Supplementary Methods

Chromosome breakage analysis (CBA)

CBA was performed by culturing T-cells from 0.5mL whole blood in 4.5 mL of RPMI 1640 containing 10% fetal bovine serum (FBS) and 60 µg/mL phytohemagglutinin (Thermo Fisher Scientific) at 37°C/5% CO_2. After 24 h, MMC was added at 50 and 100 ng/mL in the culture, and after 72 h, the cells were arrested at the metaphase by treatment with 0.02 µg/mL colcemid for 1 h. A total of 40 well-spread metaphases of the cells with normal ploidy were analysed. Ambiguous structures were excluded from the analysis. A conservative linear discriminant function that provided 100% sensitivity and 100% specificity was used to calculate the CBA scores.[71]

Isolation and culture of human dermal fibroblasts

Human dermal fibroblasts were isolated and cultured using a previously described protocol.[72] In brief, 1 mm^3 skin biopsies obtained from the patients were washed with phosphate-buffered saline (PBS; HyClone) and incubated overnight at 37°C/5% CO_2 in a 15 mL conical flask with 1 mL of a skin biopsy digestion medium containing DMEM with 20% FBS, 0.25% collagenase type I (ThermoFisher Scientific, 0.05% DNAse I (Merck, and antibiotics (100 U/mL penicillin and 100 µg/mL streptomycin). The tube was briefly vortexed the next day, and 7 mL of DMEM containing 20% FBS was added. The dissociated cells were then plated in a T25 culture flask and incubated for 3 days at 37°C (5% CO_2). The medium was replaced on the fourth day with DMEM containing 10% FBS and antibiotics. The cells were cultured with a media change on alternate days until they reached approximately 80% confluency. Cells were then passaged at 1:4 ratio using 0.05% trypsin–EDTA (Thermo Fisher Scientific). After two passages, the cells were cryopreserved.

FANCD2-Ub analysis in the PB cells and dermal fibroblasts

PBMNCs isolated from 5 mL of the blood sample by density gradient separation using Ficoll–Hypaque solution (Hyclone) were cultured in 5 mL of RPMI medium containing 10% FBS and 60 µg/mL phytohemagglutinin. After 4 days of culture, whole-cell lysates were prepared for western blot analysis. Fibroblasts were isolated from patients' skin biopsies for FANCD2 ubiquitination analysis (refer to supplementary methods). 2×10^6 fibroblasts were treated with 0.25 µg/µL MMC for 16 h, and the cells harvested by trypsin–EDTA treatment were used for preparing the cell lysates. Protein separation was performed in a 7% tris–glycine SDS-PAGE gel and transferred on a polyvinylidene difluoride membrane. Non-ubiquitinated and ubiquitinated forms of FANCD2 were probed using an anti-FANCD2 antibody (sc-20022, Santa Cruz Biotechnology) and horseradish peroxidase-conjugated anti-mouse secondary antibody (Cell Signaling) and detection was performed using an enhanced luminol-based chemiluminescence substrate (Westar Supernova, Cyanagen).

Next-generation sequencing

Exon capture for 35 samples was performed using SureSelect Human All Exon V5 (Agilent), 35 samples with xGen Exome Research Panel v1 (IDT), 42 samples with SureSelect Human All Exon V5+UTR (Agilent), 25 samples with SureSelect Human All Exon V6+UTR (Agilent) and 5 samples with a focused exome sequencing panel. All of these panels have 100% coverage for the coding sequence of FA genes. Sequencing was performed on an Illumina HiSeq X system to generate 2×150 bp sequence reads with 8–10 GB of data per sample.

Bioinformatics of exome sequencing data and variant prioritisation

Fastq-mcf (ea-utils-1.1.2-806) was utilised for trimming adapter sequences from the raw reads, and Bedtools-2.17 was used for coverage metrics. A GATK Best Practices bioinformatics analysis using Sentieon (v201808.01) was used to identify germline variants in the samples. The sequences obtained were quality-checked and aligned to the human reference genome (GRCh37/hg19) using the Sentieon aligner. The alignment files were processed using Sentieon to remove duplicates, recalibrate, and re-align indels. Gene annotation of the variants was performed using the Ensembl Variant Effect Predictor against the human gene model (release 91). Single-nucleotide variants (SNVs) with >3% overall population allele frequency in 1000Genome Phase 3[74], gnomAD (v2.1), EVS, dbSNP (v151), as well as MedGenome's Indian population database[76] and deep intronic and intergenic variants were excluded for further analysis. The variants were sequentially prioritised based on their reported status in disease databases (HGMD, ClinVar)and multiple in-silico prediction toools, such as PolyPhen-2, SIFT, MutationTaster2, LRT, EVE.[73] Copy number variants (CNVs) were detected from targeted

sequence data using ExomeDepth (v1.1.10). This algorithm detects CNVs based on comparing the read depths of the test data with the matched aggregate reference dataset. For all the samples, copy number variants (CNVs) were evaluated for identifying deletions and duplications involving one or more exons. Bayes factors of more than 15 were retained. CNVs with low statistical confidence and possibility of being false positive due to the presence of pseudogene in the genome were filtered out. The prioritised variants were classified based on ACMG guidelines. Online tools of Varsome and EVE, a recent computational method, were used together for the classification. Additionally, VarSeq 2.2.0 (Golden Helix Inc., Bozeman, MT, United States), a clinical genomics interpretation and reporting platform, was also used to analyse SNVs and CNVs in the 6 samples in which our pipeline identified heterozygous mutations and therefore could not establish the disease-causing genotypes.

Detection of deletions using multiplex ligation-dependent probe amplification (MLPA) and gene dosage analysis

Deletions in *FANCA* were identified by MLPA using SALSA MLPA P031 and P032 Probe Mixes and SALSA MLPA EK1 reagent kit (MRC Holland) and Coffalyser software (MRC Holland) as per the manufacturer's protocol. To confirm the deletion identified by CNV analysis in the *UBE2T* gene, a gene dosage quantitative multiplex fluorescent-PCR was performed as we described previously[78]. Briefly, a 327 bp fragment from exon 7 of *UBE2T* was amplified along with a short fragment of the human albumin gene (Mayuranathan et al., 2012) by multiplex PCR using fluorescently labeled forward primers **(Table S7)**, and the products were analysed by capillary electrophoresis. For accurate gene dosage analysis, 200ng of DNA and 20 cycles were used for the PCR with 0.4 µM labeled primers. The amplified products were separated by capillary electrophoresis in an ABI-3130 Genetic Analyser (Applied Biosystems), and the results were analysed by GeneMapper software version 4.0 (Applied Biosystems). The peak heights of the *UBE2T* amplicon were intranormalized by dividing them by the peak heights of the albumin gene. Subsequently, internormalization was performed by dividing the intranormalized peak height of *UBE2T* of the patient sample with that of the control sample.

Long-amplicon next-generation sequencing

The *FANCA* gene was amplified as 6 fragments, ranging from 7.15 to 15.25 kb, and *FANCG* as a single fragment of 6.73kb from the patients' DNA samples using GoTaq Long PCR Master Mix (Promega), according to the manufacturer's protocol. Primers were designed using Primer3 (v. 0.4.0) software (http://bioinfo.ut.ee/primer3-0.4.0/) to encompass all the exons and their splice sites **(Table S7)**. The amplified DNA fragments (amplicons) were analysed by 1% agarose gel electrophoresis. Based on the ethidium bromide staining intensities of the PCR amplicons, equal amounts of amplicons were pooled and sheared, and the libraries were prepared using TruSeq Nano DNA Library Preparation Kit (Illumina). The resulting libraries were sequenced on an Illumina NextSeq-500 sequencer to generate 2×150 bp reads at approximately 100× sequencing depth. The generated sequence data were analysed for quality, and a minimum of 75% of the sequenced bases (Q30) were processed to generate FastQ files. The NGS data (FastQ) were analysed by a pipeline developed in Galaxy (galaxyproject.org), as briefly outlined below. The browser extensible data (BED) file of the target region was created using BED Tools. The adapters were trimmed from the raw reads and aligned against the BED file using BWA (Galaxy version 0.7.17.1), alignment post-processing was performed by removing PCR duplicates[79], variant calling was done using FreeBayes, and then the variants were normalised using VCF ALLELIC PRIMITIVES[80]. The VCF files were further annotated using SnpSift (ref seq-hg19). The variants were also annotated using wANNOVAR to determine the mutations' functional consequences.

