Hypomorphic Mutations in the Central Fanconi Anemia Gene *FANCD2* Sustain a Significant Group of FA-D2 Patients with Severe Phenotype

Running title: FA-D2 phenotype and *FANCD2* mutations

Reinhard Kalb,[1] Kornelia Neveling,[1] Holger Hoehn,[1] Hildegard Schneider,[2] Yvonne Linka,[2] Sat Dev Batish,[3] Curtis Hunt,[4] Marianne Berwick,[4] Elsa Callén,[5] Jordi Surrallés,[5] José A. Casado,[6] Juan Bueren,[6] Ángeles Dasí,[7] Jean Soulier,[8] Eliane Gluckman,[8] C. Michel Zwaan,[9] Rosalina van Spaendonk,[10] Gerard Pals,[10] Johan P. de Winter,[10] Hans Joenje,[10] Markus Grompe,[11] Arleen D. Auerbach,[3] Helmut Hanenberg,[2, 12] and Detlev Schindler[1]

From the [1]Department of Human Genetics, University of Wurzburg, Germany; [2]Department of Pediatric Oncology, Hematology and Immunology, University of Dusseldorf, Germany; [3]Laboratory of Human Genetics and Hematology, The Rockefeller University, New York, NY; [4]Division of Epidemiology, University of New Mexico, Albuquerque, NM; [5]Department of Genetics and Microbiology, Universitat Autónoma de Barcelona, Bellaterra, Spain; [6]Hematopoietic Gene Therapy Program, CIEMAT, Madrid, Spain; [7]Unit of Pediatric Hematology, Hospital la Fe, Valencia; Spain; [8]Institut Universitaire d'Hematologie, Hopital Saint-Louis, Paris, France; [9]Department of Pediatric Hematology/Oncology, Erasmus MC-Sophia Children's Hospital, Rotterdam, The Netherlands, and Dutch Childhood Oncology Group; [10]Department of Clinical Genetics and Human Genetics, Vrije Universiteit Medical Center, Amsterdam, The Netherlands; [11]Department of Medical and Molecular Genetics, Oregon Health and Science University, Portland, OR; [12]Department of Pediatrics, Indiana University School of Medicine, Indianapolis, IN

Research grants and other financial support: Supported, in part, by grants from the 'Deutsche Fanconi-Anämie-Hilfe' (Ho.Ho., D.S.), the Schroeder-Kurth Fund (R.K., D.S.), the Senator Kurt und Inge Schuster Foundation (R.K.) and the National Institutes of Health (R37 HL32987, A.D.A.; R01 CA82678, M.B. and A.D.A.). The work of J.S. was funded by grants from the European Union (FI6R-CT-2003-508842), the Spanish Ministries of Science (SAF-2003-00328, SAF2006-03340) and Health (PI051205, G03/073), and by the La Caixa Foundation Oncology Programme. The CIEMAT has been supported by grants from Spanish Ministry of Health (G03/073), the VI Framework Program of the E.U. (CONSERT; Ref. 005242), and the Spanish Interministerial Commission for Science and Technology (SAF 2005-00058). J.P.d.W., H.J. and He.Ha. were supported by the Fanconi Anemia Research Fund and J.P.d.W. and H.J. by the Dutch Cancer Society.

Address for correspondence and reprints: Dr. Detlev Schindler, Department of Human Genetics, University of Wurzburg, Biozentrum, Am Hubland, D-97074 Wurzburg, Germany. Phone: +49 931 888 4089; FAX: +49 931 888 4069; E-mail: schindler@biozentrum.uni-wuerzburg.de

