEB treatment (B). FA cells display spontaneous chromosome fragility and consequent aberrations (black row), destined to significantly increase after exposure to DEB (ID: interstitial deletion; MF: multiradial figure; CB: chromosomal breakage). (Adapted from Auerbach, 2003 [18] ; Castella *et al*., 2011 [19]).

An accurate cellular diagnostic process could be further complicated by somatic mosaicism, estimated at around 15-25% of all FA cases [20], [21]. Somatic mosaicism may spring from reversion or compensatory mutations in either pluripotent or committed hematopoietic stem and progenitor cells (HSPCs), endowing corrected clones with a proliferative advantage and favoring their expansion over FA cells [21]–[23]. In this light, the phenomenon has been defined as a "natural gene therapy" process and recently rated as a good prognostic factor, able to predict the clinical course of FA patients and provide important insights for gene therapy recipients [24]. Somatic mosaicism is often first recognized by a divergent DEB susceptibility between proband's T lymphocytes and primary fibroblasts (PFs); while the former endure clastogenic concentrations typically toxic to FA cells, the latter, hardly ever

affected by mosaicism [25], tend to maintain the characteristic fragility [19], [20], [26]. In the past most of mosaic patients were not properly characterized because of limited percentage of resilient T cells, but optimized algorithms are now available to diagnose the condition, despite the lack of common standard criteria among centers [21], [24], [27]. Evaluations of refractory cells percentage within a single lineage (i. e., T lymphocytes) and timepoint, however, may not be representative of multilineage hematopoietic reversion; thus, genetic sequencing and assessment of distinct blood and bone marrow (BM) progenitor populations (e.g., BM-derived colony forming cells, CFCs) have been proposed as valuable tools to confirm FA mosaicism, whilst still confined to the research dimension due to the cumbersome feasibility in the medical routine [22].

FA cells also commonly manifest oxygen and aldehyde sensitivity, cell-reprogramming anomalies, increased cell death, aberrant production of hematopoietic stem cell (HSC) proapoptotic cytokines, and significant oxidative stress due to mitochondrial impairment. All these features are likely to be elicited by both DNA repair defects, and the alteration of additional FA/BRCA pathway functions, whose combination plainly account for patients' great phenotypic variability [9], [28].

1.3 FA/BRCA PATHWAY ROLES AND COMPONENTS

FA proteins cooperate in the FA/Breast cancer (BRCA) genome maintenance pathway, which coordinates the resolution of ICLs, covalent bonds between the two DNA strands impairing physiological replication-fork progression and transcriptional process [9], [29].
As illustrated in Figure 2, ICL-stalled replication forks elicit the assembling of at least eight FA proteins (FANC-A, -B, -C, -E, -F, -G, -L, -M) into a nuclear multimer ubiquitin ligase (FANCore complex) [9], [30], [31]. This latter, together with an E2-conjugase (FANCT/ UBE2T) and additional FA interacting proteins (FAAP100, FAAP24, HES1) [16], [32]–[34], mono-ubiquitinates the FANCD2-FANCI heterodimer (ID complex) (Figure 2), which in turns relocates to DNA damage in an ATR-FANCS/BRCA1-dependent manner forming microscopically visible repair foci [35], [36]. The ID complex mono-ubiquitination is a key stage for the activation of the FA/BRCA ICL repairosome [37], [38], and it is commonly disrupted in patients bearing mutations in any of the FANCore complex genes or *FANCI* [15], [16].
Upon the lesion, the ID complex fosters ICL unhooking and translesion synthesis through the recruitment of FANCP/SLX4 and FANCQ/ERCC nucleases, and REV1 and POLζ translesion polymerases, respectively (Figure 2) [39]–[42]. The second reaction elongates the DNA leading strand, which is then used as a template for homologous recombination repair (HRR) by downstream FA proteins, as FANCD1/BRCA2 and FANCS/BRCA1 (Figure 2) [29], [43]. Once genome integrity has been reestablished, USP1 deubiquitinase removes the ubiquitin from the ID complex, disabling FA functional network [44]. Indeed, the activation of the FA/BRCA repair pathway is confined to sole conditions of endogenous or exogenous DNA damage in the S-phase of the cell cycle [45].

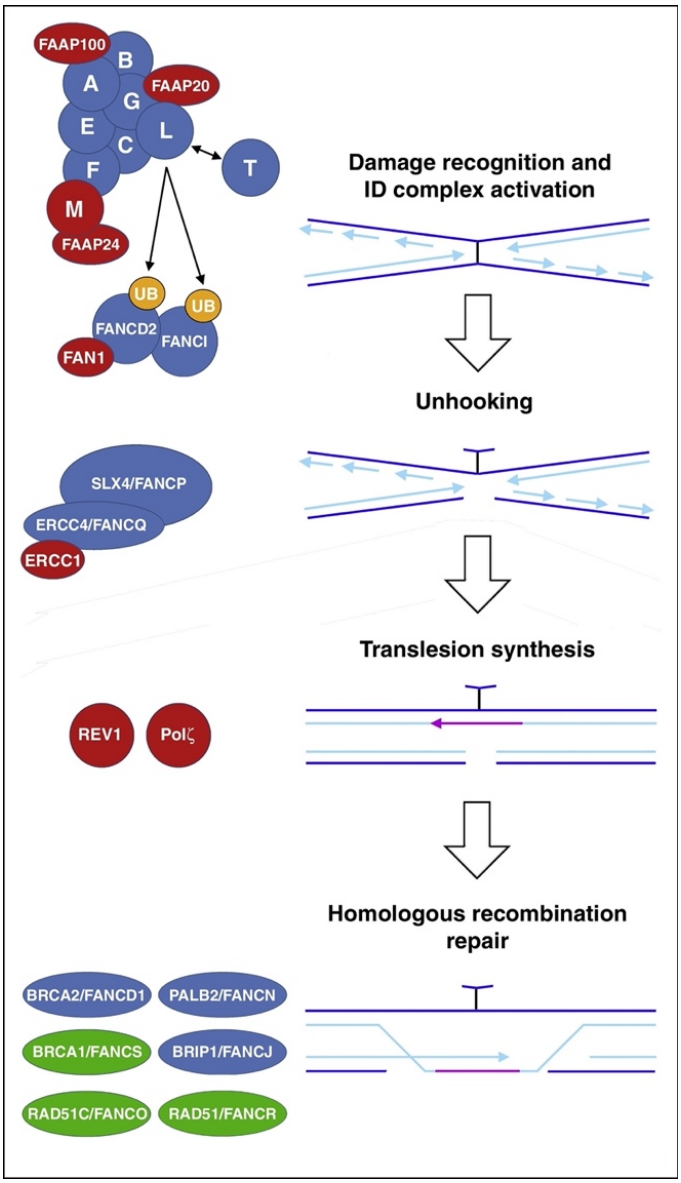


Figure 2. FA/BRCA DNA repair pathway. FANCore complex recognizes ICLs and, thus, exerts the mono-ubiquitination of the ID complex. This latter relocates to the DNA damage and, together with additional interacting proteins, stimulates ICL unhooking, translesion synthesis, and HRR. The final physiological outcome is the restoration of genome integrity. (From Bogliolo and Surrallés, 2005 [9]).

Further to their primary involvement in ICL processing, FA proteins have been associated to additional functions, encompassing mitochondrial metabolism and oxidative stress [28], [46]–[49], autophagy, inflammation [50], viral infection [51], [52], apoptosis signaling [53], telomere shortening [54], and HSCs maintenance [55]. Most of these ancillary activities are still poorly understood due to their recent discovery, but the scientific community is making great efforts to clarify FA proteins contribution in hopes of identifying novel therapeutic targets.

1.4 GENETIC HETEROGENEITY

FA is characterized by a complex genetic heterogeneity; 23 genes have thus far been associated to the disorder [1], and open clinical cases further hint at still unidentified ones [56]. Biallelic pathogenic variants affect any of 21 genes (*FANC-A*, *-C*, *-D1/BRCA2*, *-D2*, *-E*, *-F*, *-G*, *-I*, *-J/BRIP1*, *-L*, *-M*, *-N/PALB2*, *-O/RAD51C*, *-P/SLX4*, *-Q/ERCC4/XPF*, *-S/BRCA1*, *-T/UBE2T*, *-U/XRCC2*, *-V/MAD2L2, -W/RFWD3, -Y/FAAP100*), and X-linked and autosomal dominant mutations have been reported for *FANCB* and *FANCR/RAD51*, respectively [57], [58]. Moreover, monoallelic mutations in *FANCD1/BRCA2*, *FANCS/BRCA1*, *FANCJ/BRIP1*, *FANCM*, *FANCN/PALB2*, and *FANCO/RAD51C* predispose to breast and ovarian cancer, elucidating the deep connection between FA and these familial tumors, and justifying the FA/BRCA naming of the pathway [56], [59].

Stricter criteria, set before *FANCV/MAD2L2, FANCW/RFWD3,* and *FANCY/FAAP100* discovery, recognize only 15 as *bona fide* FA genes (*FANC-A*, *-B, -C*, *-D1/BRCA2*, *-D2*, *-E*, *-F*, *-G*, *-I*, *-J/BRIP1*, *-L*, *-M*, *-N/PALB2*, *-P/SLX4*, *-Q/ERCC4/XPF*, *-S/BRCA1*, *-T/UBE2T*), whereas suggest *FANCO/RAD51C, FANCR/RAD51* and *FANCS/BRCA1* inclusion into a separate category inducing a FA-like syndrome without BMF [9]. Still according to this stringent classification, *FANCM* and *FANCU/XRCC2* should be excluded from *bona fide* FA genes; biallelic mutations of *FANCM* have been found only in patients carrying variants also in another *bona fide* gene or not showing a FA phenotype [60]–[62], while homozygous *FANCU/XRCC2* mutations have been reported in a single case without further functional evidence in support of their causative role [63].

Up to 60-80% of all FA individuals harbor pathogenic variants in *FANCA*, which appears typically altered in the Mediterranean basin populations [56], [64]–[68]. *FANCC* is reported as the second most often mutated gene in the FA patients from the United States of America (USA) (10-15%) [14], the majority of whom are Ashkenazi Jews carrying c.711+4A>T founder mutation of *FANCC* [69]. However, this is not representative of most of the FA European probands, where *FANCD2* is the second mainly affected gene and, together with *FANCG*, account for around 15% of cases [56], [66]. Eventually, the other 19 genes are altered in <5% of patients [70]. A greater incidence is reported in highly inbred ethnic communities, as South Afrikaners [71], Spanish gypsies [72], Ashkenazi Jews [69], and some Tunisian [65], [73] and Italian families [64], [66] due to founder mutations [74].

1.5 FANCA: GENE, PROTEIN, VARIANT SPECTRUM

FANCA gene (MIM#607139; GenBank: NM_000135.2) maps on chromosome 16q24.3 and spans around 80 kilobase (kb), displaying a 5' GC-rich region, distinctive of the housekeeping genes, upstream the transcriptional start site [75], [76]. *FANCA* consists of 43 exons, and its open reading frame (4365 bp) encodes the FANCA protein (UniProtKB: NP_000126.2; 1,455 aa, 162.775 kDa), which exerts its main activity as a component of the FANCore complex in the nucleus [75], [77]. Accordingly, FANCA canonical isoform presents a well conserved nuclear localization signal (NLS), causing cytoplasmic retention if disrupted

[78], and five nuclear export signals (NESs) [79]. Additional functional sites include binding and/or interaction domains with other proteins (e.g., FANCG, BRCA1, FAAP20) [80]–[82], and phosphorylation-residue targets [83], [84].

FANCA is typically associated to private mutations in individual or few familial groups; as a result, the number of distinct *FANCA* alleles documented appears extremely high considering the limited amount of FA probands [56], [64]–[66], [68], [74], [85]. An exception is given by founder mutations, as the well-known exon 12-31del in South Afrikaner [71], c.295C>T (p.Q99*) in Spanish gypsy [72], and exon 15del in Tunisian communities [73].
All types of variants have been reported for *FANCA*, of which more than 25% consist of missense/nonsense mutations (Table 1) [3], [86]. According to previous studies, almost any missense or other *FANCA* not-truncating (NT) variant (i.e., in frame deletions and insertion not impairing the translational process) produces a stable, albeit not functional protein that is unable to normally translocate to the nucleus, and to activate the FA/BRCA pathway via FANCD2 mono-ubiquitination [3], [87]. Consistently, no clinical difference has been noticed between probands expressing NT mutants and FANCA-null ones [19]. The defective subcellular localization of FANCA NT proteins is still unknown. Since the variants are dispersed throughout the entire gene and rarely disrupt the NLS, inappropriate peptide folding is suspected to be the main cause [68], [78]. Moreover, FANCA nuclear relocation has been proven to strongly impact on protein function; forcing FANCA nuclear exclusion by a NES-tag is enough to compromise the whole pathway, as well as merging a N-terminal-deficient FANCA with a NLS can restore ICLs sensitivity [88]. Some FANCA NT mutants have further shown to resist clastogens exposure when expressed at higher levels than the endogenous ones [87], [88]. These findings illustrate that, in overexpressing conditions, NT proteins can not only relocate to the nucleus, but also keep a partial activity if sufficiently concentrated in the nuclear compartment. Thus, NT mutants appear as promising substrates for the activity of candidate drugs, increasing the chances of identifying new therapeutic options to rescue FA/BRCA pathway [3].
The second major category of *FANCA* variants is that of gross deletions, accounting for up to 20% all *FANCA* alleles (Table 1) [3], [66], [68], [86]. These are mainly due to long distance recombination between Alu-repeats *in cis* [74], [89], [90], and often exceed *FANCA* borders [91]; however, the adjacent genes affected are generally not responsible for any other autosomal dominant disorder, nor for any change of FA typical clinical phenotype [3].
Other recurrent *FANCA* variants encompass splicing mutations and small deletions, both found in nearly 15% of all cases (Table 1) [3], [66], [86] . Splicing variants commonly result from the disruption of conserved bases facilitating the recognition of exon-intron boundaries by the spliceosome [92], and lead to intron retention, exon skipping, and activation of cryptic sites impairing messenger RNA (mRNA) and/or protein synthesis. Small deletions, instead, are generally attributed to DNA polymerase slippage events during replication, as they often include short repetitive tracts [3], [86]. Moreover, this mutation type involves two variants frequently found among distinct ethnicities, c.1115_1118delTTGG and c.3788_3790delTCT [74], [86]. The latter is a NT variant representing the commonest *FANCA* mutation worldwide (around 20-30% of all FA alleles), with the highest percentages observed in probands from La Palma Island (Spain) and Brazil (80% and 51%, respectively) [3], [74], [86], [93].

Haplotype analysis in patients with c.3788_3790delTCT and from geographic distant areas allowed to rule out the hypothesis of a mutational hotspot, and to confirm the spreading of the variant from a common Indo-European ancestor [3].
Eventually, the remaining *FANCA* variants (regulatory substitutions, small and gross insertions/duplications, insertions and deletions (indels), complex rearrangements) are found in <8% of FA cases (Table 1).

Table 1. *FANCA* mutational spectrum. The table illustrates the mutation type and the corresponding number of all the 833 *FANCA* pathogenic variants annotated in Human Gene Mutation Database (HGMD) to date [94].

Mutation type	Total number of mutations
Missense/nonsense	300
Splicing substitutions	125
Regulatory substitutions	1
Small deletions	122
Small insertions/duplications	54
Indels	6
Gross deletions	271
Gross insertions/duplications	3
Complex rearrangements	3
TOTAL	**885**

The wide mutational spectrum and the mostly private nature of the variants of *FANCA* and the other 22 FA genes emphasize the need to adopt multiprocessing strategies for a thorough molecular diagnosis.

1.6 FA MOLECULAR DIAGNOSIS

After an early clinical suspicion, FA diagnostic process entails the investigation of the cellular phenotype via chromosome fragility test [17], [18] and, in case of positive response, the determination of the underlying gene and causative variants as ultimate confirmation.
The traditional approaches for FA variants analysis based on Sanger sequencing of candidate genes and viral complementation [66], [95] have now given way to next-generation sequencing (NGS) and orthogonal techniques [56], [66], [68], [96], [97], [1]. Consistently, FA genes have been inserted into designated [66], [96], [98] and BMF syndromes [99] gene panels, as well as in clinical and whole exome sequencing services (including ~4,000 genes involved in Online Mendelian Inheritance in Man (OMIM) [100] genetic disorders, and ~22,000 coding genes, respectively). Targeted NGS (t-NGS) methods are time and cost-effective diagnostic tools, but their target-specificity might collide with the detection of variants into genes not yet linked to FA or the brisk pace of gene discovery [56]. These limitations could be partially compensated by a high coverage of the sequences intercepted and/or periodic updates of the panels; indeed, works exploiting these

approaches succeeded in the characterization of most of the FA candidates interrogated [66], [98].

A challenge to NGS, regardless of the technology used, is given by the recognition of large genomic rearrangements, such as copy-number variants (CNVs), mainly due to short read lengths and GC-content bias [101]–[103]. This feature assumes a particular connotation within the FA context, since large deletions constitute the second major category of *FANCA* mutations [3], [66], [68], [86]. Thus, several tools to infer CNVs from NGS data have been developed, providing validation rates matching with those of the gold standards for CNVs detection (i.e., multiplex ligation-dependent probe amplification, MLPA) [56], [66], [98], [104]. These results supported the use of the NGS data produced in the diagnostic process as a first CNVs screening step, with the potential to reduce or even avoid MLPA high costs and time consumption, and enlarge the subset of genes assayed [56], [104]. Albeit time-consuming, the determination of the specific breakpoint coordinates could further provide important evidence for founder deletions [66], [68], [90], [98].

An additional hurdle to FA molecular diagnosis is the characterization of variants of unknown significance (VUSs), accounting together for about 45% of *FANCA* alterations [3], [86]. Contrary to the clear deleterious effect of CNVs, VUSs pathogenicity does require additional functional studies to be assessed [105]. Regarding unreported missense variants, this need could be addressed by investigating the genetic complementation of FA probands' cell lines [66] or cellular models carrying the target VUS [56], [68] through established tests (i.e., western blot, cell cycle arrest, and survival under MMC/DEB exposure) [95], [106]. Moreover, these strategies allowed to unveil the neutral effect of missense variants previously annotated as pathogenic, reiterating the importance of VUS functional analysis [56], [66], [68].

The impact of the splicing variants could be instead evaluated via reverse transcription polymerase chain reaction (RT-PCR) on patients' RNA [66], [68] or minigene assay, used to artificially reproduce splicing patterns in case of specimen unavailability [92]. In both situations, the effect produced could be further investigated by Sanger sequencing and/or real-time PCR (qPCR) to determine the expression levels of the aberrant transcripts.

Another bottleneck in FA molecular diagnosis is due to deep intronic variants (DIVs). Considering the location outside t-NGS and whole exome sequencing (WES) targets, DIVs presence could be deduced as a consequence of the unsuccessful identification of FA alleles, or, alternatively, determined via whole genome sequencing (WGS) or RNA sequencing (RNA-seq) [66], [68]. This last technique also allows to examine DIVs impact on splicing and/or RNA stability by evaluating transcripts quality and quantity through NGS.

Eventually, the high frequency of FA somatic patients is an additional factor that cannot be ignored in the diagnostic process. Depending on the levels and the genetic events eliciting the reversion, one pathogenic variant could be missed in probands' peripheral blood (PB) DNA [107], [108] and requires more invasive procedures, such as PFs samples, to be revealed. However, published studies successfully proved the solidity and reliability of WES approaches for the detection of FA mosaics, irrespective of the DNA source [24], [56].

A complete, prompt molecular diagnosis is of utmost importance for FA patients' appropriate clinical management, families' genetic counseling, carrier testing, prenatal and postnatal

diagnosis, donor selection for HSCT, genotype-phenotype correlations for prognostic implications, and therapeutic novelties [1], [14]. Several diagnostics and research groups worldwide have made multiple endeavors in this direction, indeed producing large collections of characterized FA probands [56], [66], [68]. Nevertheless, the identification of approved and commonly adopted standards for FA molecular diagnosis still remains a mostly unmet need.

1.7 FA THERAPIES

1.7.1 CURRENT TREATMENTS

Since the dawn of FA therapy, hematopoietic stem cell transplantation (HSCT) has remained the only curative approach for hematological defects, encompassing BMF, AML and MDS. The first effective conditioning regimen was established in 1984, requiring low-doses of cyclophosphamide and single-fraction irradiation, given patients' susceptibility to DNA cross-linkers and radiosensitivity [109], [110]. Current HSCT protocols are ideally performed before multiple transfusions and malignancies occurrence [111], and further benefit from the incorporation of the immunosuppressant fludarabine and T-cell depletion, which greatly contributed to lower both graft rejection and graft-versus-host disease (GVHD) risks [112]. Consistently, optimized regimens now provide an overall survival of up to 90% at 5 years post HSCT with sibling donors in case of BMF, and of 50% when AML and MDS are present, a rate, this latter, which still has margin for improvement [111]. As an alternative to BM, umbilical cord blood represents another valuable source of HSCs, exploited worldwide to treat FA and other blood disorders [113].

Despite the unquestionable progress of HSCT protocols in the last three decades, not all FA subjects have full access or are prepared for transplantation, nor have a compatible donor [111]. *In vitro* fertilization combined with preimplantation genetic diagnosis could represent a viable option to generate savior human leukocyte antigen-identical offspring and, thus, provide a compatible donor to cure affected siblings [114]. Where this is not possible, androgens administration (e.g., oxymetholone, oxandrolone, danazol) could be used as a palliative alternative to ameliorate low blood counts. 50-70% patients show a positive response to the treatment, but <20% tolerate it in the long term, and most of them experience side effects, as virilization and liver toxicity [115]–[117]. Beyond these problematics, the very major concern raised by HSCT is both its indirect, and direct influence in tumor solid appearance. The longer lifespan of transplanted patients is shading light on the side-effects of chemotherapeutic agents use in HSCT conditioning regimens, and GVHD. Transplanted FA subjects with GVHD indeed show an increased rate of HNSCCs and, in general, patients who underwent HSCT have a 4.4-fold higher risk of developing HNSCCs than probands who were not transplanted [118], [119]; as a result, transplanted FA patients have more than 2000-fold increased probability to suffer from HNSCCs than healthy individuals. Moreover, FA oncologic patients are not eligible for conventional cancer therapies due to their susceptibility to chemotherapy and radiation. Surgical resection is thus

the only feasible approach providing good survival at early tumor stages, but malignancies are often lately diagnosed and palliative cures, as surgery and low chemotherapy/radiation, have henceforth poor effect [4], [120].

It is therefore an open challenge to respond to the urgent need for novel chemopreventive and therapeutical strategies in FA, especially without favoring the resolution of a manifestation to the detriment of another.

1.7.2 CLINICAL TRIALS

To date over 50 FA clinical studies, of which more than 10 still active and recruiting patients [121], have been run with the aim to overcome the limitations of the currently available cures, and so prevent and effectively heal FA clinical features.

1.7.2.1 GENE THERAPY

Gene therapy (GT) provides a compelling option for the treatment of hematological defects, well on its way to supplant allogenic HSC transplantation soon. Considering somatic reverse mosaicism recall to “natural GT” [24], the *ex vivo* correction of autologous FA HSCs holds promises to rescue blood manifestations, and further abolish HSCT main issues (i.e., need for a matched donor, conditioning, rejection and GVHD risks) [122], [123]. However, despite this and the successful use of GT in other monogenic hematopoietic disorders [124]–[127], the efficacy of FA clinical trials was long confined to xenogeneic murine models only [128], [129], due to poor HSCs engraftment in patients. Several factors accounted for these negative results, such as poor numbers and quality of collected and, thus, reinfused HSCs, increased cell fragility in prolonged *ex vivo* transduction with gamma-retroviral vectors, no pre-infusion conditioning facilitating HSCs engraftment [130], [131]. Based on the conclusions of the previous studies, GT approaches for FA have been progressively optimized, reaching just in 2019 the first successful outcomes with FANCOLEN-1 (NCT03157804) conducted in FA subtype A (FA-A) probands with BMF [132]. This still active phase I/II clinical trial includes the short *ex vivo* transduction of a safe, integrative lentiviral vector for *FANCA* [133] into cells carrying $CD34^+$ HSC marker from patient’s BM or mobilized PB, then reinfused into hosts without any conditioning regimen [134]. Increases of up to 60-70% corrected cells in PB and BM at 24 months post-GT indicated stable engraftment in all candidates, with better results in those reinfused with the highest number of HSPCs. The proliferative advantage of corrected repopulating cells over FA ones was associated not with genotoxic insertions, but normalized cellular phenotypes and, consistently, no leukemogenesis events were reported [132]. These data were further linked to an arrest in BMF progression despite no marked augment in PB cell counts, suggesting patients’ hematological recovery as a slow, long-term process, on a par with the phenotypic rescue of FA mosaic individuals [21], [23]. In the wake of FANCOLEN-1, many other clinical studies have been started, aiming to augment the number of corrected HSCs to inoculate,

use non-genotoxic conditionings, and, importantly, improve the hematological response to chemotherapy in GT-treated patients in case of solid tumors [135]–[137].

1.7.2.2 BONE MARROW FAILURE AND CANCER PREVENTION

As previously mentioned, FA cells show increased oxidoreductive status imbalance due to mitophagy impairment [28], [47], [50] and hypersensitivity to proinflammatory cytokines and aldehydes [138], [139]. These observations have promoted the start of drug-based clinical trials towards reactive oxygen species (ROS) synthesis and aldehyde-induced DNA damage to prevent BMF and, potentially, solid tumor appearance. Within this framework, two active studies are centered on the usage of quercetin [140], [141], a naturally occurring flavonoid with antioxidant, chelating, and anti-inflammatory and anticancer properties [142], [143]. Quercetin viability was first tested in FA obese mice with insulin resistance, showing the reversion of both hormonal signaling and obesity phenotype, as a consequence of decreased ROS production [144]. Furthermore, early studies in FA subjects demonstrated long-term safety, good tolerance, and biologically relevant blood levels of quercetin delivery [145]. On these bases, the current trials aim to assess the efficacy of oral quercetin therapy in curtailing the progression of BMF [141], and in preventing or delaying squamous cell carcinomas (SCCs) and their associated complications in FA [140]; importantly the latter study includes both no transplanted and post-HSCT patients, with or without malignancies.
Another antioxidant drug exploited in FA clinical trials is metformin [146], used to safely treat type 2 diabetes and insulin resistance, and known to trigger kinases with a role in HSC maintenance [147]. By virtue of its aldehyde scavenger activity, metformin was also linked to enhanced hematopoiesis and delayed tumor appearance in *FancD2*-deficient mice with impaired aldehyde detoxification [148]. Consistently, the pilot study conducted at Boston Children's Hospital proved feasibility and tolerability of oral metformin application in non-diabetic FA patients, reporting an improved hematological response in a subset of them [149]. Metformin is further employed in clinical trials centered on solid tumor prevention, not excluding FA subjects enrollment [121]. As lately published, metformin administration in individuals with oral premalignant lesions led to encouraging histological response, and modulation of HNSCC-dysregulated pathways, opening up to new, possible preventive strategies in FA [150].
FA patients with oral cancer risk are admitted to participating in a clinical study focused on pembrolizumab [121], an immunotherapy antibody targeting cell programmed cell death protein (PD)-1 and approved for tumors deficient for mismatch repair [151], [152]. In a very recent publication, pembrolizumab activity was also tested in FA homologous recombination (HR)-defective malignancies, elucidating their susceptibility to immune-checkpoint blockade [153].

1.7.2.3 SOLID TUMOR TREATMENT

Clinical trials in FA have so far addressed the occurrence of solid tumors with merely prophylactic approaches, and none has yet been designed for the specific treatment of the condition. In the attempt to push in this direction, FA patients are eligible to participate in trials investigating poly (ADP-ribose) polymerase inhibitors (PARPi) delivery in individuals with HR-deficient cancers due to *BRCA1*(*FANCS*)*/BRCA2* (*FANCD1*) mutations [121]. The therapeutic triumph of PARPi resides in the exploitation of synthetic interactions (SIs) between distinct DNA repair mechanisms. Referring to the ancient Greek term "synthetic" as "combination of two entities to generate a novel one", SIs are genetic relationships where perturbations of two genes are individually tolerated, while their co-occurrence is deleterious (synthetic lethality, SL) [154]–[156] or, on the contrary, confers a proliferative advantage to cells (synthetic viability, SV) (Figure 3). Within the PARPi context, PARP repression affects the resolution of single-strand breaks (SSBs) and traps PARP enzyme at the lesion, causing DNA damage accumulation and replication interference [157]. Cells with defective BRCA1/2 proteins, involved in double-strand breaks (DSBs) repair by HR, cannot fix these ruptures nor restore replication, causing cell lethality and, thus, halting cancerous proliferation [158]–[160]. At present four PARPi (olaparib, rucaparib, talazoparib, niraparib) have received Food and Drug Administration (FDA) approval to treat breast, ovarian, pancreatic, or prostate *BRCA1/2*-mutant cancers [161]. Moreover, new research is trying to widen the spectrum of PARPi action by recapitulating *BRCA*-like impairment of HR ("*BRCA*ness") in *BRCA*-proficient tumors [162]–[164], or targeting cancers with mutations in other HR repair genes [165], [166]. However, the selectively lethal effect of PARPi still encounters some limits in FA; unlike the other oncological patients, FA subjects show biallelic mutations and HR defective pathway not only in malignant, but all body cells, thus requiring more carful toxicity controls.

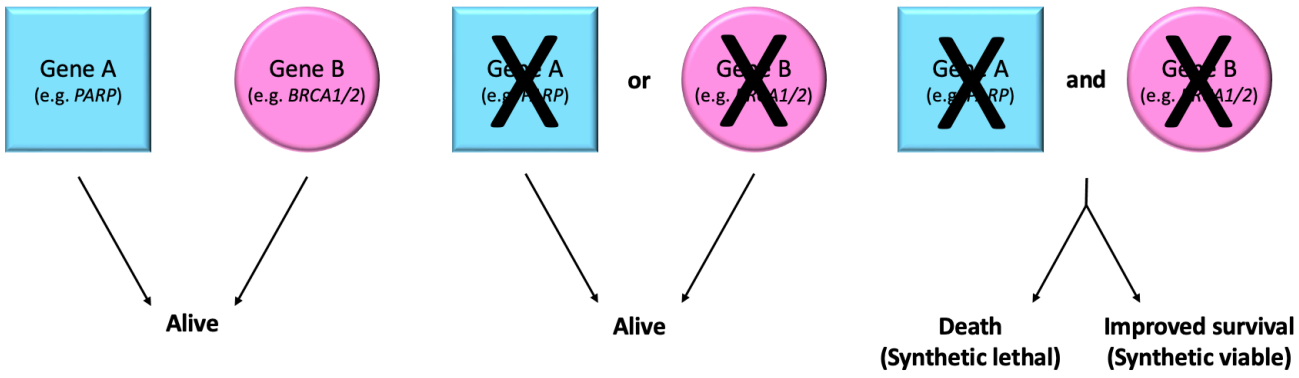


Figure 3. Schematic depiction of SIs. The expression of gene A (e.g., *PARP*) and B (e.g., *BRCA1/2*) together (left), as well as the inactivation of either alone (middle) is compatible with cell survival, while their simultaneous abrogation could compromise (synthetic lethality) or, instead, ameliorate cell fitness (synthetic viability) (right).

Optimized HSCT protocols have originally contributed to lower post-transplant mortality below 20%, and thus provide FA patients with a lifespan longer than 20 years [111], [112]

(Figure 4). However, the actual implementation of GT, non-genotoxic antitumoral molecules, new targeted and preventive approaches is foreseen to further improve FA patients' survival, achieving levels similar to healthy individuals [4] (Figure 4).

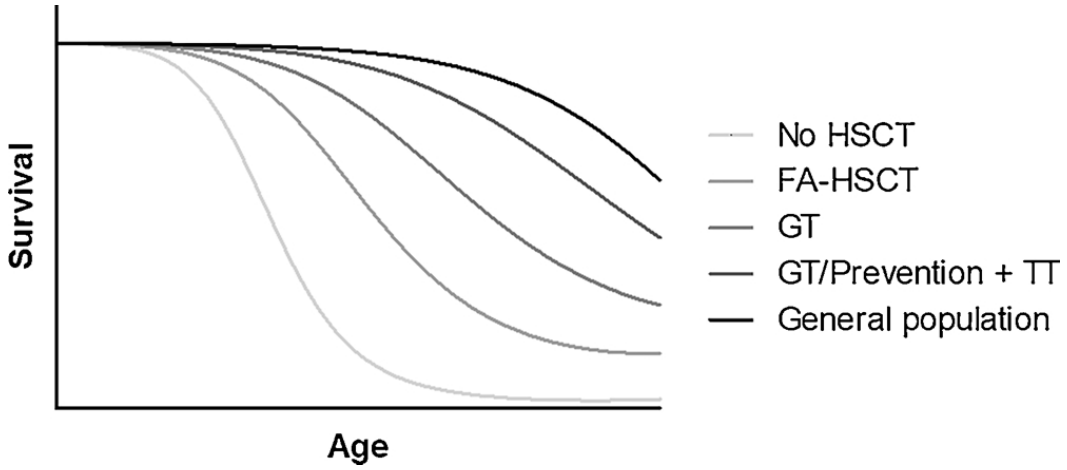


Figure 4. Estimate of FA patients' survival rates in response to novel potential therapies. FA subjects' life expectancy has significantly benefited from optimized HSCT protocols, going from a survival below 20 years (No HSCT) to a mortality reduction to less than 20% (FA-HSCT). GT approaches (GT), non-genotoxic antitumoral molecules, novel targeted (TT) and preventive treatments, as well as their combination (GT/Prevention + TT) hold the promises to further improve FA patient's quality of life, bringing it closer to that of healthy individuals (General population). (From Minguillón and Surrallés, 2018 [4]).

1.7.3 NOVEL PROMISING DIRECTIONS

1.7.3.1 CELL-BASED HIGH CONTENT DRUG SCREENINGS FOR DRUG REPURPOSING

Cell-based drug screenings, key steps of drug discovery strategies, have arose as one of the latest therapeutic promises for the still unmet burdens of FA medical treatment. Drug screenings could be divided into high throughput screenings (HTS), directed at specific pathways or functions, and high content screenings (HCS), based on phenotypic changes as readouts [167], [168]. Until recently, HTS were the prevalent drug discovery approach used, contemplating the identification of a molecular target by basic research, and its further investigation via large-scale compounds screenings in research centers or pharmaceutical industries [168], [169]. HCS, instead, have mainly caught on more recently, and entail the recognition of a disorder-specific defect with a commonly unknown target [169]. These assays are often performed in disease cell models to test candidate drugs for their ability to rescue an aberrant phenotype, and rely upon supports (i.e., microscopy or cytometry paired with robotic devices) suitable for the interrogation of high content (HC) libraries in a short time [168], [169].

A complete *de novo* drug discovery process requires on average 15 years and elevated costs, with a failure rate exceeding 95% [170]–[172]. This approach is hence hardly feasible for rare disease, which, due their little physiopathology knowledge and number of patients,

do not arouse interest from the pharmaceutical industry [168]. A possible way to overcome these hurdles is given by repurposing strategies [168], [173], successfully employed in HCS for some rare pathologies including already available drugs into the libraries screened, and exploring their therapeutic potential within a novel indication [174]–[178].

The combination of HCS and drug repositioning has also been tested in the FA field by Montanuy *et al*., who interrogated a fluorescent cell-based system with 3,082 molecules, including 1,200 FDA-approved drugs (Prestwick library), in search for candidates rescuing FA/BRCA pathway activity [169]. To achieve this purpose, they completely abolished *FANCA* expression by transcription activator-like effector nuclease (TALEN) gene editing [179] in a U2OS cell line, stably transfected with yellow fluorescent protein (YPF)-tagged FANCD2. Upon DNA damage, these cells were not able to mono-ubiquitinate FANCD2, nor to promote its relocation to fluorescent repair foci, instead visible in the nuclei of wild type (WT) cells via fluorescence microscopy (Figure 5) [169]. The model was then assayed by HCS [169], [180], and the hits identified were individually assessed for their capability to restore FANCD2 mono-ubiquitination and foci (Figure 5), and chromosomal fragility in a *FANCA*-null background. Unfortunately, all the promising candidates turned to be false positives, as any displayed high autofluorescence or FANCD2 distribution different from nuclear foci [169]. Thus, despite a mild reduction of chromosome instability, no drug was found to correct FA cellular phenotype at the conditions adopted. Nevertheless, Montanuy *et al*.'s work had the merits to illustrate the advantages and drawbacks to screen HC libraries for the development of new FA therapies, and to shift the focus of the search from inhibitors of FA/BRCA pathway [181], [182] to molecules able to recover its functionality [169].

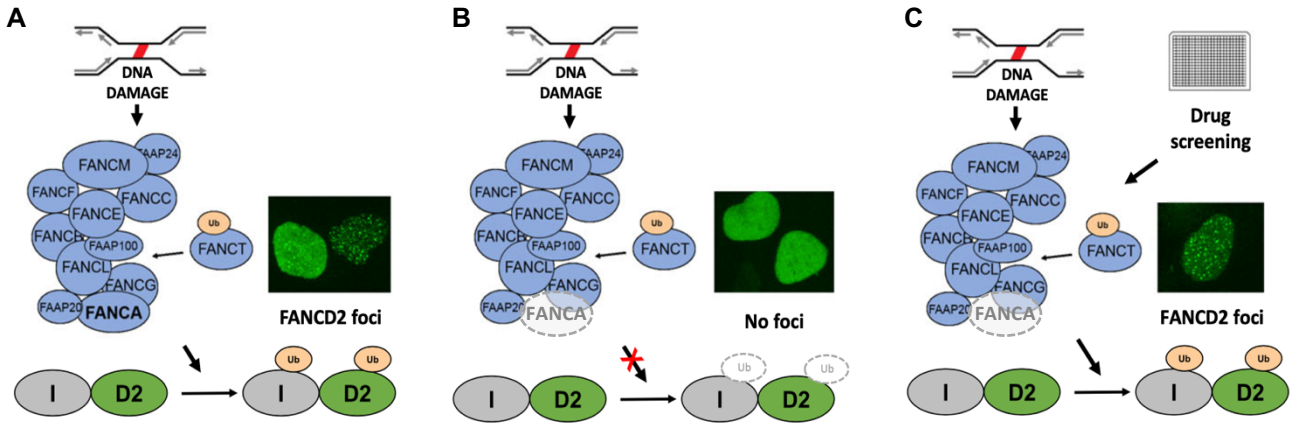


Figure 5. FA cell-based model for HCS. In WT conditions, FANCore complex is activated by DNA damage, ubiquitinates FANCD2 (FANCD2-Ub), and stimulates the formation of FANCD2 foci upon the lesion (nuclear intense spots) (A). FANCA loss prevents FANCore complex activation, compromising FANCD2-Ub and foci assembly (diffuse nuclear pattern) (B). In the HCS, YPF-FANCD2 foci restoration is the expected readout for a drug able to correct FANCA/BRCA pathway (C). (Adapted from Montanuy *et al*., 2020 [169]).

This very HCS further allowed to identify several nongenotoxic drugs with potential cancer-specific cytotoxicity, validated in PFs and HNSCC lines from FA patients [180]. Over 150 candidates tested, the two best hits were gefitinib [183] and afatinib [184], inhibitors of epidermal growth factor receptor (EGFR) kinase activity and downstream signaling,

FDA/European Medicines Agency (EMA)-approved for the treatment of non-small cell lung cancer (NSLC). Both molecules selectively inhibited HNSCC growth at low nanomolar concentrations without relevant effects on FA PFs and, used at therapeutic doses, produced no DNA damage-induced consequences (no FA pathway activation, chromosome fragility or G2/M cell-cycle arrest) on the viability of nontumor FA cells [180]. Regarding *in vivo* studies, gefitinib and afatinib were demonstrated to significantly diminish FA HNSCCs growth or size in mouse xenografts and to be well-tolerated in *Fanca*-deficient mice on chronic treatment [180]; only afatinib displayed some mild side effects, however easily managed by dose adaptation [185], [186]. Consistently, gefitinib administration to a single FA patient with EGFR-positive oral SCC led to a tumor size reduction by 80% after two months, without any documented toxicity [187].
Taken together, Montanuy *et al*.'s data provided a comprehensive preclinical analysis for gefitinib and afatinib repurposing as first, safe drugs exerting FA HNSCC-specific lethality [180]. These findings represented a real breakthrough in FA therapies, since, unlike the mainly preventive action of the available cures and clinical trials [4], both EGFR inhibitors showed potential for treating ongoing malignancies *in vivo* [180]. In addition, the two molecules did not exert any genotoxic effect on FA-deficient cells, or require combination with radio-chemotherapy, in contrast to PARPi [4] and approved antibodies for sporadic HNSCC [188], [189], respectively. Montanuy *et al*.'s work eventually led to the EMA orphan drug designation for gefitinib (EMA/OD/090/18; EU/3/18/2075) and afatinib (EMA/OD/141/18; EU/3/18/2110), with the next goal to launch a clinical trial for the treatment of HNSCC in FA [180].