Plasmids

pLX301-*FANCA*, pLX301-*FANCG* and pLX301-*FANCC* were generated by Gateway cloning using the entry vectors pENTR223.1-*FANCA* (TransOMIC Technologies), pDONR221-*FANCG* (DNASU), and pDONR221-*FANCC* (DNASU), and the destination vector pLX301 (a kind gift from David Root, Addgene ID: 25895). Gateway cloning was carried out using the LR Clonase Enzyme mix (Thermo Scientific). *FANCL* and *FANCF* cDNAs were generated from a normal individual's blood cells using PrimeScript 1st Strand cDNA Synthesis Kit (TaKaRa) and gene specific primers (**Table S7**) and amplified using PrimeSTAR GXL PCR master mix (TaKaRa) with primer sequences suitable for Gibson cloning (**Table S7**). The amplified cDNAs were cloned into pCW-Cas9 (a kind gift from Eric Lander and David Sabatini, Addgene ID: 50661) by Gibson assembly using NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs) after removing the Cas9 gene with BamHI and NheI enzymes.

Lentivirus preparation and transduction of fibroblasts

Lentiviruses were prepared using a second-generation packaging system using the envelope plasmid pMD2.G and the packaging plasmid psPAX2 (gifts from Didier Trono, Addgene IDs: 12256 and 12260, respectively). The lentiviral expression, envelope and packaging plasmids were mixed in a 2:1:1 ratio and transfected into HEK 293T cells using the TransIT-293 transfection reagent (Mirus Bio), according to the manufacturer's protocol. The supernatants containing the lentiviral particles were collected at 48, 60, and 72 h post-transfection and were concentrated using Lenti-X Concentrator (Clontech Laboratories), according to the manufacturer's protocol. After reaching 70% confluency, fibroblasts were transduced with an appropriate volume of lentiviruses in 2 mL of medium containing 6 µg/mL polybrene (Sigma–Aldrich). The medium was changed 24 h post-transduction. After 72 h, antibiotics for selecting the transduced cells were added, and the cells were cultured in this antibiotics-containing medium until the selection was complete.

Joshi G, *et al. J Med Genet* 2023; 60:801–809. doi: 10.1136/jmg-2022-108714

Supplementary Figure legends

Figure S1. Geographical distribution of the FA patients recruited in this study.

Figure S2. Pipeline for the bioinformatics analysis of WES reads. *ClinVar, OMIM, HGMD, SwissVar, 1000 Genome Phase3, gnomAD, EVS, dbSNP, 1000 Japanese Genome and internal Indian population databases were used for variant annotations. $ Nonsynonymous mutations analysed by EVE algorithm.

Figure S3. Detection of large deletions in the FA pathway genes. (A) Representative results of two FANCA deletions detected by bioinformatics analysis in two different patients and confirmation of these deletions by MLPA. FA-24 is heterozygous for a deletion from exon 21 to exon 30, and FA-11 is homozygous for a deletion from exon 32 to exon 36. Probe mixes 1 and 2 cover all the exons of *FANCA*. (B) Position of the different *FANCA* deletions identified in this study (C) Detection of *FANCT(UBE2T)* deletion by bioinformatics analysis. (D-E) Genomic quantitative-PCR for gene dosage analysis to calculate the copy number of exon 7 of the *FANCT* gene. The peak heights of the amplified product of the exon 7 of the *UBE2T* gene and the control albumin gene are shown. Arrow indicates the absence of exome 7 peak in the patient. (E) Graph showing the gene dosage of *FANCT*.

Figure S4. MLPA analysis to detect *FANCA* gene deletions detected in 19 patients.

Figure S5. Molecular analysis of *FANCL* (c.1092G>A;p.K364=) identified in our cohort of FA patients. (A) Diagrammatic representation of the location of the mutation in the *FANCL* gene, exon 13 skipping in *FANCL* mRNA, the position of the primers used for amplification of the cDNA and Sanger sequencing are shown. (B) The PCR amplification of *FANCL* cDNAs from a normal control and an FA patient with the c.1092G>A;p.K364= mutation showing the difference in molecular weights of the amplified products. (C) Sanger sequencing results showing skipping of exon 13 of *FANCL* in the patient. (D) Western blot analysis showing the restoration of FANCD2 Ub after lentiviral complementation of a 'patient's fibroblasts with wild-type *FANCL* cDNA. The results were compared with a normal control and untransduced fibroblasts from a patient with *FANCL* (c.1092G>A;p.K364=) mutation. The vector used for complementation analysis was pCW-*FANCL*, which has doxycycline-inducible *FANCL* expression.

Figure S6. Long amplicon next-generation sequencing to identify mutations in *FANCA* and *FANCG* gene (A) Diagrammatic representation of the regions of *FANCA* and *FANCG* amplified for LA-PCR-NGS. (B) Agarose gel images showing the amplified *FANCA* and *FANCG* gene fragments. *FANCA*: fragment 1- 7126 bp; fragment 2- 8595 bp; fragment 3- 9726 bp; fragment 4- 13739 bp; fragment 5- 13483 bp; fragment 6- 15253 bp; and *FANCG*: fragment 7- 6760 bp.

Figure S7. Overall schematic representation of our study. Comparison of CBA scores and FANCD2-Ub results in the 142 FA patients recruited in this study. The interpretation and primary diagnosis based on the results of CBAs and FANCD2-Ub analysis before performing exome sequencing are shown on the right-hand side. *mosaic patients; # only heterozygous variants identified by NGS; NA: not analysed.