Word counts: 5.842

Abstract

FANCD2 is an evolutionarily conserved Fanconi anemia (FA) gene that plays a central role in DNA double-strand type damage responses. Using complementation assays and immunoblotting, a consortium of American and European groups assigned 29 FA patients from 23 families and 4 additional unrelated patients to complementation group FA-D2. This amounts to 3 to 6% of FA patients registered in various datasets. Malformations are frequent in FA-D2 patients and hematological manifestations appear earlier and progress more rapidly when compared to patients from all other FA groups combined, as represented by the International Fanconi Anemia Registry, IFAR. *FANCD2* is flanked by two pseudogenes. Mutation analysis revealed the expected total of 66 mutated alleles, 34 of which result in aberrant splicing patterns. Many mutations are recurrent and have ethnic associations and shared alleles. There were no biallelic null mutations so that residual FANCD2 protein of both isotypes was observed in all patients' cell lines available. These analyses suggest that unlike in a knock-out mouse model, total absence of FANCD2 is not existing in FA-D2 patients due to constraints on viable combinations of *FANCD2* mutations. Although hypomorphic mutations are involved, the result generally is a relatively severe form of FA.

Key words: Fanconi anemia; FANCD2; Hypomorphic mutations; Splicing; Residual protein

Introduction

Fanconi anemia (FA) is a rare genome instability disorder with the variable presence of congenital malformations, progressive bone marrow failure, predisposition to malignancies, and cellular hypersensitivity towards DNA-interstrand crosslinking (ICL) agents[1]. There are at least twelve complementation groups (FA-A, B, C, D1, D2, E, F, G, I, J, L and M), each of which is associated with biallelic or hemizygous mutations in a distinct gene[2]. To date, eleven of the underlying genes have been identified, denoted *FANCA, B, C, D1/BRCA2, D2, E, F, G/XRCC9, J/BRIP1, L/PHF9 and M/HEF*[3-5]. Eight of the FA proteins (FANCA, B, C, E, F, G, L and M) and other components assemble in a nuclear complex, the FA core complex, that is required for the monoubiquitination of FANCD2 at amino acid residue K561[6,7]. Monoubiquitination occurs in response to DNA damage and during the S phase of the cell cycle[7,8]. The monoubiquitinated FANCD2 isoform (FANCD2-L as opposed to FANCD2-S) is targeted to nuclear foci containing proteins such as BRCA1, BRCA2 and RAD51 that are involved in DNA damage signaling and recombinational repair[9-12]. The precise role of FANCD2 remains unknown, but FANCD2-deficient DT40 cells show defects in homologous recombination-mediated DNA double-strand break (DSB) repair, translesion synthesis and gene conversion[9,13,14]. Therefore, FANCD2 is thought to play a central role in the maintenance of genome stability[9,14,15].

The human and murine *Fancd*2 genes show a higher degree of homology than the corresponding *Fanca*, *c*, *e*, *f* and *g* genes[16]. *Fancd2* knock-out mice suffer from perinatal lethality, microphthalmia and early epithelial cancers[17], but it remains controversial whether the murine FA-D2 phenotype generally is more severe than the corresponding murine knock-outs of the other FA genes[17,18]. Fancd2 is required for survival after DNA damage in C. elegans[19], and Fancd2-deficient zebrafish embryos display severe developmental defects due to increased apoptosis, underscoring the

importance of *Fancd2* function during vertebrate ontogenesis[20]. Finally, knock-down of drosophila Fancd2 causes pupal lethality[21]. In humans, it has been estimated that complementation group FA-D2 accounts for less than 1%[22] to 3%[23] of all FA patients. The presence of *FANCD2* pseudogenes complicating mutation analysis may explain why there has been no other report of human *FANCD2* mutations since the original description[24]. As a concerted effort among nine laboratories we present a comprehensive mutation profile of the *FANCD2* gene. We show that the FA phenotype resulting from FANCD2 deficiency is relatively severe and, in contrast to all other FA genes, (1) the mutation spectrum of *FANCD2* is dominated by splicing mutations, and (2) residual FANCD2 protein exists in all tested cell lines from FA-D2 patients, suggesting lethality of biallelic null mutations.

Patients, materials and methods

Diagnostic procedures

Anti-coagulated peripheral blood and skin biopsy samples were referred to the participating laboratories for diagnostic testing. Confirmation of the diagnosis of FA, subtyping and mutation analysis were performed with informed consent according to the Declaration of Helsinki. The study was approved by the institutional review boards of the participating centers. Clinical suspicion of FA was confirmed by the detection of cellular hypersensitivity to DNA-crosslinking agents following published procedures[25-29]. In cases with hematopoietic mosaicism, skin fibroblasts were used to confirm the diagnosis of FA.