1.7.3.2 CRISPR KNOCKOUT SCREENS FOR FA SYNTHETIC INTERACTIONS DISCOVERY

The therapeutic success of PARPi SL in *BRCA1/2*-mutant tumors has drawn attention to the discovery of new SIs to define combinatorial strategies against cancer drug resistance and relapse [190]. Consistently, there is a growing interest in adapting SL and SV principles to lower cure side-effects and treat tumors and additional features in FA, as well as in further monogenic disorders. To address this issue, CRISPR-CRISPR-associated protein 9 (Cas9) screens have emerged as the state of art to uncover novel SIs.

CRISPR-Cas9 technology derives from type II CRISPR-Cas apparatuses, prokaryotic adaptive immune systems evolved to recognize and trim invasive viral genomes and plasmids [191]–[193]. Its functioning relies upon Cas9, an endonuclease protein component, and a guide RNA (gRNA), a short oligonucleotide capable to bind its 5' end (spacer) to a complementary sequence within the target DNA, and precisely direct Cas9 cleavage *in loco* [194]–[196] (Figure 6). Another fundamental requisite for the site-specific activity of the system is the recognition of protospacer adjacent motifs (PAMs), conserved sequences flanking the foreign target [197]–[199] (Figure 6). In terms of genome editing, CRISPR-Cas9 large feasibility stems from the unique chance to program Cas9 to intercept any genomic region of interest by simply changing gRNA spacer complementarity [200]. Thus, upon the

gRNA-binding site, Cas9 introduces DSBs that, whether repaired by the error prone non-homologous end joining (NHEJ) mechanism, lead to indels (Figure 6) associated to phenotypes ranging from in-frame mutations to complete gene knockouts (KOs) [200].

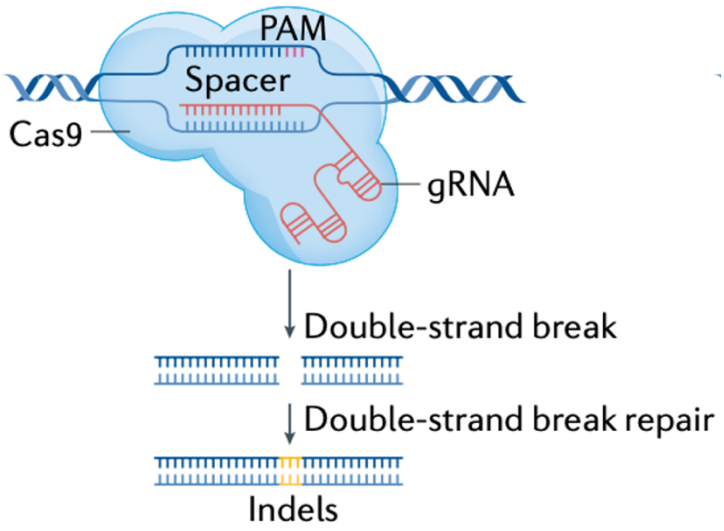


Figure 6. CRISPR-Cas9 system core components and function. Steered by a gRNA, Cas9 introduces DSBs within the target sequence flanked by PAMs. In absence of a template, DNA lesion is repaired by error-prone mechanisms (NHEJ), which result into small indels with possible KO phenotypes. (From Bock *et al*., 2022 [201]).

Moreover, CRISPR-Cas9 variants (i.e., nuclease-deactivated Cas combined with specific protein domains) have allowed to expand the genome-wide (GW) targeting potential beyond KO induction (CRISPRko) toward transcriptional control (CRISPR interference (CRISPRi) or activation (CRISPRa)) [202]–[205], RNA modulation [206]–[210], epigenome remodeling [211]–[214], base editing (CRISPR base editors) [215]–[217], and large-scale genome engineering (CRISPR prime editing) [218]–[220]. Owing to the targeting and repositioning suitability, design simplicity, effectiveness, and affordable cost, CRISPR-Cas9 technology has been exploited in a broad variety of models, and biotechnology and therapeutic branches, instead precluded to other engineered nucleases [200], [221], [222].

Forward genetic CRISPRko screens, being the mostly used CRISPR technique to date [201], [223], are meant to discover SIs through the suppression of target gene's expression. One of the first steps in the strategy includes the selection of a CRISPRko gRNA library with a GW scale activity (~20.000 gRNAs) or directed towards a smaller fraction of genes (~ 10^1-10^2 gRNAs) for unbiased or more focused approaches, respectively (Figure 7). Nowadays various tools and resources assist library design and election by sorting and ranking candidate gRNAs to optimize their efficiency and on-target action [224]–[227]; otherwise, it is possible to rely on commercial CRISPR libraries, already available and tested [228]. As a general principle, libraries are conceived to be redundant; it is advisable to use at least four independent oligonucleotides for each target *locus* to compensate gRNAs diverse efficiency and, thus, potential off-targets [226]. Libraries should also comprise gRNAs for application-specific control genes to validate the screen, as well as for safe harbor regions to monitor

DNA damage response and altered cell proliferation [201]. In most of cases, the guides are cloned into lentiviral vectors (Figure 7), which benefit of a high insert capacity, transduce both dividing and non-dividing cells, and stably integrate into the host genome.

Another key point in a CRISPR screen is the choice of a model system gathering the main biological processes and suitable for genetic screening (Figure 7). Immortalized cell lines represent a low-cost and easy to handle option with no limitation in cell number, particularly suitable for GW analyses. Nevertheless, the presence of supernumerary *loci* in genetically instable lines, such as some immortalized models, could lead to uneven CRISPR/Cas9 editing and potentially skew the final results [225], [229]. Further alternatives include primary cells, tissue-explants, organoids, or living animals, generally adopted to investigate more complex and context-dependent phenotypes [230]. To enhance results robustness and reproducibility, it is however recommendable to perform the screen in more biological replicates [226], [231] (Figure 7), as different genetic and cellular backgrounds could lead to divergent end phenotypes [232]. Alongside model selection, it should be decided whether to administer Cas9 together with the gRNAs, or to previously engineer the target cells to express the enzyme (Figure 7). Albeit requiring additional selection steps, the latter is generally favored for its greater safety- no single construct carrying both DSB-inducing components- and screening efficiency. Cas9 activity should be then validated by delivering gRNAs towards several test loci [122], [123] and assessing its editing efficiency at DNA and protein levels [233] (Figure 7).

Prior to model transduction, the viral gRNA library is titrated to meet the total amount of infectious particles required (Figure 7), based on the library size and the average number of cells per gRNA established (coverage). The library is typically delivered in bulk at low multiplicity of infection (MOI, set at 0.3- 0.5) (Figure 7), to enrich for cells integrating a single construct and avoid genetic interactions between distinct gRNAs [190], [201]. As a result, the number of viral particles multiplied by the MOI provides an estimate of the number of cells- including experimental replicates- that must be infected and cultured to maintain a congruous representativeness of the diverse gRNAs along the experiment. Once transduced, cells may be selected for antibiotic resistance during several days (~ 7 to 14 days) or the whole duration of the screen to ensure efficient KOs generation [224], [225], [229], [234] (Figure 7); their representation in the final pool will be henceforth determined by the proliferative competition with each other, fruit of the precise editing produced. Over the study, cells are also periodically split and/or harvested in independent replicates to increment end phenotypes separation, and enhance final hits percentages [226] (Figure 7). To achieve this, genomic DNA (gDNA) is extracted from different replicates, controls, and time point specimens, and polymerase chain reaction (PCR)-amplified by universal primers binding common sequences adjacent to the integrated gRNA cassettes (Figure 7). Moreover, during amplification, diverse barcodes (indexes) are added to the distinct samples to allow the simultaneous sequencing of more gRNA amplicon libraries, and the subsequent match of the single reads to the corresponding specimens [190].

Bioinformatic and statistical analyses are then run to compare the read counts of each gRNA to its representativeness within the library, and between different samples and controls [235], to calculate the enrichment/depletion scores and respective *p*-values of each gRNA/target gene [236], and to carry out screen quality control [236], [237] (Figure 7). The

outcome is a roster of candidate-gene hits, ranked according to their effect upon the phenotype of relevance. Importantly, CRISPRko screens readout for both negative and positive phenotypic selection [190], [201]; gRNAs depletion indicates genes whose repression sensitizes cells to the proliferative challenge (SL screens), while guides enrichment identifies genes that, if silenced, confer a selective advantage (SV screens).
To avoid false positives results, promising candidates should be validated by more focused screens or other small-scale assays, preferably employing complementary models and readouts [201], and performing genetic complementation as ultimate evidence [190] (Figure 7). Besides the top-raking genes, it is also advisable to test other randomly selected hits for a more representative evaluation of the screen outputs [201]. Eventually, accurate documentation of the screen workflow and data should be submitted to dedicated repositories to guarantee the reproducibility of the study conducted [201].

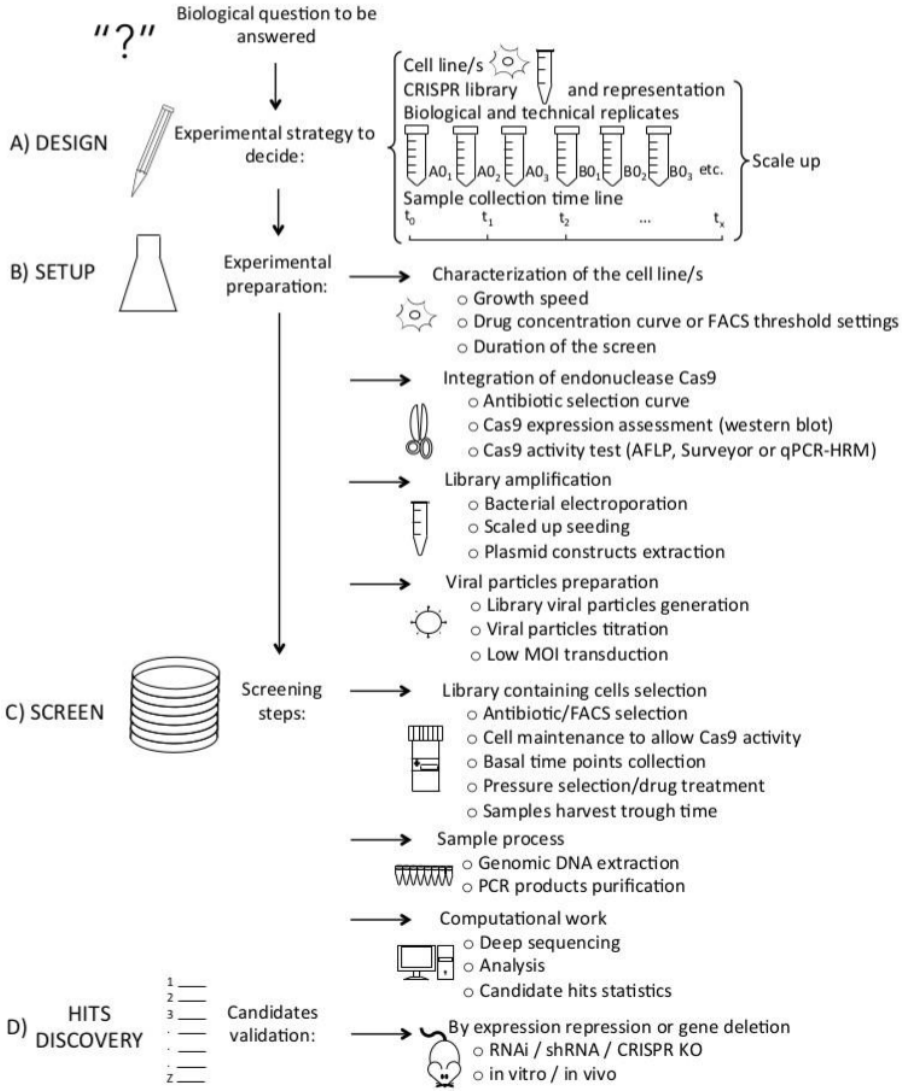


Figure 7. Schematic depiction of a CRISPRko screen workflow. A) The experimental design, the first indispensable step in a screen, should be conducted in compliance with the biological question to be answered. It encompasses the choice of the biological model, CRISPR library and its representation, and study scale up (number of biological and technical replicates, sample collection timeline). B) Previous to the actual screen

start, it is necessary to characterize the model selected and engineer it with Cas9, amplify the gRNA library, clone it into viral particles, and delivered it at low MOI. C) The experimental stages include biological model selection and maintenance for proper Cas9 activity and phenotypes representation, basal controls and samples collection over time, gDNA extraction, amplification, and sequencing, computational analysis and candidate statistics. D). Top-ranking hits should be validated by alternative approaches, such as RNA interfering (RNAi), short hairpin RNA (shRNA) and CRISPR focused KO, both *in vitro* and *in vivo*. Moreover, it is wise to further assay other randomly selected hits for a more accurate evaluation of the final results. (From Castells-Roca et al., 2021 [190]).

Richardson *et al.* demonstrated as CRISPR-Cas9 editing depends on several components of the FA/BRCA HRR pathway for single-strand template repair, but not for NHEJ [238]. FA defects thus do not obstruct the default NHEJ repair for Cas9-induced DSBs [238], [239] and, GW and focused CRISPRko approaches have consistently been exploited to uncover SIs involving FA genes.

Similarly to *BRCA1/2*-deficient tumors vulnerability to PARPi, the disruption of FA/BRCA members via a CRISPR loss-of-function (LOF) screen [240] was linked to the hypersensitization of several cancer cell lines to ataxia telangiectasia mutated (ATM) inhibitors [241]. FA and ATM pathways work in concert to activate the choline transporter-like protein (CtIP) [242], [243], which promotes the initiating step of HR, DSB end resection, and simultaneously prevents NHEJ [244]. Accordingly, the parallel inhibition of both pathways was demonstrated to significantly reduce end resection and enhance toxic NHEJ, resulting in radial chromosome accumulation and cellular death through SL [241].

Beyond the subordination to ATM-mediated repair, FA pathway loss also generates a dependency on the specific action of DNA polymerase (Pol) ι, emerged as a top hit in GW CRISPR-Cas9 fitness screens [226] on different FA KO lines [245]. The combined depletion of Pol ι and FA proteins was indeed congruent with an important attenuation of cell proliferation and survival due to heightened replicative stress, as documented by increased γH2AX foci formation and ATM/ATR activation [245], [36]. The nature of this synthetic interplay was ascribed to the lesion bypass activity of Pol ι [246], [247], assumed to compensate for replication stall by promoting fork stability and resumption in FA pathway-deficient cells [245].

Another example of FA SL interactor is aldehyde dehydrogenase 2 (ALDH2), which oxidizes aldehydes in healthy tissues [248], but is recurrently methylation-silenced in tumors [249], [250]. Via a CRISPR screen focused on the ubiquitination machinery, Yang *et al*. defined FA pathway as a biased essentiality in a subset of AML lines with low ALDH2 expression and activity [251]–[254]. This dependency hinted at a SL interaction between FA and ALDH2 proteins, indeed proved by cell sensitization and growth arrest in AMLs devoid of both FA and *ALDH2* genes [254]. FA pathway loss was therefore proposed to lead to the accumulation of aldehyde-induced ICLs in *ALDH2*-null AMLs, ultimately causing cell cycle block and death [254]. This SL relationship was further corroborated in murine models [255], and in patients with combined deficiencies of FA and *ALDH2* genes showing accelerated BMF progression [256]. Moreover, Yang *et al*.'s work pointed out a major observation: while FA predisposes to AML onset, the fitness of sporadic AMLs could hinge on the very function of FA proteins [254].

To emphasize the close link between the FA/BRCA and aldehyde metabolism pathways, Jung *et al*. carried out a CRISPRko screen with a metabolism-focused gRNA library on FANCD2-null Jurkatt cells, revealing aldehyde dehydrogenase 9 family member A1 (*ALDH9A1*) as a previously unrecognized SL gene for FANCD2 deficiency [257]. Interestingly, *ALDH2* expression was not depleted to indicate an aldehyde sensitivity typical of the cell type considered. *ALDH9A1* and *FANCD2* combined deficiency was associated to genomic instability, DNA damage-induced apoptosis and hematopoietic colony formation decrease [257]. Furthermore, *Fanca*$^{-/-}$ *Aldh9a1*$^{-/-}$ mice showed no spontaneous BMF, but a higher susceptibility to solid tumors, in particular to ovarian cancers, with aging [257]. Eventually, Jung *et al.* performed a suppression screen in *FANCD2*$^{-/-}$*ALDH9A1*$^{-/-}$ clones, which elucidated as *ATP13A3* loss improved cell survival probably by decreasing ALDH9A1 substrates and lessening DNA damage [257]–[259].
Turning to SV interactions, the exposure of *FANCC*-deficient cells to GW CRISPRko libraries [260] and insertional mutagenesis allowed to pair the lack of members of the Bloom's syndrome (BLM) complex to an augmented resistance to crosslinkers-induced damage and apoptosis [261]. Similar results were recapitulated also in *FANCI* and *FANCD2*-mutant models, but not in *BRCA1*-null cells, elucidating the specificity of the interaction to *bona fide* FA components [261]. The BLM complex indeed intervenes in the HR-mediated ICL resolution interacting with the FANCore complex via FANCM [262], and simultaneously preventing alternative NHEJ [263]. Consistently, the release of the NHEJ repair due to the absence of both complexes was suggested as the mechanism underlying the SV interaction between BLM and FA proteins. This was further supported by a decrease of DNA damage foci, but no chromosomal instability, and MMC sensitivity of *FANCC/BLM*-double KO clones upon PARPi [261]. The same work further validated the synthetic rescue of FA pathway by the nicotinamide adenine dinucleotide phosphate (NAD(P)H) quinone dehydrogenase (NQO1), which was regarded as a positive control owing to its known role in MMC bio-activation [264] and sensitization of cancer cells to MMC treatment [265].
Both BLM complex- and NQO1-SV interactions with FA were reconfirmed in genome-broad insertional mutagenesis screens performed across a panel of FA-defective cells, additionally reporting ubiquitin specific peptidase 48 (*USP48*) and late endosomal/lysosomal adaptor, MAPK and MTOR activator 1/5 (*LAMTOR1/5*) among top hit genes [266]. Ensuing studies on CRISPR-Cas9 edited *FANCC/USP48* cells correlated the loss of the ubiquitin peptidase USP48 to improved crosslinkers survival, DNA-damage removal, and chromosomal stability. This effect was attributed to an enhanced recruitment of BRCA1 and RAD51 upon the lesion, hinting at a role for USP48 in the counteraction of BRCA1 HR-activity [266], [267]. Complementary, FA-protein assembly at ICLs was assumed to tackle the onsite relocation and function of USP48 to favor HR repair [266]. All together these data endorsed USP48 intervention in the balance between different repair routes. In FA backgrounds, USP48 loss was also demonstrated not to recover FANCD2 mono-ubiquitination but, to some extent, histone H2A's one, thus possibly recognized by endonucleases solving ICLs in a FA-independent manner [266].

All together these studies emphasize the central role of CRISPRko technology in identifying novel druggable SV targets for the design of therapeutic strategies towards the entire FA

clinical spectrum, as well as targets for the treatment of sporadic cancers with alterations in a FA gene or one of its SL interactors employing inhibitors against the former or the latter, respectively.

2. AIMS

This thesis project intended to cope with the main hurdles now encountered in FA, which were addressed as follows:

- **Impairment of molecular diagnostic process due to FA vast genetic heterogeneity, multiple private and complex mutations, and high mosaicism rate**.

  (i) Development of an optimized mutation screening approach based on t-NGS and orthogonal techniques for a faster and exhaustive molecular characterization of FA patients, independently from their complementation group, inheritance pattern, variant type, and mosaicism condition.

- **Need for new systemic, and less noxious drug-based therapies, focusing on solid tumors prevention, delay and/or treatment.**

  (ii) Generation and validation of cellular models expressing three frequent FANCA NT mutants and their phenotypic correction in a HCS for drug repurposing.

  (iii) Performance of pooled GW CRISPRko screens on *FANCA*-deficient and control cellular systems to dissect novel SIs, and, thus potential druggable targets to cure FA clinical features (SV) and sporadic cancers involving mutations in FA genes or their interactors (SL).

3. MATERIALS AND METHODS

3.1 FA MOLECULAR DIAGNOSTIC STRATEGY

3.1.1 PATIENTS

All the patients assessed were provided by the Italian Association of Pediatric Hematology and Oncology (AIEOP) centers, and, according to the Declaration of Helsinki, their family members or legal guardians unanimously gave informed consent to the trials. The research and experimental protocols were endorsed by the competent Ethics Review Boards. DEB-induced chromosome fragility tests, showing positive results and thus indicating a suspected diagnosis of FA, were carried out at Gaslini Children's Hospital of Genoa.

3.1.2 CELL CULTURES

Probands' lymphoblastoid cell lines (LCLs) were generated from primary lymphocytes isolated from PB [28]. LCLs and HEK293T cells were grown in Roswell Park Memorial Institute (RPMI) 1640 medium 1X (L-Glutamine) and Dulbecco's Modified Eagle Medium (DMEM) 1X (Euroclone), respectively, with 10% Fetal Bovine Serum (FBS) and 1% penicillin- streptomycin at 37°C in a humidified atmosphere containing 5% carbonic anhydride (CO_2). Following DNA, RNA or protein extraction, cell cultures were transferred to RPMI or DMEM medium with 10% dimethyl sulfoxide (DMSO) and then frozen in liquid nitrogen for the storage.

3.1.3 DNA EXTRACTION, MUTATION SCREEENING AND VARIANTS ANALYSIS

Genomic DNA (gDNA) was extracted from patients' PB, BM and/or LCLs, and cell pellets were then resuspended in 5 ml of lysis buffer (NaCl 10 nM, Tris- HCl 10 mM (pH 7.5), SDS 0.6%, EDTA 10 mM (pH 8)) to which 1 mg/ml of proteinase K (20 mg/ml) was added before the incubation of the lysed cells at 50°C overnight (O/N). The following day 5 ml of NaCl 75 mM in cold 100% ethanol (EtOH) were added, and the solution was repeatedly inverted in order to allow DNA precipitation. After performing a centrifugation at 13000 rpm for 10', the supernatant was discarded and 500 µl of 100% EtOH. The same step was repeated, but the pellet was eventually suspended in water. The quality and the quantity of the DNA samples extracted were detected through the usage of NanoDrop-1000 spectrophotometer (Celbio), while 2 µl of each sample were loaded on 1% agarose gel in order to evaluate a possible degradation.

FA genes were analyzed using Ion PGM system for next generation sequencing (Life Technologies) [66]. *FANCA* deletion was identified analyzing Ion PGM row data [98]. Sanger sequencing PCR was carried out by Kapa 2G Fast Hot Start ReadyMix (Kapa Biosystems), and the following thermal cycle:

- Polymerase activation at 95°C for 1';
- 30 amplification cycles, further divided into:
  - ⇒ Denaturation at 95°C for 10'';
  - ⇒ Annealing at 62°C for 15'';
  - ⇒ Extension at 72°C for 4'';
- Final extension at 72°C for 30''.

Both the presence and the proper length of the amplified products were checked through agarose gel electrophoresis (Seakem). The specific primers (28F - 28R) employed for the amplification of *FANCA* variants mapping on exon 28 are reported in Table 2.

After PCR performance, DNA fragments were purified using 1.5 µl of ExoSAP-IT (Applied Biosystems) and 4.5 µl of amplified products, and the final mixture was brought to 37°C for 15' and then 85°C for 15' for enzyme activation and inhibition, respectively. Labelling procedure was thus performed by incorporating the purified products in a mixture containing 3 µl of DNA, 0.5 µl of Big Dye, 2 µl of Big Dye Buffer, 2 µl of primer forward or reverse (2 µM), specific to each fragment, and 3 µl of water for a total volume of 10.5 µl. Subsequently, the sequencing reactions were performed as follows:

- 96°C for 1';
- 26 amplification cycles, including:
  - ⇒ 96°C for 10'';
  - ⇒ 50°C for 5'';
  - ⇒ 60°C for 4''.

Labelled products were eventually purified through the usage of MicroSpin G-50 Columns (GE Healthcare) and electrophoretic run was carried out in ABI PRISM sequencer (Applied Biosystem). The analysis of the sequences deriving from direct sequencing was performed by using SeqMan application of DNASTAR's Lasergene package. Nucleotide numbering reflects *FANCA* cDNA with +1 corresponding to the A of the ATG translation initiation codon within the reference sequence (RefSeq NM_000135).

The discrimination between WT and c.2778+83C>G/ c.2778+86insT mutant allele was pursued by designing specific primers (28F_ex-28R_wt and 28F_ex-28R_mut, respectively, with 28R_mut encompassing two mismatches to anneal to the variants targeted) (Table 2) and adopting the ensuing touchdown PCR protocol:

- 96°C for 3';
- 10 amplification cycles, such as:
  - ⇒ 95°C for 15'';
  - ⇒ 68°C for 10''(reducing 0.5°C/cycle);
  - ⇒ 72°C for 1';
- 28 amplification cycles, including:
  - ⇒ 95°C for 15'';
  - ⇒ 63°C for 15'';

⇒ 72°C for 1’.

The variants identified were eventually searched in annotation databases: Single Nucleotide Polymorphism Database (dbSNP) [268]; Genome Aggregation Database (gnomAD) [269], Human Gene Mutation Database (HGMD) [94]. The in silico analyses of splicing mutations were conducted by Splice Site Prediction by Neural Network (NNSplice) [270].

3.1.4 RNA EXTRACTION AND ANALYSIS

Total RNA was extracted from patient’s PB, BM and/or LCLs by using PAXgene Blood RNA Kit IVD (QIAGEN) and High Pure RNA Isolation Kit (Roche), respectively. PB and BM samples, collected in PAXgene Blood RNA Tubes, were initially stored at -20°C, brought at 4°C the day before performing the extraction and left at room temperature (RT) for 3h before the procedure in order to increase the final yield. Both kits were employed according to manufacturer’s instructions. The quality and the quantity of the RNA samples extracted were checked as described at 3.1.3 section.

Complementary DNA (cDNA) was synthetized from 1 µg of RNA using Transcriptor First Strand cDNA Synthesis Kit (Roche), according to manufacturer’s instructions. For the RT-PCRs, prepared with KAPA2G FastHotStart (Kapa Biosystems), both oligo dT (2.5 pmol/µl) and random hexamer primers (60 pmol/µl) were employed, and the ensuing thermal cycle was applied:

- 25°C for 5’;
- 42°C for 60’;
- 70°C for 15’.

cDNA products were amplified following the procedure reported above (number of amplification cycles varying between 26 and 30 to ensure a better control of the amplicons production) and their presence and length were checked through agarose gel electrophoresis (Seakem). RT-PCR primers to verify the deletion from exon 11 to 14 (c.894_1359del) and assess c.2778+83C>G effect (9F-16R and 27F-30R, respectively) are reported in Table 2. After amplification, cDNA fragments were purified, labelled, and sequenced as explained at 3.1.3 section. The same procedure was also adopted for the RT-PCR amplicon extracted from 3% agarose gel by using QIAuick Gel Extraction Kit (QIAGEN) in line with manufacturer’s instructions. To increase the performance and purify the extraction, the products obtained were further subjected to classical PCR with the same primer pair used for RT-PCR.

3.1.4.1 MULTIPLEX LIGATION-DEPENDENT PROBE AMPLIFICATION ASSAY

SALSA MLPA P031/P032-A2 FANCA kit (MRC-Holland) was employed to perform the analysis. For each reaction, 100 ng of DNA were used and, according to the kit protocol, samples were subjected to an initial denaturation at 98°C for 5’ and then to a strict thermal

cycle. 1.5 µl of SALSA probe mix and 1.5 µl of MLPA buffer were added at a thermocycler temperature of 25°C, followed by a step at 54°C for 1'. Samples were then maintained at 60°C for 16h in order to allow probe hybridization. Ligation was achieved by lowering the temperature at 54°C and adding 32 µl of ligase-65 mix, followed by incubation at 54°C for 15' and heating at 98°C for 5'. A PCR reaction represented the final step, which was performed at a thermocycler temperature of 60°C by adding 10 µl of ligation products, 4 µl of SALSA PCR buffer and 26 µl of water (final volume of 30 µl). The reaction included 30 cycles, further divided into: DNA denaturation at 95°C for 30'', annealing of the universal primer at 60°C for 10'' and extension at 60°C for 30''. Eventually, the completion of the extension step was executed at 72°C for 20'. Samples were subjected to capillary electrophoresis with AB3130 sequencer, and the data generated were imported into GeneMapper v.4 software for both the analysis of the results and the statistical analysis.

3.1.4.2 PLASMID CONSTRUCTION AND MINIGENE ASSAY

A 533 bp cassette that contains *FANCA* exon 28 with portions of intron 27 and 28 was created by amplification of the cDNA obtained from patients' LCLs in order to obtain all the investigated haplotypes. Amplification primers (clonF-clonR, Table 2) were designed with a NdeI restriction site (underlined) at the end in order to allow the cloning into the α-globin minigene under the control of the α-globin promoter and the SV40 enhancer [92]. HEK293 cells, seeded in 6-well plate, were transfected with 3 ug of each minigene plasmid using Lipofectamine 3000 (Thermo Fisher Scientific). For the analysis of the splicing isoforms the primers used (α2F- β2R) are reported in Table 2.

Table 2. Sequence of the primers used in the FA molecular diagnostic strategy. Oligonucleotides for Sanger sequencing PCR, detection and differentiation of c.2778+83C>G/c.2778+86insT and WT alleles on gDNA (28F_ex-28R_wt and 28F_ex-28R_mut), and RT-PCR analysis of *FANCA* variants (9F-16R; 27F-30R), amplification of the minigenes carrying patients' haplotypes (clonF-clonR), and analysis of the related splicing isoforms on cDNA (α2F- β2R).

Name	Sequence (5'-3')
gDNA	
28F	GTTGATGGTCTGTTTCCACC
28R	ACCCTAGACTCGAGACGA
28F_ex	AGACCCTTGCACCTTCCTTC
28R_wt	GAAGGAACGGTCACCTACG
28R_mut	GAAGGAACGGTCACCTAACC
cDNA	
9F	TTGATGTACTGCAGAGAATGC
16R	GTGTCTTGGCCAATGAGATG
27F	AGCTGCTTATCTCCAGGCC
30R	GGAAATCCATCAGTGCGTTG
clonF	ATCATCCATATGGAATGTGGGGTTGTG
ClonR	TGTTGTCATATGGTCAAGATTCCAATC
α2F	CTTCAAGCTCCTAAGCCACTG
β2R	CACCAGGAAGTTGGTTAAATC

3.1.5 IMMUNOBLOTTING: PROTEIN EXTRACTION, SEPARETION, TRANSFER AND IMMUNODETECTION

For immunoblotting, 4×10^5 cells were seeded on 6 wells plates in duplicates and 24h later treated with 2mM HU or left untreated. After 18h, protein whole cell extracts were lysed using 200 µl 1x RIPA buffer (1 ml RIPA lysis buffer (10x, Millipore), 2 ml cOmplete ULTRA tablets protease inhibitor (5x, Roche), distilled water (dH_2O) till 10 ml) with 1:1000 benzonase (VWR International). Samples were then sonicated for 15" and 25% amplitude, centrifuged at maximum speed for 5', and quantified with Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific) according to manufacturer's instructions. Absorbance was measured at 562 nm with the Thermo Scientific Multiskan Sky (Thermo Scientific) microplate reader. Based on the quantification results and the creation of a linear standard curve, the concentration of the protein lysates was calculated to load 50 mg of protein. Subsequently, samples were mixed with 50 µl 2x Laemmli buffer containing 1:10 β-mercaptoethanol, heated at 95°C for 5' and loaded in 6% polyacrylamide gels. A voltage of 200 V was applied for the time necessary for the separation, and proteins were then transferred to a nitrocellulose membrane (Mini or Midi Nitrocellulose Transfer Packs, Bio- Rad) by using Trans-Blot® TurboTM Transfer System (Bio-Rad) with a program specific to high molecular weights. Nitrocellulose membranes were blocked with TBST (1X Tris-Buffered Saline, 0.1% Tween 20) containing 5% non-fat dried milk for 1h at RT. Membranes were then incubated with primary antibody properly diluted in blocking solution (anti-FANCA (rabbit, A301-980A, Bethyl, 1:2000), anti-FANCD2 (rabbit, Abcam, ab2187; 1:2000), anti-vinculin (rabbit, Abcam, ab155120; 1:5000)), O/N at 4°C. After two rapid washes and four washes of 5' each with TBST, membranes were incubated with the appropriate secondary antibody (anti-rabbit, goat, Bethyl, A120-101P, 1:2500) for 1h at RT. After the removal of the excess secondary antibody, membranes were developed by Immobilon® Crescendo Western HRP Substrate (Merck Millipore) (FANCA) or Pierce™ ECL Western Blotting Substrate (all the other proteins) and band signals eventually detected at the GeneGome Bio Imaging System (Syngene) by using Gene Snap Software. The densitometric analysis of anti-FANCA antibody signal was performed via ImageJ software, based on the use of vinculin for data normalization and the comparison with both positive and negative controls' results.

3.1.6 MMC SURVIVAL ASSAY

Patients' LCLs were exposed to increasing concentrations of MMC (0, 3, 10, 33 ,100, 333 nM) in culture medium. After five days, cells were resuspended in phosphate-buffered saline (PBS)–bovine serum albumin (BSA; 0.05%) with 0.5 mg/ml propidium iodide (PI), incubated for 10 min at 4°C and, their survival tested by flow cytometry based on PI exclusion test [28], [95]. The assay was performed after a few of cell passages (Time 0), and repeated a month after the establishment of the cultures (Time 1) to monitor cell survival patterns overtime.

3.2 CELLULAR MODELS FOR HIGH CONTENT DRUG SCREENING IN FA

3.2.1 CELLULAR MODELS: ESTABLISHMENT

3.2.1.1 CELL CULTURES

HEK293T cells and all FA cellular systems (Table 3), created from U2OS FANCA$^{-/-}$-YFP-FANCD2 line [169], were cultivated in DMEM high glucose (Biowest) supplemented with 10% heat inactivated FBS (Biowest) and 0.01% Plasmocin (PlasmocinTM prophylactic, ant-mpt InvivoGen) at 37°C in a humified atmosphere containing 5% CO_2, and periodically stored as indicated in *Material and Methods (3.1.2).*

Table 3. List of the FA models. For each cell line generated, the establishment characteristics are reported.

Cell line	Origin and characteristics
U2OS FANCA-/- -YFP-FANCD2	U2OS line, FANCA knocked-out by TALEN gene editing (Transposagen Biopharmaceuticals), stably transfected with pEAK8-YFP-FANCD2 plasmid for yellow fluorescent (YF) FANCD2 expression
U2OS FANCAcorrected-YFP-FANCD2	U2OS FANCA -/- YFP-FANCD2, stably transduced with PGK-FANCA.WPRE lentiviral vector for wild type (WT) FANCA expression
U2OS Arg951Gln-YFP-FANCD2	U2OS FANCA -/- YFP-FANCD2, stably transduced with PGK-FANCA.WPRE lentiviral vector mutagenized to carry c.2852G>A (p.Arg951Gln) variant
U2OS Thr1131Ala-YFP-FANCD2	U2OS FANCA -/- YFP-FANCD2, stably transduced with PGK-FANCA.WPRE lentiviral vector mutagenized to carry c.3391A>G (p.Thr1131Ala) variant
U2OS Phe1263del YFP-FANCD2	U2OS FANCA -/- YFP-FANCD2, stably transduced with PGK-FANCA.WPRE lentiviral mutagenized to carry c.3788_3790delTCT (p.Phe1263del) variant

3.2.1.2 MUTANT PLASMIDS GENERATION BY SITE-DIRECTED MUTAGENESIS

To reproduce the endogenous FANCA expression levels observed in the patients, PGK-FANCA.WPRE lentiviral vector (11606 bp), already used to successfully drive the physiological expression of FANCA protein [132], [133], was employed. Apart from c.3788_3790delTCT variant, already inserted into PGK-FANCA.WPRE vector, c.2852G>A and c.3391A>G point mutations were introduced by site directed-directed mutagenesis using the QuickChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies). The specific mutagenic primers (29Fmut-29Rmut and 34Fmut-34Rmut for c.2852G>A and c.3391A>G, respectively) (Table 4) were designed according to the web based QuickChange Primer Design program (Agilent Technologies) [271]. Sample reactions were prepared using 5 µl 10^X reaction buffer, 1 µl double stranded DNA (dsDNA) template (100 ng), 1.25 µl sense and antisense oligonucleotides (10 µM), 1 µl dNTPmix, 3 µl QuiKSolution, and dH_2O up to 50 µl per a single reaction, and finally adding 1 µl *PfuUltra* HF DNA polymerase (2.5 U/µl). The ensuing protocol was applied:

- 95°C for 1';
- 18 amplification cycles, such as:

⇒ 95°C for 50’’;
⇒ 60°C for 50’’;
⇒ 68°C for 12’ (1’/kb);

- 68°C for 7’.

3.2.1.3 PLASMID TRANSFORMATION

Following temperature cycling, amplification products were treated with 10 U/µl DpnI restriction enzyme and incubated at 37°C for 1h to digest the parental nonmutated dsDNA. Subsequently, 2 µl of DpnI-treated DNA from each sample were transferred to 45 µl of XL10-Gold ultracompetent cells, incubated on ice for 30’ and heat-pulsed at 42°C for 30’’ to allow transformation. Reactions were then incubated in 0.5 ml of preheated NZY+ broth at 37°C for 1h with shaking at 220 rpm, and 500 µl of each sample eventually split and plated on two lysogeny broth (LB) agar plates containing 1:1000 ampicillin selection. Transformation plates were incubated at 37°C for >16 h.

3.2.1.4 PLASMID AMPLIFICATION, EXTRACTION AND PURIFICATION (MINI-PREPARATION)

Per each plate, 7 colonies were picked for bacteria mini-preparations (mini-preps), consisting of 5 ml LB broth with 1:1000 ampicillin incubated overnight at 37°C and 220 rpm shaking. After 14-16h, three aliquots of 500 µl were collected from each bacterial suspension, mixed with one volume of glycerol, and stored at -80°C for further use; plasmid DNA was isolated from the remaining samples and purified by using E.Z.N.A. Plasmid DNA Mini Kit I - Plasmid DNA Extraction and Purification from *E. coli* culture (Omega Bio-tek) according to manufacturer’s instruction. Plasmids were quantified at NanodropTM 1000 Spectrophotometer (Thermo Scientific, ND1000) and diluted to reach 1 µg/µl concentration. The quality of the samples was further tested by gel electrophoresis in a 0.8% agarose gel and comparing their migration patterns to that of the WT PGK-FANCA.WPRE vector.

3.2.1.5 PLASMID SEQUENCING AND PRIMER WALKING

To verify the mutagenesis efficiency and select the properly mutagenized plasmids, 5 µl of each 100 ng/µl DNA sample, mixed with 5 µl of 5 mM specifically designed primers (Table 5), were sequenced by Macrogen EZ-sequencing service. The results were analyzed with CodonCode Aligner software, reporting at least 5 out of 14 samples carrying the mutations of interest (c.2852G>A or c.3391A>G). Furthermore, the mutant plasmids and PGK-FANCA.WPRE vector (WT control) were tested for potential aberrations within FANCA insert via primer walking strategy. The primers employed (Backbone_5’ end – Backbone 3’ end) (Table 4) were designed by GenScript DNA Sequencing Primers Design Tool [272],

considering annealing sites spaced about 400 bp a part to ensure sequence overlaps among the reads generated (read lengths of 700-800 bp).

Table 4. **Sequence of the primers used to establish HCS-suitable FA models.** Oligonucleotides employed to generate c.2852G>A and c.3391A>G PGK-FANCA.WPRE vectors by site-directed mutagenesis (29Fmut-29Rmut; 34Fmut-34Rmut), and of the sense oligonucleotides used to sequence all 43 exons of *FANCA* insert and partial backbone ends by primer walking approach (Backbone_5' end- Backbone_3' end). The name of the primers is indicative of the respective annealing sites.

Name	Sequence (5'-3')
29Fmut	ACTGGTGGAAGTCCTGCTGTTCAGTATCTGAAAGA
29Rmut	TCTTTCAGATACTGAACAGCAGGACTTCCACCAGT
34Fmut	CTGAAGAAGTGGGCAGCGATGTCCTGTGTCAGG
34Rmut	CCTGACACAGGACATCGCTGCCCACTTCTTCAG
Backbone_5' end	ACAGCGCCAGGGAGCAAT
*FANCA*_ exon 2	CTCCTGCGAAGCCATCAG
*FANCA*_ exon 6	TGGCATCTTCACGTACAAGG
*FANCA*_ exon 11	GAAGAGGTTCTTCAGTCATACCC
*FANCA*_ exon 14-15	AGACTGGTTCAAGGCCTCCT
*FANCA*_ exon 19	GTGTCCCACTTCCTCCCC
*FANCA*_ exon 23-24*	ACATTGCTGTGGACCTCCTG
*FANCA*_ exon 27*	CAAGTTTTCTTCCCAGTCACG
*FANCA*_ exon 30**	CCATTCTTGTCAACGCACTG
*FANCA*_ exon 33-34**	GAGAAACTTCTGCTCCCACG
*FANCA*_ exon37	AGCTGGACTGCGAGAGAGAG
*FANCA*_ exon 41	CTGGCAGGAGCCTGGAGC
Backbone_3' end	TAGTTCAGGTGTATTGCCACAA

*, **: primers to assess the presence of c.2852G>A and c.3391A>G variants, respectively.