Figure S1

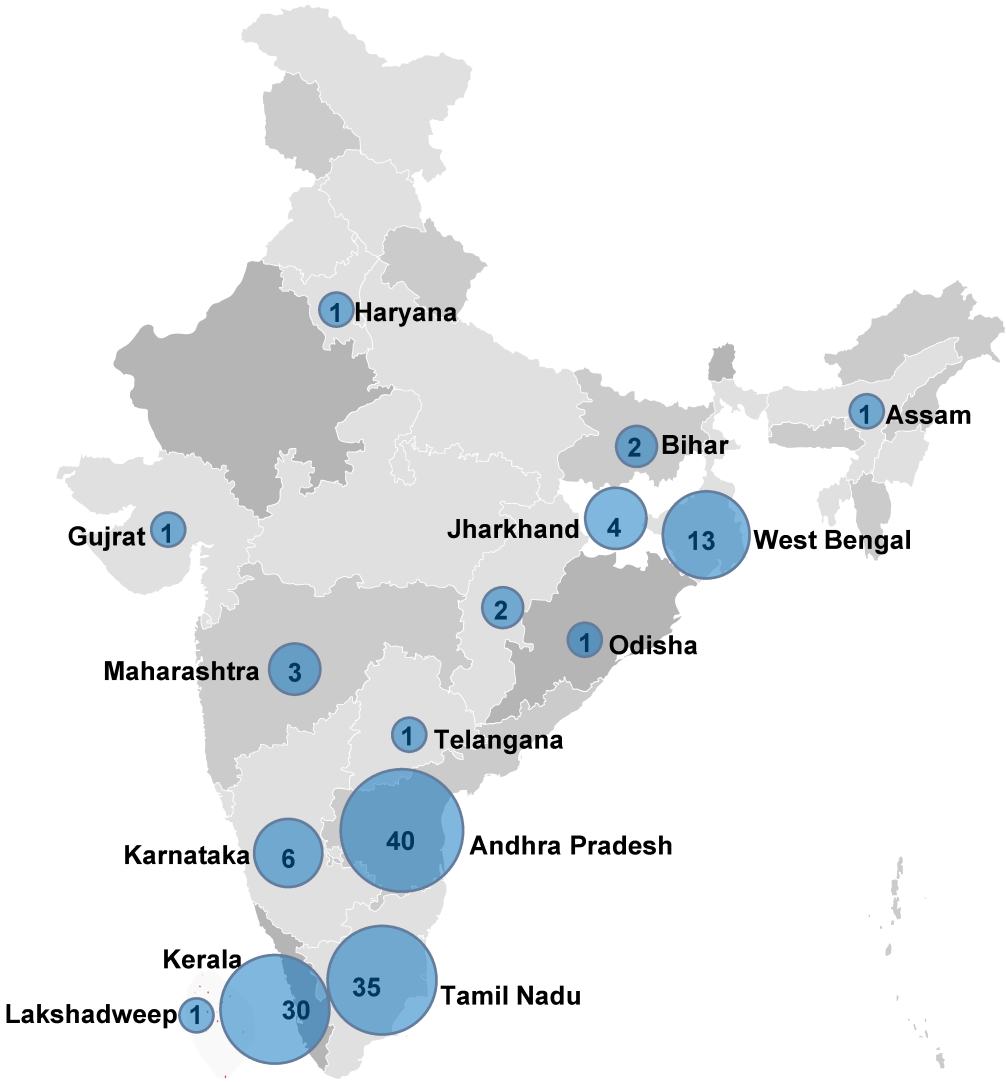

1 Haryana
1 Assam
2 Bihar
Gujrat 1
Jharkhand 4
13 West Bengal
2
1 Odisha
Maharashtra 3
1 Telangana
40 Andhra Pradesh
Karnataka 6
Kerala
35 Tamil Nadu
Lakshadweep 1
30

Figure S2

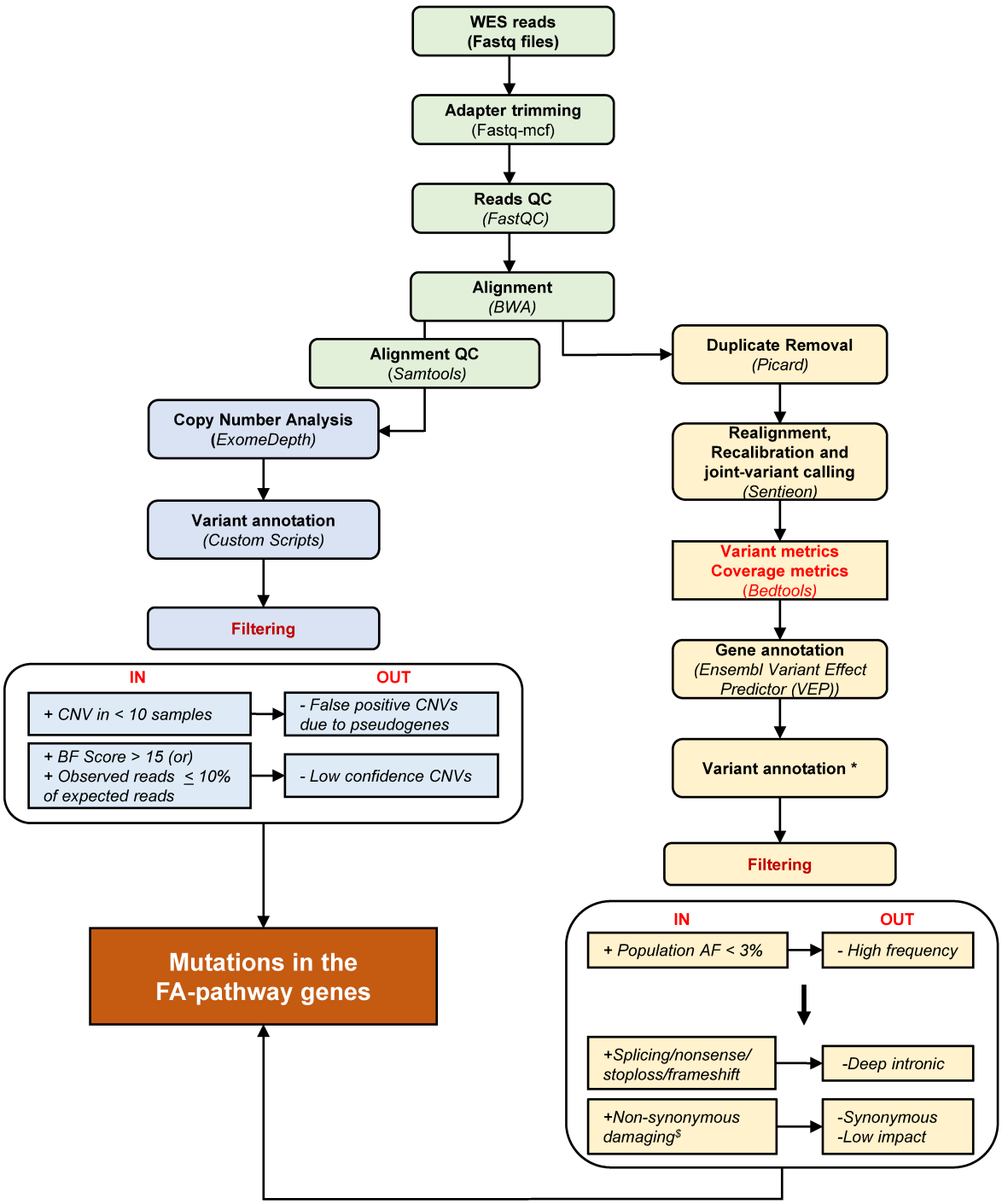

WES reads (Fastq files)
Adapter trimming (Fastq-mcf)
Reads QC (FastQC)
Alignment (BWA)
Alignment QC (Samtools)
Duplicate Removal (Picard)
Copy Number Analysis (ExomeDepth)
Realignment, Recalibration and joint-variant calling (Sentieon)
Variant annotation (Custom Scripts)
Variant metrics Coverage metrics (Bedtools)
Filtering
Gene annotation (Ensembl Variant Effect Predictor (VEP))
IN
OUT
+ CNV in < 10 samples
- False positive CNVs due to pseudogenes
+ BF Score > 15 (or) + Observed reads ≤ 10% of expected reads
- Low confidence CNVs
Variant annotation *
Filtering
IN
OUT
+ Population AF < 3%
- High frequency
+Splicing/nonsense/ stoploss/frameshift
-Deep intronic
+Non-synonymous damaging$
-Synonymous -Low impact
Mutations in the FA-pathway genes