Patient statistics

A total of 29 fully informative FA-D2 patients (no. 1-29) were included in the present genotype-phenotype study. A fetal case (no. 19) and five patients with hematopoietic mosaicism (no. 3, 14, 15, 25 and 26) were excluded from clinical follow-up studies. Four additional FA-D2 patients (no. 30-33) with incomplete clinical data are not part of the phenotype analysis, but results concerning their mutations are shown as indicated in the text, tables and figures.

For calculations of cumulative incidence, the following three end points were evaluated: time to bone marrow failure (BMF; hematological onset, defined as cell count of at least one lineage constantly below normal range), period from BMF to hematological stem cell transplantation (HSCT), and time to HSCT. Kaplan-Meier estimates were computed for the length of overall survival. Birth was taken as the date of FA onset for all these calculations. Comparisons were made to a body of FA patients in the IFAR as previously reported[30] by means of log-rank test statistics.

Multivariate and competing-risk analyses were not possible due to the limited number of informative patients.

Cell culture

Epstein-Barr virus (EBV)-transformed lymphoblastoid cell lines (LCLs) were established using cyclosporin A as previously described[31]. All blood-derived cell cultures were maintained in RPMI 1640 medium with GlutaMAX (Gibco) supplemented with 15% fetal bovine serum (FBS; Sigma). Fibroblast strains were established using standard cell culture procedures and propagated in Earle's MEM with GlutaMAX (Gibco) and 15% FBS. All cultures were kept in high humidity incubators in an atmosphere of 5% (v/v) CO_2 and, in case of fibroblasts, 5% (v/v) O_2 by replacing ambient air with nitrogen[32]. Mitomycin C (MMC) treatments were for 48 h at 12 ng/ml (fibroblasts) or 15 ng/ml (LCLs) to cause cell cycle arrest, or for 12 h at 100 ng/ml to induce FANCD2-L. In some cases, RNA stabilization was achieved by cycloheximide (CHX) added to cell cultures at a final concentration of 250 μg/ml 4.5 h prior to RNA isolation.

Retroviral complementation

For construction of the D2-IRES-neo retroviral expression vector S11FD2IN, the D2-IRES-puro plasmid pMMP-FANCD2[24] was cut using Sal I. The ends were blunted and the fragment containing the FANCD2 ORF was cut out with EcoRI and ligated into S11IN cut with BamHI, blunted and also cut again with EcoRI (Figure 1A and Supplementary Figure S1). S11 vectors are based on the spleen focus forming virus and are derived from the GR plasmid[33]. Sequencing of the retroviral plasmid S11FD2IN revealed three reported polymorphisms in the FANCD2 ORF, c.1122A>G, c.1509C>T, c.2141C>T[24] and another silent base substitution, c.3978C>T. Stable

retroviral packaging cells were generated by infection of PG13 cells and selection in G418 (Sigma) as previously described[34]. In addition, enhanced green fluorescent protein (GFP) and FANCA cDNAs were separately cloned into the vector S11IN (designated S11EGIN and S11FAIN; Figure 1A) for transduction of the cells under study, with EGFP serving to monitor complete selection and FANCA serving as negative complementation control.

Retroviral transduction of cultured cells follwed published protocols[35,36]. Selection of transduced cells was in G418 (Sigma) at a final concentration of 0.8 to 1.2 mg/ml for about 10 days. Transduced cells were analyzed for their sensitivity to MMC using flow cytometry to assess survival rates and cell cycle arrest[36,37].

Immunoblotting

FANCD2 immunoblotting was performed as first described[7] with minor modifications. Detection was by the chemiluminescence technique using standard ECL reagent (Amersham) or SuperSignalWestFemto (Pierce).

Mutation and haplotype characterization

Primers used for cDNA amplification are shown in Supplementary Table S1A, those additionally used for cDNA sequencing are shown in Supplementary Table S1B.