3.2.1.6 PLASMID AMPLIFICATION, EXTRACTION AND PURIFICATION (MAXI-PREPARATION)

To amplify c.2852G>A and c.3391A>G mutant plasmids, 500 µl of corresponding bacterial glycerol stocks were used for maxi-preparations (maxi-preps), consisting of 250 ml LB broth with 1:1000 ampicillin incubated overnight at 37°C and 220 rpm shaking. After 14-16h, plasmid DNA was isolated and purified by using NucleoBond Plasmid DNA purification (Macherey-Nagel) according to manufacturer's instruction. Plasmids were quantified and diluted to reach 1µg/µl concentration, and their quality was tested by gel electrophoresis in a 0.8% agarose gel. To reconfirm the presence of the target mutations, DNA samples were sequenced by Macrogen EZ-sequencing service using the starred primers of Table 4.

3.2.1.7 LENTIVIRAL PARTICLES PRODUCTION AND CONCENTRATION

To generate the infectious lentiviral particles, $4x10^6$ HEK293T packaging cells were seeded in a 10 cm diameter dish (3 dishes per each lentivirus to produce) and, after 24h, incubated in fresh medium supplemented with 100 mM chloroquine. Solution A of Calphos-kit infection was then preprepared employing 10 µl Transfer plasmid (10 µg), 6.5 µl psAX2 packaging plasmid (6.5 µg), 3.5 µl ENV lenti VSV G envelope plasmid (3.5 µg), 87 µl Calcium solution (Calphos), and dH_2O up to 700 µl per a single reaction. One volume of solution B was next added dropwise, and the final infection solution incubated for 15' at RT, vortexed and released dropwise onto HEK293T cultured cells. After 24h medium was changed, and lentivirus-containing supernatant was collected at 48h and 72h. Whole supernatant was then centrifuged in Millipore Ultra-15 centrifugal filter units at 4000 rpm during 30' at 6 °C. Concentrated lentiviral particles were eventually stored at -80 °C until further use.

3.2.1.8 CELL TRANSDUCTION

$2x10^5$ U2OS $FANCA^{-/-}$ -YFP-FANCD2 cells for each cellular model to establish were seeded in a well of a 6-well plate and incubated O/N. Culture medium was then changed with fresh one supplemented with 1:1000 polybrene and, after 15', 1:50 lentiviral particles were administered. Infected cells were incubated for 24h to allow proper transduction, and medium was subsequently replaced to remove the lentiviral excess and avoid potential cytotoxic effects. The cell line established were progressively expanded in culture to allow the start of the functional assays.

3.2.2 CELLULAR MODELS: FUNCTIONAL VALIDATION

3.2.2.1 CELL SORTING

Once established, the models were checked for YFP-FANCD2 fluorescence by analyzing $3x10^5$ cells in 300 µl 1x PBS with 1% FBS at the MACSQuant VYB flow cytometer (Miltenyi Biotec). $10x10^6$ cells per each system were then collected in 1 mL 1x PBS with 1%FBS, filtered and selected for fluorescence with the BD FACSAria II cell sorter (BD Bioscience). In brief, a single-cell suspension passes through a fluid system and is hit by lasers, which change the charge of the cell-containing droplets depending on the cell fluorescence. This shift is used to divert the positively charged droplets (containing fluorescent cells), from those with a negative charge (containing non fluorescent cells), allowing to collect and regrow the former ones and to discard the others. Fluorescent cells were periodically tested and sorted to ensure fluorescence maintenance over time (at least 85% fluorescent cells in the final pool).

3.2.2.2 IMMUNOBLOTTING: PROTEIN EXTRACTION, SEPERATION, TRANSFER AND IMMUNODETECTION

Total protein extracts were lysed, quantified, prepared for separation, and immunoblotting as reported in *Material and Methods (3.1.4).* Protein immunodetection was operated consistently by using the following antibodies: anti-FANCA (rabbit, A301-980A, Bethyl, 1:1000), anti-FANCD2 (rabbit, Abcam, ab2187; 1:2000), anti-vinculin (rabbit, Abcam, ab155120; 1:5000) as loading control, anti-rabbit (goat, Bethyl, A120-101P, 1:2500).

3.2.2.3 SUBCELLULAR PROTEIN FRACTIONATION

To investigate FANCA subcellular localization and capability to enter the nucleus, 8×10^5 cells were seeded on 6 wells plates, and sample preparation and extraction was performed by NE-PERTM Nuclear and Cytoplasmic Extraction Reagents (Thermo Scientific). In brief, 20 mg of dry cellular pellet were resuspended in 200 µl of CERI reagent supplemented with protease inhibitor, vortexed for 15" and incubated on ice for 10'. 5 µl of CERII reagent were then added, and samples were vortexed for 5', incubated on ice for 1', vortexed again and centrifuged for 5' at maximum speeded. Supernatant (cytoplasmic extract) was collected, while the insoluble fraction, containing nuclei, was suspended in 100 µl of NER reagent supplemented with protease inhibitor. Samples were vortexed for 15", placed on iced and vortexed for 15" every 10' for a total of 40'. After a centrifugation at maximum speed for 10', supernatant (nuclear extract) was collected. Both cytoplasmic and nuclear fractions were sonicated, quantified (CERI + CERII and NER used as blanks for cytoplasmic and nuclear samples, respectively) and prepared for separation (4-10% gradient polyacrylamide gel) and immunoblotting as reported in *Material and Methods (3.1.4).* Protein immunodetection was operated as previously described by using the antibodies indicated in *Material and Methods (3.2.2.2)* and anti-ORC2 (mouse, Abcam, ab31930; 1:2500) and anti-mouse (goat, Bethyl, A90-116P, 1:2500). FANCA nuclear and cytoplasmic fractions were quantified by using ImageJ software and operating an intrasample normalization based on the quantification of both nuclear (ORC2) and cytoplasmic (GAPDH) loading controls from the same specimen.

3.2.2.4 CELL SURVIVAL BY SULFORHODAMINE B (SRB) COLORIMETRIC ASSAY

Survival assays were performed following an adaptation of the protocol established by Vichai, & Kirtikara [273]. Cells were exposed to four DEB concentrations (50, 100, 150, 200 ng/ml) for a 120h period (t=144h), and a 24h time-point post-seeding was performed for normalization purposes (t=24h). For t=144h, $1x10^3$ cells were seeded in 96 well plates establishing 6 repetitions for each DEB dose to be tested, treated 24h later, incubated for 120 h, and then fixed with 10% trichloroacetic acid (TCA) solution overnight at 4°C. For t=24h, 1x103 cells were seeded in a 96 well plate in 10 repetitions and, 24h after seeding,

left untreated and fixed with TCA solution. Subsequently, t=144h and t=24h plates were rinsed two times with slow-running tap water and once with distilled water and let air-dry at RT. Cells were then stained with SRB solution for 1h at RT with mild shacking, rinsed three times with 1% acid acetic solution to discard unbound dye and let air-dry at RT. To allow SRB release and solubilization, samples were resuspended with 10mM Tris base solution for OD measurement at 510 nm by the Thermo Scientific Multiskan Sky (Thermo Scientific) microplate reader. GraphPad Prism tool (version 8.4.0, Dotmatics) was used for the statistical analysis of the results of four independent technical replicates. Data were expressed as means ±SEM and their parametrical distribution was assessed by Shapiro-Wilk test. The means of the samples of interest were then compared to that of the positive control (FANCA YFP-FANCD2 cell line) using two-way Anova and Dunnett's multiple comparisons tests, and considering P-values <0.05 for statistical significance.

3.2.2.5 G2/M CELL CYCLE ARREST

$2.5x10^4$ cells were seeded in 6 well plates in duplicates, 24h later treated with 50 or 100 ng/ml DEB or not treated and left in culture for 72h. Cells were trypsinized and resuspended in the staining solution (10 µl sodium citrate dehydrate (340 mM), 10 µl Triton X-100 (10%), 20 µl RNase (10mg/ml), 5 µl propidium iodide (10 mg/ml), 200 µl FBS (20%) and dH_2O up to 1 ml. Samples were eventually analyzed at MACSQuant VYB flow cytometer (Miltenyi Biotec), acquiring 15,000 live event cells based on the DNA content. The statical analysis of the results was performed with GraphPad Prism tool (version 8.4.0, Dotmatics) as reported above, considering three independent technical replicates.

3.2.2.6 FLUORESCENT FOCI ANALYSIS BY CONFOCAL MICROSCOPY

$2×10^5$ cells were plated in duplicate in coverslips in 6 wells plates, and 24h later treated with 2 mM HU or left untreated. After 24h, samples were washed with PBS, fixed with 4% paraformaldehyde (PFA) in PBS for 15', permeabilized with 0.5% TX-100 solution in PBS for 10 min, stained with 4'-6'-diamidino-2-phenylindole (DAPI), and foci imaged at confocal fluorescent microscope.

3.3 GENOME-WIDE CRISPR KNOCKOUT SCREENS FOR SYNTHETIC INTERACTIONS WITH FA DEFICIENCY

3.3.1 CELL MODELS CREATION AND MAINTENANCE

h-TERT RPE-1 $p53^{-/-}$ line, kindly provided by Prof. Daniel Durocher (Lunenfeld-Tanenbaum Research Institute), was engineered for Cas9 expression by using lentiCas9-Blast plasmid (#52962, Addgene) [274], and then subjected to blasticidin selection to keep only the properly integrating clones. h-TERT RPE-1 $Cas9^{+/+}p53^{-/-}$ cells were subsequently infected with lentiviral particles generated with the transfer plasmids of Table 4 to establish RPE-1 $NT^{-/-}$ green fluorescent protein (GFP) and RPE-1 $FANCA^{-/-}$ GFP models. Transduction efficiency was assessed by checking GFP-signal at the MACSQuant VYB flow cytometer (Miltenyi Biotec), and positive clones were sorted for fluorescence with the BD FACSAria II cell sorter (BD Bioscience) and regrown. Gene KOs production was further proved by Sanger sequencing, and FANCA loss, and thus Cas9 on-target activity via immunoblotting. Cells were then maintained in DMEM (Thermo Fisher Scientific) supplemented with 10% heat inactivated FBS (Thermo Fisher Scientific) and 1% penicillin-streptomycin (Sigma-Aldrich) at 37°C with 5% CO_2, and periodically stored as indicated in *Material and Methods (3.1.2).*

Table 4. Transfer plasmids used to produce the lentiviral particles for KO models establishment. Both vectors were generated by mutagenizing pLKO5.sgRNA.EFS.GFP (#57822, Addgene) [275]. pLKO5.NT.GFP and pLKO5.FANCA.GFP were exploited to establish RPE-1 NT-/- GFP and RPE-1 FANCA-/- GFP systems, respectively.

Plasmid	Generation and characteristics
pLKO5.NT.GFP	pLKO5.sgRNA.EFS.GFP with sgRNA spacer substituted by non-human gRNA target , GFP reporter
pLKO5.FANCA.GFP	pLKO5.sgRNA.EFS.GFP with sgRNA spacer substituted by gRNA for *fanca* target, GFP reporter

sgRNA: single guide RNA, deriving from the fusion of the trans-activating CRISPR RNA (tracrRNA) and CRISPR RNA (crRNA) duplex within a single molecule; EFS: elongation factor 1alpha binding sequence.

3.3.2 LIBRARIES GENERATION AND AMPLIFICATION, AND EVALUTION OF TRANSFORMATION EFFICIENCY

To perform the screens, we employed two lentiviral GW KO pooled libraries of gRNAs; the commercially available Toronto KnockOut v3 (TKOv3) library [276], and a collection (from now on in-house library) just fine-tuned at Prof. Manuel Kaulich's laboratory (Institute of Biochemistry II, University Hospital Frankfurt Goethe University), created following the protocol of the "covalently-closed-circular synthesized (3Cs) gRNA library generation technology" by Wegner *et al*. [277], [278].

Both libraries were amplified electroporating 6 µg of DNA into 400 µl electrocompetent *E. coli (*10-beta, New England Biolabs, C3020K) with a Bio-Rad Gene Pulser (resistance 200 Ω, capacity 25 F, voltage 2.5 kV*).* Cells were immediately rescued in 25 ml of pre-warmed super optimal broth with catabolite repression (SOC) medium, incubated for 30' at 37°C and 200 rpm, and eventually transferred into 400 ml of LB medium with 100 µg/ml ampicillin and shaken at 37°C O/N. To maintain a library representation of at least 1,000-fold, the number of transformants was assessed by making serial dilutions of electroporated bacteria (from 10^{-1} to 10^{-8}) in PB, then plated in triplicates onto LB agar plates with 100 µg/ml ampicillin, and incubated at 37°C for >16 h. The next day, the transformation efficiency observed was calculated as the product of the number of colonies/ml by the corresponding coefficient of dilution and the final volume of electroporated bacteria suspension. To keep library diversity, the transformation efficiency had to be at least 100-fold higher than the library complexity.

3.3.3 LENTIVIRAL PARTICLES PRODUCTION AND TITER

24h before transfection, HEK293T cells, grown in 10 cm dishes and at about 90-100% confluence, were split 1:2 in 15 cm dishes in order to prepare at least 4 dishes/library. To treat them, transfection medium containing 2 ml cultured volume Opti-MEM I (Thermo Fisher Scientific), 210 µl GeneJuice Transfection reagent (Merck), 54 µl pPAX (500ng/µl), 20 µl pMD2 (500 ng/µl) and 33 µg DNA (TKOv3 or in-house library plasmid) was prepared per each 15 cm dish. The mixture was then incubated for 30' at RT and added dropwise to the cells, with no medium switching. Lentiviral supernatant was harvested after 48h and stored at -80°C.
To determine the lentiviral titer, RPE1 puromycin-sensitive cells were seeded in a 6 well plate with 50,000 cells/well. After 24h, cells were transduced with 8 µg/ml polybrene (Sigma-Aldrich) and a series of viral supernatant dilutions (from 10^{2} to 10^{-7} and an untreated well, as control) to provide puromycin resistance in case of proper infection. After 72h, cells were subjected to 2.5 µg/ml puromycin selection for a duration of 2 weeks, and then the colonies established counted per viral dilution. The final lentiviral titer was established based on the number of colonies in the highest dilution normalized for the volume used to grow the cells.

3.3.4 CELLULAR MODELS TRANSDUCTION AND SELECTION

Prior to start the transduction, the number of lentiviral particles needed per each screen was calculated as the ratio of the library diversity by the desired coverage, and the volume containing them as the ratio of the number of lentiviral particles by the titer obtained (Table 5). Similarly, the number of cells to transduce at the beginning of a screen was established by dividing the number of lentiviral particles by the MOI, set at 0.5 for all experiments (Table 5). On the first day of transduction, RPE-1 $NT^{-/-}$ GFP and RPE-1 $FANCA^{-/-}$ GFP cells were mixed with polybrene (final concentration of 8 µg/ml), and the volume of lentiviral supernatant was next added. Subsequently, the suspension was split into a number of total

flasks consistent with that of the cells/flask desired (Table 5). After 1-2 days and 3-5 days two puromycin selection steps were performed.

Table 5. Conditions adopted for the transduction of the cell models. The table reports the precise values of each parameter considered to perform the screen based on the usage of the TKOv3 library (middle) or the in-house one (left) and calculated as described above.

	TKOv3 library screen	In-house library screen
Library size	~80,000	~85,000
Coverage	250	100
MOI	0.5	0.5
Total cell number	40x10^6	18x10^6
Cells/Flask (T175)	2x10^6	1x10^6
Number of flasks	20	18
Virus titer (particles/ml)	0.5x10^7	4.5x10^6
Total viral particles needed	20x10^6	9x10^6
Virus volume (ml)	4	20

3.3.5 REPLICATES ESTABLISHMENT, AND CELL DOUBLINGS AND LIBRARY REPRESENTATION EVALUATION

After selection, RPE-1 $NT^{-/-}$ GFP and RPE-1 $FANCA^{-/-}$ GFP cells were left to grow and, at 80-90% confluence (day 6-7 from transduction), they were collected from all the flasks, mixed together, and counted to determine the number of cell doublings and the respective library representation. These two parameters were calculated as the logarithmic ratio of the number of cells harvested by that of cells infected, and the ratio of the number of cells harvested by the size of the library, respectively. Thereafter, cells were reseeded to establish two biological replicates per line (each made up of 9 and 10 flasks for the in-house and TKOv3 screen, respectively) to keep and treat separately. The same procedure was repeated twice at 80-90% confluence of the cells of the biological replicates (day 12-14 and 17-21 from transduction).

3.3.6 CELLS HARVERSTING, GENOMIC DNA ISOLATION AND PREPERATION FOR SEQUENCING

At 10 cell doubling (day 17-21 from transduction), considered enough to obtain a significant phenotypic fold change, cells were harvested, washed with PBS, and spun down. To isolate gDNA via the desalting method, pellets were resuspended in 12 ml of TEX buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA pH 7.9, 0.5% SDS), and 1 ml of RNase (20 mg/ml) and the incubated O/N at 37°C at constant shaking. Subsequently, 1 ml of proteinase K was added, and samples were incubated as previously done. The following day, 4 ml of NaCl (5M) were added to the cells, yet completely degraded, and the mixture was shacked vigorously and incubated at 4°C for 1h. Samples were then centrifuged at maximum speed for 1h at 4°C, the supernatant was collected, supplemented with 24 ml of ice cold 100% ethanol, vortexed

shortly and incubated at -20°C O/N. Specimens were thereafter centrifuged at maximum speed for 1h at 4°C, supernatant was removed, and 14 ml of ice cold 70% ethanol were added to precipitate DNA, mixing well by inversion. After an additional centrifugation step at maximum speed for 1h at 4°C, supernatant was discarded, and the gDNA-containing pellet was dried at RT and subsequently dissolved in 1-5 ml of sterile H_2O, depending on its size. gDNA was quantified and stored at -20°C until util use.

To prepare samples for sequencing and maintain the coverage of the libraries, the amount of gDNA required was calculated based on the number of cells transduced multiplied by of 6.6 pg (average quantity of gDNA per diploid cell) [279]. The amount of gDNA established was then split in PCR reactions, each with 2.5 µg of gDNA (recommended template/PCR reaction limit), 2.5 µl PCR primers containing Illumina adaptors and barcodes (Table 6), 25 µl Next High-Fidelity 2× PCR Master Mix (New England Biolabs), and H_2O up to 50 µl of final volume. Thermal cycler parameters were set as follows:

- 98°C for 5';
- 30 amplification cycles, such as:
  - ⇒ 98°C for 50'';
  - ⇒ 65°C for 45';
  - ⇒ 72°C for 1';
- 72°C for 5'.

A test PCR was previously performed to assay protocol conditions, primers efficiency and specificity, and gDNA quality and integrity. Amplified products were then analyzed by gel electrophoresis on a 1.5% TAE/agarose gel (100 V, 45'-1h), and purified using GeneJet Gel Extraction Kit (Thermo Fisher Scientific), according to manufacturer's instructions. In brief, 600 µl of binding buffer and 5 µl of sodium acetate (3 M, Sigma-Aldrich) were added to PCR products, mixed, and applied to two purification columns and centrifuged for 3' at 460 RFC twice to maximize the final yield. After two wash steps and a centrifugation of 3' at maximum speed, gDNA samples were eluted in 50 µl pre-warmed H_2O and quantified at Qubit (Life Technologies Invitrogen). Subsequently, they were diluted according to Illumina protocol, all set to a final concentration of 15 nM in a total volume of 20 µl, and sequenced onto a NovaSeq Illumina sequencer (Azenta/Genewiz) considering 250 coverage, and $2x10^7$ read counts.

Table 6. Primers used to sequence the gDNA samples of the TKOv3 and in-house library screens. F1-F8 and pefor#17-24: pool of forward primers for all the specimens from TKOv3 and in-house library screens, respectively. pe rev#3.1-8, pe rev#4.1-8, pe rev#5.1-8, pe rev#6.1-8, pe rev#7.1-8: pool of barcode-tagged reverse primers (*), each one used for a distinct sample of the TKOv3 library screen; pe rev#7.1-8 oligonucleotides were used to sequence the library itself. R026-29: barcode-tagged reverse primers (*), each one employed for a different sample of the in-house library screen. Stagger sequences are indicated in italics (N: any nucleotide), barcodes in bold, and priming sites in lowercase.

	Name	Full sequence (5'-3')
		TKOv3 library screen
Forward primers pool	F1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT*N* TAAGGCGAtcttgtggaaaggacgaggtaccg
	F2	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT*NN* TAAGGCGAtcttgtggaaaggacgaggtaccg
	F3	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT*NNN* TAAGGCGAtcttgtggaaaggacgaggtaccg
	F4	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT*NNNN* TAAGGCGAtcttgtggaaaggacgaggtaccg
	F5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT*NNNNN* TAAGGCGAtcttgtggaaaggacgaggtaccg
	F6	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT*NNNNNN* TAAGGCGAtcttgtggaaaggacgaggtaccg
	F7	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT*NNNNNNN* TAAGGCGAtcttgtggaaaggacgaggtaccg
	F8	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT*NNNNNNNN* TAAGGCGAtcttgtggaaaggacgaggtaccg
Reverse primers pool*	pe rev#3.1	CAAGCAGAAGACGGCATACGAGAT**CATGATCG**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNNNN* ggactagccttattttaacttgctatttctagctc
	pe rev#3.2	CAAGCAGAAGACGGCATACGAGAT**CATGATCG**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNNN* ggactagccttattttaacttgctatttctagctc
	pe rev#3.3	CAAGCAGAAGACGGCATACGAGAT**CATGATCG**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNN* ggactagccttattttaacttgctatttctagctc
	pe rev#3.4	CAAGCAGAAGACGGCATACGAGAT**CATGATCG**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NN* ggactagccttattttaacttgctatttctagctc
	pe rev#3.5	CAAGCAGAAGACGGCATACGAGAT**CATGATCG**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*N* ggactagccttattttaacttgctatttctagctc
	pe rev#3.6	CAAGCAGAAGACGGCATACGAGAT**CATGATCG**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNNNNN* ggactagccttattttaacttgctatttctagctc
	pe rev#3.7	CAAGCAGAAGACGGCATACGAGAT**CATGATCG**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNNNNNN* ggactagccttattttaacttgctatttctagctc
	pe rev#3.8	CAAGCAGAAGACGGCATACGAGAT**CATGATCG**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNNNNNNN* ggactagccttattttaacttgctatttctagctc
	pe rev#4.1	CAAGCAGAAGACGGCATACGAGAT**CGTTACCA**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNNNN* ggactagccttattttaacttgctatttctagctc
	pe rev#4.2	CAAGCAGAAGACGGCATACGAGAT**CGTTACCA**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNNN* ggactagccttattttaacttgctatttctagctc
	pe rev#4.3	CAAGCAGAAGACGGCATACGAGAT**CGTTACCA**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNN* ggactagccttattttaacttgctatttctagctc
	pe rev#4.4	CAAGCAGAAGACGGCATACGAGAT**CGTTACCA**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NN* ggactagccttattttaacttgctatttctagctc
	pe rev#4.5	CAAGCAGAAGACGGCATACGAGAT**CGTTACCA**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*N* ggactagccttattttaacttgctatttctagctc
	pe rev#4.6	CAAGCAGAAGACGGCATACGAGAT**CGTTACCA**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNNNNN* ggactagccttattttaacttgctatttctagctc
	pe rev#4.7	CAAGCAGAAGACGGCATACGAGAT**CGTTACCA**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNNNNNN* ggactagccttattttaacttgctatttctagctc
	pe rev#4.8	CAAGCAGAAGACGGCATACGAGAT**CGTTACCA**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNNNNNNN* ggactagccttattttaacttgctatttctagctc
	pe rev#5.1	CAAGCAGAAGACGGCATACGAGAT**ATGCGCAG**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNNNN* ggactagccttattttaacttgctatttctagctc
	pe rev#5.2	CAAGCAGAAGACGGCATACGAGAT**ATGCGCAG**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNNN* ggactagccttattttaacttgctatttctagctc
	pe rev#5.3	CAAGCAGAAGACGGCATACGAGAT**ATGCGCAG**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNN* ggactagccttattttaacttgctatttctagctc
	pe rev#5.4	CAAGCAGAAGACGGCATACGAGAT**ATGCGCAG**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NN* ggactagccttattttaacttgctatttctagctc
	pe rev#5.5	CAAGCAGAAGACGGCATACGAGAT**ATGCGCAG**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*N* ggactagccttattttaacttgctatttctagctc
	pe rev#5.6	CAAGCAGAAGACGGCATACGAGAT**ATGCGCAG**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNNNNN* ggactagccttattttaacttgctatttctagctc
	pe rev#5.7	CAAGCAGAAGACGGCATACGAGAT**ATGCGCAG**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNNNNNN* ggactagccttattttaacttgctatttctagctc
	pe rev#5.8	CAAGCAGAAGACGGCATACGAGAT**ATGCGCAG**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNNNNNNN* ggactagccttattttaacttgctatttctagctc
	pe rev#6.1	CAAGCAGAAGACGGCATACGAGAT**TAGCGCTC**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNNNN* ggactagccttattttaacttgctatttctagctc
	pe rev#6.2	CAAGCAGAAGACGGCATACGAGAT**TAGCGCTC**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNNN* ggactagccttattttaacttgctatttctagctc
	pe rev#6.3	CAAGCAGAAGACGGCATACGAGAT**TAGCGCTC**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNN* ggactagccttattttaacttgctatttctagctc
	pe rev#6.4	CAAGCAGAAGACGGCATACGAGAT**TAGCGCTC**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NN* ggactagccttattttaacttgctatttctagctc
	pe rev#6.5	CAAGCAGAAGACGGCATACGAGAT**TAGCGCTC**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*N* ggactagccttattttaacttgctatttctagctc
	pe rev#6.6	CAAGCAGAAGACGGCATACGAGAT**TAGCGCTC**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNNNNN* ggactagccttattttaacttgctatttctagctc
	pe rev#6.7	CAAGCAGAAGACGGCATACGAGAT**TAGCGCTC**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNNNNNN* ggactagccttattttaacttgctatttctagctc
	pe rev#6.8	CAAGCAGAAGACGGCATACGAGAT**TAGCGCTC**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNNNNNNN* ggactagccttattttaacttgctatttctagctc
	pe rev#7.1	CAAGCAGAAGACGGCATACGAGAT**ACTGAGCG** GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNNNN* ggactagccttattttaacttgctatttctagctc
	pe rev#7.2	CAAGCAGAAGACGGCATACGAGAT**ACTGAGCG** GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNNN* ggactagccttattttaacttgctatttctagctc
	pe rev#7.3	CAAGCAGAAGACGGCATACGAGAT**ACTGAGCG** GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNN* ggactagccttattttaacttgctatttctagctc
	pe rev#7.4	CAAGCAGAAGACGGCATACGAGAT**ACTGAGCG** GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NN* ggactagccttattttaacttgctatttctagctc
	pe rev#7.5	CAAGCAGAAGACGGCATACGAGAT**ACTGAGCG** GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*N* ggactagccttattttaacttgctatttctagctc
	pe rev#7.6	CAAGCAGAAGACGGCATACGAGAT**ACTGAGCG** GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNNNNN* ggactagccttattttaacttgctatttctagctc
	pe rev#7.7	CAAGCAGAAGACGGCATACGAGAT**ACTGAGCG** GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNNNNNN* ggactagccttattttaacttgctatttctagctc
	pe rev#7.8	CAAGCAGAAGACGGCATACGAGAT**ACTGAGCG** GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*NNNNNNNN* ggactagccttattttaacttgctatttctagctc
		In house-library screen
Forward primers pool	pe for#17	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT*N* tgcgccgcttgggtacctcG
	pe for#18	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT*NN* tgcgccgcttgggtacctcG
	pe for#19	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT*NNN* tgcgccgcttgggtacctcG
	pe for#20	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT*NNNN* tgcgccgcttgggtacctcG
	pe for#21	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT*NNNNN* tgcgccgcttgggtacctcG
	pe for#22	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNtgcgccgcttgggtacctcG
	pe for#23	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT*NNNNNNN* tgcgccgcttgggtacctcG
	pe for#24	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT*NNNNNNNN* tgcgccgcttgggtacctcG
Reverse primers*	R026	CAAGCAGAAGACGGCATACGAGAT**CGATCAGT**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*t* gctgtttccagcatagctcttaaac
	R027	CAAGCAGAAGACGGCATACGAGAT**TGCAGCTA**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*at* gctgtttccagcatagctcttaaac
	R028	CAAGCAGAAGACGGCATACGAGAT**TCGACGTC**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*gat* gctgtttccagcatagctcttaaac
	R029	CAAGCAGAAGACGGCATACGAGAT**AGCGATAG**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*cgat* gctgtttccagcatagctcttaaac

3.3.7 SEQUENCING DATA QUALITY CONTROL, READ COUNT TABLE CREATION, AND ENRICHMENT ANALYSIS

Bioinformatics analysis was performed along the lines of Diehl *et al*.'s work [279]. Bcl2fastq v2.19.1.403 (Illumina) was used to process and demultiplex raw sequencing data. Cutadapt 2.8 and Botwie2 2.3.0 were exploited to trim reads, and to truncate them to 20 nucleotides and align them to the corresponding gRNA library with no mismatches allowed, respectively. The cumulative distribution of the sequencing reads was plotted as a Lorenz curve and the area under the curve determined, allowing to examine the uniformity of the library distribution. The density of the read counts of each library was plotted and the 90^{th} percentile divided by the 10^{th} percentile to evaluate the library distribution skew. The quality score of or biased library screening was assessed by applying Cohen's d statistics and measuring the separation of mean logarithmic fold change (LFC) values of on-targeting sequences and sequencing targeting core essential genes, as described in Kim & Hart's work [280]. Pearson's correlation of the normalized read counts was used establish pairwise sample correlations, visualized as hierarchically clustered heat maps [281].
MAGeCK enrichment analysis were conducted with median or total normalization of read counts, discarding gRNAs having zero counts in the control samples [282]. Data were down-sampled, generating a series of read count tables containing 1-16 random gRNAs combinations per gene. The cutoff filters used to determine the statical significance of the gRNA combinations were set to false discovery rates (FDR) ≤ 10% and LFC ≤ -0.5 [283], [284].

4. RESULTS AND DISCUSSION

4.1 COMPREHENSIVE OVERVIEW OF FA MOLECULAR DIAGNOSTIC STRATEGY

Whilst observing the guidelines for FA diagnosis and clinical management [1], the present strategies for patients' molecular characterization are mostly shaped on the expertise of the single analytical laboratories and, hence, tend to differentially address the multiple obstacles mentioned in *Introduction (1.6)*. The genetic analysis, however, constitutes an integral part of the FA diagnostic path, performed before or after the chromosome breakage evaluation depending on the case considered [1].
Our algorithm for FA testing, perfected on the basis of the knowledge gained over the last five years (over 150 FA diagnosed patients within 2018-2022), is schematically illustrated in Figure 8. In particular, for the objectives of this thesis work, major attention will be given to the lower section, including targeted sequencing and in-depth analysis for the detection and validation of both FA mutant alleles (Figure 8).

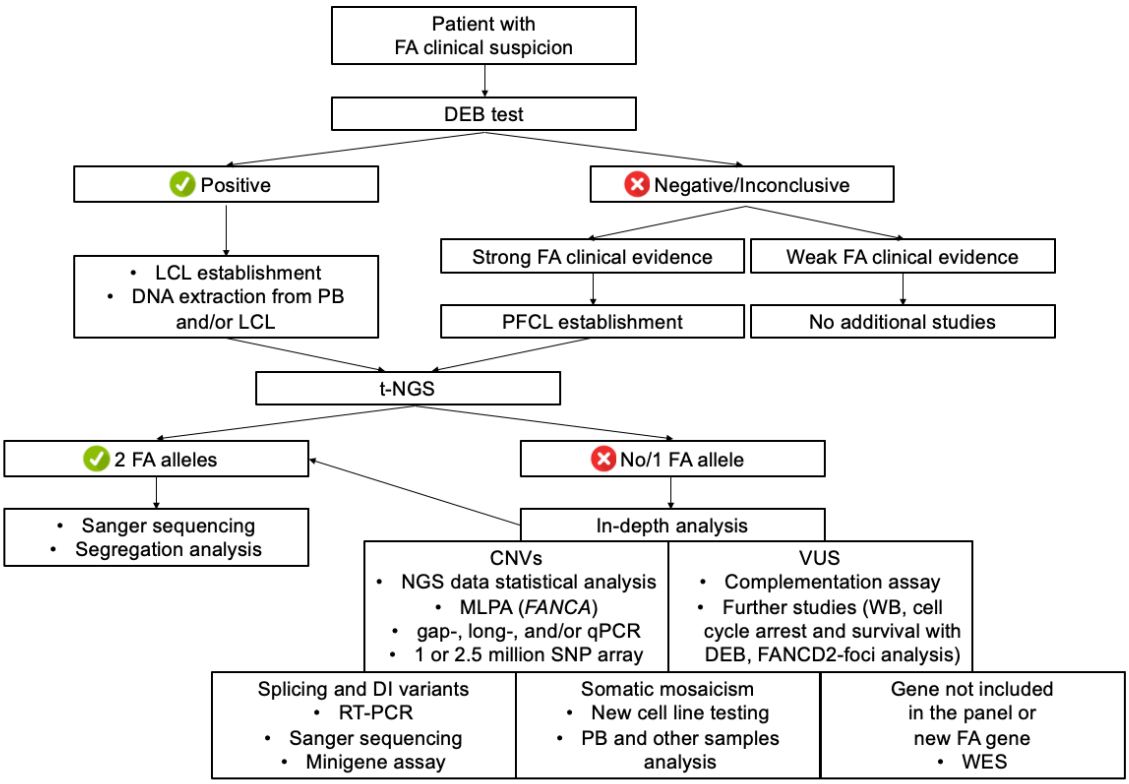


Figure 8. Schematic depiction of the FA diagnostic algorithm adopted. A DEB test is carried out on the PB cells of a patient with a clinical suspicion of FA for chromosome fragility confirmation. A negative or equivocal test response leads to the hypothesis of somatic mosaicism, followed by the establishment of a primary fibroblast cell line (PFCL) and its testing. In case of positive result, t-NGS is performed using a FA gene panel to identify the disease-causing variants, potentially validated by additional functional analysis.

The diagnostic process involves the initial recruitment of individuals with a clinical suspicion of FA, followed by the performance of a DEB test on patient's PB for evidence of chromosomal fragility (Figure 8). In case of positive result, PB DNA is extracted to keep a biological sample not subjected to reversion overtime, and, whether possible, proband's LCL is established guaranteeing a virtually unlimited supply of material (e.g., DNA, RNA, proteins) for next analysis (Figure 8). This last procedure also contains future requests of specimens from patients, as they frequently experience pancytopenia and/or become unsuitable for genetic testing after HSCT. Conversely, a negative DEB test removes the need for additional studies, unless there is strong clinical evidence of FA. In this context, or in presence of inconclusive results, PFs testing is carried out to examine the hypothesis of somatic mosaicism, using the same techniques adopted for LCL (Figure 8).
Once ascertained a chromosome breakage profile within the FA range, genetic testing is performed to determine the precise disease-causing gene and related variants (Figure 8). The mutation screening strategy described in this thesis relies upon the targeted sequencing approach by Ion Torrent platform. This NGS technology allows examining 16 out of 23 FA known genes split into 639 amplicons of 125-225 bp each, and provides 96% coverage of the coding exons sequences, together with nearly 30 bp of the flanking introns. The sensitivity and specificity ensured are respectively around 95.58% and 100% at least for *FANCA*, *FANCC* e *FANCG* genes [98]. Nevertheless, the design does not encompass regions variably dispersed through the sequences intercepted, within a range of 0.03%- 10% depending on the target, and the 6 remaining FA genes (*FANCS*/*BRCA1, FANCO/RAD51C, FANCV/MAD2L2, FANCW/RFWD3, FANCT/UBE2T, FANCU/XRCC2, FANCY/FAAP100)*. The exclusion of these latter is due to both their discovery after the design of the mutation panel, and detection in only a few families, without affecting the informativity of the analysis. The characterization of both FA mutant alleles is eventually followed by variants confirmation in the proband via Sanger sequencing, and, whether possible, segregation analysis within the family (Figure 8). Negative or incomplete molecular test results, associated with suggestive clinical phenotype and/or positive DEB test, instead lead to in-depth analyses to unveil the pathogenic variant/s as reported in Figure 8.

- CNVs characterization: gross deletions and duplications. We developed a statistical analysis to infer CNVs by easily reprocessing the row NGS data generated during the same mutation screening process. We exploited the amplicon read depth, and performed a two-step normalization (including both an intra- and inter-sample normalization) to account for sample and sequencing runs variations [98]. We eventually assumed amplicons average values lower than 0.7 and upper than 1.3 as indicative of deletions and duplications, respectively.
  We further validated the NGS outcomes by commercial MPLA kit for *FANCA,* gap- and/or long-range PCR to narrow down the breakpoints, or qPCR. Where possible, large CNVs (>5-6 exons) were additionally ascertained through single nucleotide polymorphism (SNP) array technologies based on 1 or 2.5 million SNP probes.
- Evaluation of VUS pathogenicity. We focused on assessing and/or ascertaining the deleterious effect of VUS of *FANCA*, as the most frequent mutated FA gene worldwide. We introduced VUS in WT *FANCA* cDNA by site-directed mutagenesis, and then transduced both WT and mutant lentiviral vectors in FANCA-null cells to test their

capability to complement the aberrant phenotype through more in-depth studies (western blot, cell cycle arrest and survival under DEB exposure, FANCD2-foci formation via fluorescence microscopy).

- Investigation of splicing variants impact (including DIVs). We parallelly carried out RT-PCR and Sanger sequencing on proband and WT control's RNAs, monitoring the presence of aberrant splicing profiles. In a few cases, we also established minigene systems to recapitulate the splicing patterns, and univocally determine the pathogenicity of the variant detected. Thanks to the inclusion of exon-flanking intronic regions within the panel, in our experience we successfully identified DIVs without resorting to WGS or RNA-seq.
- Somatic mosaicism detection. To rule out the reversion of the single cell line examined, new lines were established and tested at distance of time. Moreover, whether possible, we analysed the DNA from patient's PB and/or additional biological specimens to dissect the *in vivo* occurrence of potential *de novo* variants.
- Involvement of genes not included in the panel or new FA genes. In case of indecisive molecular test results, we evaluated this possibility through WES analysis performance.

4.1.1 FA PATIENTS' MOLECULAR CHARACTERIZATION

Over the realization of this thesis work, the application of our diagnostic algorithm (Figure 8) on AIEOP FA families allowed us to accomplish the molecular characterization of both FA alleles in 14 out of 14 probands (Table 7), proving the full mutation detection success of our integrated approach. Consistently with the literature [3], [66], FA-A emerged has the predominant complementation group in our case study (12/14 patients, 85.7%), followed by FA-C and FA-D2 (both 1/14 patients, 7.1%) (Table 7). Moreover, we observed a surprisingly elevated rate of compound heterozygote probands (13/14 individuals, 93.0%), suggesting a FA incidence within the population higher than that documented (Table 7).
In terms of mutation analysis, we described 23 genetically distinct alleles, such as (Table 7):

- 19 *FANCA* alleles: 5 large deletions (2 frameshift and 3 in frame), 4 splicing variants (3 frameshift and a in frame), 4 nonsense variants, 3 small deletions (2 frameshift and an in frame one), 2 missense variants, a small frameshift indel.
  Among these, we reported 12 already known FA-causing mutations [94], and 3 novel anomalies (1 small deletion (frameshift), and 2 splicing variants (frameshift) (Table 7). In most of cases we could not ascertain the previous identification of *FANCA* gross deletions, as generally annotated without their precise breakpoints [94]; nevertheless, we confirmed their considerably high rate in our records (5/19 alleles, 26.3%), in line with the literature [3], [66], [94]. Eventually, we explored further c.(1470+1_1471-1)_(1626+1_1627-1)del in FA2, FA11 (Table 2) and other two already characterized patients, unveiling common breakpoint coordinates and, thus, postulating a possible founder effect of this mutation.
- 2 *FANCC* alleles: a reported nonsense variant, and a novel frameshift splicing mutation.
- 2 *FANCD2* alleles: 2 already documented splicing variants with a frameshift effect.

Table 7. FA gene variants identified in 14 recruited families. Within the "ATG" start codon, "A" is considered as the +1 nucleotide of *FANCA*, *FANCC* and *FANCD2* genes. The reference gene sequences are reported in GenBank (FANCA: NM_000135.4; NP_000126.2. FANCC: NM_000136.2; NP_000127.2. FANCD2: NM_001018115.1; NP_001018125.1).