Figure S3

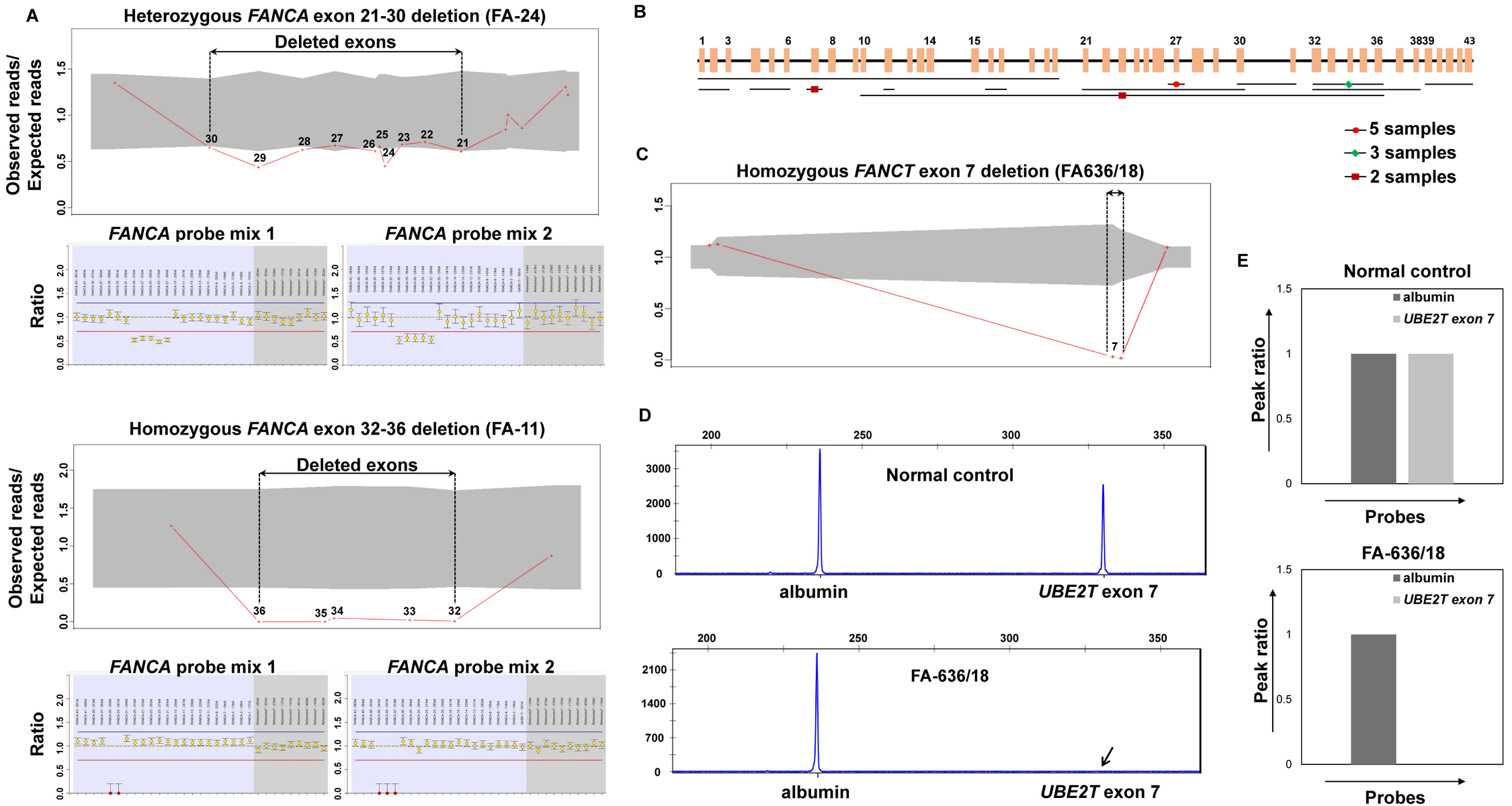
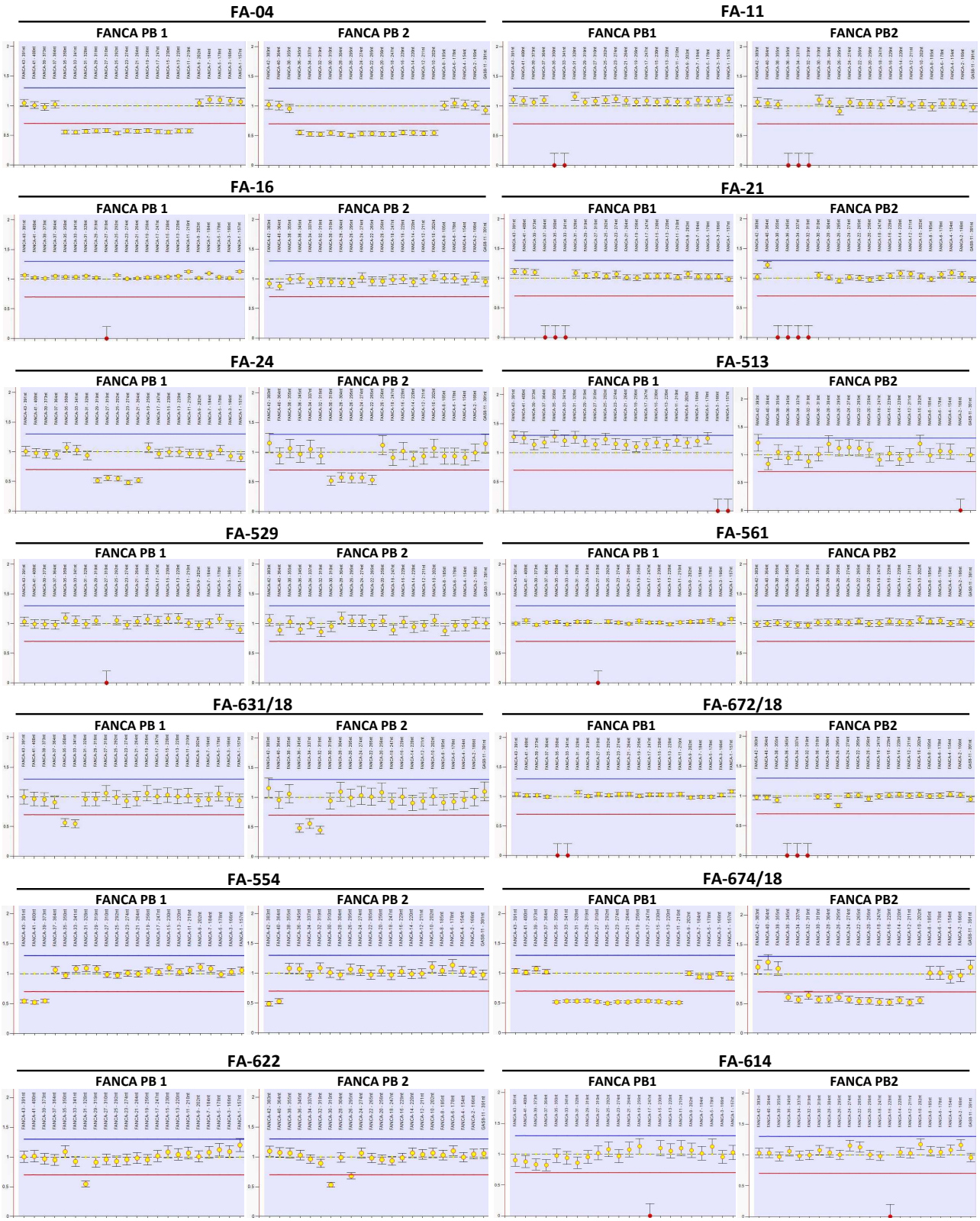

A
Heterozygous *FANCA* exon 21-30 deletion (FA-24)
Deleted exons
Observed reads/ Expected reads
FANCA probe mix 1
FANCA probe mix 2
Ratio
Homozygous *FANCA* exon 32-36 deletion (FA-11)
B
5 samples
3 samples
2 samples
C
Homozygous *FANCT* exon 7 deletion (FA636/18)
D
Normal control
albumin
UBE2T exon 7
FA-636/18
E
Normal control
albumin
UBE2T exon 7
Peak ratio
Probes
FA-636/18