A total of 15 large amplicons (superamplicons) that are unique to certain regions of the functional *FANCD2* gene were generated to serve as templates in place of genomic DNA. The primers used for superamplifications and the sizes of the superamplicons are shown in Supplementary Table S2A. Genomic primers for the amplification of all *FANCD2* exons and adjacent intron regions and their sizes are displayed in Supplementary Table S2B. Additional genomic mutation-specific primers are given in Supplementary Table S2C.

For haplotyping, four microsatellite markers in the vicinity up- and downstream of *FANCD2* on chromosome 3 were studied as detailed in the Supplementary methods. Primers used for microsatellite amplifications are specified in Supplementary Table S3.

Results

Assignment to and frequency of group FA-D2

Figure 1B-E demonstrates our strategy for the assignment of cultured FA cells to group FA-D2. Cell cycle analysis was used to ensure MMC sensitivity by G2 phase arrest of the LCLs (Figure 1C, lane 2)[27-29], while the apparent absence of FANCD2 bands on standard exposure immunoblots suggested their belonging to group D2 (Figure 1B, lane 2)[38]. Transduction of putative D2 LCLs with *FANCD2* cDNA using the retroviral vector S11FD2IN restored FANCD2 expression and function as reflected by the emergence of both FANCD2 isoforms (FANCD2-S and -L; Figure 1B, lanes 3); simultaneously, the MMC sensitivity of transduced cells returned to normal control cell levels as evidenced by the reduction of G2 phase cell cycle fractions (Figure 1C, lane 3; Figure 1B and C, lanes 1). Transduction of D2 LCLs with *GFP* or *FANCA* did not result in the restoration of either FANCD2 isoform nor in a normalization of G2 phase arrest, as exemplified *for FANCA* using S11FAIN in Figure 1B and C, lanes 4. In case of suspected hematopoietic mosaicism, cultured fibroblasts were assayed using a corresponding strategy (Figure 1D and E). As shown in Supplementary Table S4, only a few patients were assigned to group FA-D2 by primary mutation analysis, including four affected siblings of four different index patients, and an unrelated deceased patient with only DNA available.

In the North American IFAR collection, of 630 classified FA patients 18 were assigned to D2. Within in the period of study, seven fully informative of them could be included in the present cohort (no. 19-25); another one is among the four additional patients (no. 32). Among the patients referred to the two German labs, 15/243 FA patients were D2. These data suggest that the proportion of complementation group

D2 among two larger series of FA patients may be more frequent than previously reported[2,22,23].

Clinical data of FA-D2 patients

Including, where possible, information from a prenatal case (no. 19), malformations in the present cohort of 29 FA-D2 patients with adequate clinical information were of the following types and frequencies (Supplementary Table S5): 25/28 (89%) had microcephaly, 25/29 (86%) (intrauterine) growth retardation, 21/28 (75%) anomalies of skin pigmentation, 21/29 (72%) radial ray defects, 17/28 (61%) microphtalmia, 10/28 (36%) renal anomalies, 9/28 (32%) malformations of the external ear, 9/29 (31%) anomalies of the brain (including 5/29 or 17% with hydrocephalus), 7/28 (25%) hypogonadism or other genital anomalies, 4/28 (14%) anomalies of the heart and 4/28 (14%) malformations of the GI tract. Of note was a high proportion of FA-D2 patients with psychomotor retardation and attention deficit/hyperactivity disorder (8/28, 29%). Dysplasia and dislocation of the hip (6/28, 21%) also were relatively common. VACTERL-like association (1/28, holoprosencephaly (1/28), the Karthagener syndrome (1/28) and the Michelin tire baby syndrome (2/28) were noted as distinct disorders occurring in some FA-D2 patients. With the exception of single families, there was no general tendency of our families with FA-D2 offspring for increased rates of spontaneous abortions. Among the 28 fully informative FA-D2 patients, there was only a single malignancy (AML) during observation and there was no apparent overrepresentation of malignancies in the parents or grandparents of the D2 patients in our cohort.