Patient ID	Variant							
	FA Gene	Exon/Intron	cDNA (FANCA NM_000135.2)	Protein (FANCA NP_000126.2)	Status	Segregation	Classification	Reference
FA1*	FANCA	e11-e14	c.894_1359del	p.Phe299Profs*72	Hetero	P	Large deletion (frameshift)	
		i28	c.2778+83C>G	p.His926fs*955	Hetero	M	Splicing (frameshift)	Savino et al., 1997
FA2	FANCA	e6-e14	c.(522+1_523-1)_(1359+1_1360-1)del	p.Ser175_Lys453del	Hetero	M	Large deletion (in frame)	
		e16-e17	c.(1470+1_1471-1)_(1626+1_1627-1)del	p.Val491_Glu542del	Hetero	P	Large deletion (in frame)	
FA3 - FA4	FANCA	e13	c.1115_1118del	p.Val372Alafs*42	Hetero	M	Small deletion (frameshift)	FABCC, 1996
		e32	c.3109_3137delinsG	p.Pro1037Alafs*14	Hetero	P	Small indel (frameshift)	De Rocco et al., 2014
FA5	FANCA	i29	c.2852+1dupG	p.Gln952Alafs*10	Homo	P/M	Splicing (frameshift)	Novel
FA6	FANCA	e18	c.1645C>T	p.Gln549*	Hetero	P	Nonsense	De Rocco et al., 2014
		e38	c.3788_3790del	p.Phe1263del	Hetero	M	Small deletion (in frame)	Castella et al., 2011
FA7	FANCA	e29	c.2851C>T	p.Arg951Trp	Hetero	P	Missense	Bottega et al., 2018
		e42	c.4237_4238del	p.Lys1413Glufs*11	Hetero	M	Small deletion (frameshift)	Novel
FA8	FANCA	e4-e8	c.(283+1_284-1)_(792+1_793-1)del	p.Ser96Hisfs*3	Hetero	P	Large deletion (frameshift)	
		i7	c.709+1G>A	p.250_251insGlyAlaPheMetThrArgCysGlyPheLeu	Hetero	M	Splicing (in frame)	Levran et al., 1997
FA9	FANCA	i29	c.2852+1dupG	p.Gln952Alafs*10	Hetero	P	Splicing (frameshift)	Novel
		e42	c.4198C>T	p.Arg1400Cys	Hetero	M	Missense	Levran et al., 2005
FA10	FANCA	e8	c.790C>T	p.Gly1123*	Hetero	P	Nonsense	Mattioli et al., 2014
		e34	c.3367G>T	p.Gln264*	Hetero	M	Nonsense	De Rocco et al., 2014
FA11	FANCA	e16-e17	c.(1470+1_1471-1)_(1626+1_1627-1)del	p.Val491_Glu542del	Hetero	M	Large deletion (in frame)	
		e28-e31	c.(2601+1_2602-1)_(3066+1_3067-1)del	p.Phe868_Gln1022del	Hetero	P	Large deletion (in frame)	
FA12	FANCA	i24	c.2223-138A>G	ND	Hetero	P	Splicing (frameshift)	Bouchlaka et al., 2003
		e28	c.2749C>T	p.Arg917*	Hetero	M	Nonsense	Solanki et al., 2016
FA13	FANCC	e2	c.37C>T	p.Gln13*	Hetero	NA	Nonsense	Verlander et al., 1994
		i6	c.522-1G>C	ND	Hetero	NA	Splicing (frameshift)	Novel
FA14	FANCD2	i15	c.1278+6T>C	p.Phe380_Val427del	Hetero	M	Splicing (frameshift)	Chandrasekharappa et al., 2013
		i17	c.3777+5G>A	p.His1229Phefs*22	Hetero	P	Splicing (frameshift)	Bogliolo et al., 2020

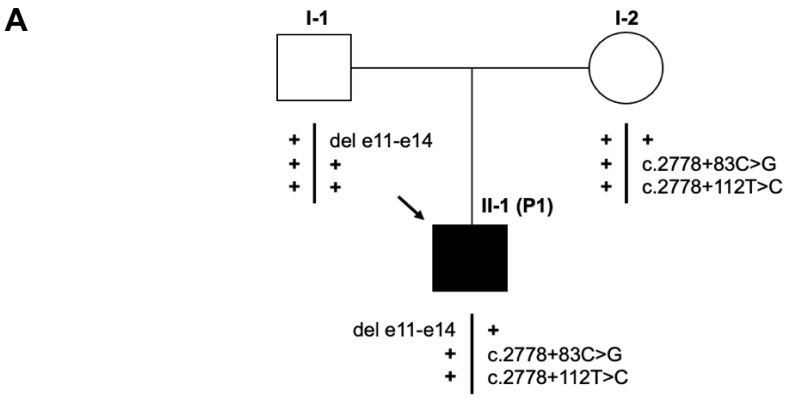
NA: not available data; *: mosaic patient; P: paternal segregation; M: maternal segregation.

4.1.2 PATIENT FA1: AN EMBLEMATIC CASE OF FA COMPLEX MOLECULAR DIAGNOSIS

In the present section we chose to delve into the case of patient FA1 (Table 7; from now on P1), who, owing to his variants (a splicing mutation and a large intragenic deletion) and mosaic status, provides a prime example of the feasibility and effectiveness of our integrated strategy for a complete, definite molecular diagnosis of FA.

4.1.2.1 T-NGS: VARIANTS DETECTION AND CNVs INFERENCE

P1 propositus (II-1, Figure 9A), the seven-year-old child of a non-related couple of Italian descent, was suspected of FA on the basis of the clinical symptoms (white blood cells $6.7x10^9$/l, neutrophils 2.5 $x10^9$/l, hemoglobin 13.3 g/dl, platelets 95 $x10^9$/l, indicative of mild thrombocytopenia) and chromosome fragility test positivity.
The early diagnostic hypothesis was molecularly explored via t-NGS on P1's PB DNA, which revealed two heterozygous variants of *FANCA* intron 28 (c.2778+83C>G and c.2778+112T>C) confirmed in the same proband and his mother (I-2, Figure 9A) by Sanger sequencing (Figure 9B). On the paternal allele (I-1, Figure 9A), the analysis of the t-NGS amplicon coverage (*Results and Discussion (4.1))* led us to postulate an intragenic deletion of *FANCA* exons 11-14 (c.894_1359del, Figure 10A), later ascertained at RNA level through MLPA (Figure 10B) and RT-PCR followed by Sanger sequencing (284bp; Figure 10C).



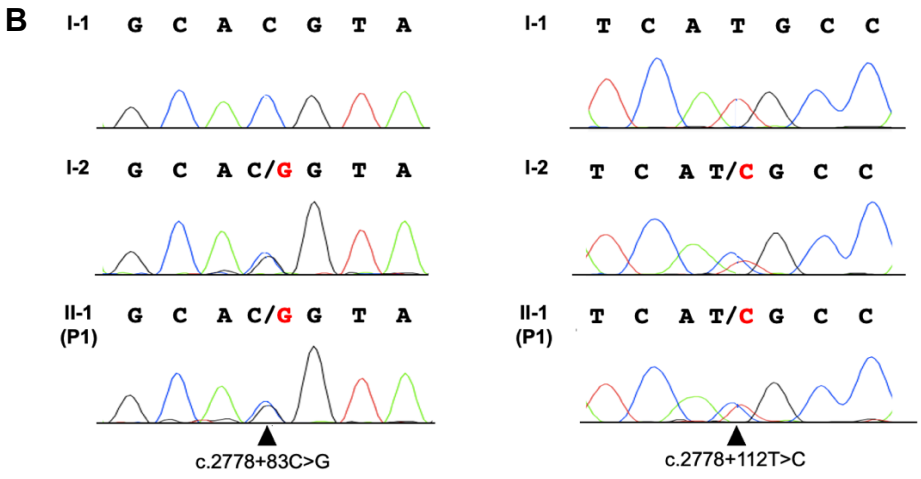
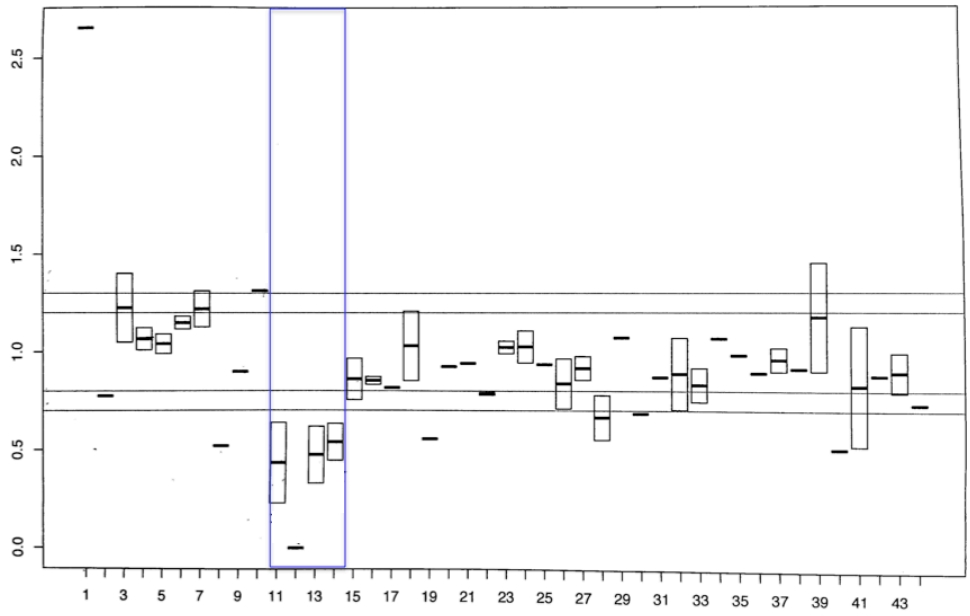


Figure 9. Identification of c.2778+83C>G and c.2778+112T>C variants and the deletion of exons 11-14 (del e11-e14) in *FANCA* gene. A) Patient's familial pedigree: the black arrow shows the proband (II-1, P1) and plus symbols (+) represent the wild type (WT) alleles. B) Electropherograms of the family members' variants confirmed by Sanger sequencing. Black arrows indicate c.2778+83C>G (left) and c.2778+112T>C (right) substitutions in heterozygous status.

A



B

SALSA mix 1

Gene-Exon	Chr.band	P1	C1	C2
FANCA-43	16q24.3	0,98	1,02	0,97
FANCA-41	16q24.3	0,88	1,01	1,07
FANCA-39	16q24.3	0,99	0,97	1,05
FANCA-37	16q24.3	1,01	0,99	0,95
FANCA-35	16q24.3	1,01	0,99	0,96
FANCA-33	16q24.3	**0**	**0**	**0**
FANCA-31	16q24.3	1,01	0,99	0,98
FANCA-29	16q24.3	1	0,95	**0**
FANCA-27	16q24.3	1	0,95	**0**
FANCA-25	16q24.3	1,01	0,98	**0**
FANCA-23	16q24.3	1	0,99	**0**
FANCA-21	16q24.3	1,03	1	**0**
FANCA-19	16q24.3	1	0,99	**0**
FANCA-17	16q24.3	1	**0,5** ?	**0**
FANCA-15	16q24.3	1,01	0,99	0,99
FANCA-13	16q24.3	**0,6** ?	1,01	1
FANCA-11	16q24.3	**0,56**?	0,99	1,01
FANCA-9	16q24.3	1,01	1,05	0,94
FANCA-7	16q24.3	1,04	0,96	0,87
FANCA-5	16q24.3	1,06	1,01	0,95
FANCA-3	16q24.3	1,13	0,94	0,96
FANCA-1	16q24.3	1	1	0,91
Reference*	02q36.1	0,97	1,02	1,03
Reference*	03q27.1	1	0,96	1
Reference*	04q31.21	0,98	1,08	1,02
Reference*	05q14.3	1,12	0,98	0,98
Reference*	08q24.22	1,02	1,01	0,99
Reference*	09q33.1	0,97	1,03	1,04
Reference*	13q14.3	0,99	1,01	1
Reference*	14q23.3	1,12	0,99	0,99
Reference*	18q21.2	0,91	1	1,03
Reference*	20p13	1,07	1	0,97
probe quality:		**OK**	**OK**	**OK**

SALSA mix 2

Gene-Exon	Chr.band	P2	C1	C2
FANCA-42	16q24.3	0,98?	1	1,13?
FANCA-40	16q24.3	0,92?	1	1,07?
FANCA-38	16q24.3	0,95?	1	1,01?
FANCA-36	16q24.3	0,97?	1	1,08?
FANCA-34	16q24.3	0,97?	1	1,19?
FANCA-32	16q24.3	1 ?	1	1,01?
FANCA-30	16q24.3	1,07?	1	1,07?
FANCA-28	16q24.3	1,09?	1	**0**
FANCA-26	16q24.3	0,98?	1	**0**
FANCA-24	16q24.3	1,11?	1	**0**
FANCA-22	16q24.3	1,15?	1	**0**
FANCA-20	16q24.3	1,07?	1	**0**
FANCA-18	16q24.3	1,05?	1	**0**
FANCA-16	16q24.3	1,12?	1	**0**
FANCA-14	16q24.3	**0,57**?	1	1,14?
FANCA-12	16q24.3	**0,56**?	1	1,03?
FANCA-10	16q24.3	1,13?	1	1,09?
FANCA-8	16q24.3	1,08?	1	0,98?
FANCA-6	16q24.3	1,11?	1	0,97?
FANCA-4	16q24.3	1,16?	1	1,01?
FANCA-2	16q24.3	1,11?	1	0,96?
GAS8-11	16q24.3	0,86?	1	1 ?
Reference*	01p22.1	0,99?	1	0,91?
Reference*	06p12.2	0,87?	1	1 ?
Reference*	07q22.1	1,03?	1	1,05?
Reference*	08q24.22	0,91?	1	0,85?
Reference*	09q22.32	1 ?	1	1,01?
Reference*	10q25.1	1 ?	1	1,01?
Reference*	12q24.31	1,12?	1	1 ?
Reference*	13q14.3	0,96?	1	0,99?
Reference*	14q11.2	0,92?	1	1,06?
Reference*	16p13.3	1,09?	1	0,92?
Reference*	17p13.1	1,14?	1	1 ?
probe quality:		**OK**	**OK**	**OK**

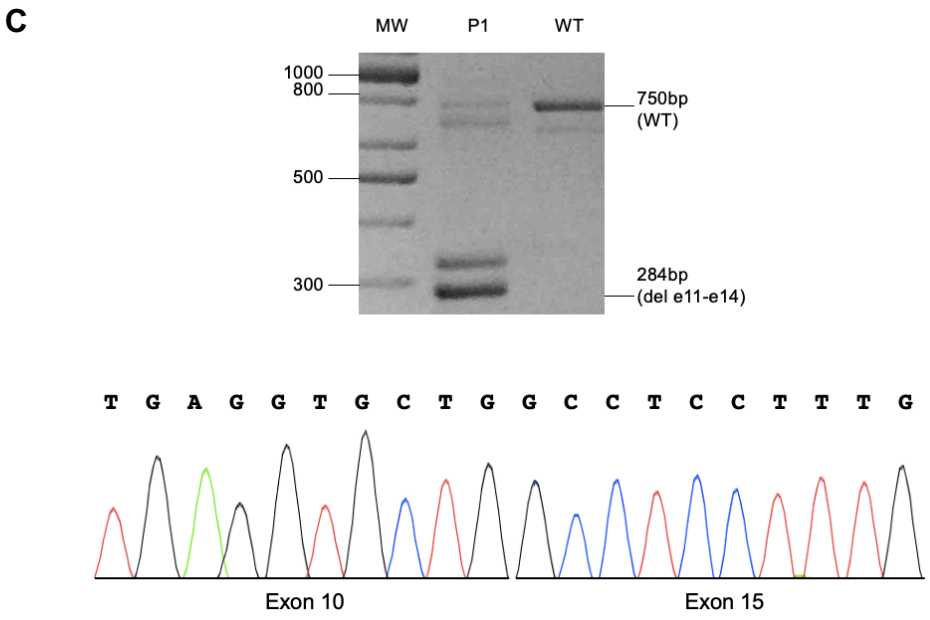


Figure 10. Prediction and characterization of exon 11-14 deletion. A) t-NGS amplicon coverage analysis: the average of the amplicons of exons 11-14 (blue box) is lower than 0.7 (disomic condition). B) MLPA assay. SALSA mix 1 and SALSA mix 2 are the mixtures of probes targeting the even and the odd exons (chromosome location beside), respectively. C1 is the negative control without deletions. while C2 the positive one carrying a homozygous deletion (involving exons from 16 to 29). C) RT-PCR performed with primers 9F and 16R on the proband (P1) and WT (above). Further to the expected fragment of 750 bp, P1 showed a 284 bp band hypothesized to be the deleted product. The other bands detected in the samples were likely to represent alternative splicing isoforms or non-specific products. Sanger sequencing analysis of 284 nucleotides band gel extraction shows exon 10 to be immediately followed by exon 15, confirming exons 11-14 deletion (below).

4.1.2.2 *FANCA* TRANSVERSIONS INVESTIGATION: BIOINFORMATICS PREDICTIONS AND C.2778+83C>G EFFECT ON RNA SPLICING

A closer examination of P1's *FANCA* transversions revealed that c.2778+83C>G was previously reported in another Italian case [64] and annotated in gnomAD ("uncertain significance", minor allele frequency (MAF): 0.000004012) and in HGMD ("disease causing"). Moreover, according to NNSplice software, it was predicted to generate a cryptic donor splice site (SS, gcagca**g**gtaggtga) at position 83 of intron 28, favoring its recognition over the canonical one (tgttcacgtaggtga) (Table 8). Conversely, c.2778+112T>C substitution, documented in gnomAD (MAF: 0.0002739) and dbSNP (rs140073727), was foretold not to exert any effect on splicing, not even in combination with c.2778+83C>G (Table 8).

Table 8. Splicing scores of c.2778+83C>G and c.2778+112T>C calculated with NNSplice. The scores are reported both for the single FANCA variants identified in intron 28, as well as combined as found in P1.

FANCA NM_000135		Possible donor site (underlined)*		Splicing site score**
FANCA variants	Wild type	Canonical	tgttcac**gt**aggtga	0,68
		Cryptic site	gcagcac**gt**aggtga	0,47
	c.2778+83C>G	Canonical	tgttcac**gt**aggtga	0,68
		Cryptic site	gcagca**g****gt**aggtga	**0,96**
	c.2778+112T>C	Canonical	tgttcac**gt**aggtga	0,68
		Cryptic site	gcagcac**gt**aggtga	0,47
	c.2778+83C>G c.2778+112T>C	Canonical	tgttcac**gt**aggtga	0,68
		Cryptic site	gcagca**g****gt**aggtga	**0,96**

*c.2778+83C>G is indicated in bold red. Canonical and cryptic site refer to c.2778+1_2778+2GT and c.2778+84_2778+85GT, respectively. **In colored are indicated the preferentially used site.

To assess c.2778+83C>G splicing impact, we thus performed a RT-PCR on the total RNA of the LCLs established from our proband (P1) and the previously reported patient (P2) [64], used as a control (Figure 11A, B). In this latter, c.2778+83C>G was already demonstrated to activate a cryptic "GT" donor SS, resulting into the synthesis of both WT mRNA and an alternative transcript with the retention of the first 83 nucleotides of intron 28 (Figure 11A). In line with the literature [64], in P2 we observed four bands, such as the WT (403bp) and mutant products (486 bp, including 83 bp of intronic retention) and their alternatively spliced forms (521 bp and 604 bps) encompassing exon 29a (118 bp) and with an uncertain biological significance (Figure 11B). In P1, instead, we noticed only two bands being the WT amplicon (403 bp) and a novel blurry product (559 bp), absent in the other samples (Figure 11B) and later identified as another *FANCA* alternative spliced form with exon 28a (156 bp) by Sanger sequencing. Of note, we could not detect any product with the expected intronic inclusion.

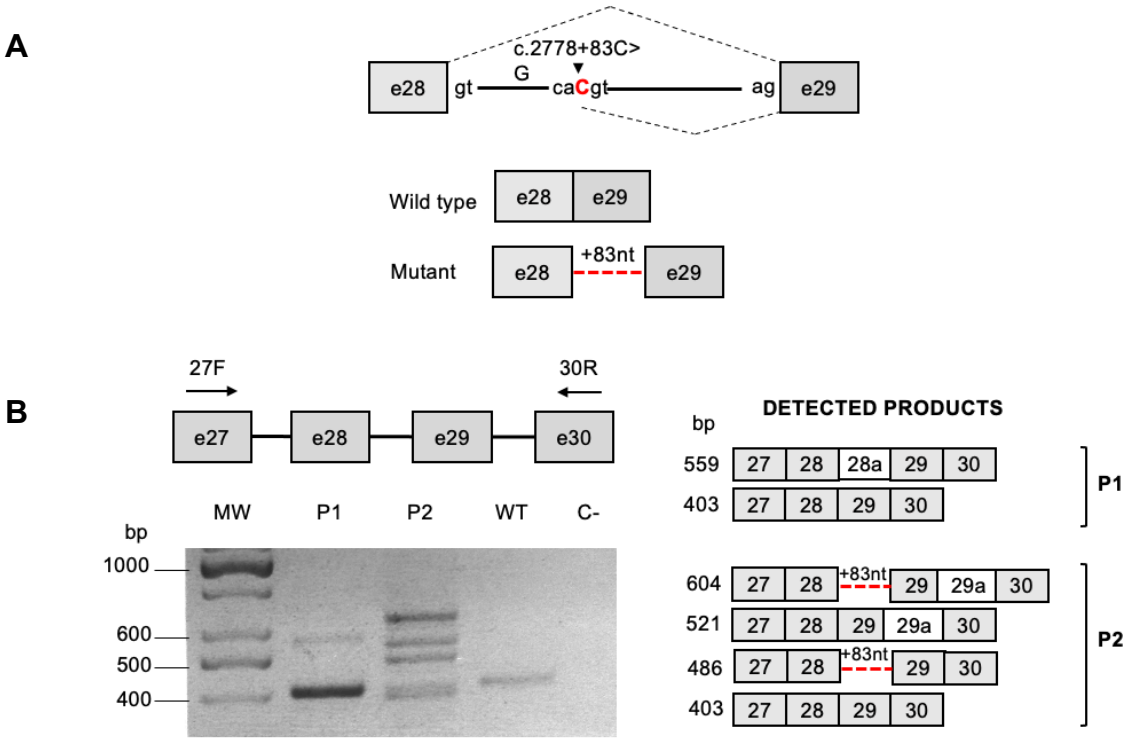


Figure 11. Analysis of c.2778+83C>G. A) Schematic depiction of c.2778C>G reported splicing effect. Grey boxes indicate exons and the continuous line the intron. Nucleotides involved in the mutations are in red. The black dotted lines represent the normal splicing pattern (above) and the aberrant effect on the process (below), respectively. B) RT-PCR carried out with primers 27F and 30R (black arrows). The schematic depiction on the right shows the expected RT-PCR products (dark grey boxes: exons; light gray boxes: alternatively spliced exons with an unknown biological effect, such as exon 28a (156 bp) between exons 28-29 or exon 29a (118bp) between exons 29-30; red dotted lines: inclusion of the first 83 nucleotides of intron 28). C: negative control.

In trying to explain the distinct RT-PCR profiles observed in patients' lines, we assumed that a mutational event could have arisen in P1's LCL restoring the physiological splicing pattern. We hence sequenced exon 28 and its flanking regions from the gDNA of both individuals' cultured cells; while in P2 we ascertained c.2778+83C>G and reported the unmentioned 2778+112T>C variant [64], in P1 we unveiled c.2778+86insT, a *de novo* thymine insertion (Figure 12), not found in the gDNA from proband's PB and BM, nor from his parents' PB. However, we could reconfirm c.2778+86insT also in another P1's LCL established two years apart (data not shown), suggesting its onset *in vivo*.

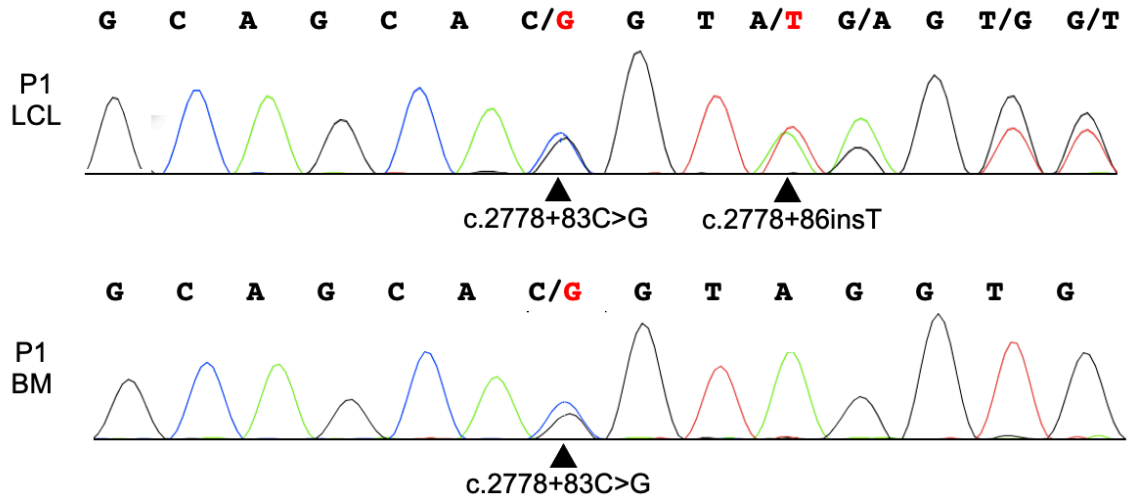


Figure 12. Validation of c.2778+83C>G on the DNA from the first LCL established (P1 LCL) (above) and BM of P1 (below). Black arrows indicate c.2778+83C>G in both lines and, the novel c.2778+86insT insertion in the only LCL.

The initial difficulties in identifying the *de novo* variant in P1's PB and BM samples via t-NGS and Sanger sequencing, respectively, should not be interpreted as an exceptional case, nor undermine the strength of our mutation detection strategy. It is indeed important to consider that patient's body cells are not exposed to the same stress conditions as the cultured cells, such as those to induce the fast selection and survival of the sole clones with a potential proliferative advantage. Moreover, the hematopoiesis stage of emergence of c.2778+86insT is another factor to contemplate; it certainly involved B lymphocytes - used to establish P1's LCLs where the mutation was first found -, but there is no evidence for the other cell types present in PB and BM, districts with limited B cell representativeness in both numerical and DNA contribution terms.

4.1.2.3 c.2778+86insT COMPENSATORY EFFECT ON SPLICING

The next step to elucidate c.2778+86insT possible involvement in P1's LCL splicing rescue and *in vivo* emergence entailed the interrogation of NNSplice software, which linked the conjunction of the variant and c.2778+83C>G to a decreased strength of the cryptic SS at the intronic position 83 (Table 9).

Table 9. Splicing scores calculated for c.2778+83C>G and c.2778+86insT with NNSplice. The scores are reported both for the single *FANCA* variants identified in intron 28, as well as combined as detected in P1. *In bold red and bold green are indicated the c.2778+83C>G and c.2778+86insT, respectively.

FANCA NM_000135		Possible donor site (underlined)*		Splicing site score**
FANCA variants	Wild type	Canonical	tgttcac**gt**aggtga	0,68
		Cryptic site	gcagcac**gt**aggtga	0,47
	c.2778+83C>G	Canonical	tgttcac**gt**aggtga	0,68
		Cryptic site	gcagca**g****gt**aggtga	**0,96**
	c.2778+86insT	WT site	tgttcac**gt**aggtga	0,68
		Cryptic site	gcagca**g****gt**t aggtga	0,78
	c.2778+83C>G c.2778+86insT	WT site	tgttcac**gt**aggtga	0,68
		Cryptic site	gcagca**g****gt**t aggtga	0,78

**In colored are indicated the preferentially used site.

Therefore, to dissect the influence of c.2778+83C>, c.2778+86insT, c.2778+112T>C and their combinations on pre-mRNA processing, we established a minigene model where we subcloned *FANCA* exon 28 and adjacent portions of intron 27 and 28, with the latter in its WT or patients' mutant forms [285] (Figure 13A). Owing to the amplification of cDNA from patients' LCLs for the creation of the *FANCA* cassette, the analysis further served as a proof of concept to validate the in *cis* arrangement of c.2778+83C>G and c.2778+86insT in P1. Minigenes of the three different haplotypes of intron 28 were transiently transfected, and the related splicing patterns evaluated by RT-PCR (Figure 13B), which unveiled the expression of two transcripts resulting from the competition between the canonical (TCACgtaggt; 507 bp) and cryptic (gcacgtaggt; 590 bp) "gt" donor SS (Table 10). GCTT WT haplotype was associated to intron 28 removal in almost 90% of cases, while GGTC mutant set (c.2778+83C> and c.2778+112T>C of P2's LCL and P1's PB) showed a preferential recognition of the cryptic SS, matched to a relevant decrease of WT transcript levels (29%) (Figure 13B, Table 10). GGTTC haplotype (c.2778+83C>, c.2778+86insT, c.2778+112T>C of P1's LCL), instead, restored an almost complete splicing of intron 28 (86%), definitively proving the compensatory effect of c.2778+86insT variant (Figure 13B, Table 10).

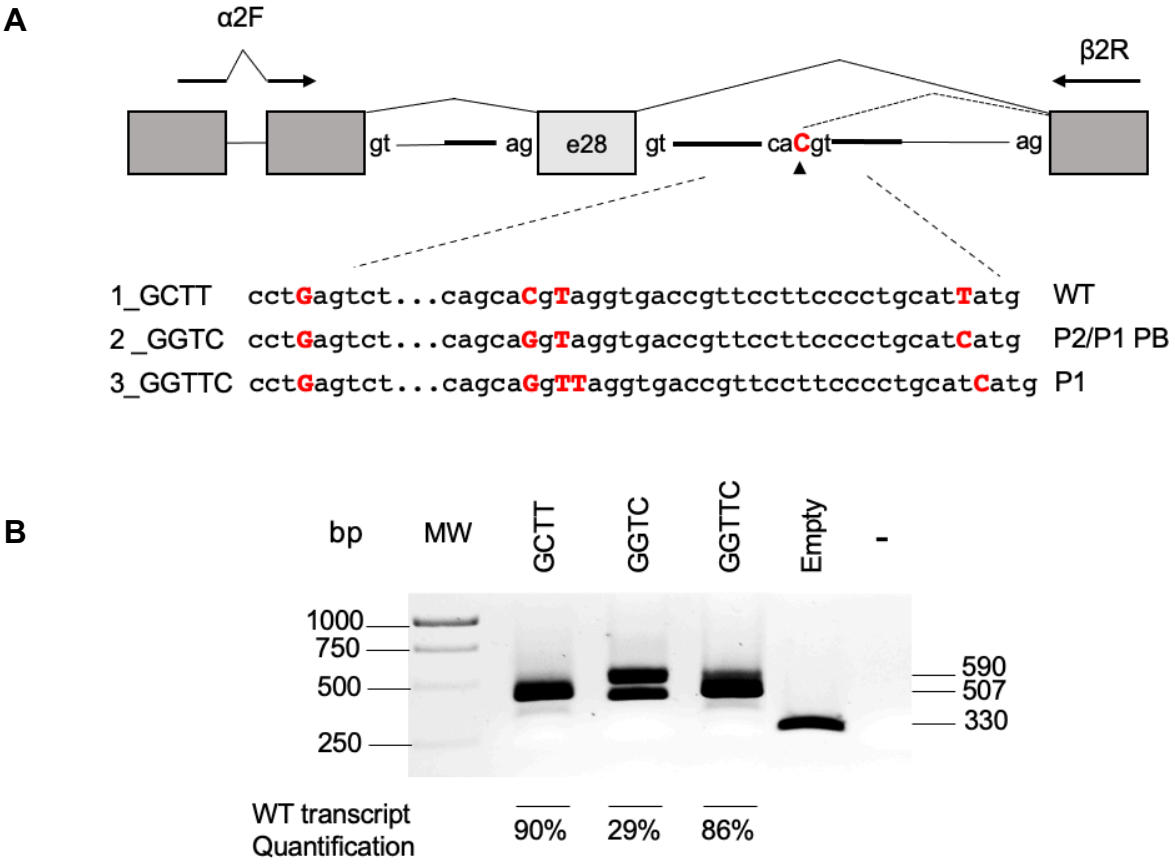


Figure 13. Schematic representation of the minigene system expressing patients' haplotypes (left) and analysis of the associated splicing patterns (right). Boxes represent exons, arrows indicate the primers used for the PCR analysis. Mutated nucleotides are indicated in red bold (left). The percentages of the WT transcript expression are reported below and calculated based on the mean±st.dev of three independent experiments. MW: molecular weight (marker).

Table 10. Splicing scores calculated for c.2778+83C>G, c.2778+86insT and c.2778+112T>C with NNSplice. The scores are reported both for the single *FANCA* variants identified in intron 28, as well as combined as detected in patients.

Aplotype ID		Patient	Intron 28 position		Possible donor site (underlined)*		Splicing site score**
FANCA haplotype***	1 GCTT	WT	+83	C	WT site	*tgttcac**gt**aggtga*	0,68
			+86	T	Cryptic site	*gcagcac**gt**aggtga*	0,47
			+112	T			
	2 GGTC	P2	+83	G	WT site	*tgttcac**gt**aggtga*	0,68
			+86	T	Cryptic site	*gcagca**g gt**aggtga*	**0,96**
			+112	C			
	3 GGTTC	P1	+83	G	WT site	*tgttcac**gt**aggtga*	0,68
			+86	TT	Cryptic site	*gcagca**g gt**t aggtga*	0,78
			+112	C			

*In bold red and bold green are indicated the c.2778+83C>G and c.2778+86insT, respectively. **In colored are indicated the preferentially used site. ***In red are indicated nucleotide changes with respect to the WT.

4.1.2.4 c.2778+86insT REVERSION OF FA CELLULAR PHENOTYPE

In light of these results, we compared P1 and P2's LCL cellular features to inspect the possible phenotypic outcomes of the compensatory c.2778+86insT variant.
Since both affected individuals were compound heterozygous for c.2773+83C>G and a large *FANCA* deletion (encompassing exons 11-14 in P1 and 11-31 in P2), only the allele with the intronic mutation was expected to contribute to the protein synthesis. We detected FANCA (163 kDa) in both patients, albeit at lower levels in comparison to the WT control; in particular, we assessed a 38.6% proteins expression in P1 against a 15.4% in P2, corroborating the role of c.2778+86insT in the restauration of the canonical donor SS activity (Figure 14A).
We then evaluated whether FANCA residual expression could provide for FANCD2 mono-ubiquitination (FANCD2-Ub) maintenance. We reported detectable levels of FANCD2-Ub (175 kDa) in both untreated and treated cells with HU in P1 but not in P2 (Figure 14A), suggesting that about 40% of FANCA was enough to preserve the mono-ubiquitination. Conversely, when FANCA expression was as reduced as in P2, only no FANCD2-Ub (166 kDa) was detected (Figure 14A).
Assuming therefore that P1's LCL was a mosaic for c.2778+86insT variant, and the clones carrying it could have some proliferative advantages, we eventually examined the proficiency of patients' lines to survive different concentrations of MMC (3-100 nM) over time (Time 0: few cell passages; Time 1: after one moth of culture) (Figure 14B). P2's cells showed a MMC sensitivity comparable to that of the FANCA-/- control at both time points; on the contrary, P1's LCL survival profile shifted from a status intermediate between the WT and FANCA-/- lines at time 0 to a completely restored one at Time 1 (Figure 14B), bringing about c.2778+86insT capability to revert FA cellular phenotype.

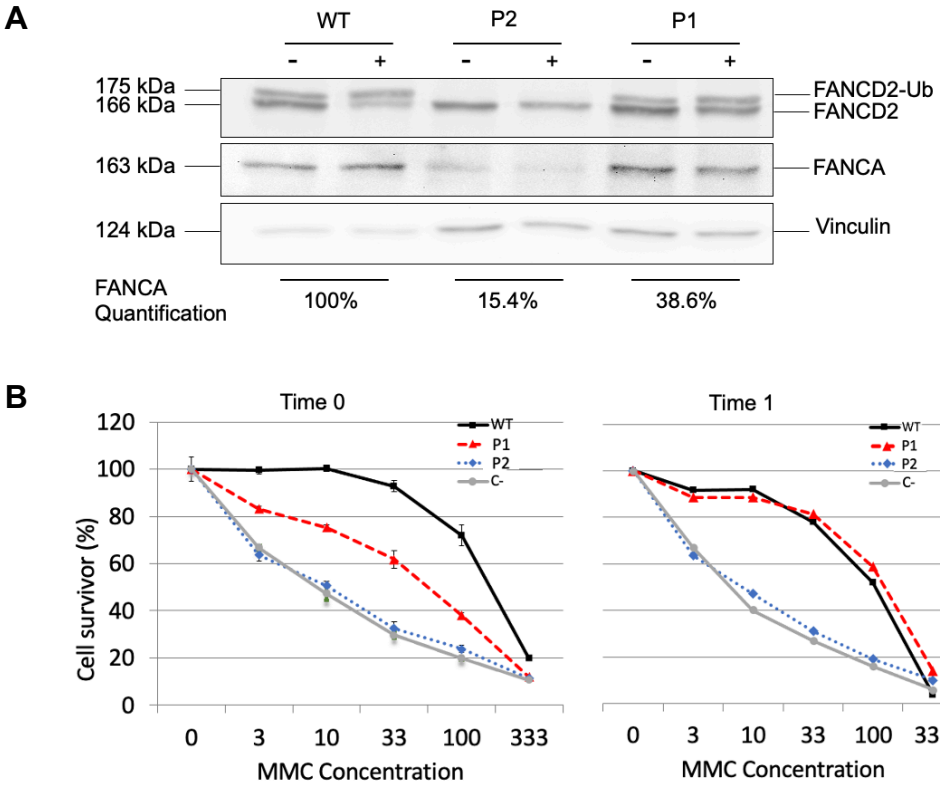


Figure 14. Effects of c.2778+86insT in P1' LCLs. A) Evaluation of FANCA expression (below) and FANCD2 mono-ubiquitination (above) in P1, P2 and WT (positive control) in absence (-) or presence (+) of 2mM HU. Vinculin was used as loading control. FANCA protein quantification was calculated based on the average of the values obtain from three independent technical replicates, comparison with the positive control, and Vinculin for data normalization. B) MMC resistance of LCLs from P1, P2 and WT after a few of cell passages (Time 0) and after one month of culture under stress condition (Time 1). C- represents a FANCA-/- patient used as negative control. Turning to Time 0, three experimental replications were carried out and the standard deviation is reported.

In line with the ameliorated cellular features, five years after diagnosis P1 only displayed a moderate hematopoietic deterioration (white blood cells 4.6 $x10^9$/l, neutrophils 1.2 $x10^9$/l, hemoglobin 12.3 g/dl, mean corpuscular volume 103 fl, platelets 50 $x10^{9/}$/l; no BM failure, nor cancer development), despite reaching a generally critical age for FA patients. Taken together, these findings hinted at a somatic mosaicism condition, and an ongoing natural gene therapy *in vivo* with major implications for his clinical progress and management.

4.2 FA CELLULAR MODELS FOR DRUG REPOSITIONING HIGH CONTENT DRUG SCREENING

HSCT latest enhancements have allowed FA transition from a pediatric to an all-age group-affecting syndrome, exacerbating adult patients' vulnerability to solid tumor development [6], [10]. This phenomenon, along with the cancer proclivity inherent to the disorder, and the few medical options for FA oncology patients, emphasizes the demand for new pharmacological approaches to treat FA at a more systemic level.

As described in *Introduction (1.7.3.1)*, Montanuy et al. sought to respond to the need creating a *FANCA*-null cellular system to monitor FA/BRCA pathway reactivation in a HCS context, exploiting fluorescent FANCD2 foci as readout (Figure 15) [169]. Regardless of the partial results achieved, this study was essential to provide us with all the expertise required to design the concept herein illustrated; we generated novel HCS-suitable cell models with the same objective to restore FA/BRCA activity but, differently from the previous one, not by supplying FANCA absence, and instead correcting the phenotype of already known FANCA NT mutants (Figure 15). We decided to direct our attention on this specific aberration category for the ensuing reasons:

- *FANCA* represents the most frequently altered FA gene worldwide, and about one-fourth of its variants are classified as NT [3], [86];
- Almost all missense and other NT *FANCA* mutations produce stable, albeit defective proteins, some of which were proved to rescue ICL sensitivity when overexpressed [87], [88].

Considering this, we deem the insertion of another possible drug-target (i.e., stable, but defective FANCA NT mutant) as a key factor for the potential success of our cellular models in identifying new therapeutic molecules for FA/BRCA pathway recovery.