Figure S4


FA-04
FANCA PB 1
FANCA PB 2
FA-11
FANCA PB1
FANCA PB2
FA-16
FANCA PB 1
FANCA PB 2
FA-21
FANCA PB1
FANCA PB2
FA-24
FANCA PB 1
FANCA PB 2
FA-513
FANCA PB 1
FANCA PB2
FA-529
FANCA PB 1
FANCA PB 2
FA-561
FANCA PB 1
FANCA PB2
FA-631/18
FANCA PB 1
FANCA PB 2
FA-672/18
FANCA PB1
FANCA PB2
FA-554
FANCA PB 1
FANCA PB 2
FA-674/18
FANCA PB1
FANCA PB2
FA-622
FANCA PB 1
FANCA PB 2
FA-614
FANCA PB1
FANCA PB2

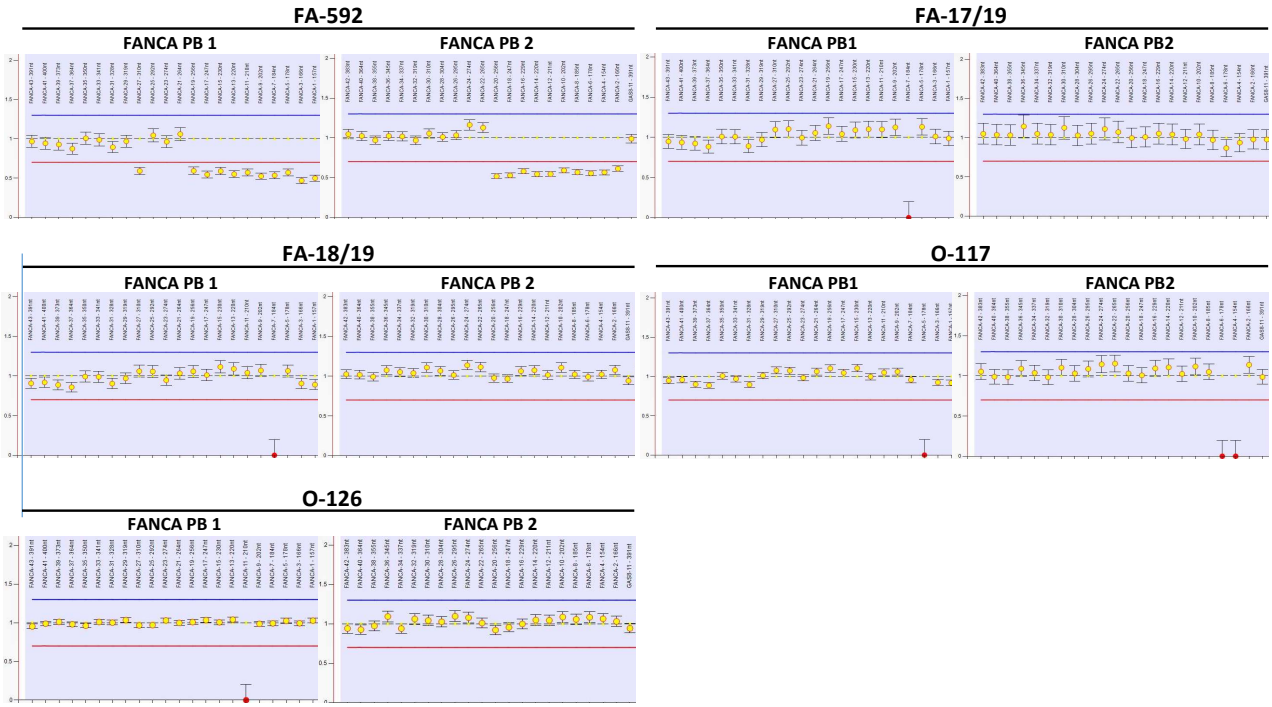

FA-592
FANCA PB 1
FANCA PB 2
FA-17/19
FANCA PB1
FANCA PB2
FA-18/19
FANCA PB 1
FANCA PB 2
O-117
FANCA PB1
FANCA PB2
O-126
FANCA PB 1
FANCA PB 2

Figure S5

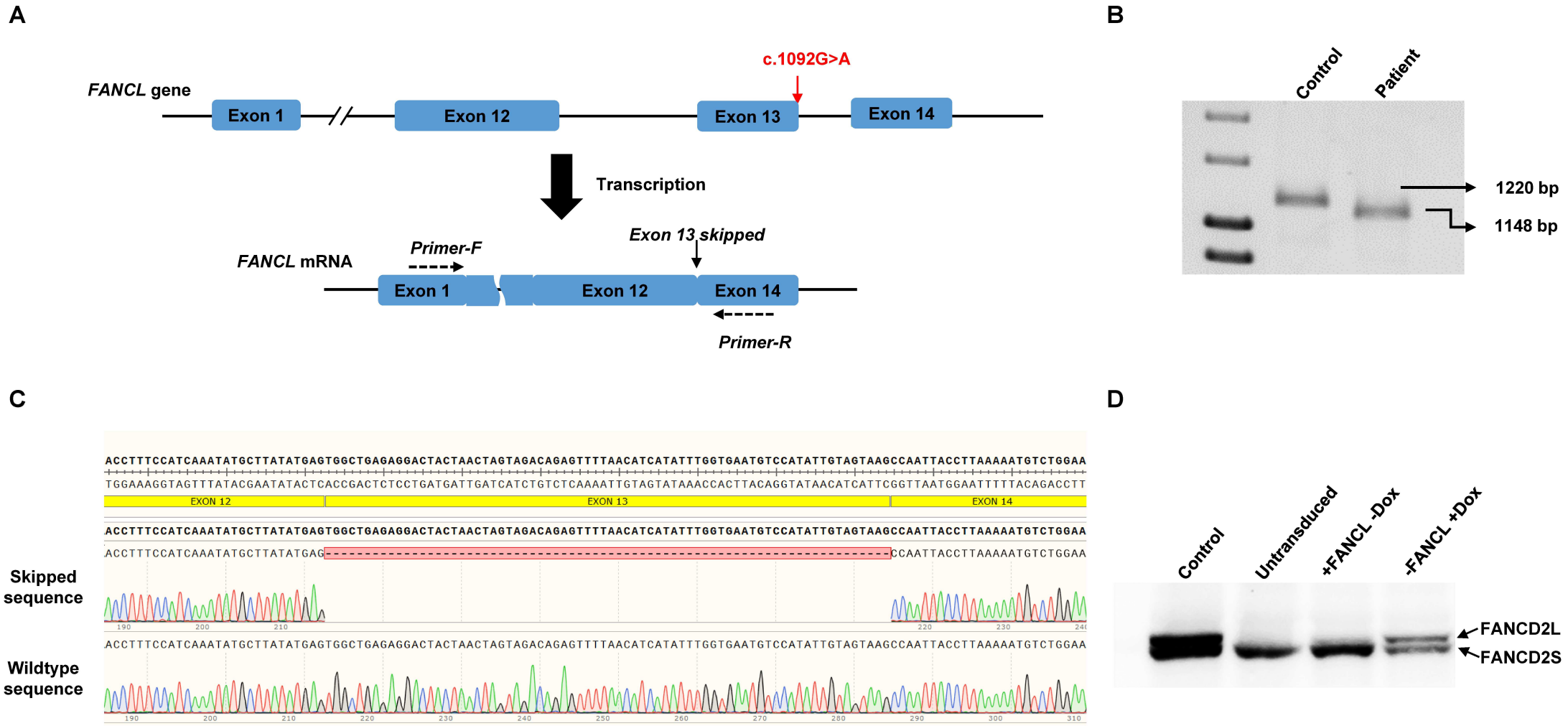
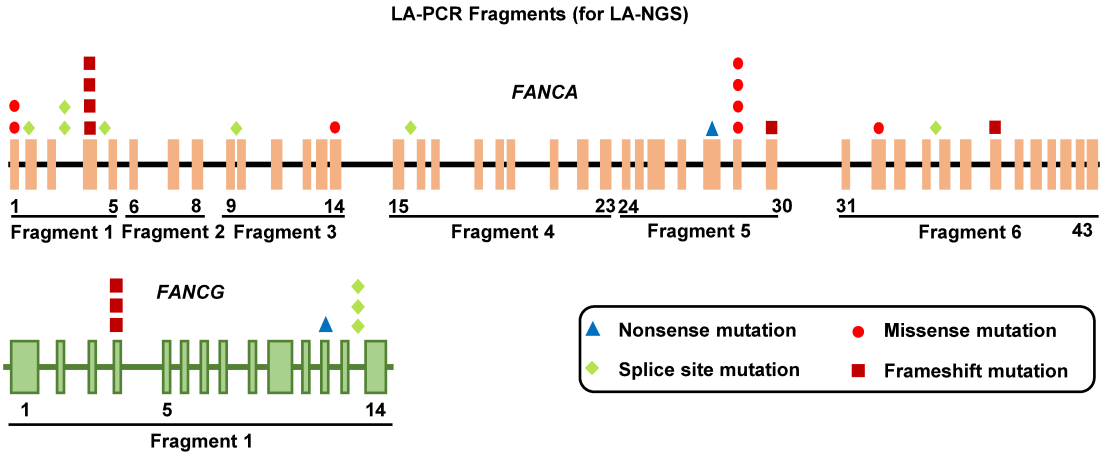