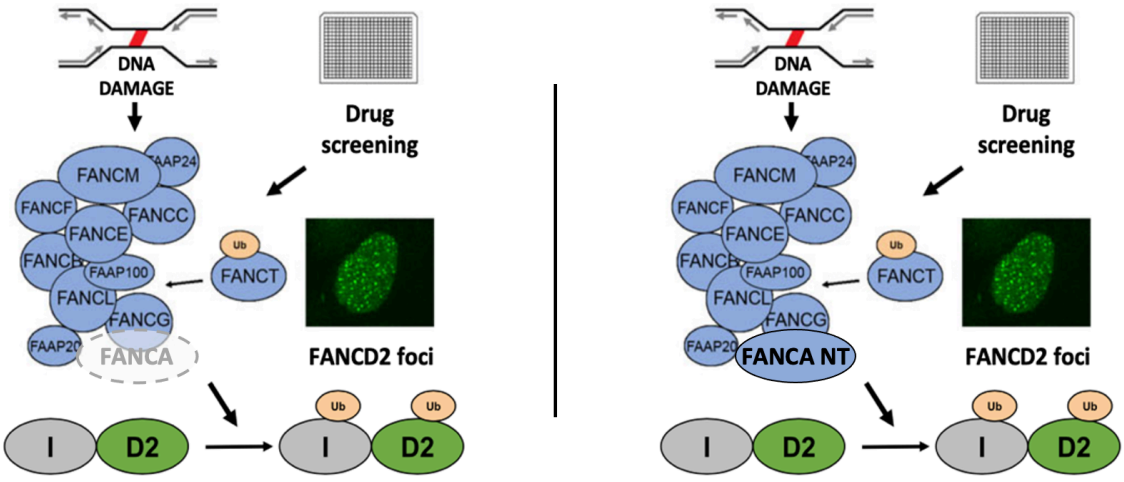


Figure 15. Comparison between the design of Montanuy *et al.* and our systems for HCS. Montanuy et al.'s attempt entailed the recognition of a drug capable to reinstate FA/BRCA pathway activity within a FANCA-null background (left), unfortunately leading to scarce results. Here, we prose a renew version of this system, which, relying upon the insertion and phenotypic correction of stable FANCA NT mutants (right), will increase our chances to discover novel molecules for FA treatment. (Adapted from Montanuy *et al*., 2020 [169]).

4.2.1 CANDIDATE VARIANTS SELECTION AND MODELS GENERATION

The first step to develop our cell-based systems involved the selection of *FANCA* candidate variants, which were chosen based on (i) their predictions not to impair the translational process (implicit in the "nontruncating" definition), (ii) relatively high frequency among reported FA patients [86], [269], (iii) and functional evidence supporting their deleterious effect [3], [28], [87]. The variants identified to meet the inclusion criteria encompassed (reference to FANCA NM_000135.4, NP_000126.2) (Table 11):

- c.2852G>A (p.Arg951Gln) and c.3391A>G (p.Thr1131Ala) missense mutations;
- c.3788_3790delTCT (p.Phe1263del) in-frame deletion.

All of them were annotated in the main FA reference databases and search engines consulted [58], [269], [286], and rated as likely-/pathogenic (Table 11). They also appeared to often occur in distinct ethnic groups (c.2852G>A total allele frequency (AF): 0.00008326; c.3391A>G total AF: 0.00006010; c.3788_3790delTCT total AF: 0.00009929) and, regardless of their diverse frequency, were found in several patients of the FA databases at our disposal (c.2852G>A and c.3391A>G: 5 alleles (5 individuals) and 1 allele (single individual) in the Italian cohort, respectively; c.3788_3790delTCT: 47 alleles, of which 36 (26 subjects) in the Spanish case study and 11 (8 subjects) in the Italian one) (Table 11) [56], [66]. Consistently with its high AF and numerous cases documented, c.3788_3790delTCT, also named "Brazilian variant", is the world most frequent FA pathogenic alteration, present in around 20-30% of all FA-A subjects [3], [86], [93]. Thus, in the event of identifying a molecule able to reestablish FA/BRCA pathway function, there would be a fair amount of FA patients suitable for a pharmacological treatment overall.

In terms of functional validation, both Arg951Gln and Phe1263del mutants were tested for their stability, cytoplasmic retention, and FANCD2 mono-ubiquitination failure, asserting their pathogenic role (Table 11) [19], [28]. Conversely, Thr1131Ala displayed a nearly wild-type activity when inserted in *FANCA*-null fibroblasts [87]. This phenomenon was attributed to a possible reduction of FANCA endogenous expression due to an increased protein degradation, and masked by the enforced overexpression conditions, or to a defective pre-mRNA splicing process that was not further explored [87]. A third hypothesis considered this variant as a benign polymorphism, albeit it was not detected in healthy individuals, and further works sustained its pathogenicity based on the inheritance pattern and screening of normal alleles (Table 11) [74]. By virtue of this "borderline" behavior, we regarded Thr1131Ala as a valuable candidate for our cellular models; its potential to recover a WT function could increase the chances to find a drug rescuing its phenotype, and, thus, the entire FA/BRCA ICL repairosome.

Table 11. Characteristics of the candidate variants selected. The three mutations chosen were all NT alterations of *FANCA* (NM_000135.4, NP_000126.2), frequently found among annotated FA probands, and with a proven deleterious role.

Variant	Type	Rockefeller DB	Varsome	Total AF GnomAD	Total AF GnomAD	Total allele count DBs	Pathogenicity evidence
c.2852G>A (p.Arg951Gln)	Missense	Yes (Lilely-/Pathogenic)	Yes (Uncertain significance)	0.00008326	European (non-Finnish), East Asian, Other	Spanish: / Italian: 5 (5 patients)	Yes
c.3391A>G (p.Thr1131Ala)	Missense	Yes (Pathogenic)	Yes (Pathogenic)	0.00006010	European (Finnish/non-Finnish), Latino/Admixed American, African/African American	Spanish : / Italian : 1 (1 patient)	Yes*
c.3788_3790delTCT (p.Phe1263del)	In-frame deletion	Yes (Pathogenic)	Yes (Pathogenic)	0.00009929	European (non-Finnish), Latino/Admixed American, Other	Spanish: 36 (29 patients) Italian: 11 (8 patients)	Yes

AF: allele frequency; /: no allele reported; *: pathogenicity evidence based on pedigree analyses and normal allele screenings.

After variants selection, we proceeded to the establishment of the models in the perspective of reproducing endogenous expression levels of *FANCA* like those observed in FA probands. In particular, we employed PGK-FANCA.WRPE lentiviral vector, already tested to successfully drive FANCA physiological levels [132], [133], where we introduced the single patient-derived variants by site directed mutagenesis [56], [287] (Figure 16). WT and mutant *FANCA* constructs were then separately transduced into Montanuy et al.'s system (U2OS FANCA$^{-/-}$-YFP-FANCD2) (Figure 16) [169], generating clonal cell cultures to be validate by functional studies.

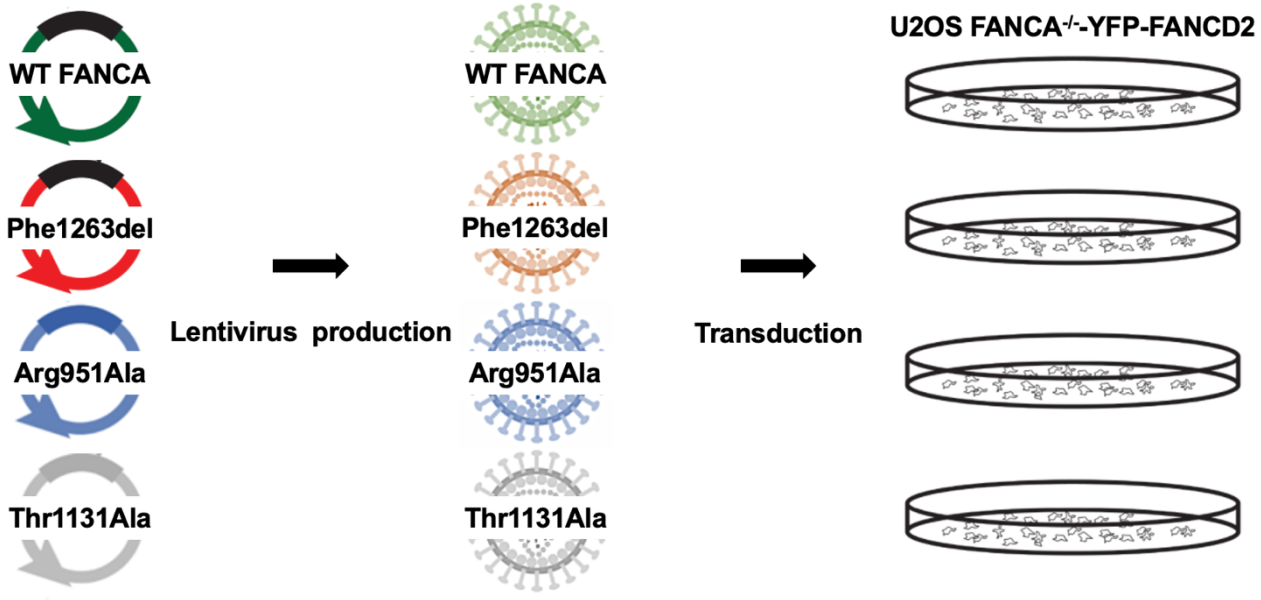


Figure 16. Schematic representation of the sequential steps for establishing WT FANCA cellular model (control) and three FANCA NT variants' ones (Arg951Gln, Thr1131Ala, Phe1263del). The alterations planned to test were inserted into PGK-FANCA.WRPE, WT FANCA-expressing lentiviral vector, by site directed mutagenesis (left). Lentiviral vectors carrying WT FANCA or the NT variants were produced (middle), and subsequently transduced into the already available U2OS FANCA-/--YFP-FANCD2 model [169] (right), unable to mono-ubiquitinate FANCD2 or form fluorescent foci due to FANCA loss.

4.2.2 FUNCTIONAL VALIDATION OF FANCA NONTRUNCATING MUTANTS

4.2.2.1 FANCA PROTEIN EXPRESSION, SUBCELLULAR LOCALIZATION AND FUNCTIONING

To ensure the recapitulation of FA patients' cellular phenotypes, we functionally assessed the three NT mutants, starting from the investigation of FANCA protein synthesis.
In line with the literature [3], [87], we overall confirmed FANCA stable presence (163 kDa) (Figure 17), albeit at different ratios depending on the sample considered. Interestingly, the systems created tend to display FANCA expression levels higher than the WT cells, used as positive control (Figure 17). We interpreted this observation as a possible result of the diverse genetic backgrounds of the lines under comparison: while FANCA^{corrected} (positive control), Arg951Gln, Thr1131Ala, Phe1263del models were all derived from the insertion of WT or modified PGK-FANCA.WRPE vector into Montanuy et al.'s system (employed as negative control), the WT line did not receive any manipulation. Considering this, we regarded FANCA^{corrected} line as a more appropriate reference for our models than the WT one, which we removed as control in the following functional studies.

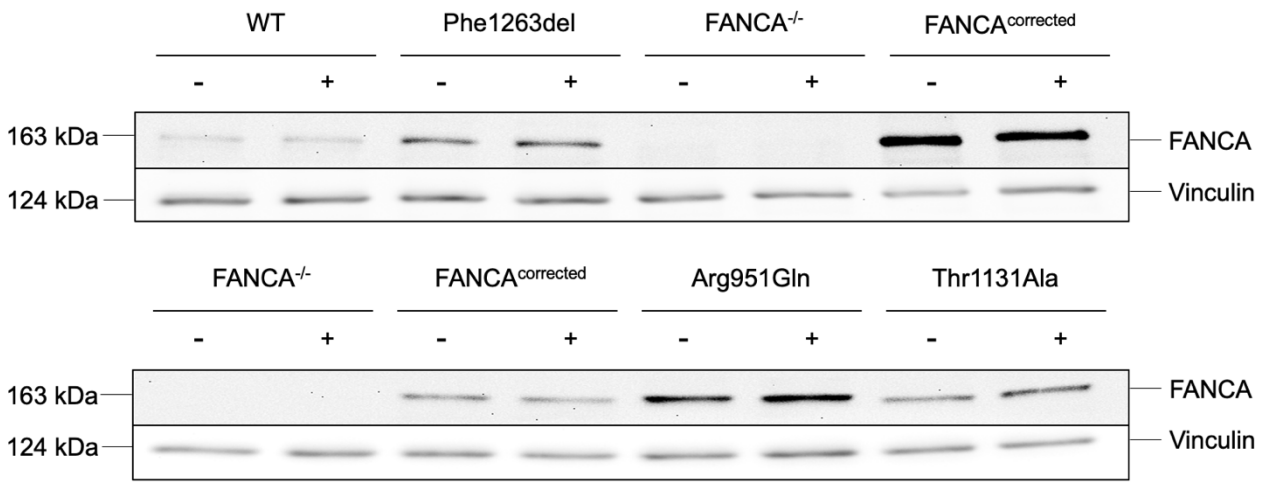


Figure 17. Evaluation of FANCA expression. FANCA was stably produced in all positive controls (U2OS WT (WT), U2OS FANCA$^{-/-}$-YFP-FANCD2+PGK-FANCA.WRPE (FANCA^{corrected})) and NT mutant models generated (U2OS FANCA$^{-/-}$-YFP-FANCD2+PGK-FANCA.WRPE_Phe1263del (above) and Arg951Gln/Thr1131Ala (below)), albeit at different levels, and absent in the negative control (U2OS FANCA$^{-/-}$-YFP-FANCD2). Vinculin was used as loading control. – and +: samples untreated or treated with 2mM HU, respectively.

Once ascertained FANCA stable expression, we focused on the exploration of its subcellular localization. In Phe1263del model, FANCA cytoplasmic percentage was predominant in comparison to its nuclear level (94% versus 6%) (Figure 18), and thus consistent with the FANCA NT pattern expected. Conversely, in both remaining mutant systems we reported a FANCA cytosolic concentration of about 40%, even exceeding that of the FANCA^{corrected} line (Figure 18).

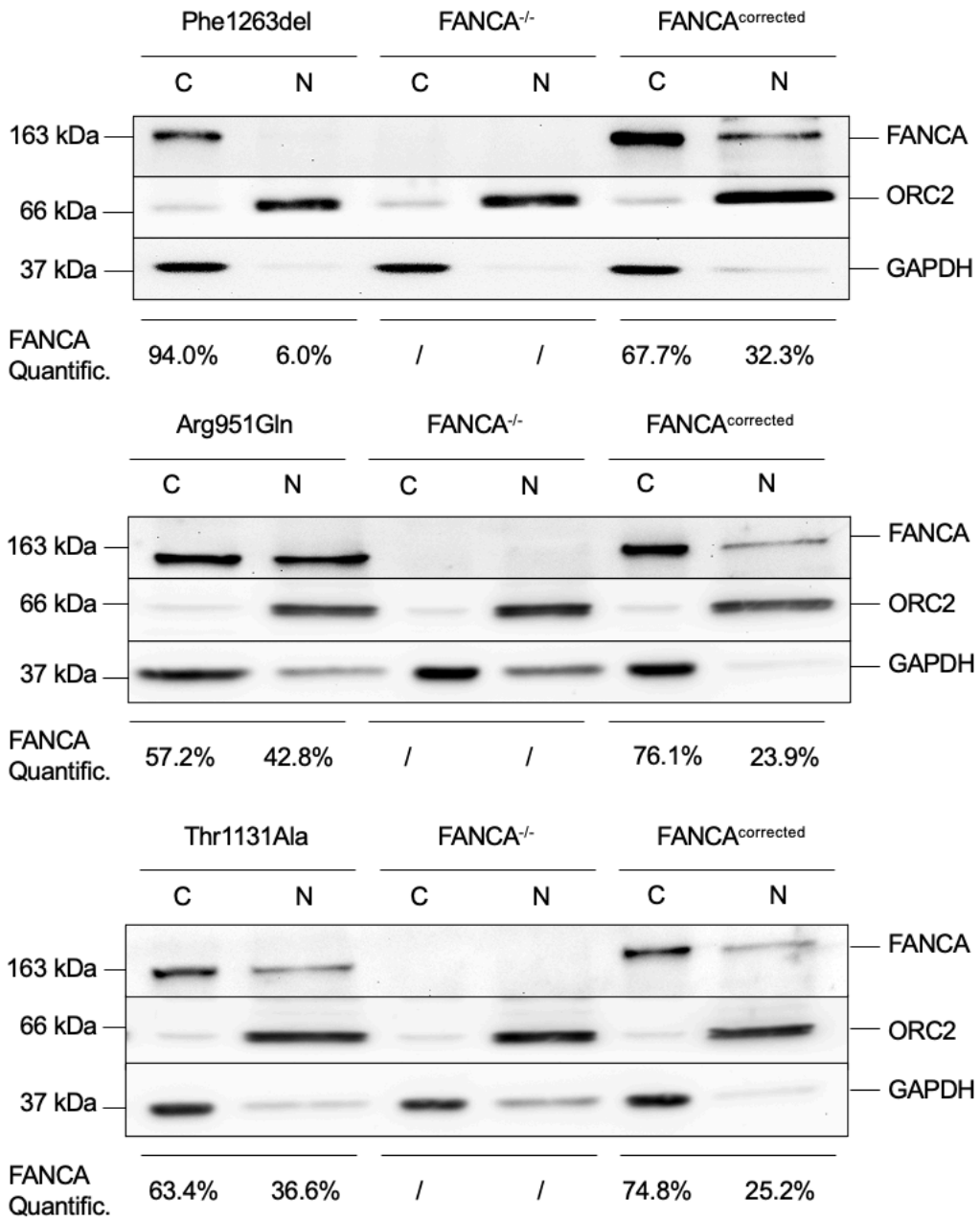


Figure 18. Assessment of FANCA localization via subcellular protein fractionation. Only Phe1263del model showed a negligible concentration of FANCA at nuclear level (above), while both Arg951Gln (middle) and Thr1131Ala (below) mutants were associated to FANCA nuclear percentages like those of the FANCA^{corrected} control. All the samples were treated with 2mM HU to stimulate FANCA relocation to the nuclear compartment. GADPH and ORC2 were used as loading controls for the cytosolic and nuclear fraction, respectively. FANCA protein quantification was calculated operating an intrasample normalization based on the quantification of both same specimen-loading controls from three independent technical replicates and considering possible contaminations between the two fractions (e.g., Arg951Gln C and N samples). C: cytoplasmic fraction; N: nuclear fraction (including soluble and insoluble fractions).

We therefore evaluated the effect of the diverse FANCA nuclear percentages reported on the capability of the mutants to preserve FANCD2-Ub. Along with FANCA$^{-/-}$ control, Phe1263del line was the only failing to promote FANCD2-Ub (FANCD2-Ub: 175 kDa; YFP-FANCD2-Ub: 201 kDa) (Figure 19). The other NT models, instead, showed mono-

ubiquitination ratios of both FANCD2 isoforms compatible with those of the FANCA^{corrected} control (Figure 19), indicating that 25%-40% of FANCA nuclear retention was sufficient to maintain FANCD2-Ub. Moreover, we explained the higher intensity of FANCD2-Ub bands compared to YFP-FANCD2-Ub ones as a consequence of the possible competition for mono-ubiquitination between the two isoforms, favoring the modification of the former (Figure 19). Alternatively, we posited this phenomenon to be due to the higher expression of the endogenous protein, the bigger size of exogenous one, or a perturbation of the antigen-antibody interaction by the fluorescent tag.
Taken together, these preliminary results were describing a FA phenotype for the sole Phe1263del, and elucidating a WT-like behavior for the other NT mutants.

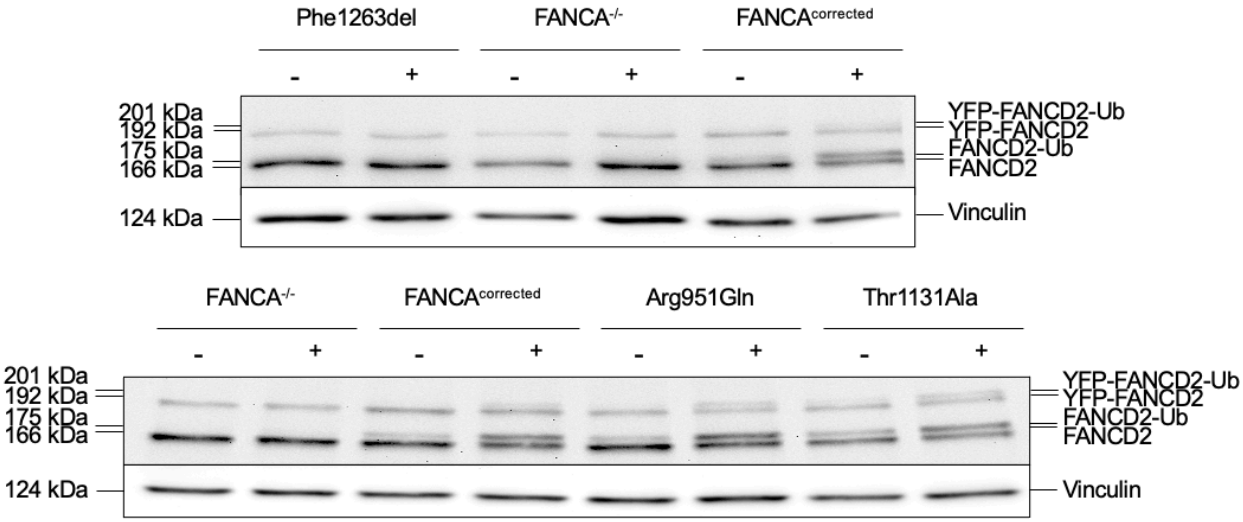


Figure 19. Analysis of FANCD2 mono-ubiquitination. Consistently with the amount of FANCA concentrated at nuclear level, Phe1263del mutant was not able to promote FANCD2-Ub (above), or YFP-FANCD2, while the others showed WT-like mono-ubiquitination profiles (below). Vinculin was employed as loading control. – and +: samples untreated or treated with 2mM HU, respectively.

4.2.2.2 FA PATHWAY ACTIVITY: DEB SENSITIVITY, G2/M CELL CYCLE BLOCK, FANCD2 FOCI FORMATION

In the perspective to deepen the phenotypic investigation of the three FANCA mutants, we carried out cell survival assays to scrutiny their resilience to increasing doses of DEB (50-200 ng/ml) (Figure 20). In line with the previous findings, Phe1263del sensitivity matched up with that of the FANCA$^{-/-}$ control, and significantly diverged from the FANCA^{corrected} profile already at low drug concentrations (≥ 100 ng/ml); it further showed a halved survival rate at around 100 ng/ml DEB, reaching a viability of about 38% at the highest dosage (200 ng/ml) (Figure 20). On the contrary, both Arg951 and Thr1131Ala displayed a sensitivity trend compatible to that of the FANCA^{corrected} control, with a survival variation of only ~30% between the untreated condition and the exposure to DEB maximal dose (Figure 20).

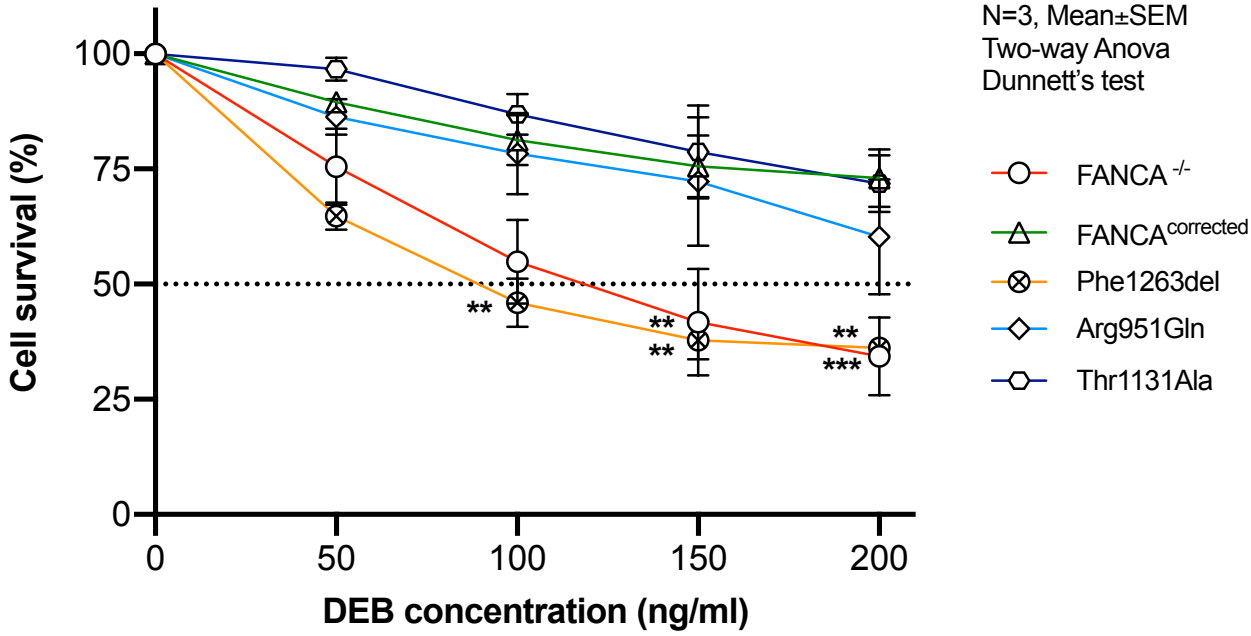
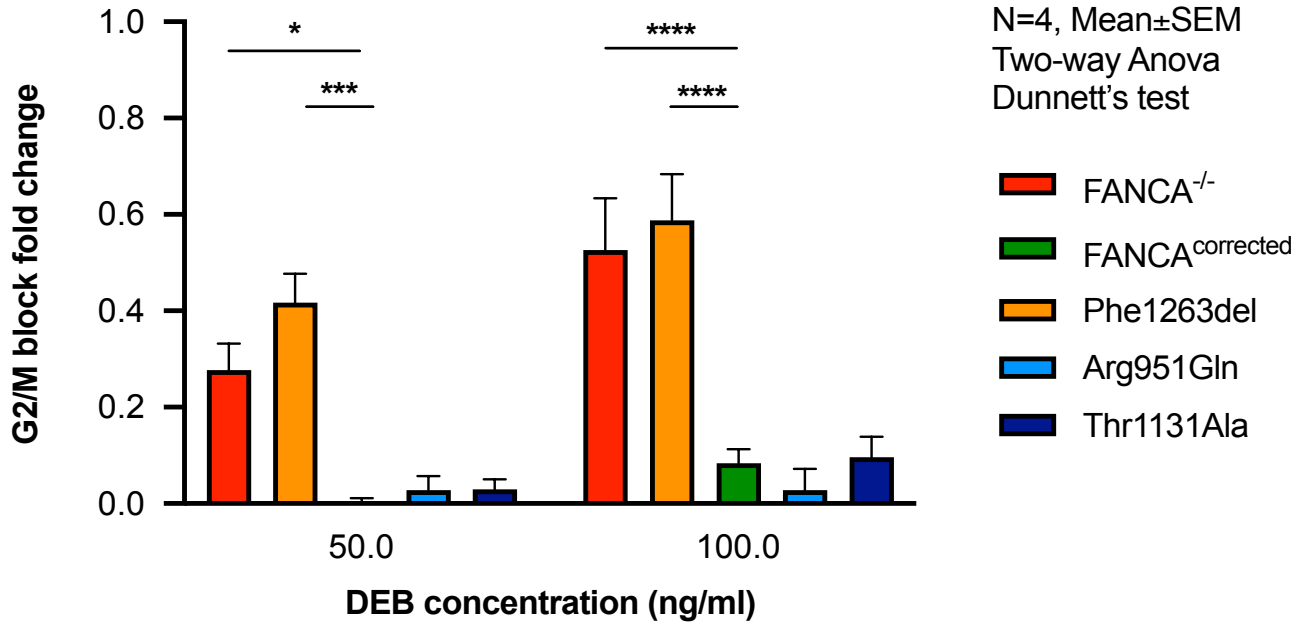


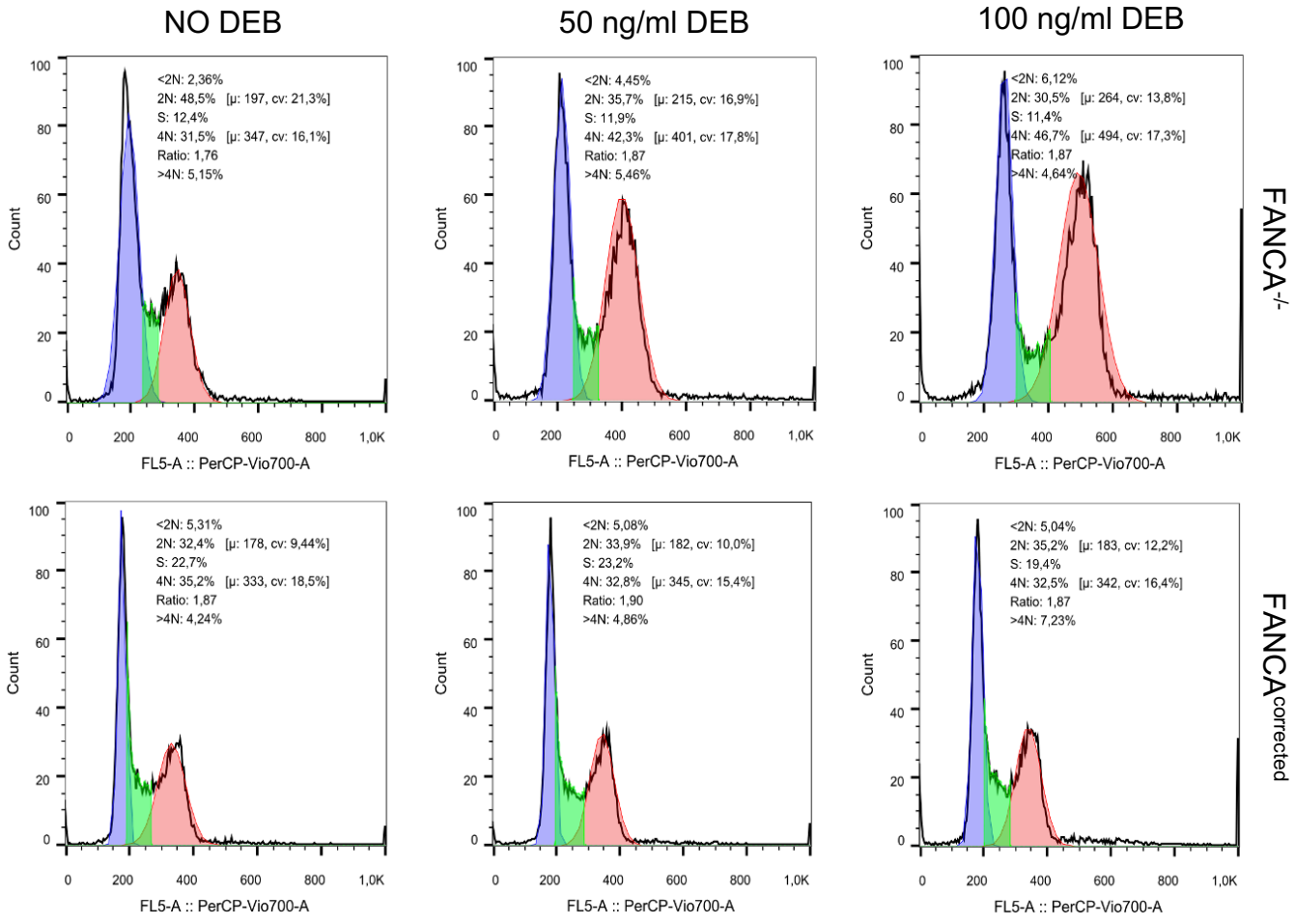
Figure 20. Cell survival of FANCA NT lines in response to DEB-induced DNA damage. The graph describes the percentage cell survival rate of the negative (FANCA$^{-/-}$) and positive (FANCA^{corrected}) controls and the FANCA NT models (Phe1263del, Arg951Gln, Thr1131Ala) under the exposure to increasing doses of DEB (0, 50, 100, 150, 200 ng/ml). Three independent technical replicates were carried out, each one including 10 repetitions without DEB treatment (0 ng/ml) and 6 repetitions exposed for 144h to each dose tested (50, 10, 150, 200 ng/ml) per any sample. Data are expressed as mean± SEM. The statistical analysis was conducted by applying Shapiro-Wilk test for data normalization, and two-way Anova and Dunnett's test for multiple data comparisons (FANCA^{corrected} vs any sample, P<0.05 (*), P<0.01(**), P<0.001 (***)).

We also challenged the potential of the DEB-treated mutants to overcome the G2/M cell checkpoint, which generally blocks FA cells cycle due to the excessive accumulation of DNA damage. Once again, Phe1263del exhibited a behavior consistent with that of the FANCA$^{-/-}$ line (Figure 21A), entailing a remarkable arrest at the G2 phase under DEB exposure compared to the baseline (Figure 21B). The other mutants, instead, presented a WT-like progression of the cell cycle (Figure 21A), which indeed remained nearly the same in spite of the treatment with DEB (Figure 21B).

A


1.0
0.8
0.6
0.4
0.2
0.0
G2/M block fold change
*
***
****
****
50.0
100.0
DEB concentration (ng/ml)
N=4, Mean±SEM
Two-way Anova
Dunnett's test
FANCA-/-
FANCAcorrected
Phe1263del
Arg951Gln
Thr1131Ala

B


NO DEB
50 ng/ml DEB
100 ng/ml DEB
<2N: 2,36%
2N: 48,5% [μ: 197, cv: 21,3%]
S: 12,4%
4N: 31,5% [μ: 347, cv: 16,1%]
Ratio: 1,76
>4N: 5,15%
<2N: 4,45%
2N: 35,7% [μ: 215, cv: 16,9%]
S: 11,9%
4N: 42,3% [μ: 401, cv: 17,8%]
Ratio: 1,87
>4N: 5,46%
<2N: 6,12%
2N: 30,5% [μ: 264, cv: 13,8%]
S: 11,4%
4N: 46,7% [μ: 494, cv: 17,3%]
Ratio: 1,87
>4N: 4,64%
FANCA-/-
<2N: 5,31%
2N: 32,4% [μ: 178, cv: 9,44%]
S: 22,7%
4N: 35,2% [μ: 333, cv: 18,5%]
Ratio: 1,87
>4N: 4,24%
<2N: 5,08%
2N: 33,9% [μ: 182, cv: 10,0%]
S: 23,2%
4N: 32,8% [μ: 345, cv: 15,4%]
Ratio: 1,90
>4N: 4,86%
<2N: 5,04%
2N: 35,2% [μ: 183, cv: 12,2%]
S: 19,4%
4N: 32,5% [μ: 342, cv: 16,4%]
Ratio: 1,87
>4N: 7,23%
FANCAcorrected
Count
FL5-A :: PerCP-Vio700-A

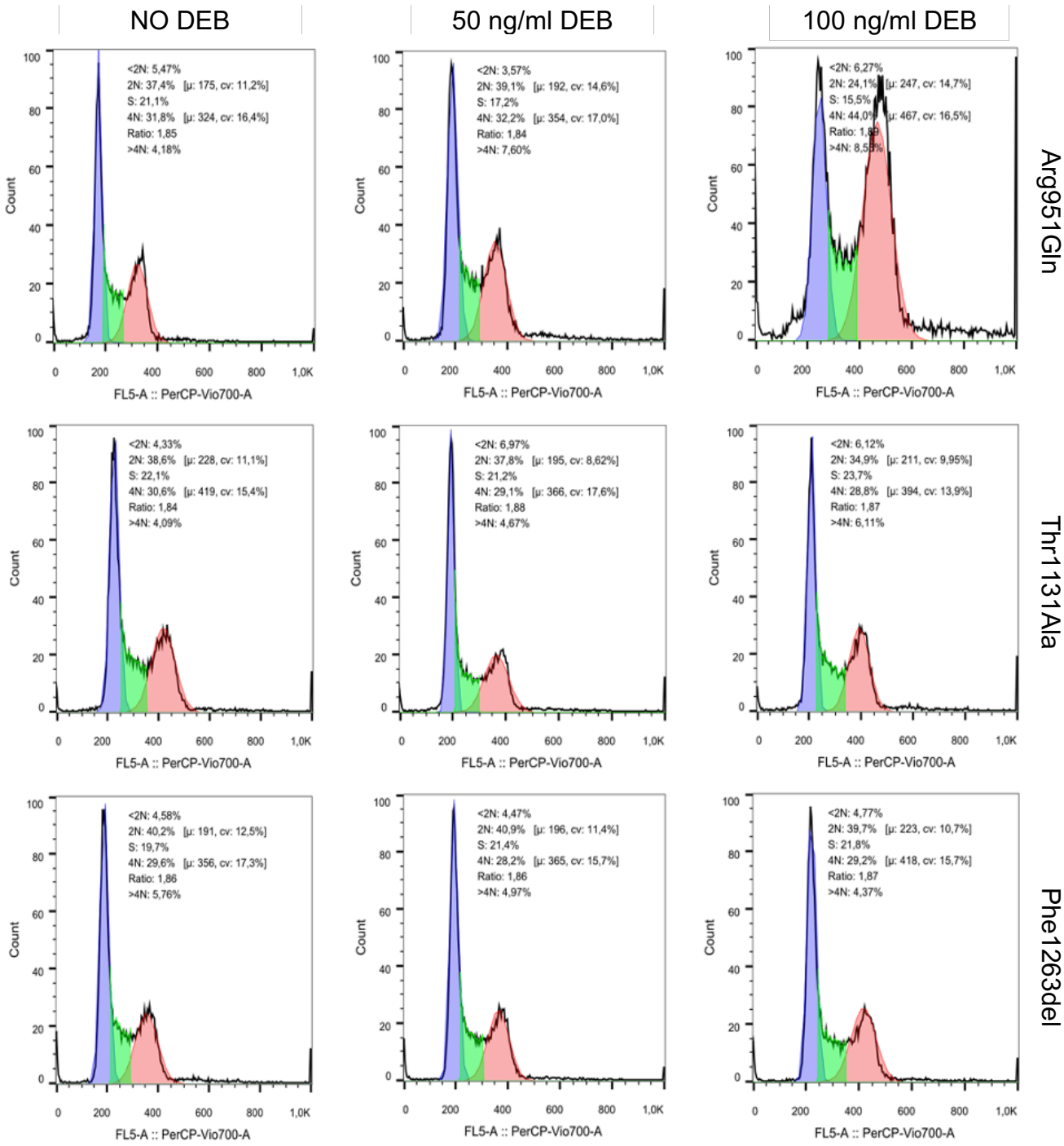


Figure 21. G2/M arrest in FANCA NT lines under DEB exposure. A) The histogram represents the G2/M block fold change (FC) (= (events frequency with DEB treatment – events frequency without DEB treatment)/ (events frequency without DEB)) – parameter mitigating the G2/M block differences of the same sample in distinct replicates- of the negative ($FANCA^{-/-}$) and positive ($FANCA^{corrected}$) controls and the FANCA NT models (Phe1263del, Arg951Gln, Thr1131Ala) under the exposure to increasing doses of DEB (50 and 100 ng/ml). Four independent technical replicates were carried out, each one including 2 repetitions without DEB treatment and exposed for 72h to each dose tested (50 and 100 ng/ml) per any sample. Data are expressed as mean± SEM. The statistical analysis was conducted by applying Shapiro-Wilk test for data normalization, and two-way Anova and Dunnett's test for multiple data comparisons ($FANCA^{corrected}$ vs any sample, P<0.05 (*), P<0.01(**), P<0.001 (***), P<0.0001(****)). B) The graphs illustrate the percentage count of cells in G1 (violet),

S (green) and G2 (red) phases determined by using the emission spectrum of the PerCP-Vio700A dye for flow cytometry. 2N: diploid DNA content of the cells in G1 stage; 4N: tetraploid DNA content of the cells in G2 stage.

We ultimately examined the ability of the mutants to induce the formation of YFP-FANCD2 foci upon DNA damage, core feature of the models generated since conceived as the HCS readout. In line with the already described phenotypes, Phe1263del cells were the only unable to promote the relocation of YFP-FANCD2 to repair foci, manifesting a nuclear diffuse pattern also on HU treatment (Figure 22). By contrast, both Arg951 and Thr1131Ala systems displayed perfectly visible fluorescent foci, appearing as intense spots in the nuclei and even clearer under HU exposure (Figure 22).

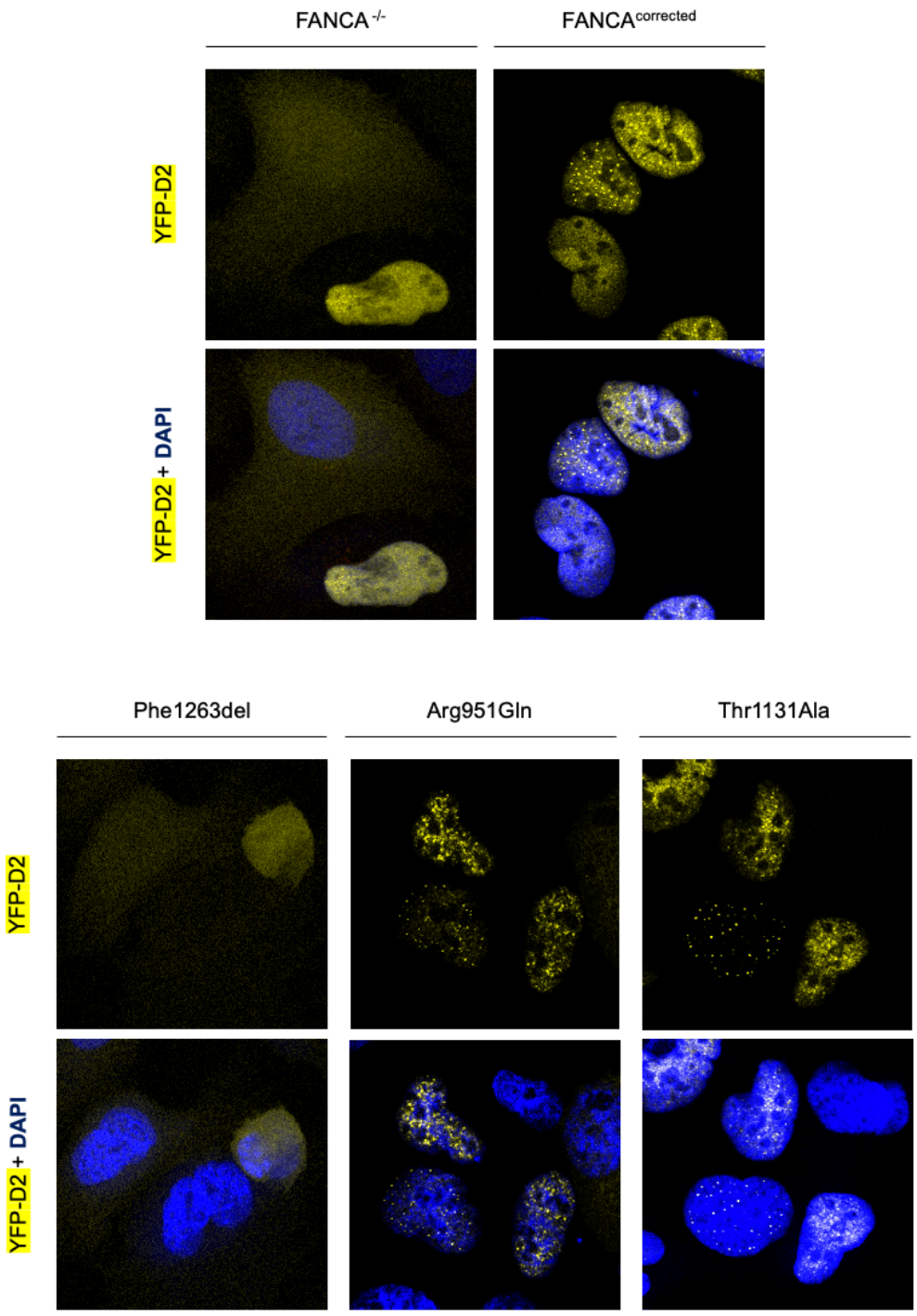


Figure 22. FANCA NT mutants' capability to promote YFP-FANCD2 foci formation. YFP-FANCD2 fluorescence (yellow) was analyzed by microscopy after exposure of the controls (FANCA-/- and FANCA corrected) and the samples (Phe1263del, Arg951Gln, Thr1131Ala) to 2mM Hu for 72h to induce DNA damage. Upper images show only YFP-FANCD2 signal, while the lower ones its overlay with DAPI (nuclei staining). In case of FA/BRCA2 pathway functioning, YFP-FANCD2 foci are seen as bright spots within the cell nuclei; conversely, YFP-FANCD2 foci cannot be detected, exhibiting a dispersed nuclear pattern.

Considering the results achieved, we decided to keep the sole Phe1263del mutant as a valuable HCS candidate, since it is the only one to faithfully mimic FA patients' phenotype in the experimental conditions adopted, and, thus, effective for the detection of a possibly therapeutic molecule. Contrariwise, the WT-like characteristics of Arg951 and Thr1131Ala models let us to discarding the two mutants from further analysis. Seeking potential explanations of their behaviors, we also postulated that the background of our systems, different from patients' endogenous one, could elicit the synthesis of some FANCA NT proteins to such an extent as to allow their entrance to the nucleus and, thus, rescue the pathway activity. Alternatively, we hypothesized the two variants to affect splicing, to be hypomorphic for DNA repair or, despite their current classification, variants in linkage disequilibrium with the actual deleterious mutations.

4.3 GENOME-WIDE CRISPR KNOCKOUT SCREENS FOR SYNTHETIC INTERACTIONS IN FA CELLULAR MODELS

4.3.1 SCREEN DESIGN AND SETUP: CELLULAR SYSTEMS CREATION, gRNA LIBRARIES SELECTION AND PRODUCTION

Our second strategy for the pharmacological management of solid tumors in FA relied on the GW CRISPRko technology that, based upon the rationale to abrogate the expression of virtually any genetic locus, allowed us to conduct a systemic, unbiased exploration of the SIs for FA deficiency.

In this perspective, we initially transduced the h-TERT RPE-1 Cas9$^{+/+}$p53$^{-/-}$ line with lentiviral gRNA expression vectors towards *FANCA* or a non-human non-target (NT) sequence to generate FANCA KO (RPE-1 FANCA$^{-/-}$ GFP) and control (RPE-1 NT$^{-/-}$ GFP) models amenable for gene editing, respectively. Interferences of the integrated gRNA cassettes or vector sequences on the NGS screening step were ruled out by amplifying cell systems' gDNA with NGS primers and comparing their electrophoretic profiles to that of a positive control. We chose *FANCA* as FA target gene in consideration of its clinical relevance [3], while we introduced a NT sequence into the vector for the control system to generate a reference model with a comparable genetic background, and to avert Cas9 random cleavage activity. Moreover, we employed the h-TERT RPE-1 cell line since its immortalized status makes it ideal for GW screens requiring several millions of cells, and its genetic stability (near-diploid human line with a modal chromosome number of 46 in 90% of cells [288]) tailors it to even, effective CRISPR editing. We confirmed proper cell infection checking GFP-signal, reporter of the transfection vectors, through flow cytometry, and singled out the sole green-fluorescent clones via cell sorting (Figure 23A). We further ascertained gene KOs generation by Sanger sequencing and FANCA expression abolishment, and thus Cas9 on-target activity trough immunoblotting (Figure 23B).

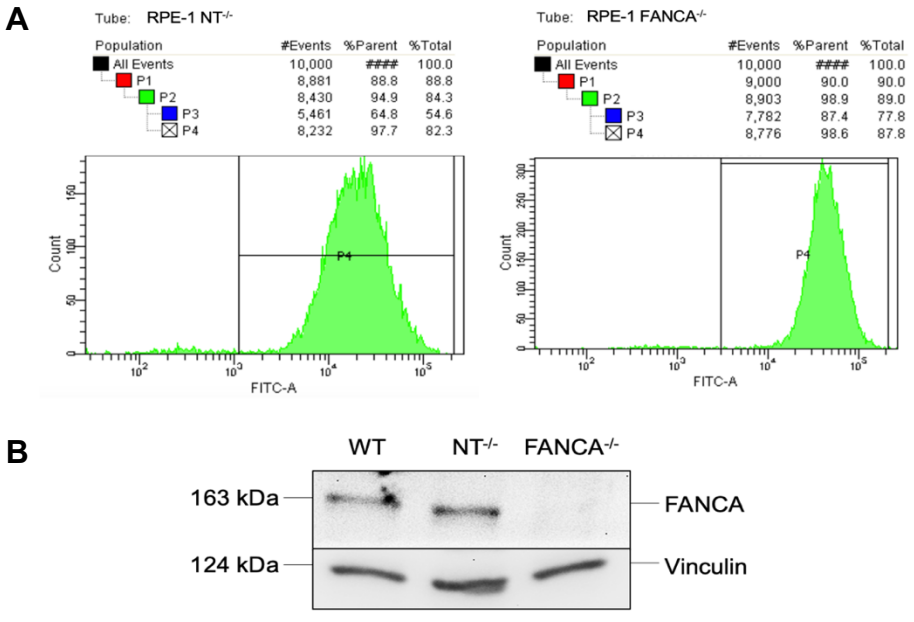


Figure 23. Generation and validation of the cellular models for CRISPRko screens. A) Proper cell infection and, thus, control (RPE1 $NT^{-/-}$) and FANCA KO (RPE-1 $FANCA^{-/-}$) systems establishment were assessed by measuring the GFP-signal of 10,000 cells (events) with the FICT-A dye for flow cytometry. Only the GFP-positive clones with the highest fluorescence (P4) were selected and kept for further analysis. P1: live cells; P2: singlets; P3: GFP-positive clones. B) In line with the expectations, FANCA protein expression was observed in the RPE-1 $Cas9^{+/+}p53^{-/-}$ (WT) and RPE-1 $NT^{-/-}$ GFP ($NT^{-/-}$) lines, while lost in the RPE-1 $FANCA^{-/-}$ GFP ($FANCA^{-/-}$) model confirming Cas9 on-target activity. Vinculin was used as loading control.

To investigate the possible impact of genetic disruptions on FA cells viability, we employed two lentiviral GW KO pooled libraries, such as the commercially available Toronto KnockOut v3 (TKOv3) library [276] and a newly generated in-house one. Both of them are designed to be redundant (4 independent guides/gene) to compensate for gRNAs diverse efficiency and/or potential off-target activity [226]. In addition, they include control gRNAs towards safe harbor loci or other genomic regions with no expected editing effect to account for DNA damage response and aspecific cell proliferation reduction due to CRISPRko [201].
In more detail, the TKOv3 commercial library, characterized by a U6 promoter and a WT gRNA scaffold, contains 70,948 gRNAs intercepting 18,053 human genes with further 142 control guides for EGFP, LacZ and luciferase [276]. As documented by the broad literature on it, TKOv3 guarantees an excellent performance, and is widely exploited in laboratories specialized in CRISPR screens [226], [276]; however, since synthetized at the turn of 2017 and 2019, it does not encompass all the genes annotated thus far, nor benefits of the most recent library design optimizations. For this reason, we decided to further employ a just fine-tuned in-house library, tested in a preliminary version on 17 different lines (h-TERT RPE-1 $Cas9^{+/+}p53^{-/-}$ and pulmonary cancer cells). This library benefits of 7SK promoter and a spCas9 gRNA scaffold V2, and has been designed based on the GenomeCRISPR database [289], providing the performance data of around 700,000 sgRNAs assayed in 421 distinct human cell lines, and the CRISPick tool for gRNA design [240], [290], ranking and picking the best CRISPR guides for the sequences of interest. Eventually, the in-house library comprises 83,549 gRNAs towards all the genes reported to date (20,996 genes), thus partially overlapping TKOv3 targets. As a result, the comparison of the outcomes of the screens performed with the two libraries on the *FANCA* KO and control models will contribute to enhance the reproducibility and robustness of our SL or SV candidate hits.

4.3.2 SCREEN STEPS: FROM CELLULAR MODELS TRANSDUCTION TO GDNA SAMPLES PREPERATION FOR NGS

Following models systems and guides setup, we transduced both FANCA KO and control lines with the two titrated libraries (day 0), administered in bulk and at low MOI (0.5) to guarantee the integration and, thus, the perturbation of a single gRNA per each cell infected (Figure 24). We next performed two antibiotic treatment steps- spaced at least a couple days apart (~day 1-2 and 3-5) - to pick the sole properly mutagenized clones (Figure 24). Particularly, this was the unique selective pressure applied during the whole screen; indeed,

we did not introduce any further biological challenge, such that the representation of each end phenotype will be only determined by the proliferative competition between cells, directly dependent on the precise gene KO induced.

Over the screens, we periodically split cells at ~80-90% confluence (day 6-7 and 12-14) to establish independent biological duplicates, and thereby increase the separation between divergent end phenotypes (phenotypic fold change, PFC) for more reliable results (Figure 24). Every time we reached this point, we also monitored the number of cell doublings, and the library coverage to keep 250x (TKOv3 library) or 100x (in-house library) representation per gRNA through the experiments (Figure 25 A, B). The overall data elucidated a quicker proliferation rate of the lines infected with the in-house library compared to the TKOv3 one, as well as a slightly faster growth pace of the FANCA KO model over the control (Figure 25 A, B), possibly due to the altered cell cycle regulation caused by *FANCA* loss. We also reported a system-independent decrease of both parameters around the second week from the infection (day 6-14) owing to the antibiotic selection and the emergence of the SL effects at the phenotypic level (Figure 25A, B). However, population doublings and gRNAs representation values rose again between day 13-21 because of the proliferative advantage of the positively selected clones (Figure 25A, B).
Once reached about 10 cell doublings, considered enough to obtain a significant PFC, we ended the screens by harvesting all biological replicates, and isolating, quantifying, and preparing the gDNA for massive parallel sequencing (Figure 24).

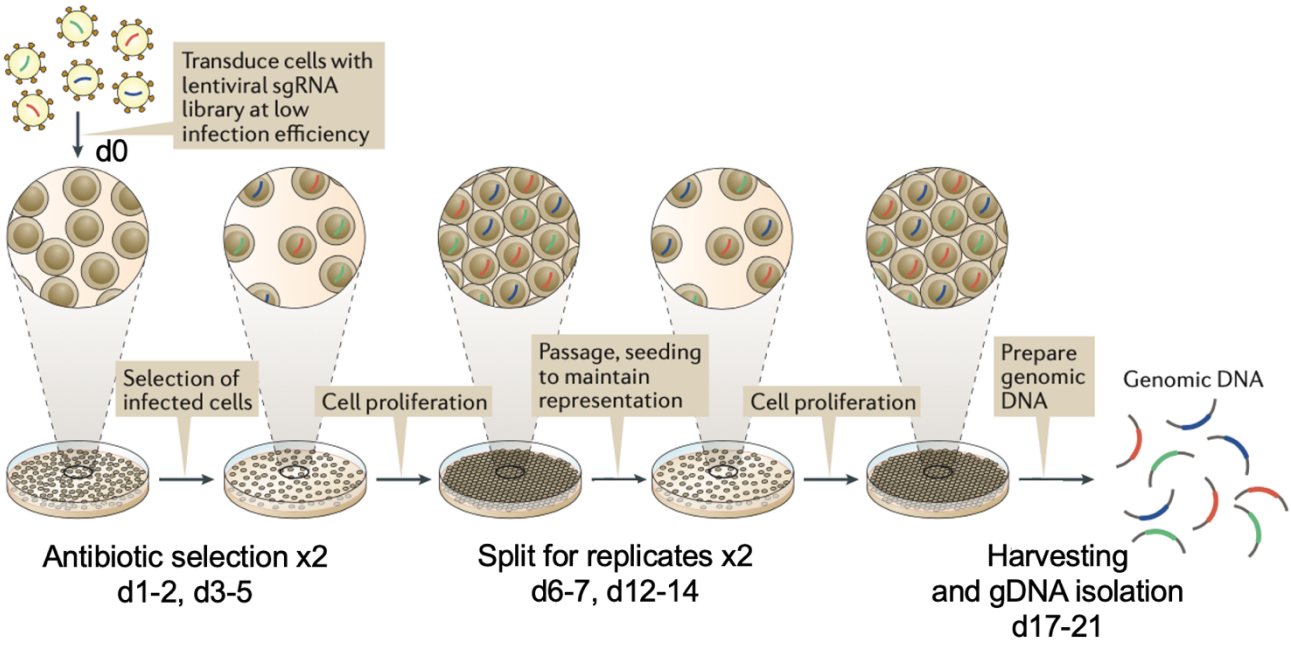
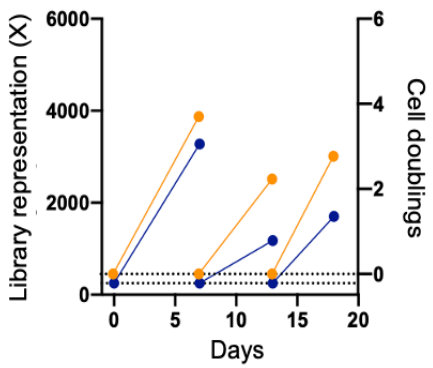
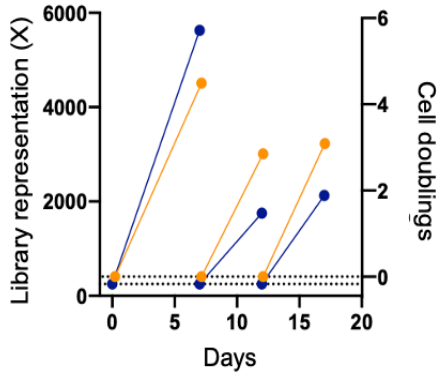


Figure 24. Schematic of the CRISPRko screen workflow adopted. Cellular models were initially transduced with the lentiviral KO pooled libraries (TKOv3 and in-house library) at low MOI (day (d) 0), and then subjected to two antibiotic selection steps (d1-2, and d3-5). Over the screens, cells were let to grow without applying any biological challenge, but the proliferative competition between the differently edited clones. Once reached ~80-90% confluence, cells were passed and seeded back to establish independent biological replicates (d6-7, and d12-14), and thus augment the PFC; at this points, cell doublings and library coverage were monitored to keep representation. Around 10 cell doublings (d17-21), screens were concluded through biological replicates harvesting, and gDNA extraction and preparation for NGS. Adapted from [291].

A

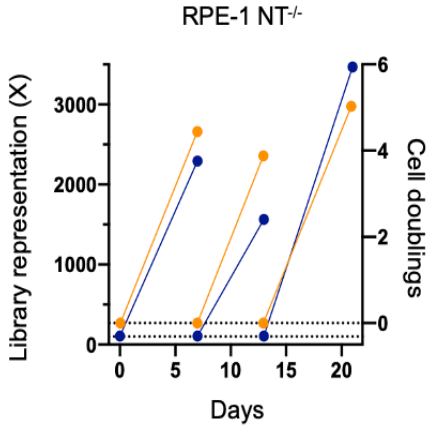


Days	Library representation (x)	N° Cells doublings
d0-d7	3275	3.7
d7-d13*	1175	2.2
d13-d18	1700	2.8
Approximate cell doubling time: 44.5 h		

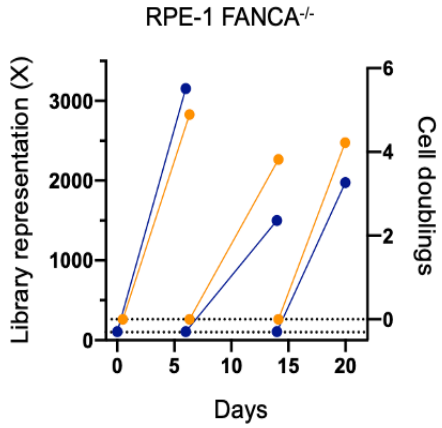


Days	Library representation (x)	N° Cells doublings
d0-d7	3153	4.5
d7-d12	1500	2.9
d12-d17	1977	3.0
Approximate cell doubling time: 39.6		

B



Days	Library representation (x)	N° Cells doublings
d0-d7	2294	4.4
d7-d13	1564	3.9
d13-d21	3471	5.0
Approximate cell doubling time: 37.5 h		



Days	Library representation (x)	N° Cells doublings
d0-d6	3153	4.9
d6-d14*	1500	3.8
d14-d20	1977	4.4
Approximate cell doubling time: 32 h		

Figure 25. Monitoring of cell doublings and library representation of the TKOv3 (A) and in-house (B) screens on the RPE-1 NT$^{-/-}$ and RPE-1 FANCA$^{-/-}$ models over time. Cell doublings and library coverage data are in orange and purple, respectively, and their specific values reported in the tables below the schematics. The approximate cell doubling time was estimated based on the days of culturing and the relative number of cell divisions; the data of RPE-1 FANCA$^{-/-}$ d7-d12 in the in-house screen and RPE-1 NT$^{-/-}$ d7-d13 in the TKOv3 screen (*) were omitted due to confluence-induced proliferation halt, skewing the parameters considered.

Prior to the amplification of the screens-derived samples for NGS, we performed a test PCR to assess gDNA quality and integrity, and the protocol to adopt. Moreover, we proved the efficiency and specificity of the NGS primer sets, which included:

- Same pool of forward primers for all the samples from TKOv3 or in-house screen. This consisted of 8 distinct oligonucleotides with differential staggers, which, introducing spacers

of variable length, produce shifts in the sequenced regions and, thus, contribute to enhance the heterogeneity of the final reads.
- Single (in-house screen) or pooled (TKOv3 screen) barcode-tagged reverse primers different for any specimen from the same screen. Since carrying barcode sequences, these oligonucleotides allow to distinguish the NGS reads produced and match them with their corresponding sample.
To maintain the coverage of the libraries, we analyzed an amount of gDNA congruent to that contained within the number of clones transduced, considering an average quantity of gDNA per diploid cell of 6.6 pg [279]. We next split the required gDNA amount into distinct PCR reactions in such a way not to exceed 2.5 µg template/PCR reaction [279], and combined them in a single mix at the end of the amplification. We eventually subjected a representative volume of each PCR sample to electrophoresis, followed by gDNA purification via gel extraction and quantification, and NGS analysis. Particularly, this last step required the use of equimolar samples; expressing the exact number of molecules in a volume, molarity guarantees an even representation of each template, regardless of its size on the sequencer flow-cell, and the obtainment of the reads number ($2x10^7$ reads) set on the basis of the library complexity (~80,000 guides) and the desired NGS coverage (250x).

4.3.3 BIOINFORMATICS ANALYSIS AND HITS EVALUATION

At present, we have already obtained and examined the NGS results of the specimens from the in-house screen, while the sequencing of the TKOv3 screen samples is still underway.
Before delving into the description of the bioinformatic analysis and candidate hits found, it is important to recall that a cornerstone of CRISPRko screens is the dual function of the gRNAs, being both carriers and quantitative readouts of the precise genetic perturbation introduced. They indeed determine which gene must be knocked out, and, simultaneously, changes in their number correspond to variations in the levels of the target genes and associated phenotypes in the final cell pool.
Consistently, as last screening step, we performed the sequencing-based counting of any gRNA within the experimental duplicates and compared its read counts to those of the same guide from the library - sequenced likewise and used as reference- and the replicates of the other cellular system. This procedure allowed us to determine the scores and p-values of each guide/target, confirming the dropout of Hart *et al*.'s essentials, set of vital genes for all human cells [226], and good range of hits FC through the comparison of models' screen results to the in-house library's ones (Figure 26 A). Moreover, it led to the identification of significantly depleted and enriched hits of the FANCA KO system over the library and the NT model, indicative of potential SL and SV genetic interactions, respectively (Figure 26 B, C).

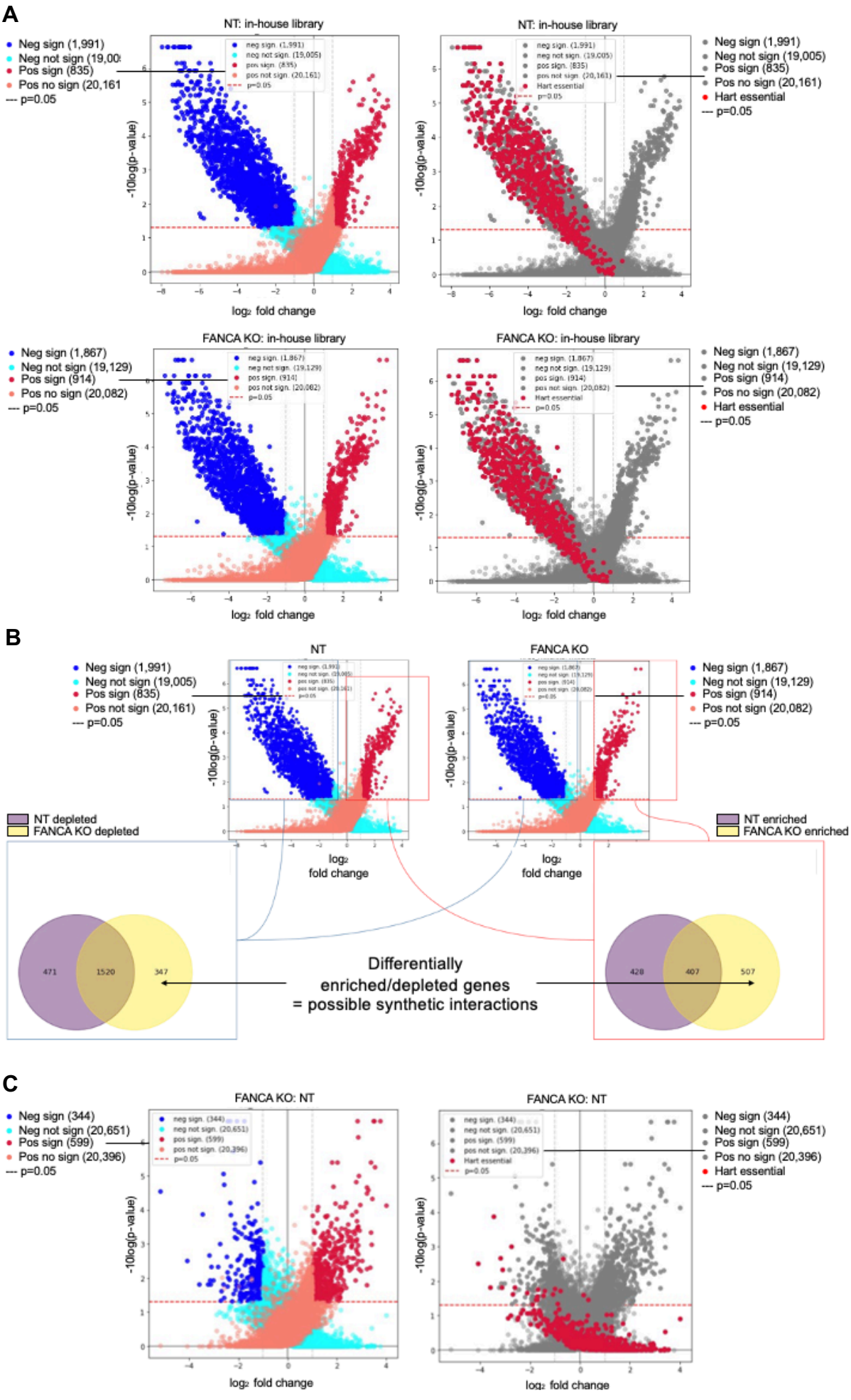

A
NT: in-house library
Neg sign (1,991)
Neg not sign (19,005)
Pos sign (835)
Pos no sign (20,161)
Hart essential
p=0.05
-10log(p-value)
log2 fold change
FANCA KO: in-house library
Neg sign (1,867)
Neg not sign (19,129)
Pos sign (914)
Pos no sign (20,082)
B
NT
FANCA KO
NT depleted
FANCA KO depleted
NT enriched
FANCA KO enriched
471
1520
347
428
407
507
Differentially
enriched/depleted genes
= possible synthetic interactions
C
FANCA KO: NT
Neg sign (344)
Neg not sign (20,651)
Pos sign (599)
Pos no sign (20,396)

Figure 26. Genetic interactions with FA pathway. A) The iterative MAGeCK analysis of the screen performed with the in-house library describes the comparison of the results of the NT (above) and FANCA KO (below) replicates clusters to those of the library (proliferation control). Significantly negative (neg sign, P<0.05; NT: 1,991 of 19,005; FANCA KO: 1,867 of 19,129) and positive (pos sign, P<0.05; NT: 835 of 20,161; FANCA KO: 914 of 20,082) hits are indicated in dark blue and red, respectively (above/below left). Hart et al.'s essential core genes are highlighted in dark red (above/below right). Visualized are MAGeCK-derived log_2 fold changes and p-values. B) MAGeCK analysis displays the genes found as depleted and enriched in the NT (above left) and FANCA KO (above right) replicates clusters. Combinations of the significantly negative (NT: 835 of 20,161; FANCA KO: 914 of 20,082) genes are indicated in blue, while the positive (NT: 1,991 of 19,005; FANCA KO: 1,867 of 19,129) in red. The sets provide a direct comparison of the depleted (below left) and enriched (below right) hits of the FANCA KO replicates cluster (yellow) versus the NT one (purple), where the nonoverlapping regions correspond to differentially represented targets and, thus, potential genetic interactions. B) MAGeCK analysis shows the direct comparison between FANCA KO and NT screen results. Significantly negative (344 of 20,651) and positive (599 of 20,396) hits are reported in dark blue and red, respectively (left). Hart et al.'s essential core genes [226] are in in dark red (right).

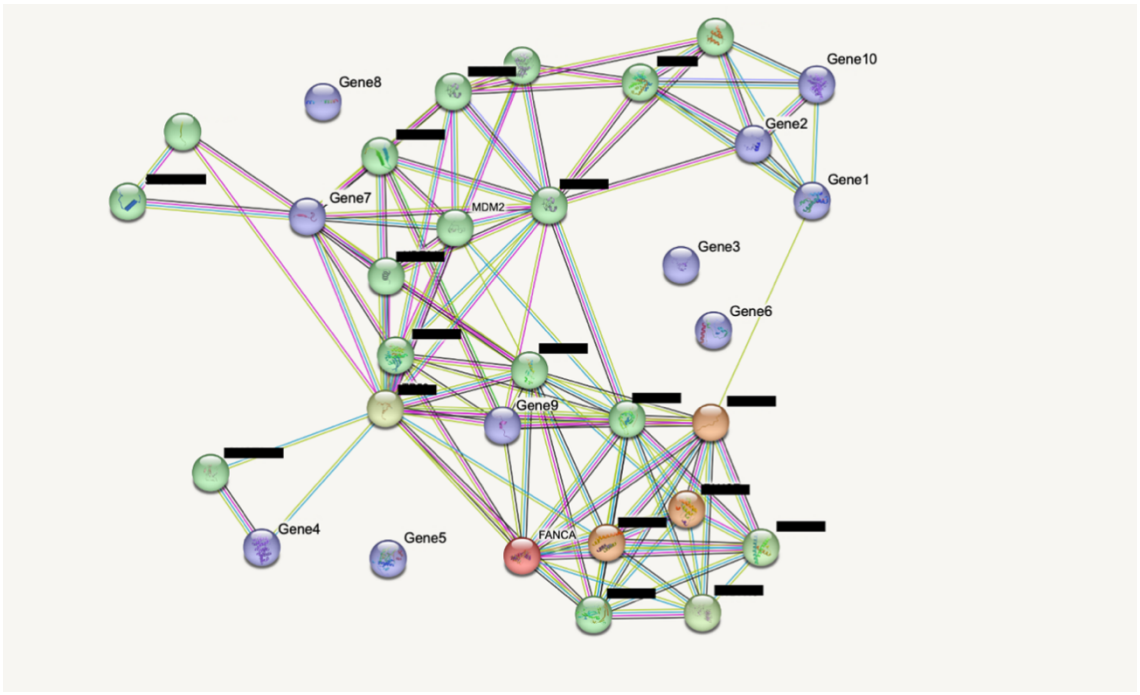
Based on these findings, we were able to compile a roster of preliminary gene hits, ranked according to their potential to hinder cell proliferation (SL interactions) or, on the contrary, enhance clonal expansion (SV interactions) within the FA context. Features of the top 10 entries for gRNAs depletion or enrichment are indicated in Table 12, and the known or predicted interactions of these candidates with FANCA or between them illustrated in Figure 27 [292].

Table 12. Depleted and enriched top 10 candidates from the in-house screen. The hits found are reported for their role/biological process, involvement in tumorigenesis, and already documented SIs with FANCA or other FA genes.

Gene hits	Role/Biological process	Known tumorigenesis involvement	Known SIs with FANCA or FA genes
Depleted top 10 candidates			
Gene1	Transcription elongation factor	Yes	No
Gene2	Transcription elongation factor	Yes	No
Gene3	Transcription factor	Yes	No
Gene4	mRNA deadenylase	No	No
Gene5	Protein phosphatese	No	No
Gene6	E3 ubiquitin-protein ligase	No	No
Gene7	Tumor suppressor phosphatase involved in genome integrity maintenance	Yes	Yes
Gene8	Histone chaperone	No	No
Gene9	Endonuclease involved in DNA damage response	Yes	Yes
Gene10	Cyclin	No	No
Enriched top 10 candidates			
Gene1	Guanine nucleotide exchange factor	No	No
Gene2	Actin-capping protein	Yes	No
Gene3	Cellular myosin	Yes	No
Gene4	Component of a nucleosome associated complex	No	No
Gene5	Subunit of the ER membrane protein complex	No	No
Gene6	Integral membrane component	No	No
Gene7	Substrate-specific adapter of a E3 ubiquitin ligase complex	Yes	No
Gene8	RNA binding protein	No	No
Gene9	snRNP-binding protein	Yes	No
Gene10	Subunit of a chaperone complex	No	No

RNAP II: RNA polymerase II; ER: endoplasmic reticulum; snRNP: small nuclear ribonucleoprotein.

A


Gene8
Gene10
Gene2
Gene1
Gene7
MDM2
Gene3
Gene6
Gene9
Gene4
Gene5
FANCA

B

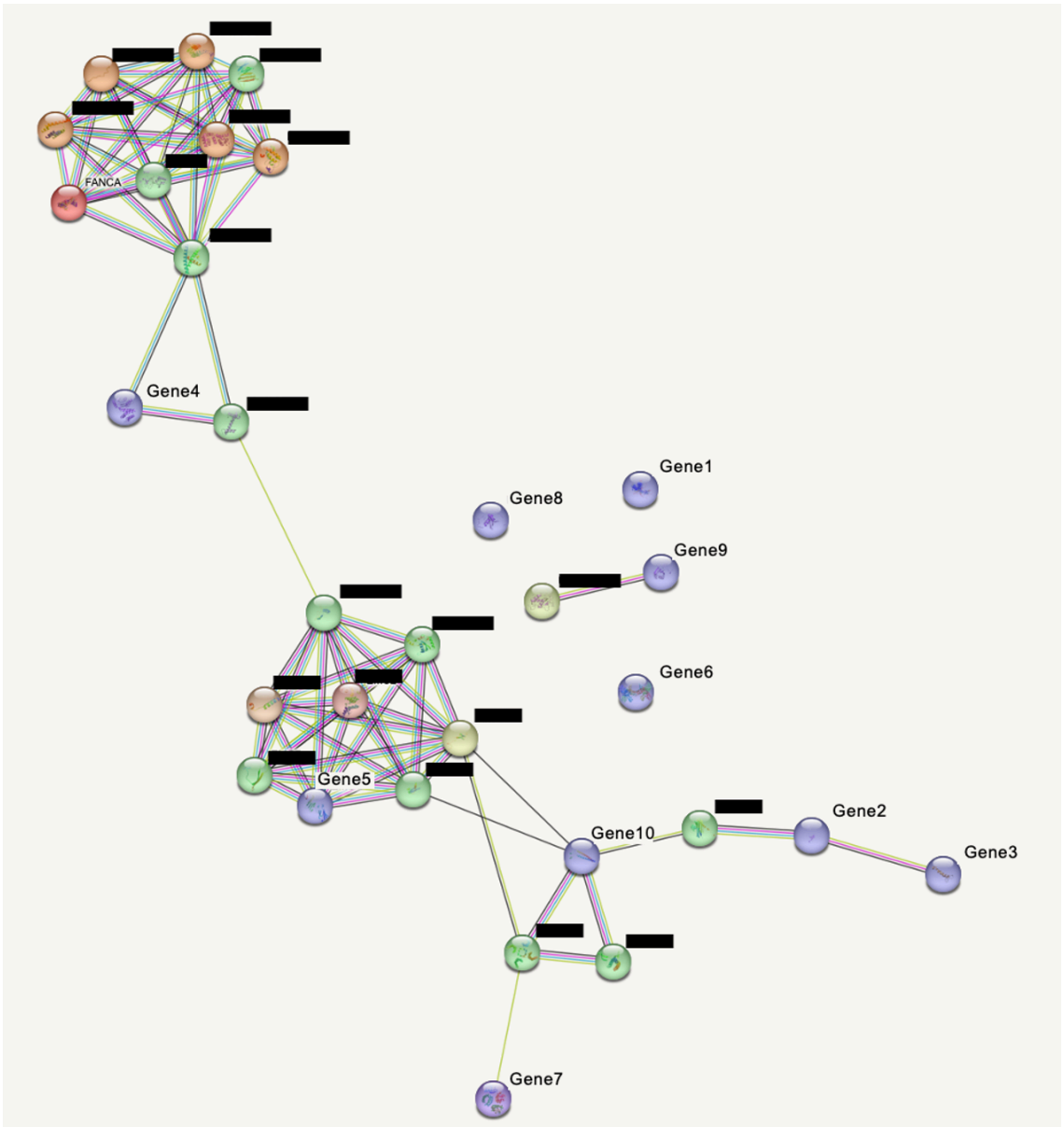

FANCA
Gene4
Gene1
Gene8
Gene9
Gene6
Gene5
Gene10
Gene2
Gene3
Gene7

Figure 27. STRING network of the predicted associations between FANCA and the depleted (A) and enriched (B) top 10 hits Entries and functional associations are represented as nodes and edges, respectively. FANCA is indicated as a red node, FA genes as orange, top 10 candidates as purple, and other interactors as green nodes. The final position of the nodes is computed to minimize the energy of the system; the physical distance between two of them along an edge has no meaning. Edges are depicted as lines colored depending on the association entity (light blue: database evidence; purple: experimental evidence; green: neighborhood evidence; red: fusion evidence; blue: concurrence evidence; yellow: textmining evidence; black: co-expression evidence) [292].

To sum up, among the best entries, we mainly observed genes implicated in replication, transcription, translation, chromatin remodeling, gene expression control, cellular response to xenobiotics, DNA damage repair, cell cycle checkpoint regulation, ubiquitination, and tumorigenesis; some of these were further known to be included within the same pathway and/or interact between each other. Moreover, we found candidates already annotated for their cooperation with the FA pathway and/or with documented SIs with FA genes, corroborating our initial data. We instead ascribed the lack of previously reported targets (*Introduction 1.7.3.2*) to the different conditions adopted in the studies compared to ours (e.g., CRISRP-Cas9 technology, screen focus, gRNA library, model system, biological challenge introduced).
The following stage will entail the examination of the NGS results of the TKOv3 screen, as illustrated above, and their comparison to those derived from the in-house one to select the most promising candidates for further functional validation and, thus, confirmation of their SI with the FA pathway.

5. CONCLUSIONS

FA is a disease that due to its many heritability patterns, wide genetic heterogeneity, private and complex variants, and notable mosaicism incidence, poses some very current challenges, first of all to provide a comprehensive molecular diagnosis.
Our prime attention was therefore devoted to the design and optimization of a mutation screening strategy, grounded in the conjunction of t-NGS and orthogonal techniques to test variant pathogenicity. This integrated line enabled us to successfully diagnose the totality of recruited patients from three different complementation groups (FA-A, FA-C, FA-D2), and to define 23 genetically distinct alleles, further elucidating a novel founder mutation. We also fine-tuned a statistical analysis to infer CNVs from the same t-NGS data [98]; for all the gross deletions predicted, we reported detection rates perfectly matching with those of the gold standards, carried out in parallel to assess the reliability of our method. Thus, based on these outcomes, we considered CNVs confirmation by additional assays to be redundant and needlessly time-consuming in case of well-characterized rearrangements via t-NGS.
Moreover, our strategy proved crucial for the thorough molecular characterization of patient P1, who alone exemplified some of the main obstacles encountered in FA diagnosis. We indeed ascertained his status of compound heterozygous for a large intragenic deletion, foretold by our CNVs analysis and then experimentally validated, and the already known c.2778+83C>G mutation of *FANCA*, directly identified by t-NGS despite its intronic location. Failing to mimic c.2778+83C>G documented splicing effect [64] via RT-PCR on P1's LCL, we sequenced proband's line DNA, and identified the *de novo* c.2778+86insT variant, not found in parents' PB and predicted to compensate 2778+83C>G. We indeed elucidated its capability to repristinate the physiological splicing mechanism within a minigene model, and in P1's LCL we linked a ~40% expression of FANCA to FANCD2-Ub and WT survival under clastogen exposure, implying a proliferative advantage of c.2778+86insT reverted clones at least *in vitro.* In line with these findings, P1 showed a mild somatic phenotype and a slow hematopoietic deterioration overtime, possibly due to c.2778+86insT emergence *in vivo*. We however reported this mutation only in two LCLs established two years apart, and not in proband's PB or BM specimens likely because of a limited percentage of cells carrying it. The lack of suitable BM material further precluded the chance to search the variant in distinct cell populations, and define the stage of the hematopoietic hierarchy of its emergence. Nevertheless, our findings highlight the relevance to investigate the natural "gene" therapy effect of c.2778+86insT or similar events for new personalized therapeutic strategies (e.g., genome editing with CRISPR/Cas9 system, synthesis of novel drugs acting on splicing), enabling more accurate treatments and decision plans, and especially sparing FA patients from the possible side effects of HSCT.