A
FANCL gene
Exon 1
Exon 12
Exon 13
Exon 14
c.1092G>A
Transcription
FANCL mRNA
Primer-F
Exon 13 skipped
Exon 1
Exon 12
Exon 14
Primer-R
B
Control
Patient
1220 bp
1148 bp
C
Skipped sequence
Wildtype sequence
EXON 12
EXON 13
EXON 14
D
Control
Untransduced
+FANCL -Dox
-FANCL +Dox
FANCD2L
FANCD2S

Figure S6

A


LA-PCR Fragments (for LA-NGS)
FANCA
1
5
6
8
9
14
15
23
24
30
31
43
Fragment 1
Fragment 2
Fragment 3
Fragment 4
Fragment 5
Fragment 6
FANCG
1
5
14
Fragment 1
Nonsense mutation
Missense mutation
Splice site mutation
Frameshift mutation

B

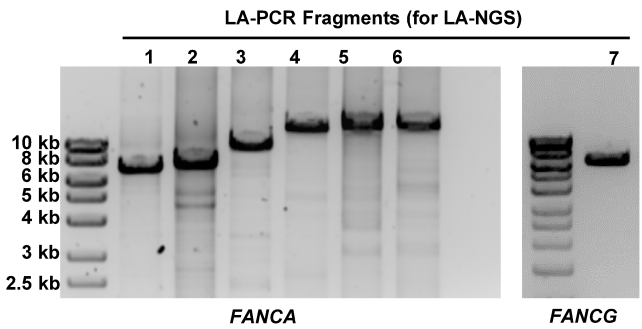

LA-PCR Fragments (for LA-NGS)
1
2
3
4
5
6
7
10 kb
8 kb
6 kb
5 kb
4 kb
3 kb
2.5 kb
FANCA
FANCG

Figure S7

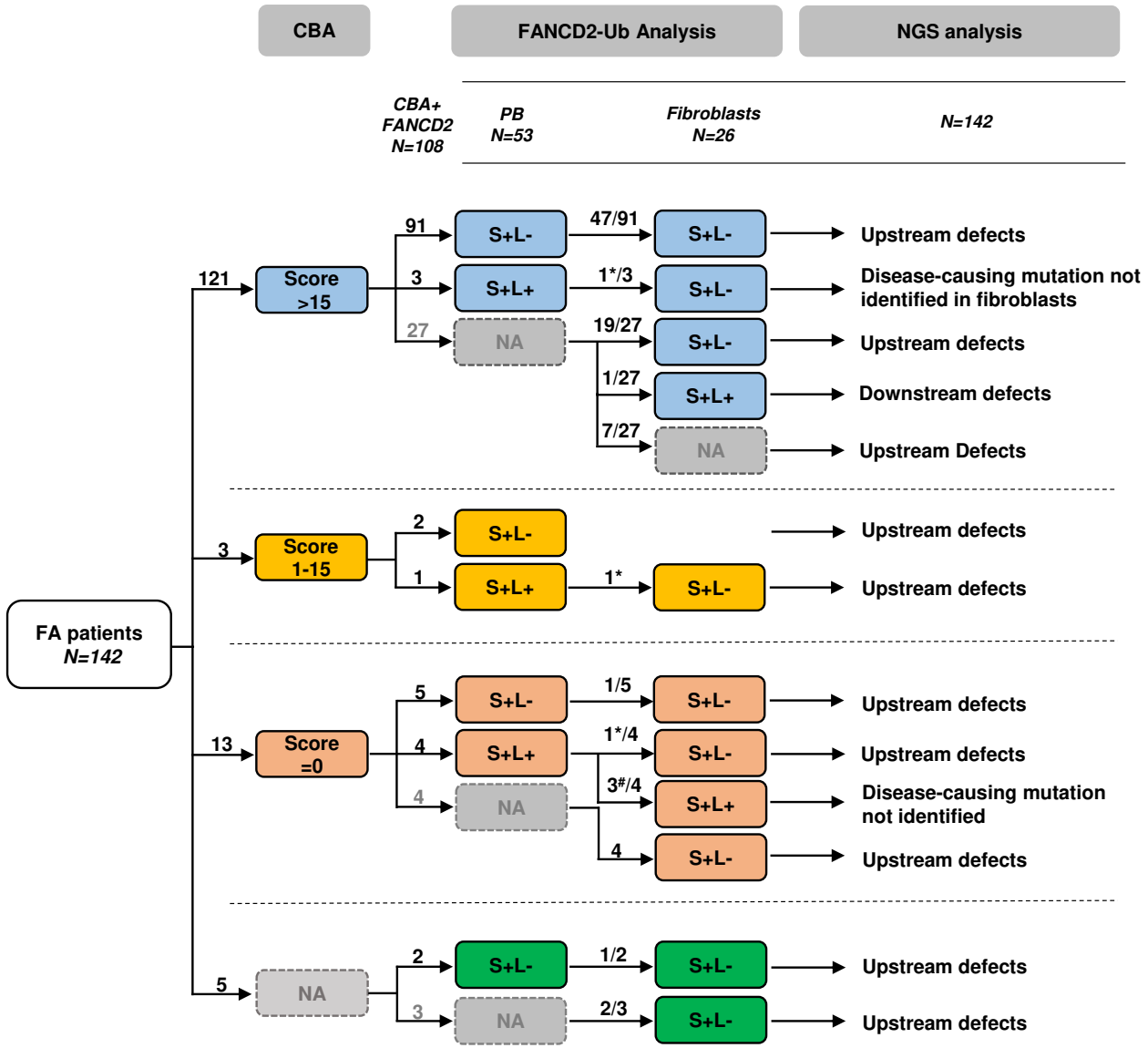

CBA
FANCD2-Ub Analysis
NGS analysis
CBA+ FANCD2 N=108
PB N=53
Fibroblasts N=26
N=142
FA patients N=142
121
Score >15
91
S+L-
47/91
S+L-
Upstream defects
3
S+L+
1*/3
S+L-
Disease-causing mutation not identified in fibroblasts
27
NA
19/27
S+L-
Upstream defects
1/27
S+L+
Downstream defects
7/27
NA
Upstream Defects
3
Score 1-15
2
S+L-
Upstream defects
1
S+L+
1*
S+L-
Upstream defects
13
Score =0
5
S+L-
1/5
S+L-
Upstream defects
4
S+L+
1*/4
S+L-
Upstream defects
4
NA
3#/4
S+L+
Disease-causing mutation not identified
4
S+L-
Upstream defects
5
NA
2
S+L-
1/2
S+L-
Upstream defects
3
NA
2/3
S+L-
Upstream defects