Optimized HSCT protocols have allowed a greater number of FA subjects to reach adulthood, but at the cost of exposing them to an increased solid tumor risk added up to the implicit cancer susceptibility of the disorder. Moreover, all the current therapies for FA focus

on the recovery of the sole hematological phenotype, while no cure tackles the “solid tumor dilemma” recognized as the main reason of patients’ death nowadays.
To address the need of new systemic drug-based approaches for FA, we thus embraced a dual strategy. Along the lines of Montanuy *et al*.’s work [169], we initially established cellular models suitable for HCS, but this time we enhanced them with the insertion of three frequent FANCA NT mutants (Arg951Gln, Thr1131Ala, Phe1263del) to boost the chances to find molecule/s reactivating FA pathway via their phenotypic correction. In this perspective, we transduced lentiviral vectors carrying the variants of interest into U2OS FANCA$^{-/-}$-YFP-FANCD2 line by Montanuy *et al*. [169], deficient in FANCD2-Ub and fluorescent foci upon DNA damage. We next proceeded to the functional validation of the models generated to assess their capability to mimic patients’ cellular phenotypes. Despite confirming FANCA stable expression in all lines as expected [3], [87], we reported no FANCD2-Ub in the sole Phe1263del model, while ubiquitination ratios compatible to those of the FANCA^{corrected} control for the other two mutants. These values were indeed proportional to the percentage amount of FANCA located at nuclear level, negligible in Phe1263del system and close to 40% in the others. We consistently linked Phe1263del mutant to patterns of cell survival and G2/M block under DEB exposure matching with those of the FANCA$^{-/-}$ control, and again Arg951Gln and Thr1131Ala to WT-like behaviors. Moreover, we demonstrated as only Phe1263del line failed to promote the formation of YFP-FANCD2 foci, which instead were clearly detectable in the nuclei of the remaining ones. Taken together, our results accented the unlikelihood to distinguish between Arg951Gln and Thr1131Ala WT-like phenotypes within our cellular models and their actual correction within the HCS, leading us to question their characterization or pathogenic role, and to exclude them from further studies. We instead showed the highly frequent Phe1263del mutant [3] to be a profitable candidate for the HCS, which we will conduct exploiting YFP- FANCD2 foci formation as readout as previously done [169], [180]. The pharmacological correction of the variant would provide a sizeable subset of patients with an unprecedented possibility for a new and less toxic therapy towards all FA clinical features, with particular regard to solid tumors. Moreover, the identification of FDA-approved drug/s, included among the candidates, able to correct Phe1263del phenotype would ensure a novel personalized medical approach with an immediate translational applicability.
To continue to touch upon novel options for a systemic treatment in FA and, in addition, for sporadic tumors with mutations in a FA gene or one of its interactors, we also embarked on the investigation of SIs with FA deficiency by GW CRISPRko screens. In detail, we separately transduced the Cas9-expressing FANCA KO line and its control with two GW KO gRNAs libraries, such as the commercial TKOv3 and an in-house one, taking advantage of their target overlap to increase the reliability of the final hits. After infection, we performed antibiotic selection to preserve the only transduced cells, and did not introduce any further biological challenge over the experiments; our objective was indeed to examine the representativeness of each end phenotype resulting from the proliferative competition between cells and, thus, the specific gene KO produced in them. We periodically split the cells in independent biological replicates, and monitored the number of cell doublings and library coverage to keep the desired gRNAs/phenotypes representation, ending the screens at ~10 cell divisions considered enough for a significant PFC. Subsequently, we extracted

the gDNA from all the replicates, and amplified it with barcode-tagged primers for massive parallel sequencing. Now, we have already performed the bioinformatic and statical analyses of the in-house screen, based on gRNA sequencing-based counting and inter-sample comparison, and obtained a preliminary list of depleted (SL) and enriched (SV) entries. Among the top hits belonging to the two subsets, we predominantly identified genes with a role in replication, transcription, translation, chromatin organization, gene expression, cell response to clastogens and DNA repair, checkpoints control, and carcinogenesis. We also reported candidates already known for their synthetic association with FA pathway and/or interacting with it, further proving the quality of our data. We will then contrast these results with the TKOv3 screen ones, just finished to be sequenced, to identify joint candidates with similar ranking positions, prioritizing those involved in the same pathway and with commercially available drugs/inhibitors. As a final step, we will validate the best hits under the exposure of distinct biological challenges and/or in different models to verify the existence of SIs with FA deficiency, and thus dissect novel druggable targets for FA therapy (SV genes) and extend the potential of the inhibitors of FA pathway or its interactors to treat sporadic cancers within the general population (SL genes).

To sum up, this thesis project sough to lay the foundations for the possible resolution of the main limitations now met in FA molecular diagnosis and therapeutics. We developed an integrated mutation screening approach, providing a rapid and exhaustive characterization of all patients enrolled, independently from their complementation group, mutation type, and mosaicism status. In parallel, we established FA cellular models suitable for HCS and CRISPRko screens to correct highly frequent variants and find candidates for new SIs, respectively, generating knowledge to be translated in the clinical practice for the prevention and/or treatment of any FA feature.

.

.

6. BIBLIOGRAPHY

[1] “Fanconi Anemia Research Fund.” https://www.fanconi.org/clinical-care/clinical-care-guidelines-home (accessed Aug. 12, 2022).

[2] “Fanconi G. Famili¨are infantile perniziosaartige An¨amie (pernizi¨oses Blutbild und Konstitution). Jahrbuch f¨ur Kinderheilkunde und physische 1927;117:257e80.

[3] M. Castella *et al.*, “Origin, functional role, and clinical impact of Fanconi anemia FANCA mutations,” *Blood*, vol. 117, no. 14, pp. 3759–3769, Apr. 2011, doi: 10.1182/blood-2010-08-299917.

[4] J. Minguillón and J. Surrallés, “Therapeutic research in the crystal chromosome disease Fanconi anemia,” *Mutat. Res. Genet. Toxicol. Environ. Mutagen.*, vol. 836, no. Pt A, pp. 104–108, Dec. 2018, doi: 10.1016/j.mrgentox.2018.05.012.

[5] A. M. Risitano, S. Marotta, R. Calzone, F. Grimaldi, and A. Zatterale, “Twenty years of the Italian Fanconi Anemia Registry: where we stand and what remains to be learned,” *Haematologica*, vol. 101, no. 3, pp. 319–327, Mar. 2016, doi: 10.3324/haematol.2015.133520.

[6] B. P. Alter, N. Giri, S. A. Savage, and P. S. Rosenberg, “Cancer in the National Cancer Institute inherited bone marrow failure syndrome cohort after fifteen years of follow-up,” *Haematologica*, vol. 103, no. 1, pp. 30–39, Jan. 2018, doi: 10.3324/haematol.2017.178111.

[7] D. I. Kutler *et al.*, “A 20-year perspective on the International Fanconi Anemia Registry (IFAR),” *Blood*, vol. 101, no. 4, pp. 1249–1256, Feb. 2003, doi: 10.1182/blood-2002-07-2170.

[8] P. S. Rosenberg, M. H. Greene, and B. P. Alter, “Cancer incidence in persons with Fanconi anemia,” *Blood*, vol. 101, no. 3, pp. 822–826, Feb. 2003, doi: 10.1182/blood-2002-05-1498.

[9] M. Bogliolo and J. Surrallés, “Fanconi anemia: a model disease for studies on human genetics and advanced therapeutics,” *Curr. Opin. Genet. Dev.*, vol. 33, pp. 32–40, Aug. 2015, doi: 10.1016/j.gde.2015.07.002.

[10] D. I. Kutler *et al.*, “Natural history and management of Fanconi anemia patients with head and neck cancer: A 10-year follow-up,” *The Laryngoscope*, vol. 126, no. 4, pp. 870–879, Apr. 2016, doi: 10.1002/lary.25726.

[11] B. D. Solomon *et al.*, “An approach to the identification of anomalies and etiologies in neonates with identified or suspected VACTERL (vertebral defects, anal atresia, tracheo-esophageal fistula with esophageal atresia, cardiac anomalies, renal anomalies, and limb anomalies) association,” *J. Pediatr.*, vol. 164, no. 3, pp. 451-457.e1, Mar. 2014, doi:

10.1016/j.jpeds.2013.10.086.

[12] B. P. Alter and N. Giri, “Thinking of VACTERL-H? Rule out Fanconi Anemia according to PHENOS,” *Am. J. Med. Genet. A.*, vol. 170, no. 6, pp. 1520–1524, Jun. 2016, doi: 10.1002/ajmg.a.37637.

[13] B. Altintas, N. Giri, L. J. McReynolds, A. Best, and B. P. Alter, “Genotype-phenotype and outcome associations in patients with Fanconi anemia: The National Cancer Institute cohort,” *Haematologica*, Apr. 2022, doi: 10.3324/haematol.2021.279981.

[14] A. D. Auerbach, “Fanconi anemia and its diagnosis,” *Mutat. Res.*, vol. 668, no. 1–2, pp. 4–10, Jul. 2009, doi: 10.1016/j.mrfmmm.2009.01.013.

[15] M. Levitus, H. Joenje, and J. P. de Winter, “The Fanconi anemia pathway of genomic maintenance,” *Cell. Oncol. Off. J. Int. Soc. Cell. Oncol.*, vol. 28, no. 1–2, pp. 3–29, 2006, doi: 10.1155/2006/974975.

[16] W. Wang, “Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins,” *Nat. Rev. Genet.*, vol. 8, no. 10, pp. 735–748, Oct. 2007, doi: 10.1038/nrg2159.

[17] A. D. Auerbach, “A test for Fanconi’s anemia,” *Blood*, vol. 72, no. 1, pp. 366–367, Jul. 1988.

[18] A. D. Auerbach, “Diagnosis of fanconi anemia by diepoxybutane analysis,” *Curr. Protoc. Hum. Genet.*, vol. Chapter 8, p. Unit 8.7, Jul. 2003, doi: 10.1002/0471142905.hg0807s37.

[19] M. Castella *et al.*, “Chromosome fragility in patients with Fanconi anaemia: diagnostic implications and clinical impact,” *J. Med. Genet.*, vol. 48, no. 4, pp. 242–250, Apr. 2011, doi: 10.1136/jmg.2010.084210.

[20] J. R. Lo Ten Foe *et al.*, “Somatic mosaicism in Fanconi anemia: molecular basis and clinical significance,” *Eur. J. Hum. Genet. EJHG*, vol. 5, no. 3, pp. 137–148, Jun. 1997.

[21] J. Soulier *et al.*, “Detection of somatic mosaicism and classification of Fanconi anemia patients by analysis of the FA/BRCA pathway,” *Blood*, vol. 105, no. 3, pp. 1329–1336, Feb. 2005, doi: 10.1182/blood-2004-05-1852.

[22] J. J. Gregory *et al.*, “Somatic mosaicism in Fanconi anemia: evidence of genotypic reversion in lymphohematopoietic stem cells,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, no. 5, pp. 2532–2537, Feb. 2001, doi: 10.1073/pnas.051609898.

[23] M. Gross *et al.*, “Reverse mosaicism in Fanconi anemia: natural gene therapy via molecular self-correction,” *Cytogenet. Genome Res.*, vol. 98, no. 2–3, pp. 126–135, 2002, doi: 10.1159/000069805.

[24] M. J. Ramírez *et al.*, “Natural gene therapy by reverse mosaicism leads to improved hematology in Fanconi anemia patients,” *Am. J. Hematol.*, vol. 96, no. 8, pp. 989–999, Aug.

2021, doi: 10.1002/ajh.26234.

[25] R. S. Asur *et al.*, “Somatic mosaicism of an intragenic FANCB duplication in both fibroblast and peripheral blood cells observed in a Fanconi anemia patient leads to milder phenotype,” *Mol. Genet. Genomic Med.*, vol. 6, no. 1, pp. 77–91, Nov. 2017, doi: 10.1002/mgg3.350.

[26] J. H. Fargo, A. Rochowski, N. Giri, S. A. Savage, S. B. Olson, and B. P. Alter, “Comparison of chromosome breakage in non-mosaic and mosaic patients with Fanconi anemia, relatives, and patients with other inherited bone marrow failure syndromes,” *Cytogenet. Genome Res.*, vol. 144, no. 1, pp. 15–27, 2014, doi: 10.1159/000366251.

[27] R. Kalb *et al.*, “Hypomorphic mutations in the gene encoding a key Fanconi anemia protein, FANCD2, sustain a significant group of FA-D2 patients with severe phenotype,” *Am. J. Hum. Genet.*, vol. 80, no. 5, pp. 895–910, May 2007, doi: 10.1086/517616.

[28] R. Bottega *et al.*, “Hypomorphic FANCA mutations correlate with mild mitochondrial and clinical phenotype in Fanconi anemia,” *Haematologica*, vol. 103, no. 3, pp. 417–426, Mar. 2018, doi: 10.3324/haematol.2017.176131.

[29] M. C. Kottemann and A. Smogorzewska, “Fanconi anaemia and the repair of Watson and Crick DNA crosslinks,” *Nature*, vol. 493, no. 7432, pp. 356–363, Jan. 2013, doi: 10.1038/nature11863.

[30] J. M. Kim, Y. Kee, A. Gurtan, and A. D. D’Andrea, “Cell cycle-dependent chromatin loading of the Fanconi anemia core complex by FANCM/FAAP24,” *Blood*, vol. 111, no. 10, pp. 5215–5222, May 2008, doi: 10.1182/blood-2007-09-113092.

[31] S. Shakeel *et al.*, “Structure of the Fanconi anaemia monoubiquitin ligase complex,” *Nature*, vol. 575, no. 7781, pp. 234–237, Nov. 2019, doi: 10.1038/s41586-019-1703-4.

[32] Y. J. Machida *et al.*, “UBE2T is the E2 in the Fanconi anemia pathway and undergoes negative autoregulation,” *Mol. Cell*, vol. 23, no. 4, pp. 589–596, Aug. 2006, doi: 10.1016/j.molcel.2006.06.024.

[33] C. S. Tremblay *et al.*, “HES1 is a novel interactor of the Fanconi anemia core complex,” *Blood*, vol. 112, no. 5, pp. 2062–2070, Sep. 2008, doi: 10.1182/blood-2008-04-152710.

[34] G.-L. Moldovan and A. D. D’Andrea, “How the fanconi anemia pathway guards the genome,” *Annu. Rev. Genet.*, vol. 43, pp. 223–249, 2009, doi: 10.1146/annurev-genet-102108-134222.

[35] I. Garcia-Higuera *et al.*, “Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway,” *Mol. Cell*, vol. 7, no. 2, pp. 249–262, Feb. 2001, doi: 10.1016/s1097-2765(01)00173-3.

[36] M. Bogliolo *et al.*, “Histone H2AX and Fanconi anemia FANCD2 function in the same

pathway to maintain chromosome stability," *EMBO J.*, vol. 26, no. 5, pp. 1340–1351, Mar. 2007, doi: 10.1038/sj.emboj.7601574.

[37] E. Rajendra *et al.*, "The genetic and biochemical basis of FANCD2 monoubiquitination," *Mol. Cell*, vol. 54, no. 5, pp. 858–869, Jun. 2014, doi: 10.1016/j.molcel.2014.05.001.

[38] P. Swuec *et al.*, "The FA Core Complex Contains a Homo-dimeric Catalytic Module for the Symmetric Mono-ubiquitination of FANCI-FANCD2," *Cell Rep.*, vol. 18, no. 3, pp. 611–623, Jan. 2017, doi: 10.1016/j.celrep.2016.11.013.

[39] W. Niedzwiedz, G. Mosedale, M. Johnson, C. Y. Ong, P. Pace, and K. J. Patel, "The Fanconi anaemia gene FANCC promotes homologous recombination and error-prone DNA repair," *Mol. Cell*, vol. 15, no. 4, pp. 607–620, Aug. 2004, doi: 10.1016/j.molcel.2004.08.009.

[40] M. Räschle *et al.*, "Mechanism of replication-coupled DNA interstrand crosslink repair," *Cell*, vol. 134, no. 6, pp. 969–980, Sep. 2008, doi: 10.1016/j.cell.2008.08.030.

[41] P. Knipscheer *et al.*, "The Fanconi Anemia Pathway Promotes Replication-Dependent DNA Interstrand Cross-Link Repair," *Science*, vol. 326, no. 5960, pp. 1698–1701, Dec. 2009, doi: 10.1126/science.1182372.

[42] H. Kim, K. Yang, D. Dejsuphong, and A. D. D'Andrea, "Regulation of Rev1 by the Fanconi Anemia Core Complex," *Nat. Struct. Mol. Biol.*, vol. 19, no. 2, pp. 164–170, Jan. 2012, doi: 10.1038/nsmb.2222.

[43] H. Walden and A. J. Deans, "The Fanconi anemia DNA repair pathway: structural and functional insights into a complex disorder," *Annu. Rev. Biophys.*, vol. 43, pp. 257–278, 2014, doi: 10.1146/annurev-biophys-051013-022737.

[44] A. F. Alpi and K. J. Patel, "Monoubiquitylation in the Fanconi anemia DNA damage response pathway," *DNA Repair*, vol. 8, no. 4, pp. 430–435, Apr. 2009, doi: 10.1016/j.dnarep.2009.01.019.

[45] C. Jacquemont and T. Taniguchi, "The Fanconi anemia pathway and ubiquitin," *BMC Biochem.*, vol. 8, no. 1, p. S10, Nov. 2007, doi: 10.1186/1471-2091-8-S1-S10.

[46] G. Pagano *et al.*, "From clinical description, to in vitro and animal studies, and backward to patients: oxidative stress and mitochondrial dysfunction in Fanconi anemia," *Free Radic. Biol. Med.*, vol. 58, pp. 118–125, May 2013, doi: 10.1016/j.freeradbiomed.2013.01.015.

[47] P. Shyamsunder *et al.*, "Impaired mitophagy in Fanconi anemia is dependent on mitochondrial fission," *Oncotarget*, vol. 7, no. 36, pp. 58065–58074, Aug. 2016, doi: 10.18632/oncotarget.11161.

[48] R. Sumpter and B. Levine, "Emerging functions of the Fanconi anemia pathway at a glance," *J. Cell Sci.*, vol. 130, no. 16, pp. 2657–2662, Aug. 2017, doi: 10.1242/jcs.204909.

[49] S. Ravera, C. Dufour, P. Degan, and E. Cappelli, “Fanconi anemia: from DNA repair to metabolism,” *Eur. J. Hum. Genet. EJHG*, vol. 26, no. 4, pp. 475–476, Apr. 2018, doi: 10.1038/s41431-017-0046-6.

[50] R. Sumpter and B. Levine, “Novel functions of Fanconi anemia proteins in selective autophagy and inflammation,” *Oncotarget*, vol. 7, no. 32, pp. 50820–50821, Aug. 2016, doi: 10.18632/oncotarget.10970.

[51] H. Karttunen *et al.*, “Co-opting the Fanconi anemia genomic stability pathway enables herpesvirus DNA synthesis and productive growth,” *Mol. Cell*, vol. 55, no. 1, pp. 111–122, Jul. 2014, doi: 10.1016/j.molcel.2014.05.020.

[52] S. L. Sauter *et al.*, “Oral human papillomavirus is common in individuals with Fanconi anemia,” *Cancer Epidemiol. Biomarkers Prev.*, vol. 24, no. 5, pp. 864–872, May 2015, doi: 10.1158/1055-9965.epi-15-0097-t.

[53] M. D. Tischkowitz and S. V. Hodgson, “Fanconi anaemia,” *J. Med. Genet.*, vol. 40, no. 1, pp. 1–10, Jan. 2003, doi: 10.1136/jmg.40.1.1.

[54] H. Hanson, C. G. Mathew, Z. Docherty, and C. Mackie Ogilvie, “Telomere shortening in Fanconi anaemia demonstrated by a direct FISH approach,” *Cytogenet. Cell Genet.*, vol. 93, no. 3–4, pp. 203–206, 2001, doi: 10.1159/000056985.

[55] Y. Li, S. Amarachintha, A. F. Wilson, X. Li, and W. Du, “Persistent response of Fanconi anemia haematopoietic stem and progenitor cells to oxidative stress,” *Cell Cycle Georget. Tex*, vol. 16, no. 12, pp. 1201–1209, Jun. 2017, doi: 10.1080/15384101.2017.1320627.

[56] M. Bogliolo *et al.*, “Optimised molecular genetic diagnostics of Fanconi anaemia by whole exome sequencing and functional studies,” *J. Med. Genet.*, vol. 57, no. 4, pp. 258–268, Apr. 2020, doi: 10.1136/jmedgenet-2019-106249.

[57] P. A. Mehta and C. Ebens, *Fanconi Anemia*. University of Washington, Seattle, 2021. Accessed: Aug. 15, 2022. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK1401/

[58] “The Rockefeller University » Fanconi Anemia Mutation Database.” https://www2.rockefeller.edu/fanconi/ (accessed Aug. 15, 2022).

[59] N. G. Howlett *et al.*, “Biallelic inactivation of BRCA2 in Fanconi anemia,” *Science*, vol. 297, no. 5581, pp. 606–609, Jul. 2002, doi: 10.1126/science.1073834.

[60] A. R. Meetei *et al.*, “A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M,” *Nat. Genet.*, vol. 37, no. 9, Art. no. 9, Sep. 2005, doi: 10.1038/ng1626.

[61] T. R. Singh *et al.*, “Impaired FANCD2 monoubiquitination and hypersensitivity to camptothecin uniquely characterize Fanconi anemia complementation group M,” *Blood*, vol.

114, no. 1, pp. 174–180, Jul. 2009, doi: 10.1182/blood-2009-02-207811.

[62] E. T. Lim *et al.*, “Distribution and Medical Impact of Loss-of-Function Variants in the Finnish Founder Population,” *PLOS Genet.*, vol. 10, no. 7, p. e1004494, Jul. 2014, doi: 10.1371/journal.pgen.1004494.

[63] H. E. Shamseldin, M. Elfaki, and F. S. Alkuraya, “Exome sequencing reveals a novel Fanconi group defined by XRCC2 mutation,” *J. Med. Genet.*, vol. 49, no. 3, pp. 184–186, Mar. 2012, doi: 10.1136/jmedgenet-2011-100585.

[64] M. Savino *et al.*, “Mutations of the Fanconi anemia group A gene (FAA) in Italian patients.,” *Am. J. Hum. Genet.*, vol. 61, no. 6, pp. 1246–1253, Dec. 1997.

[65] C. Bouchlaka *et al.*, “Fanconi anemia in Tunisia: high prevalence of group A and identification of new FANCA mutations,” *J. Hum. Genet.*, vol. 48, no. 7, Art. no. 7, Jul. 2003, doi: 10.1007/s10038-003-0037-z.

[66] D. De Rocco *et al.*, “Molecular analysis of Fanconi anemia: the experience of the Bone Marrow Failure Study Group of the Italian Association of Pediatric Onco-Hematology,” *Haematologica*, vol. 99, no. 6, pp. 1022–1031, Jun. 2014, doi: 10.3324/haematol.2014.104224.

[67] A. Palovcak, W. Liu, F. Yuan, and Y. Zhang, “Maintenance of genome stability by Fanconi anemia proteins,” *Cell Biosci.*, vol. 7, no. 1, p. 8, Feb. 2017, doi: 10.1186/s13578-016-0134-2.

[68] D. C. Kimble *et al.*, “A comprehensive approach to identification of pathogenic FANCA variants in Fanconi anemia patients and their families,” *Hum. Mutat.*, vol. 39, no. 2, pp. 237–254, 2018, doi: 10.1002/humu.23366.

[69] D. I. Kutler and A. D. Auerbach, “Fanconi anemia in Ashkenazi Jews,” *Fam. Cancer*, vol. 3, no. 3, pp. 241–248, Sep. 2004, doi: 10.1007/s10689-004-9565-8.

[70] A. T. Wang and A. Smogorzewska, “SnapShot: Fanconi anemia and associated proteins,” *Cell*, vol. 160, no. 1–2, pp. 354-354.e1, Jan. 2015, doi: 10.1016/j.cell.2014.12.031.

[71] A. J. Tipping *et al.*, “Molecular and genealogical evidence for a founder effect in Fanconi anemia families of the Afrikaner population of South Africa,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, no. 10, pp. 5734–5739, May 2001, doi: 10.1073/pnas.091402398.

[72] E. Callén *et al.*, “A common founder mutation in FANCA underlies the world’s highest prevalence of Fanconi anemia in Gypsy families from Spain,” *Blood*, vol. 105, no. 5, pp. 1946–1949, Mar. 2005, doi: 10.1182/blood-2004-07-2588.

[73] A. Amouri *et al.*, “High frequency of exon 15 deletion in the FANCA gene in Tunisian patients affected with Fanconi anemia disease: implication for diagnosis,” *Mol. Genet. Genomic Med.*, vol. 2, no. 2, pp. 160–165, Mar. 2014, doi: 10.1002/mgg3.55.

[74] O. Levran, R. Diotti, K. Pujara, S. D. Batish, H. Hanenberg, and A. D. Auerbach,

“Spectrum of sequence variations in the FANCA gene: an International Fanconi Anemia Registry (IFAR) study,” *Hum. Mutat.*, vol. 25, no. 2, pp. 142–149, Feb. 2005, doi: 10.1002/humu.20125.

[75] L. Ianzano *et al.*, “The Genomic Organization of the Fanconi Anemia Group A (FAA) Gene,” *Genomics*, vol. 41, no. 3, pp. 309–314, May 1997, doi: 10.1006/geno.1997.4675.

[76] “FANCA FA complementation group A [Homo sapiens (human)] - Gene - NCBI.” https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=2175 (accessed Aug. 16, 2022).

[77] “FANCA - Fanconi anemia group A protein - Homo sapiens (Human) | UniProtKB | UniProt.” https://www.uniprot.org/uniprotkb/O15360/entry#names_and_taxonomy (accessed Aug. 14, 2022).

[78] J. Lightfoot, N. Alon, L. Bosnoyan-Collins, and M. Buchwald, “Characterization of regions functional in the nuclear localization of the Fanconi anemia group A protein,” *Hum. Mol. Genet.*, vol. 8, no. 6, pp. 1007–1015, Jun. 1999, doi: 10.1093/hmg/8.6.1007.

[79] M. Ferrer, J. A. Rodríguez, E. A. Spierings, J. P. de Winter, G. Giaccone, and F. A. E. Kruyt, “Identification of multiple nuclear export sequences in Fanconi anemia group A protein that contribute to CRM1-dependent nuclear export,” *Hum. Mol. Genet.*, vol. 14, no. 10, pp. 1271–1281, May 2005, doi: 10.1093/hmg/ddi138.

[80] Q. Waisfisz *et al.*, “A physical complex of the Fanconi anemia proteins FANCG/XRCC9 and FANCA,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 96, no. 18, pp. 10320–10325, Aug. 1999.

[81] A. Folias *et al.*, “BRCA1 interacts directly with the Fanconi anemia protein FANCA,” *Hum. Mol. Genet.*, vol. 11, no. 21, pp. 2591–2597, Oct. 2002, doi: 10.1093/hmg/11.21.2591.

[82] A. M. Ali *et al.*, “FAAP20: a novel ubiquitin-binding FA nuclear core-complex protein required for functional integrity of the FA-BRCA DNA repair pathway,” *Blood*, vol. 119, no. 14, pp. 3285–3294, Apr. 2012, doi: 10.1182/blood-2011-10-385963.

[83] T. Otsuki *et al.*, “Phosphorylation of Fanconi Anemia Protein, FANCA, Is Regulated by Akt Kinase,” *Biochem. Biophys. Res. Commun.*, vol. 291, no. 3, pp. 628–634, Mar. 2002, doi: 10.1006/bbrc.2002.6504.

[84] N. B. Collins *et al.*, “ATR-dependent phosphorylation of FANCA on serine 1449 after DNA damage is important for FA pathway function,” *Blood*, vol. 113, no. 10, pp. 2181–2190, Mar. 2009, doi: 10.1182/blood-2008-05-154294.

[85] H. Yagasaki, S. Hamanoue, T. Oda, T. Nakahata, S. Asano, and T. Yamashita, “Identification and characterization of novel mutations of the major Fanconi anemia gene FANCA in the Japanese population,” *Hum. Mutat.*, vol. 24, no. 6, pp. 481–490, 2004, doi: 10.1002/humu.20099.

[86] “The FANCA gene homepage - Global Variome shared LOVD.” https://databases.lovd.nl/shared/genes/FANCA (accessed May 08, 2022).

[87] D. Adachi *et al.*, “Heterogeneous activation of the Fanconi anemia pathway by patient-derived FANCA mutants,” *Hum. Mol. Genet.*, vol. 11, no. 25, pp. 3125–3134, Dec. 2002, doi: 10.1093/hmg/11.25.3125.

[88] A. Mankad *et al.*, “Natural gene therapy in monozygotic twins with Fanconi anemia,” *Blood*, vol. 107, no. 8, pp. 3084–3090, Apr. 2006, doi: 10.1182/blood-2005-07-2638.

[89] N. V. Morgan, A. J. Tipping, H. Joenje, and C. G. Mathew, “High frequency of large intragenic deletions in the Fanconi anemia group A gene,” *Am. J. Hum. Genet.*, vol. 65, no. 5, pp. 1330–1341, Nov. 1999, doi: 10.1086/302627.

[90] E. Callén *et al.*, “Quantitative PCR analysis reveals a high incidence of large intragenic deletions in the FANCA gene in Spanish Fanconi anemia patients,” *Cytogenet. Genome Res.*, vol. 104, no. 1–4, pp. 341–345, 2004, doi: 10.1159/000077513.

[91] E. K. Flynn *et al.*, “Comprehensive analysis of pathogenic deletion variants in Fanconi anemia genes,” *Hum. Mutat.*, vol. 35, no. 11, pp. 1342–1353, Nov. 2014, doi: 10.1002/humu.22680.

[92] C. Mattioli *et al.*, “Unusual splice site mutations disrupt FANCA exon 8 definition,” *Biochim. Biophys. Acta*, vol. 1842, no. 7, pp. 1052–1058, Jul. 2014, doi: 10.1016/j.bbadis.2014.03.014.

[93] N. Magdalena *et al.*, “Frequency of Fanconi anemia in Brazil and efficacy of screening for the FANCA 3788-3790del mutation,” *Braz. J. Med. Biol. Res.*, vol. 38, pp. 669–673, May 2005, doi: 10.1590/S0100-879X2005000500003.

[94] “HGMD® home page.” https://www.hgmd.cf.ac.uk/ac/index.php (accessed Aug. 17, 2022).

[95] J. Antonio Casado *et al.*, “A comprehensive strategy for the subtyping of patients with Fanconi anaemia: conclusions from the Spanish Fanconi Anemia Research Network,” *J. Med. Genet.*, vol. 44, no. 4, pp. 241–249, Apr. 2007, doi: 10.1136/jmg.2006.044719.

[96] N. Ameziane *et al.*, “Diagnosis of fanconi anemia: mutation analysis by next-generation sequencing,” *Anemia*, vol. 2012, p. 132856, 2012, doi: 10.1155/2012/132856.

[97] S. C. Chandrasekharappa *et al.*, “Massively parallel sequencing, aCGH, and RNA-Seq technologies provide a comprehensive molecular diagnosis of Fanconi anemia,” *Blood*, vol. 121, no. 22, pp. e138-148, May 2013, doi: 10.1182/blood-2012-12-474585.

[98] E. Nicchia *et al.*, “Identification of point mutations and large intragenic deletions in Fanconi anemia using next-generation sequencing technology,” *Mol. Genet. Genomic Med.*, vol. 3, no. 6, pp. 500–512, Jul. 2015, doi: 10.1002/mgg3.160.

[99] I. Ghemlas *et al.*, “Improving diagnostic precision, care and syndrome definitions using comprehensive next-generation sequencing for the inherited bone marrow failure syndromes,” *J. Med. Genet.*, vol. 52, no. 9, pp. 575–584, Sep. 2015, doi: 10.1136/jmedgenet-2015-103270.

[100] “Home - OMIM.” https://www.omim.org/ (accessed Nov. 28, 2022).

[101] S. M. Teo, Y. Pawitan, C. S. Ku, K. S. Chia, and A. Salim, “Statistical challenges associated with detecting copy number variations with next-generation sequencing,” *Bioinforma. Oxf. Engl.*, vol. 28, no. 21, pp. 2711–2718, Nov. 2012, doi: 10.1093/bioinformatics/bts535.

[102] J. M. Moreno-Cabrera *et al.*, “Evaluation of CNV detection tools for NGS panel data in genetic diagnostics,” *Eur. J. Hum. Genet.*, vol. 28, no. 12, Art. no. 12, Dec. 2020, doi: 10.1038/s41431-020-0675-z.

[103] G. Demidov, M. Sturm, and S. Ossowski, “ClinCNV: multi-sample germline CNV detection in NGS data,” *bioRxiv*, p. 2022.06.10.495642, Jan. 2022, doi: 10.1101/2022.06.10.495642.

[104] S. Lauhasurayotin *et al.*, “Reanalysing genomic data by normalized coverage values uncovers CNVs in bone marrow failure gene panels,” *NPJ Genomic Med.*, vol. 4, 2019, doi: 10.1038/s41525-019-0104-9.

[105] S. Richards *et al.*, “Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology,” *Genet. Med. Off. J. Am. Coll. Med. Genet.*, vol. 17, no. 5, pp. 405–424, May 2015, doi: 10.1038/gim.2015.30.

[106] J. Soulier, “Fanconi anemia,” *Hematol. Am. Soc. Hematol. Educ. Program*, vol. 2011, pp. 492–497, 2011, doi: 10.1182/asheducation-2011.1.492.

[107] J. R. Lo Ten Foe *et al.*, “Somatic mosaicism in Fanconi anemia: molecular basis and clinical significance,” *Eur. J. Hum. Genet. EJHG*, vol. 5, no. 3, pp. 137–148, Jun. 1997.

[108] R. Kalb, K. Neveling, I. Nanda, D. Schindler, and H. Hoehn, “Fanconi anemia: causes and consequences of genetic instability,” *Genome Dyn.*, vol. 1, pp. 218–242, 2006, doi: 10.1159/000092510.

[109] A. D. Auerbach, B. Adler, R. J. O’Reilly, D. Kirkpatrick, and R. S. Chaganti, “Effect of procarbazine and cyclophosphamide on chromosome breakage in Fanconi anemia cells: relevance to bone marrow transplantation,” *Cancer Genet. Cytogenet.*, vol. 9, no. 1, pp. 25–36, May 1983, doi: 10.1016/0165-4608(83)90021-3.

[110] E. Gluckman, R. Berger, and J. Dutreix, “Bone marrow transplantation for Fanconi anemia,” *Semin. Hematol.*, vol. 21, no. 1, pp. 20–26, Jan. 1984.

[111] C. L. Ebens, M. L. MacMillan, and J. E. Wagner, “Hematopoietic cell transplantation

in Fanconi anemia: current evidence, challenges and recommendations," *Expert Rev. Hematol.*, vol. 10, no. 1, pp. 81–97, Jan. 2017, doi: 10.1080/17474086.2016.1268048.

[112] M. L. MacMillan, M. R. Hughes, S. Agarwal, and G. Q. Daley, "Cellular Therapy for Fanconi Anemia: The Past, Present, and Future," *Biol. Blood Marrow Transplant.*, vol. 17, no. 1, pp. S109–S114, Jan. 2011, doi: 10.1016/j.bbmt.2010.11.027.

[113] M. C. Yoder, "Cord blood banking and transplantation: advances and controversies," *Curr. Opin. Pediatr.*, vol. 26, no. 2, pp. 163–168, Apr. 2014, doi: 10.1097/MOP.0000000000000065.

[114] Y. Verlinsky, S. Rechitsky, W. Schoolcraft, C. Strom, and A. Kuliev, "Preimplantation diagnosis for Fanconi anemia combined with HLA matching," *JAMA*, vol. 285, no. 24, pp. 3130–3133, Jun. 2001, doi: 10.1001/jama.285.24.3130.

[115] C. Masserot-Lureau *et al.*, "Incidence of liver abnormalities in Fanconi anemia patients," *Am. J. Hematol.*, vol. 87, no. 5, pp. 547–549, May 2012, doi: 10.1002/ajh.23153.

[116] K. Scheckenbach *et al.*, "Treatment of the bone marrow failure in Fanconi anemia patients with danazol," *Blood Cells. Mol. Dis.*, vol. 48, no. 2, pp. 128–131, Feb. 2012, doi: 10.1016/j.bcmd.2011.11.006.

[117] S. R. Rose *et al.*, "Oxandrolone for the treatment of bone marrow failure in Fanconi anemia," *Pediatr. Blood Cancer*, vol. 61, no. 1, pp. 11–19, Jan. 2014, doi: 10.1002/pbc.24617.

[118] P. S. Rosenberg, B. P. Alter, G. Socié, and E. Gluckman, "Secular trends in outcomes for Fanconi anemia patients who receive transplants: implications for future studies," *Biol. Blood Marrow Transplant. J. Am. Soc. Blood Marrow Transplant.*, vol. 11, no. 9, pp. 672–679, Sep. 2005, doi: 10.1016/j.bbmt.2005.05.007.

[119] P. Anur *et al.*, "Late effects in patients with Fanconi anemia following allogeneic hematopoietic stem cell transplantation from alternative donors," *Bone Marrow Transplant.*, vol. 51, no. 7, pp. 938–944, Jul. 2016, doi: 10.1038/bmt.2016.32.

[120] A. C. Birkeland *et al.*, "Postoperative clinical radiosensitivity in patients with fanconi anemia and head and neck squamous cell carcinoma," *Arch. Otolaryngol. Head Neck Surg.*, vol. 137, no. 9, pp. 930–934, Sep. 2011, doi: 10.1001/archoto.2011.154.

[121] "Home - ClinicalTrials.gov." https://www.clinicaltrials.gov/ (accessed May 22, 2022).

[122] R. A. Morgan, D. Gray, A. Lomova, and D. B. Kohn, "Hematopoietic Stem Cell Gene Therapy: Progress and Lessons Learned," *Cell Stem Cell*, vol. 21, no. 5, pp. 574–590, Nov. 2017, doi: 10.1016/j.stem.2017.10.010.

[123] E. Verhoeyen, F. J. Roman-Rodriguez, F.-L. Cosset, C. Levy, and P. Rio, "Gene Therapy in Fanconi Anemia: A Matter of Time, Safety and Gene Transfer Tool Efficiency," *Curr. Gene Ther.*, vol. 16, no. 5, pp. 297–308, 2017, doi:

10.2174/1566523217666170109114309.

[124] A. Aiuti *et al.*, “Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott-Aldrich syndrome,” *Science*, vol. 341, no. 6148, p. 1233151, Aug. 2013, doi: 10.1126/science.1233151.

[125] A. Biffi *et al.*, “Lentiviral hematopoietic stem cell gene therapy benefits metachromatic leukodystrophy,” *Science*, vol. 341, no. 6148, p. 1233158, Aug. 2013, doi: 10.1126/science.1233158.

[126] F. Eichler *et al.*, “Hematopoietic Stem-Cell Gene Therapy for Cerebral Adrenoleukodystrophy,” *N. Engl. J. Med.*, vol. 377, no. 17, pp. 1630–1638, Oct. 2017, doi: 10.1056/NEJMoa1700554.

[127] A. A. Thompson *et al.*, “Gene Therapy in Patients with Transfusion-Dependent β-Thalassemia,” *N. Engl. J. Med.*, vol. 378, no. 16, pp. 1479–1493, Apr. 2018, doi: 10.1056/NEJMoa1705342.

[128] F. J. Molina-Estevez *et al.*, “Lentiviral-Mediated Gene Therapy in Fanconi Anemia-A Mice Reveals Long-Term Engraftment and Continuous Turnover of Corrected HSCs,” *Curr. Gene Ther.*, vol. 15, no. 6, pp. 550–562, 2015, doi: 10.2174/1566523215666150929110903.

[129] P. Río *et al.*, “Engraftment and in vivo proliferation advantage of gene-corrected mobilized CD34+ cells from Fanconi anemia patients,” *Blood*, vol. 130, no. 13, pp. 1535–1542, Sep. 2017, doi: 10.1182/blood-2017-03-774174.

[130] P. F. Kelly *et al.*, “Stem cell collection and gene transfer in Fanconi anemia,” *Mol. Ther. J. Am. Soc. Gene Ther.*, vol. 15, no. 1, pp. 211–219, Jan. 2007, doi: 10.1038/sj.mt.6300033.

[131] J. M. Liu *et al.*, “Engraftment of hematopoietic progenitor cells transduced with the Fanconi anemia group C gene (FANCC),” *Hum. Gene Ther.*, vol. 10, no. 14, pp. 2337–2346, Sep. 1999, doi: 10.1089/10430349950016988.

[132] P. Río *et al.*, “Successful engraftment of gene-corrected hematopoietic stem cells in non-conditioned patients with Fanconi anemia,” *Nat. Med.*, vol. 25, no. 9, pp. 1396–1401, Sep. 2019, doi: 10.1038/s41591-019-0550-z.

[133] A. González-Murillo *et al.*, “Development of lentiviral vectors with optimized transcriptional activity for the gene therapy of patients with Fanconi anemia,” *Hum. Gene Ther.*, vol. 21, no. 5, pp. 623–630, May 2010, doi: 10.1089/hum.2009.141.

[134] J. Sevilla, “Clinical Trial Phase I / II to Evaluate the Safety and Efficacy of the Infusion of Autologous CD34 + Cells Transduced With a Lentiviral Vector Carrying the Gene FANCA in Patients With FA Subtype A (FANCOLEN-1),” clinicaltrials.gov, Clinical trial registration NCT03157804, Nov. 2020. Accessed: May 24, 2022. [Online]. Available: https://clinicaltrials.gov/ct2/show/NCT03157804

[135] A. Czechowicz *et al.*, “Changing the Natural History of Fanconi Anemia Complementation Group-À with Gene Therapy: Early Results of U.S. Phase I Study of Lentiviral-Mediated Ex-Vivo FANCA Gene Insertion in Human Stem and Progenitor Cells,” *Biol. Blood Marrow Transplant.*, vol. 26, no. 3, pp. S39–S40, Mar. 2020, doi: 10.1016/j.bbmt.2019.12.106.