Table S2: FANCD2-Ub status, genotypes and clinical characteristics of the patients with very low CBA scores (0-10)

ID	CBA Score[#]	FANCD2 Ub		DNA change	Gene	Zygosity	Skeletal abnormalities	Facial dysmorphism	Skin changes	Other abnormalities
		PB	Fib							
FA-07	0	FANCD2-Ub-	NA	NC_000002.11:g.58387243C>T	*FANCL*	Homozygous	NA	MCH, frontal blossing	HP, HPOP	NA
FA-29	0	FANCD2-Ub+	FANCD2-Ub+	NC_000003.11:g.10076177A>C	*FANCD2*	Heterozygous	NA	NA	NA	HSK, kidney defects
FA-39	0	FANCD2-Ub+	FANCD2-Ub+	NC_000009.11:g.97873815G>A	*FANCC*	Heterozygous	Thenar hypoplasia	NA	HP	NA
FA-523*	0	FANCD2-Ub+	FANCD2-Ub-	NC_000002.11:g.58387243C>T	*FANCL*	Homozygous	NA	MCH	CAL, HP	NA
FA-525	0	NA	FANCD2-Ub-	NC_000002.11:g.58387243C>T	*FANCL*	Homozygous	Syndactyly, Thumb placed distally	NA	HP, HPOP	NA
FA-528	0	FANCD2-Ub+	FANCD2-Ub+	NC_000017.10:g.79516335G>T	*C17orf70*	Heterozygous	SS	MCH	NA	NA
FA-532	0	NA	FANCD2-Ub-	NC_000002.11:g.58387243C>T	*FANCL*	Homozygous	NA	DF	HP, HPOP	NA
FA-535	0	FANCD2-Ub-	NA	NC_000002.11:g.58387243C>T	*FANCL*	Homozygous	SS	MGH	HP	NA
FA-538	0	FANCD2-Ub-	NA	NC_000002.11:g.58387243C>T	*FANCL*	Homozygous	NA	MCH	CAL	NA
FA-541	0	NA	FANCD2-Ub-	NC_000006.11:g.35423588G>A	*FANCE*	Heterozygous	NA	NA	NA	NA
FA-561	0	FANCD2-Ub-	NA	NC_000016.9:g.89833551_89833647del	*FANCA*	Homozygous	NA	Hypertelorism	HP	NA
FA-554	4.1	FANCD2-Ub-	NA	NC_000016.9:g.89857810C>G NC_000016.9:g.89803957_89806507del	*FANCA*	Compound heterozygous	Bilateral clinodactyly	NA	CAL, HP, Pallor	NA
FA-650/18	4.1	FANCD2-Ub-	NA	NC_000009.11:g.35079204dup	*FANCG*	Homozygous	SS	NA	HP	NA
FA01/19*	8.2	FANCD2-Ub+	FANCD2-Ub-	NC_000002.11:g.58387243C>T	*FANCL*	Homozygous	NA	NA	CAL, HP	NA

SS: short stature; DF: dysmorphic face; HP: hyperpigmentation; MCH: microcephaly; HPOP: hypopigmentation; MGH: micrognathia; CAL: café au lait; HSK: horse shoe kidney.

* mosaicism detected by comparing the FANCD2-Ub analysis in the PB cells and fibroblasts.

[#] CBA score: chromosome breakage analysis score

Table S4: Heterozygous mutations in 6 FA patients

FA ID	Gene	Mutation	VarSome results	EVE results	Type of mutations	CBA score	FANCD2 ubiquitination		Physical anomalies
							T-cells	Fibroblasts	
FA-29	*FANCD2*	NC_000003.11:g.10076177A>C	Likely benign	Pathogenic	Missense	0	FANCD2-Ub+	FANCD2-Ub+	HSK, kidney defects
FA-518	*FANCC*	NC_000009.11:g.97873815G>A	Likely benign	Benign	Missense	0	FANCD2-Ub+	FANCD2-Ub+	Thenar hypoplasia, HP
FA-519	*FANCD2*	NC_000003.11:g.10083368C>T	Pathogenic	NA (nonsense variant)	Nonsense	62.7	FANCD2-Ub-	FANCD2-Ub-	SS, Absent simian crease, TH, TF, HP
FA-528	*C17orf70*	NC_000017.10:g.79516335G>T	Likely benign	NA	Missense	0	FANCD2-Ub+	FANCD2-Ub+	SS, MCH
FA-541	*FANCE*	NC_000006.11:g.35423588G>A	Likely benign	VUS	Missense	0	NA	FANCD2-Ub-	NA
FA-32/19	*FANCA*	NC_000016.9:g.89869731A>T	VUS	Pathogenic	Missense	22.1	FANCD2-Ub-	NA	Musculoskeletal abnormalities, Hypertelorism, MCH, HAP, CAL

HSK: horse shoe kidney; AE: abnormal eyes; MCH: microcephaly; HP: hyperpigmentation; SS: short stature; TH: thumb hypoplasia ;TF: triangular facies

Table S5: Mutations identified in *FANCA* and *FANCG* genes by LA-NGS

FA ID	Gene	Mutation	Type of mutations	Site of mutation
FA-31	*FANCA*	NC_000016.9:g.89882944C>G	Splice site	Intron 1
FA-521	*FANCA*	NC_000016.9:g.89877481T>G	Splice site	Intron 3
FA-500/ FA-05	*FANCA*	NC_000016.9:g.89857866C>T	Missense	Exon 14
FA-503/ FA-15	*FANCA*	NC_000016.9:g.89851261C>T	Splice site	Intron 15
FA-505/ FA-09	*FANCA*	NC_000016.9:g.89877340_89877341del	Splice site	Intron 4
		NC_000016.9:g.89877448del	Frameshift deletion	Exon 4
FA-515/ FA-03	*FANCA*	NC_000016.9:g.89813237A>C	Splice site	Intron 34
		NC_000016.9:g.89828423T>G	Missense	Exon 29
		NC_000016.9:g.89882965;C>T	Missense	Exon 1
FA-551/ FA-18	*FANCA*	NC_000016.9:g.89828357C>G	Missense	Exon 29
FA-555/ FA-516	*FANCA*	NC_000016.9:g.89825022del	Frameshift insertion	Exon 30
FA-568	*FANCA*	NC_000016.9:g.89828423T>G	Missense	Exon 29
FA-574/ FA-511	*FANCA*	NC_000016.9:g.89831327G>A	Nonsense	Exon 28
FA-20/19	*FANCA*	NC_000016.9:g.89877481T>G	Splice site	Intron 3
FA-27/19	*FANCA*	NC_000016.9:g.89866011A>G	Splice site	Intron 9
FA-36/19	*FANCG*	NC_000009.11:g.35074215T>G	Splice site	Intron 13
FA-37/19	*FANCG*	NC_000009.11:g.35074215T>G	Splice site	Intron 13
FA-573	*FANCA*	NC_000016.9:g.89877448del	Frameshift Deletion	Exon 4
		NC_000016.9:g.89883022A>T	Missense	Exon 1
FA-580	*FANCG*	NC_000009.11:g.35078222_35078223delinsAGCAGTT	Frameshift insertion deletion	Exon 4
FA-581	*FANCA*	NC_000016.9:g.89809216_89809219del	Frameshift Deletion	Exon 37
FA-584	*FANCA*	NC_000016.9:g.89877327_89877344del	Frameshift Deletion	Exon 4, Intron 4
FA-588	*FANCG*	NC_000009.11:g.35074215T>G	Splice site	Intron 13
FA-591	*FANCA*	NC_000016.9:g.89866011A>G	Splice site	Intron 9
FA-593	*FANCA*	NC_000016.9:g.89828358G>A	Missense	Exon 29
FA-598	*FANCG*	NC_000009.11:g.35075059G>A	Nonsense	Exon 12
FA-599	*FANCA*	NC_000016.9:g.89877327_89877344del	Frameshift Deletion	Exon 4, Intron 4
FA-622	*FANCA*	NC_000016.9:g.89816231_89816232del	Frameshift deletion	Exon 32

Table S6: List of primers used in this study

Purpose	Target Gene	Forward Primer (5'-3')	Reverse primer (5'-3')
Long amplicons for NGS	*FANCA*	TGGTCTCTTCAGGACCAACC	TGCGATAAGCCAAGATAGCA
		GATCTGGATGGAGGCAACAGAGT	CCCAACAAGAGATGACCGGATAC
		GGGCTTTGTTTGAGGAAGTCTGTT	CGAGAAGGTGAGCTTTCTGTACCA
		TACCCCTAAGGATCCCAAAAGGAT	AGAAGGCTCCATGCGTCTAA
		GCCTCGACTGGTCTAGAAGTCC	TGCCCAGGATCTACTAGGCCATTT
		AGGGGCAGGGTAAACAATGTGAGT	CCCTCAAGTACCACATGACCAAAC
	FANCG	GCTCGACAGTGAGCAGAGAAAGGAT	CTTTCACCCTGTACCCACACAGACA
Gene dosage PCR	*UBE2T*	TTGATAATCTACCAGAGGCT	GGCCAGTTTACTCCCAGACA
Primer for cDNA synthesis of *FANCL* gene	*FANCL*	-	TTAAGTTTCCAGCTCTTCAC
Primer for detection of mutation (c.1092G>A;p.K364=) in *FANCL*		CATCCCTTACTCTTGCAAAACA	TCACCTAGGAAATCTAGAAAAGGA
Amplification of *FANCL* cDNA		GTCTAGAGCTTTTCTGTGTT	TTAAGTTTCCAGCTCTTCAC
Primer for cDNA synthesis of *FANCF* gene	*FANCF*	-	ATACTTTGGACACACGAAGG
Gibson cloning of *FANCF* gene into pCW vector	*FANCF*	tggctagcGTAGGGCCTTCGCGCAC	ccggatccCATATATTTGGTGAGAACATTGTAATTTTCATTTTGTAAAC
Gibson cloning *FANCL* gene into pCW vector	*FANCL*	cagatcgcctggagaattggGTTTCTCCGGACTTCGAG	aaggcgcaaccccaaccccgCGAAATGTTGTATTCTTATTTCAGTG

Table S7: The genotypes of the FA patients selected for complementation analysis

ID	Gene	Zygosity	DNA change	Type of mutation	Varsome	EVE
FA-01	*FANCG*	Homozygous	NC_000009.11:g.35074215T>G	Splice site	Pathogenic	NA
FA-03	*FANCA*	Compound heterozygous	NC_000016.9:g.89813237A>C	Splice site	Pathogenic	NA
			NC_000016.9:g.89828423T>G	Missense	VUS	Pathogenic
FA-03/19	*FANCI*	Compound heterozygous	NC_000015.9:g.89804822del	Frameshift deletion	Pathogenic	NA
			NC_000015.9:g.89848828_89848832del	Intronic splice acceptor variant	VUS	NA
FA-04	*FANCA*	Compound heterozygous	NC_000016.9:g.89882396T>C	Splice site	Pathogenic	NA
			NC_000016.9:g.89811367_89865640del	Deletion	Pathogenic	NA
FA-06	*FANCA*	Homozygous	NC_000016.9:g.89828423T>G	Missense	Likely pathogenic	Pathogenic
FA-12	*FANCA*	Compound heterozygous	NC_000016.9:g.89809286_89809290dup	Frameshift insertion	Likely pathogenic	NA
			NC_000016.9:g.89857867G>A	Missense	Likely pathogenic	Pathogenic
FA-18	*FANCA*	Homozygous	NC_000016.9:g.89828357C>G	Missense	Likely pathogenic	Pathogenic
FA-24	*FANCA*	Compound heterozygous	NC_000016.9:g.89809302C>T	Nonsense	Pathogenic	NA
			NC_000016.9:g.89828358_89842224del	Deletion	Pathogenic	NA
FA-26	*FANCG*	Homozygous	NC_000009.11:g.35076427_35076431del	Splice site	VUS	NA
FA-30	*FANCF*	Homozygous	NC_000011.9:g.22646233T>C	Stop loss	VUS	NA
FA-31	*FANCA*	Homozygous	NC_000016.9:g.89882944C>G	Splice site	Pathogenic	NA
FA-36	*FANCG*	Homozygous	NC_000009.11:g.35075650_35075651del	Frameshift deletion	Pathogenic	NA
FA-38	*FANCG*	Compound heterozygous	NC_000009.11:g.35078223A>G	Missense	VUS	Pathogenic
			NC_000009.11:g.35076856_35076857del	Frameshift deletion	VUS	NA
FA-40	*FANCA*	Homozygous	NC_000016.9:g.89877448del	Frameshift deletion	Pathogenic	NA
FA-517	*FANCC*	Homozygous	NC_000009.11:g.97873912C>A	Nonsense	Pathogenic	NA
FA-527	*FANCC*	Homozygous	NC_000009.11:g.97864081T>G	Missense	VUS	Pathogenic
FA-531	*FANCG*	Homozygous	NC_000009.11:g.35076026C>G	Splice site	Pathogenic	NA
FA-533	*FANCA*	Compound heterozygous	NC_000016.9:g.89858399_89858400del	Frameshift deletion	VUS	NA
			NC_000016.9:g.89806402C>T	Missense	VUS	Pathogenic
FA-543	*FANCL*	Homozygous	NC_000002.11:g.58387243C>T	Exonic splice donor variant	Pathogenic	NA
FA-641/18	*FANCF*	Homozygous	NC_000011.9:g.22647316A>C	Missense	VUS	VUS

NA- not applicable

VUS- Variants of uncertain significance

Table S8: Calculation of sensitivity of CBA and FANCD2 Ub analysis with NGS as gold standard

		CBA		
		Disease		
Test	**Present**	**n**	**Absent**	**n**
Positive	True Positive	119	False positive	2
Negative	False Negative	13	True Negative	3
Sensitivity	91.50%			
Specificity	60%			

		FANCD2 Ub analysis		
		Disease		
Test	**Present**	**n**	**Absent**	**n**
Positive	True Positive	130	False positive	2
Negative	False Negative	2	True Negative	3
Sensitivity	97.00%			
Specificity	60%			

CBA and FANCD2 Ub analysis is compared to NGS