[136] Rocket Pharmaceuticals Inc., “A Phase 2 Clinical Trial to Evaluate the Efficacy of the Infusion of Autologous CD34+ Cells Transduced With a Lentiviral Vector Carrying the FANCA Gene in Pediatric Subjects With Fanconi Anemia Subtype A,” clinicaltrials.gov, Clinical trial registration NCT04248439, Nov. 2020. Accessed: May 24, 2022. [Online]. Available: https://clinicaltrials.gov/ct2/show/NCT04248439

[137] Rocket Pharmaceuticals Inc., “A Phase II Clinical Trial to Evaluate the Efficacy of the Infusion of Autologous CD34+ Cells Transduced With a Lentiviral Vector Carrying the FANCA Gene (Orphan Drug) in Patients With Fanconi Anemia Subtype A,” clinicaltrials.gov, Clinical trial registration NCT04069533, Oct. 2020. Accessed: May 24, 2022. [Online]. Available: https://clinicaltrials.gov/ct2/show/NCT04069533

[138] J. I. Garaycoechea, G. P. Crossan, F. Langevin, M. Daly, M. J. Arends, and K. J. Patel, “Genotoxic consequences of endogenous aldehydes on mouse haematopoietic stem cell function,” *Nature*, vol. 489, no. 7417, pp. 571–575, Sep. 2012, doi: 10.1038/nature11368.

[139] W. Du, O. Erden, and Q. Pang, “TNF-α signaling in Fanconi anemia,” *Blood Cells. Mol. Dis.*, vol. 52, no. 1, pp. 2–11, Jan. 2014, doi: 10.1016/j.bcmd.2013.06.005.

[140] Children’s Hospital Medical Center, Cincinnati, “Quercetin Chemoprevention for Squamous Cell Carcinoma in Patients With Fanconi Anemia,” clinicaltrials.gov, Clinical trial registration NCT03476330, Feb. 2022. Accessed: May 26, 2022. [Online]. Available: https://clinicaltrials.gov/ct2/show/NCT03476330

[141] Children’s Hospital Medical Center, Cincinnati, “Quercetin in Children With Fanconi Anemia; a Pilot Study,” clinicaltrials.gov, Clinical trial registration NCT01720147, Jan. 2022. Accessed: May 26, 2022. [Online]. Available: https://clinicaltrials.gov/ct2/show/NCT01720147

[142] S. D. Skaper, M. Fabris, V. Ferrari, M. Dalle Carbonare, and A. Leon, “Quercetin protects cutaneous tissue-associated cell types including sensory neurons from oxidative stress induced by glutathione depletion: cooperative effects of ascorbic acid,” *Free Radic. Biol. Med.*, vol. 22, no. 4, pp. 669–678, 1997, doi: 10.1016/s0891-5849(96)00383-8.

[143] M. Kobori, S. Masumoto, Y. Akimoto, and Y. Takahashi, “Dietary quercetin alleviates diabetic symptoms and reduces streptozotocin-induced disturbance of hepatic gene expression in mice,” *Mol. Nutr. Food Res.*, vol. 53, no. 7, pp. 859–868, Jul. 2009, doi: 10.1002/mnfr.200800310.

[144] J. Li *et al.*, “Fanconi anemia links reactive oxygen species to insulin resistance and

obesity," *Antioxid. Redox Signal.*, vol. 17, no. 8, pp. 1083–1098, Oct. 2012, doi: 10.1089/ars.2011.4417.

[145] P. A. Mehta *et al.*, "Quercetin: A Novel Targeted Chemoprevention for Patients with Fanconi Anemia (FA)," *Blood*, vol. 130, no. Supplement 1, p. 1178, Dec. 2017, doi: 10.1182/blood.V130.Suppl_1.1178.1178.

[146] E. F. MD, "Pilot Study of Metformin for Patients With Fanconi Anemia," clinicaltrials.gov, Clinical trial registration NCT03398824, Oct. 2020. Accessed: May 26, 2022. [Online]. Available: https://clinicaltrials.gov/ct2/show/NCT03398824

[147] L. He, "Metformin and Systemic Metabolism," *Trends Pharmacol. Sci.*, vol. 41, no. 11, pp. 868–881, Nov. 2020, doi: 10.1016/j.tips.2020.09.001.

[148] Q.-S. Zhang *et al.*, "Metformin improves defective hematopoiesis and delays tumor formation in Fanconi anemia mice," *Blood*, vol. 128, no. 24, pp. 2774–2784, Dec. 2016, doi: 10.1182/blood-2015-11-683490.

[149] J. Pollard *et al.*, "Metformin for Treatment of Cytopenias in Children and Young Adults with Fanconi Anemia," *Blood Adv.*, p. bloodadvances.2021006490, May 2022, doi: 10.1182/bloodadvances.2021006490.

[150] J. S. Gutkind *et al.*, "Inhibition of mTOR signaling and clinical activity of metformin in oral premalignant lesions," *JCI Insight*, vol. 6, no. 17, Sep. 2021, doi: 10.1172/jci.insight.147096.

[151] D. T. Le *et al.*, "PD-1 Blockade in Tumors with Mismatch-Repair Deficiency," *N. Engl. J. Med.*, vol. 372, no. 26, pp. 2509–2520, Jun. 2015, doi: 10.1056/NEJMoa1500596.

[152] L. Marcus, S. J. Lemery, P. Keegan, and R. Pazdur, "FDA Approval Summary: Pembrolizumab for the Treatment of Microsatellite Instability-High Solid Tumors," *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.*, vol. 25, no. 13, pp. 3753–3758, Jul. 2019, doi: 10.1158/1078-0432.CCR-18-4070.

[153] M. A. Villalona-Calero *et al.*, "Pembrolizumab activity in patients with Fanconi anemia repair pathway competent and deficient tumors," *Biomark. Res.*, vol. 10, no. 1, p. 39, Jun. 2022, doi: 10.1186/s40364-022-00386-0.

[154] Th. Dobzhansky, "Genetics of Natural Populations. Xiii. Recombination and Variability in Populations of Drosophila Pseudoobscura," *Genetics*, vol. 31, no. 3, pp. 269–290, May 1946.

[155] J. L. Hartman, B. Garvik, and L. Hartwell, "Principles for the buffering of genetic variation," *Science*, vol. 291, no. 5506, pp. 1001–1004, Feb. 2001, doi: 10.1126/science.291.5506.1001.

[156] C. Boone, H. Bussey, and B. J. Andrews, "Exploring genetic interactions and networks with yeast," *Nat. Rev. Genet.*, vol. 8, no. 6, pp. 437–449, Jun. 2007, doi:

10.1038/nrg2085.

[157] P. McGlynn and R. G. Lloyd, “Recombinational repair and restart of damaged replication forks,” *Nat. Rev. Mol. Cell Biol.*, vol. 3, no. 11, pp. 859–870, Nov. 2002, doi: 10.1038/nrm951.

[158] A. G. Patel, J. N. Sarkaria, and S. H. Kaufmann, “Nonhomologous end joining drives poly(ADP-ribose) polymerase (PARP) inhibitor lethality in homologous recombination-deficient cells,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 8, pp. 3406–3411, Feb. 2011, doi: 10.1073/pnas.1013715108.

[159] P. Warrener *et al.*, “Synthetic lethality of PARP inhibition in BRCA-network disrupted tumor cells is associated with interferon pathway activation and enhanced by interferon-γ,” *Apoptosis Int. J. Program. Cell Death*, vol. 17, no. 7, pp. 691–701, Jul. 2012, doi: 10.1007/s10495-012-0707-4.

[160] D. Hühn, H. A. Bolck, and A. A. Sartori, “Targeting DNA double-strand break signalling and repair: recent advances in cancer therapy,” *Swiss Med. Wkly.*, vol. 143, p. w13837, 2013, doi: 10.4414/smw.2013.13837.

[161] O. of the Commissioner, “U.S. Food and Drug Administration,” *FDA*, May 18, 2022. https://www.fda.gov/home (accessed May 29, 2022).

[162] A. Cerrato, F. Morra, and A. Celetti, “Use of poly ADP-ribose polymerase [PARP] inhibitors in cancer cells bearing DDR defects: the rationale for their inclusion in the clinic,” *J. Exp. Clin. Cancer Res. CR*, vol. 35, no. 1, p. 179, Nov. 2016, doi: 10.1186/s13046-016-0456-2.

[163] D. Kamel, C. Gray, J. S. Walia, and V. Kumar, “PARP Inhibitor Drugs in the Treatment of Breast, Ovarian, Prostate and Pancreatic Cancers: An Update of Clinical Trials,” *Curr. Drug Targets*, vol. 19, no. 1, pp. 21–37, 2018, doi: 10.2174/1389450118666170711151518.

[164] R. Abbotts, A. J. Dellomo, and F. V. Rassool, “Pharmacologic Induction of BRCAness in BRCA-Proficient Cancers: Expanding PARP Inhibitor Use,” *Cancers*, vol. 14, no. 11, Art. no. 11, Jan. 2022, doi: 10.3390/cancers14112640.

[165] M. Rose, J. T. Burgess, K. O’Byrne, D. J. Richard, and E. Bolderson, “PARP Inhibitors: Clinical Relevance, Mechanisms of Action and Tumor Resistance,” *Front. Cell Dev. Biol.*, vol. 8, p. 564601, Sep. 2020, doi: 10.3389/fcell.2020.564601.

[166] S. J. Taylor, M. J. Arends, and S. P. Langdon, “Inhibitors of the Fanconi anaemia pathway as potential antitumour agents for ovarian cancer,” *Explor. Target. Anti-Tumor Ther.*, vol. 1, no. 1, pp. 26–52, Feb. 2020, doi: 10.37349/etat.2020.00003.

[167] F. Sams-Dodd, “Drug discovery: selecting the optimal approach,” *Drug Discov. Today*, vol. 11, no. 9–10, pp. 465–472, May 2006, doi: 10.1016/j.drudis.2006.03.015.

[168] F. Bellomo, D. L. Medina, E. De Leo, A. Panarella, and F. Emma, “High-content drug

screening for rare diseases," *J. Inherit. Metab. Dis.*, vol. 40, no. 4, pp. 601–607, Jul. 2017, doi: 10.1007/s10545-017-0055-1.

[169] H. Montanuy *et al.*, "High content drug screening for Fanconi anemia therapeutics," *Orphanet J. Rare Dis.*, vol. 15, no. 1, p. 170, Jun. 2020, doi: 10.1186/s13023-020-01437-1.

[170] S. M. Paul *et al.*, "How to improve R&D productivity: the pharmaceutical industry's grand challenge," *Nat. Rev. Drug Discov.*, vol. 9, no. 3, pp. 203–214, Mar. 2010, doi: 10.1038/nrd3078.

[171] S. Pushpakom *et al.*, "Drug repurposing: progress, challenges and recommendations," *Nat. Rev. Drug Discov.*, vol. 18, no. 1, Art. no. 1, Jan. 2019, doi: 10.1038/nrd.2018.168.

[172] M. Schlander, K. Hernandez-Villafuerte, C.-Y. Cheng, J. Mestre-Ferrandiz, and M. Baumann, "How Much Does It Cost to Research and Develop a New Drug? A Systematic Review and Assessment," *PharmacoEconomics*, vol. 39, no. 11, pp. 1243–1269, Nov. 2021, doi: 10.1007/s40273-021-01065-y.

[173] T. T. Ashburn and K. B. Thor, "Drug repositioning: identifying and developing new uses for existing drugs," *Nat. Rev. Drug Discov.*, vol. 3, no. 8, pp. 673–683, Aug. 2004, doi: 10.1038/nrd1468.

[174] J. Almaça *et al.*, "High-content siRNA screen reveals global ENaC regulators and potential cystic fibrosis therapy targets," *Cell*, vol. 154, no. 6, pp. 1390–1400, Sep. 2013, doi: 10.1016/j.cell.2013.08.045.

[175] B. Cautain *et al.*, "Identification of the Lipodepsipeptide MDN-0066, a Novel Inhibitor of VHL/HIF Pathway Produced by a New Pseudomonas Species," *PLoS ONE*, vol. 10, no. 5, p. e0125221, May 2015, doi: 10.1371/journal.pone.0125221.

[176] S. Cappato *et al.*, "Correction: High-throughput screening for modulators of ACVR1 transcription: discovery of potential therapeutics for fibrodysplasia ossificans progressiva," *Dis. Model. Mech.*, vol. 9, no. 9, p. 1067, Sep. 2016, doi: 10.1242/dmm.027573.

[177] F. Bellomo *et al.*, "Drug Repurposing in Rare Diseases: An Integrative Study of Drug Screening and Transcriptomic Analysis in Nephropathic Cystinosis," *Int. J. Mol. Sci.*, vol. 22, no. 23, Art. no. 23, Jan. 2021, doi: 10.3390/ijms222312829.

[178] S. Shah, M. M. Dooms, S. Amaral-Garcia, and M. Igoillo-Esteve, "Current Drug Repurposing Strategies for Rare Neurodegenerative Disorders," *Front. Pharmacol.*, vol. 12, p. 768023, Dec. 2021, doi: 10.3389/fphar.2021.768023.

[179] M. Boettcher and M. T. McManus, "Choosing the Right Tool for the Job: RNAi, TALEN, or CRISPR," *Mol. Cell*, vol. 58, no. 4, pp. 575–585, May 2015, doi: 10.1016/j.molcel.2015.04.028.

[180] H. Montanuy *et al.*, "Gefitinib and Afatinib Show Potential Efficacy for Fanconi

Anemia-Related Head and Neck Cancer," *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.*, vol. 26, no. 12, pp. 3044–3057, Jun. 2020, doi: 10.1158/1078-0432.CCR-19-1625.

[181] D. Chirnomas *et al.*, "Chemosensitization to cisplatin by inhibitors of the Fanconi anemia/BRCA pathway," *Mol. Cancer Ther.*, vol. 5, no. 4, pp. 952–961, Apr. 2006, doi: 10.1158/1535-7163.MCT-05-0493.

[182] C. Jacquemont, J. A. Simon, A. D. D'Andrea, and T. Taniguchi, "Non-specific chemical inhibition of the Fanconi anemia pathway sensitizes cancer cells to cisplatin," *Mol. Cancer*, vol. 11, no. 1, p. 26, Apr. 2012, doi: 10.1186/1476-4598-11-26.

[183] D. Kazandjian, G. M. Blumenthal, W. Yuan, K. He, P. Keegan, and R. Pazdur, "FDA Approval of Gefitinib for the Treatment of Patients with Metastatic EGFR Mutation-Positive Non-Small Cell Lung Cancer," *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.*, vol. 22, no. 6, pp. 1307–1312, Mar. 2016, doi: 10.1158/1078-0432.CCR-15-2266.

[184] T. Vavalà, "Role of afatinib in the treatment of advanced lung squamous cell carcinoma," *Clin. Pharmacol. Adv. Appl.*, vol. 9, pp. 147–157, 2017, doi: 10.2147/CPAA.S112715.

[185] Z. Yang *et al.*, "Comparison of gefitinib, erlotinib and afatinib in non-small cell lung cancer: A meta-analysis," *Int. J. Cancer*, vol. 140, no. 12, pp. 2805–2819, Jun. 2017, doi: 10.1002/ijc.30691.

[186] N. Sharma and S. Graziano, "Overview of the LUX-Lung clinical trial program of afatinib for non-small cell lung cancer," *Cancer Treat. Rev.*, vol. 69, pp. 143–151, Sep. 2018, doi: 10.1016/j.ctrv.2018.06.018.

[187] H. S. Jung *et al.*, "Gefitinib trial in a fanconi's anemia patient with multiple squamous cell carcinomas and hepatocellular carcinoma," *Cancer Res. Treat.*, vol. 37, no. 6, pp. 370–373, Dec. 2005, doi: 10.4143/crt.2005.37.6.370.

[188] J. Moreira, A. Tobias, M. P. O'Brien, and M. Agulnik, "Targeted Therapy in Head and Neck Cancer: An Update on Current Clinical Developments in Epidermal Growth Factor Receptor-Targeted Therapy and Immunotherapies," *Drugs*, vol. 77, no. 8, pp. 843–857, May 2017, doi: 10.1007/s40265-017-0734-0.

[189] R. H. Lee, H. Kang, S. S. Yom, A. Smogorzewska, D. E. Johnson, and J. R. Grandis, "Treatment of Fanconi Anemia–Associated Head and Neck Cancer: Opportunities to Improve Outcomes," *Clin. Cancer Res.*, vol. 27, no. 19, pp. 5168–5187, Oct. 2021, doi: 10.1158/1078-0432.CCR-21-1259.

[190] L. Castells-Roca, E. Tejero, B. Rodríguez-Santiago, and J. Surrallés, "CRISPR Screens in Synthetic Lethality and Combinatorial Therapies for Cancer," *Cancers*, vol. 13, no. 7, p. 1591, Mar. 2021, doi: 10.3390/cancers13071591.

[191] F. J. M. Mojica, C. Díez-Villaseñor, J. García-Martínez, and E. Soria, "Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements," *J.*

Mol. Evol., vol. 60, no. 2, pp. 174–182, Feb. 2005, doi: 10.1007/s00239-004-0046-3.

[192] K. S. Makarova, N. V. Grishin, S. A. Shabalina, Y. I. Wolf, and E. V. Koonin, “A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action,” *Biol. Direct*, vol. 1, p. 7, Mar. 2006, doi: 10.1186/1745-6150-1-7.

[193] R. Barrangou *et al.*, “CRISPR provides acquired resistance against viruses in prokaryotes,” *Science*, vol. 315, no. 5819, pp. 1709–1712, Mar. 2007, doi: 10.1126/science.1138140.

[194] R. Barrangou and L. A. Marraffini, “CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity,” *Mol. Cell*, vol. 54, no. 2, pp. 234–244, Apr. 2014, doi: 10.1016/j.molcel.2014.03.011.

[195] J. Bondy-Denomy and A. R. Davidson, “To acquire or resist: the complex biological effects of CRISPR-Cas systems,” *Trends Microbiol.*, vol. 22, no. 4, pp. 218–225, Apr. 2014, doi: 10.1016/j.tim.2014.01.007.

[196] J. van der Oost, E. R. Westra, R. N. Jackson, and B. Wiedenheft, “Unravelling the structural and mechanistic basis of CRISPR-Cas systems,” *Nat. Rev. Microbiol.*, vol. 12, no. 7, pp. 479–492, Jul. 2014, doi: 10.1038/nrmicro3279.

[197] H. Deveau *et al.*, “Phage response to CRISPR-encoded resistance in Streptococcus thermophilus,” *J. Bacteriol.*, vol. 190, no. 4, pp. 1390–1400, Feb. 2008, doi: 10.1128/JB.01412-07.

[198] P. Horvath *et al.*, “Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus,” *J. Bacteriol.*, vol. 190, no. 4, pp. 1401–1412, Feb. 2008, doi: 10.1128/JB.01415-07.

[199] F. J. M. Mojica, C. Díez-Villaseñor, J. García-Martínez, and C. Almendros, “Short motif sequences determine the targets of the prokaryotic CRISPR defence system,” *Microbiol. Read. Engl.*, vol. 155, no. Pt 3, pp. 733–740, Mar. 2009, doi: 10.1099/mic.0.023960-0.

[200] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity,” *Science*, vol. 337, no. 6096, pp. 816–821, Aug. 2012, doi: 10.1126/science.1225829.

[201] C. Bock *et al.*, “High-content CRISPR screening,” *Nat. Rev. Methods Primer*, vol. 2, no. 1, Art. no. 1, Feb. 2022, doi: 10.1038/s43586-021-00093-4.

[202] A. Chavez *et al.*, “Highly efficient Cas9-mediated transcriptional programming,” *Nat. Methods*, vol. 12, no. 4, pp. 326–328, Apr. 2015, doi: 10.1038/nmeth.3312.

[203] L. R. Polstein and C. A. Gersbach, “A light-inducible CRISPR-Cas9 system for control

of endogenous gene activation," *Nat. Chem. Biol.*, vol. 11, no. 3, pp. 198–200, Mar. 2015, doi: 10.1038/nchembio.1753.

[204] J. G. Zalatan *et al.*, "Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds," *Cell*, vol. 160, no. 1–2, pp. 339–350, Jan. 2015, doi: 10.1016/j.cell.2014.11.052.

[205] Y. Gao, X. Xiong, S. Wong, E. J. Charles, W. A. Lim, and L. S. Qi, "Complex transcriptional modulation with orthogonal and inducible dCas9 regulators," *Nat. Methods*, vol. 13, no. 12, pp. 1043–1049, Dec. 2016, doi: 10.1038/nmeth.4042.

[206] D. B. T. Cox *et al.*, "RNA editing with CRISPR-Cas13," *Science*, vol. 358, no. 6366, pp. 1019–1027, Nov. 2017, doi: 10.1126/science.aaq0180.

[207] M. Gapinske *et al.*, "CRISPR-SKIP: programmable gene splicing with single base editors," *Genome Biol.*, vol. 19, no. 1, p. 107, Aug. 2018, doi: 10.1186/s13059-018-1482-5.

[208] S. Konermann, P. Lotfy, N. J. Brideau, J. Oki, M. N. Shokhirev, and P. D. Hsu, "Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors," *Cell*, vol. 173, no. 3, pp. 665-676.e14, Apr. 2018, doi: 10.1016/j.cell.2018.02.033.

[209] O. O. Abudayyeh *et al.*, "A cytosine deaminase for programmable single-base RNA editing," *Science*, vol. 365, no. 6451, pp. 382–386, Jul. 2019, doi: 10.1126/science.aax7063.

[210] M. Du, N. Jillette, J. J. Zhu, S. Li, and A. W. Cheng, "CRISPR artificial splicing factors," *Nat. Commun.*, vol. 11, no. 1, p. 2973, Jun. 2020, doi: 10.1038/s41467-020-16806-4.

[211] I. B. Hilton *et al.*, "Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers," *Nat. Biotechnol.*, vol. 33, no. 5, pp. 510–517, May 2015, doi: 10.1038/nbt.3199.

[212] D. Y. Kwon, Y.-T. Zhao, J. M. Lamonica, and Z. Zhou, "Locus-specific histone deacetylation using a synthetic CRISPR-Cas9-based HDAC," *Nat. Commun.*, vol. 8, p. 15315, May 2017, doi: 10.1038/ncomms15315.

[213] P. Stepper *et al.*, "Efficient targeted DNA methylation with chimeric dCas9-Dnmt3a-Dnmt3L methyltransferase," *Nucleic Acids Res.*, vol. 45, no. 4, pp. 1703–1713, Feb. 2017, doi: 10.1093/nar/gkw1112.

[214] X. Xu *et al.*, "High-fidelity CRISPR/Cas9- based gene-specific hydroxymethylation rescues gene expression and attenuates renal fibrosis," *Nat. Commun.*, vol. 9, no. 1, p. 3509, Aug. 2018, doi: 10.1038/s41467-018-05766-5.

[215] A. C. Komor *et al.*, "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity," *Sci. Adv.*, vol. 3, no. 8, p. eaao4774, Aug. 2017, doi: 10.1126/sciadv.aao4774.

[216] K. A. Molla and Y. Yang, "CRISPR/Cas-Mediated Base Editing: Technical Considerations and Practical Applications," *Trends Biotechnol.*, vol. 37, no. 10, pp. 1121–

1142, Oct. 2019, doi: 10.1016/j.tibtech.2019.03.008.

[217] L. W. Koblan *et al.*, “Efficient C•G-to-G•C base editors developed using CRISPRi screens, target-library analysis, and machine learning,” *Nat. Biotechnol.*, vol. 39, no. 11, pp. 1414–1425, Nov. 2021, doi: 10.1038/s41587-021-00938-z.

[218] A. V. Anzalone *et al.*, “Search-and-replace genome editing without double-strand breaks or donor DNA,” *Nature*, vol. 576, no. 7785, pp. 149–157, Dec. 2019, doi: 10.1038/s41586-019-1711-4.

[219] A. V. Anzalone *et al.*, “Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing,” *Nat. Biotechnol.*, vol. 40, no. 5, pp. 731–740, May 2022, doi: 10.1038/s41587-021-01133-w.

[220] T. Jiang, X.-O. Zhang, Z. Weng, and W. Xue, “Deletion and replacement of long genomic sequences using prime editing,” *Nat. Biotechnol.*, vol. 40, no. 2, pp. 227–234, Feb. 2022, doi: 10.1038/s41587-021-01026-y.

[221] M. Bibikova, K. Beumer, J. K. Trautman, and D. Carroll, “Enhancing gene targeting with designed zinc finger nucleases,” *Science*, vol. 300, no. 5620, p. 764, May 2003, doi: 10.1126/science.1079512.

[222] M. J. Moscou and A. J. Bogdanove, “A simple cipher governs DNA recognition by TAL effectors,” *Science*, vol. 326, no. 5959, p. 1501, Dec. 2009, doi: 10.1126/science.1178817.

[223] “GenomeCRISPR,” *GenomeCRISPR*. http://www.genomecrispr.org/ (accessed Jun. 26, 2022).

[224] O. Shalem *et al.*, “Genome-scale CRISPR-Cas9 knockout screening in human cells,” *Science*, vol. 343, no. 6166, pp. 84–87, Jan. 2014, doi: 10.1126/science.1247005.

[225] T. Wang, J. J. Wei, D. M. Sabatini, and E. S. Lander, “Genetic screens in human cells using the CRISPR-Cas9 system,” *Science*, vol. 343, no. 6166, pp. 80–84, Jan. 2014, doi: 10.1126/science.1246981.

[226] T. Hart *et al.*, “Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens,” *G3 Bethesda Md*, vol. 7, no. 8, pp. 2719–2727, Aug. 2017, doi: 10.1534/g3.117.041277.

[227] “CRISPick.” https://portals.broadinstitute.org/gppx/crispick/public (accessed Jun. 26, 2022).

[228] “Addgene: Pooled Libraries.” https://www.addgene.org/pooled-library/ (accessed Jun. 26, 2022).

[229] T. Hart *et al.*, “High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities,” *Cell*, vol. 163, no. 6, pp. 1515–1526, Dec. 2015, doi: 10.1016/j.cell.2015.11.015.

[230] M. Kuhn, A. J. Santinha, and R. J. Platt, “Moving from in vitro to in vivo CRISPR screens,” *Gene Genome Ed.*, vol. 2, p. 100008, Dec. 2021, doi: 10.1016/j.ggedit.2021.100008.

[231] C. Pacini *et al.*, “Integrated cross-study datasets of genetic dependencies in cancer,” *Nat. Commun.*, vol. 12, no. 1, p. 1661, Mar. 2021, doi: 10.1038/s41467-021-21898-7.

[232] F. M. Behan *et al.*, “Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens,” *Nature*, vol. 568, no. 7753, pp. 511–516, Apr. 2019, doi: 10.1038/s41586-019-1103-9.

[233] A. H. Smits *et al.*, “Biological plasticity rescues target activity in CRISPR knock outs,” *Nat. Methods*, vol. 16, no. 11, pp. 1087–1093, Nov. 2019, doi: 10.1038/s41592-019-0614-5.

[234] J. Joung *et al.*, “Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening,” *Nat. Protoc.*, vol. 12, no. 4, pp. 828–863, Apr. 2017, doi: 10.1038/nprot.2017.016.

[235] S. Bodapati, T. P. Daley, X. Lin, J. Zou, and L. S. Qi, “A benchmark of algorithms for the analysis of pooled CRISPR screens,” *Genome Biol.*, vol. 21, no. 1, p. 62, Mar. 2020, doi: 10.1186/s13059-020-01972-x.

[236] W. Li *et al.*, “Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR,” *Genome Biol.*, vol. 16, p. 281, Dec. 2015, doi: 10.1186/s13059-015-0843-6.

[237] T. Hart, K. R. Brown, F. Sircoulomb, R. Rottapel, and J. Moffat, “Measuring error rates in genomic perturbation screens: gold standards for human functional genomics,” *Mol. Syst. Biol.*, vol. 10, p. 733, Jul. 2014, doi: 10.15252/msb.20145216.

[238] C. D. Richardson *et al.*, “CRISPR-Cas9 genome editing in human cells occurs via the Fanconi anemia pathway,” *Nat. Genet.*, vol. 50, no. 8, pp. 1132–1139, Aug. 2018, doi: 10.1038/s41588-018-0174-0.

[239] H. J. van de Vrugt *et al.*, “Effective CRISPR/Cas9-mediated correction of a Fanconi anemia defect by error-prone end joining or templated repair,” *Sci. Rep.*, vol. 9, no. 1, Art. no. 1, Jan. 2019, doi: 10.1038/s41598-018-36506-w.

[240] J. G. Doench *et al.*, “Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9,” *Nat. Biotechnol.*, vol. 34, no. 2, pp. 184–191, Feb. 2016, doi: 10.1038/nbt.3437.

[241] M.-Y. Cai *et al.*, “Cooperation of the ATM and Fanconi Anemia/BRCA Pathways in Double-Strand Break End Resection,” *Cell Rep.*, vol. 30, no. 7, pp. 2402-2415.e5, Feb. 2020, doi: 10.1016/j.celrep.2020.01.052.

[242] M. L. Duquette *et al.*, “CtIP Is Required to Initiate Replication-Dependent Interstrand

Crosslink Repair," *PLoS Genet.*, vol. 8, no. 11, p. e1003050, Nov. 2012, doi: 10.1371/journal.pgen.1003050.

[243] J. E. Yeo, E. H. Lee, E. A. Hendrickson, and A. Sobeck, "CtIP mediates replication fork recovery in a FANCD2-regulated manner," *Hum. Mol. Genet.*, vol. 23, no. 14, pp. 3695–3705, Jul. 2014, doi: 10.1093/hmg/ddu078.

[244] R. Ceccaldi, B. Rondinelli, and A. D. D'Andrea, "Repair Pathway Choices and Consequences at the Double-Strand Break," *Trends Cell Biol.*, vol. 26, no. 1, pp. 52–64, Jan. 2016, doi: 10.1016/j.tcb.2015.07.009.

[245] R. Wang *et al.*, "DNA polymerase ι compensates for Fanconi anemia pathway deficiency by countering DNA replication stress," *Proc. Natl. Acad. Sci.*, vol. 117, no. 52, pp. 33436–33445, Dec. 2020, doi: 10.1073/pnas.2008821117.

[246] S. Naldiga *et al.*, "Error-prone replication of a 5-formylcytosine-mediated DNA-peptide cross-link in human cells," *J. Biol. Chem.*, vol. 294, no. 27, pp. 10619–10627, Jul. 2019, doi: 10.1074/jbc.RA119.008879.

[247] J. McIntyre, "Polymerase iota - an odd sibling among Y family polymerases," *DNA Repair*, vol. 86, p. 102753, Feb. 2020, doi: 10.1016/j.dnarep.2019.102753.

[248] C.-H. Chen, J. C. B. Ferreira, E. R. Gross, and D. Mochly-Rosen, "Targeting aldehyde dehydrogenase 2: new therapeutic opportunities," *Physiol. Rev.*, vol. 94, no. 1, pp. 1–34, Jan. 2014, doi: 10.1152/physrev.00017.2013.

[249] J. Barretina *et al.*, "The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity," *Nature*, vol. 483, no. 7391, pp. 603–607, Mar. 2012, doi: 10.1038/nature11003.

[250] M. Ghandi *et al.*, "Next-generation characterization of the Cancer Cell Line Encyclopedia," *Nature*, vol. 569, no. 7757, pp. 503–508, May 2019, doi: 10.1038/s41586-019-1186-3.

[251] R. Ceccaldi, P. Sarangi, and A. D. D'Andrea, "The Fanconi anaemia pathway: new players and new functions," *Nat. Rev. Mol. Cell Biol.*, vol. 17, no. 6, pp. 337–349, Jun. 2016, doi: 10.1038/nrm.2016.48.

[252] R. M. Meyers *et al.*, "Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells," *Nat. Genet.*, vol. 49, no. 12, pp. 1779–1784, Dec. 2017, doi: 10.1038/ng.3984.

[253] J. M. Dempster *et al.*, "Extracting Biological Insights from the Project Achilles Genome-Scale CRISPR Screens in Cancer Cell Lines." bioRxiv, p. 720243, Jul. 31, 2019. doi: 10.1101/720243.

[254] Z. Yang *et al.*, "Transcriptional Silencing of ALDH2 Confers a Dependency on Fanconi Anemia Proteins in Acute Myeloid Leukemia," *Cancer Discov.*, vol. 11, no. 9, pp.

2300–2315, Sep. 2021, doi: 10.1158/2159-8290.CD-20-1542.

[255] F. Langevin, G. P. Crossan, I. V. Rosado, M. J. Arends, and K. J. Patel, “Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice,” *Nature*, vol. 475, no. 7354, pp. 53–58, Jul. 2011, doi: 10.1038/nature10192.

[256] A. Hira *et al.*, “Variant ALDH2 is associated with accelerated progression of bone marrow failure in Japanese Fanconi anemia patients,” *Blood*, vol. 122, no. 18, pp. 3206–3209, Oct. 2013, doi: 10.1182/blood-2013-06-507962.

[257] M. Jung *et al.*, “ALDH9A1 Deficiency as a Source of Endogenous DNA Damage that Requires Repair by the Fanconi Anemia Pathway.” bioRxiv, p. 2022.05.26.493623, May 27, 2022. doi: 10.1101/2022.05.26.493623.

[258] A. E. Pegg, “Toxicity of polyamines and their metabolic products,” *Chem. Res. Toxicol.*, vol. 26, no. 12, pp. 1782–1800, Dec. 2013, doi: 10.1021/tx400316s.

[259] N. N. Hamouda *et al.*, “ATP13A3 is a major component of the enigmatic mammalian polyamine transport system,” *J. Biol. Chem.*, vol. 296, p. 100182, Jun. 2021, doi: 10.1074/jbc.RA120.013908.

[260] N. E. Sanjana, O. Shalem, and F. Zhang, “Improved vectors and genome-wide libraries for CRISPR screening,” *Nat. Methods*, vol. 11, no. 8, pp. 783–784, Aug. 2014, doi: 10.1038/nmeth.3047.

[261] M. Moder *et al.*, “Parallel genome-wide screens identify synthetic viable interactions between the BLM helicase complex and Fanconi anemia,” *Nat. Commun.*, vol. 8, no. 1, Art. no. 1, Nov. 2017, doi: 10.1038/s41467-017-01439-x.

[262] A. J. Deans and S. C. West, “FANCM Connects the Genome Instability Disorders Bloom’s Syndrome and Fanconi Anemia,” *Mol. Cell*, vol. 36, no. 6, pp. 943–953, Dec. 2009, doi: 10.1016/j.molcel.2009.12.006.

[263] A. Grabarz *et al.*, “A role for BLM in double-strand break repair pathway choice: prevention of CtIP/Mre11-mediated alternative nonhomologous end-joining,” *Cell Rep.*, vol. 5, no. 1, pp. 21–28, Oct. 2013, doi: 10.1016/j.celrep.2013.08.034.

[264] D. Siegel, C. Yan, and D. Ross, “NAD(P)H:quinone oxidoreductase 1 (NQO1) in the sensitivity and resistance to antitumor quinones,” *Biochem. Pharmacol.*, vol. 83, no. 8, pp. 1033–1040, Apr. 2012, doi: 10.1016/j.bcp.2011.12.017.

[265] K. Mikami, M. Naito, A. Tomida, M. Yamada, T. Sirakusa, and T. Tsuruo, “DT-diaphorase as a critical determinant of sensitivity to mitomycin C in human colon and gastric carcinoma cell lines,” *Cancer Res.*, vol. 56, no. 12, pp. 2823–2826, Jun. 1996.

[266] G. Velimezi *et al.*, “Map of synthetic rescue interactions for the Fanconi anemia DNA repair pathway identifies USP48,” *Nat. Commun.*, vol. 9, no. 1, Art. no. 1, Jun. 2018, doi: 10.1038/s41467-018-04649-z.

[267] M. Uckelmann *et al.*, “USP48 restrains resection by site-specific cleavage of the BRCA1 ubiquitin mark from H2A,” *Nat. Commun.*, vol. 9, no. 1, Art. no. 1, Jan. 2018, doi: 10.1038/s41467-017-02653-3.

[268] “Home - SNP - NCBI.” https://www.ncbi.nlm.nih.gov/snp/ (accessed Sep. 28, 2022).

[269] “gnomAD.” https://gnomad.broadinstitute.org/ (accessed Sep. 28, 2022).

[270] “BDGP: Splice Site Prediction by Neural Network.” https://www.fruitfly.org/seq_tools/splice.html (accessed Sep. 28, 2022).

[271] “QuickChange Primer Design.” https://www.agilent.com/store/primerDesignProgram.jsp (accessed Oct. 12, 2022).

[272] “GenScript Sequencing Primer Design.” https://www.genscript.com/tools/dna-sequencing-primer-design (accessed Oct. 12, 2022).

[273] V. Vichai and K. Kirtikara, “Sulforhodamine B colorimetric assay for cytotoxicity screening,” *Nat. Protoc.*, vol. 1, no. 3, pp. 1112–1116, 2006, doi: 10.1038/nprot.2006.179.

[274] N. E. Sanjana, O. Shalem, and F. Zhang, “Improved vectors and genome-wide libraries for CRISPR screening,” *Nat. Methods*, vol. 11, no. 8, pp. 783–784, Aug. 2014, doi: 10.1038/nmeth.3047.

[275] D. Heckl *et al.*, “Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing,” *Nat. Biotechnol.*, vol. 32, no. 9, pp. 941–946, Sep. 2014, doi: 10.1038/nbt.2951.

[276] B. Mair *et al.*, “Essential Gene Profiles for Human Pluripotent Stem Cells Identify Uncharacterized Genes and Substrate Dependencies,” *Cell Rep.*, vol. 27, no. 2, pp. 599-615.e12, Apr. 2019, doi: 10.1016/j.celrep.2019.02.041.

[277] M. Wegner *et al.*, “Circular synthesized CRISPR/Cas gRNAs for functional interrogations in the coding and noncoding genome,” *eLife*, vol. 8, p. e42549, Mar. 2019, doi: 10.7554/eLife.42549.

[278] M. Wegner, K. Husnjak, and M. Kaulich, “Unbiased and Tailored CRISPR/Cas gRNA Libraries by SynthesizingCovalently-closed-circular (3Cs) DNA,” *Bio-Protoc.*, vol. 10, no. 1, p. e3472, Jan. 2020, doi: 10.21769/BioProtoc.3472.

[279] V. Diehl *et al.*, “Minimized combinatorial CRISPR screens identify genetic interactions in autophagy,” *Nucleic Acids Res.*, vol. 49, no. 10, pp. 5684–5704, Jun. 2021, doi: 10.1093/nar/gkab309.

[280] E. Kim and T. Hart, “Improved analysis of CRISPR fitness screens and reduced off-target effects with the BAGEL2 gene essentiality classifier,” *Genome Med.*, vol. 13, no. 1, p. 2, Jan. 2021, doi: 10.1186/s13073-020-00809-3.

[281] P. Weber, J. K. Pauling, M. List, and J. Baumbach, “BALSAM—An Interactive Online

Platform for Breath Analysis, Visualization and Classification," *Metabolites*, vol. 10, no. 10, p. 393, Oct. 2020, doi: 10.3390/metabo10100393.

[282] W. Li *et al.*, "MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens," *Genome Biol.*, vol. 15, no. 12, 2014, doi: 10.1186/s13059-014-0554-4.

[283] S. H. Ong, Y. Li, H. Koike-Yusa, and K. Yusa, "Optimised metrics for CRISPR-KO screens with second-generation gRNA libraries," *Sci. Rep.*, vol. 7, no. 1, Art. no. 1, Aug. 2017, doi: 10.1038/s41598-017-07827-z.

[284] T. Gonatopoulos-Pournatzis *et al.*, "Genetic interaction mapping and exon-resolution functional genomics with a hybrid Cas9-Cas12a platform," *Nat. Biotechnol.*, vol. 38, no. 5, pp. 638–648, May 2020, doi: 10.1038/s41587-020-0437-z.

[285] F. Pagani *et al.*, "New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in CFTR exon 12," *Hum. Mol. Genet.*, vol. 12, no. 10, pp. 1111–1120, May 2003, doi: 10.1093/hmg/ddg131.

[286] "VarSome The Human Genomics Community." https://varsome.com/ (accessed Oct. 06, 2022).

[287] M. Bogliolo *et al.*, "Biallelic truncating FANCM mutations cause early-onset cancer but not Fanconi anemia," *Genet. Med. Off. J. Am. Coll. Med. Genet.*, vol. 20, no. 4, pp. 458–463, Apr. 2018, doi: 10.1038/gim.2017.124.

[288] "hTERT RPE-1 - CRL-4000 | ATCC." https://www.atcc.org/products/crl-4000 (accessed Oct. 17, 2022).

[289] B. Rauscher, F. Heigwer, M. Breinig, J. Winter, and M. Boutros, "GenomeCRISPR - a database for high-throughput CRISPR/Cas9 screens," *Nucleic Acids Res.*, vol. 45, no. D1, pp. D679–D686, Jan. 2017, doi: 10.1093/nar/gkw997.

[290] K. R. Sanson *et al.*, "Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities," *Nat. Commun.*, vol. 9, no. 1, Art. no. 1, Dec. 2018, doi: 10.1038/s41467-018-07901-8.

[291] J. G. Doench, "Am I ready for CRISPR? A user's guide to genetic screens," *Nat. Rev. Genet.*, vol. 19, no. 2, pp. 67–80, Feb. 2018, doi: 10.1038/nrg.2017.97.

[292] D. Szklarczyk *et al.*, "STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D607–D613, Jan. 2019, doi: 10.1093/nar/gky1131.