Median age at diagnosis of these FA-D2 patients was 4 y and 5 mo (n=29). Excluding the fetal case (no. 19) and five mosaic patients (no. 3, 14, 15, 25 and 26), the median age of transfusion dependency was 10 y 10 mo (n=23). Figure 2

compares the progressive hematological course and the outcome of our group of FA-D2 patients to previously reported altogether 754 North American IFAR patients[39]. BMF in our D2 group (n=23) occurred at an earlier age (median D2 4 y vs. IFAR 6 y 7 mo, p=0.001; Figure 2A), and the period from BMF to HSCT was shorter (D2 n(HSCT)=9, median 5 y 6 mo. vs. IFAR n(HSCT)= 218, median 11 y 4 mo; p<0.08; Figure 2B). HSCT in our D2 patients was earlier than in the IFAR patients of combined groups (median D2 10 y 11 mo vs. IFAR 27 y 11 mo, p<0.01; Figure 2C). 9/23 FA-D2 patients of our cohort had HSCT. Kaplan-Meier estimates (Figure 2D) suggest that our D2 patients (n=23) may have a shorter overall lifespan as their survival curve falls below that of the IFAR patients after age 9 y; however, the difference of median survival (D2 11 y 4 mo vs. IFAR 24 y 3 mo) was not significant due to only two non-mosaic D2 patients who reached adulthood.

FANCD2 and the *FANCD2* pseudogenes

BLAT searches identified two pseudogene regions: *FANCD2-P1* spanning 16 kb, located about 24 kb upstream of *FANCD2*, and *FANCD2-P2* spanning 31.9 kb, located about 1.76 Mb downstream of *FANCD2* (Figure 3A). *P1* and *P2* are in the same orientation as the functional gene. They are characterized by high homology with certain *FANCD2* exons and have retained ordered arrays of their exon equivalents. On the other hand, the exon replicas of *FANCD2-P1* and *FANCD2-P2* have acquired numerous deletions and insertions. Striking sequence similarity of the D2 pseudogenes extends into some *FANCD2* introns, prominently the regions of IVS21-IVS26. Thus, *P1* and *P2* reveal recognizable patterns of conserved gene structure (Figure 3B). *FANCD2-P1* is a rough copy of the front portion of *FANCD2* including, with intermittent gaps, the region of exons 1-18 (homology with *FANCD2* exons 1, 12 to 16 and the 3´ portion of exon 18). The region upstream of *FANCD2-P1*

shares homology with the putative *FANCD2* promoter predicted within the CpG-rich region of approximately 800 bp upstream of the start codon of the functional gene. The corresponding region upstream of *P1* is interrupted by an *AluY* element. *FANCD2-P2* is an approximate match of the middle portion of *FANCD2* spanning, also with gaps, the region of exons 12 through 28 (homology with *FANCD2* exons 12 to 14 and 17 to 28).

Mutations in *FANCD2*

Unique amplification of the functional *FANCD2* gene using primers excluding pseudogene sequences resulted in 15 superamplicons (Figure 3C) that were used for genomic mutation screens. Studies at the RNA level were implemented to guide the genomic analyses. All mutations identified and their predicted consequences at the protein level are compiled in Table 1. The distribution of the mutations among the individual patients is shown in Supplementary Table S4.

Mutations affecting pre-mRNA splicing

In PBLs, LCLs and cultured fibroblasts from normal controls, two species of *FANCD2* cDNAs were consistently detected by sequence analysis of the regions corresponding to exon 22 (Figure 4A and B) and exons 15-17 (data not shown) due to low-level skipping of these exons, consistent with FANCD2 RNA being subject to alternative splicing. This finding was confirmed by mRNA stabilization via CHX treatments of cultured cells, which resulted in a relative increase of the alternatively spliced mRNA species (Figure 4A and C) implying instability of the alternatively spliced FANCD2 mRNAs.

Without CHX treatments, cell lines from patients 2, 8, 9, 10, 14, 15 and 20 in our cohort displayed almost equal levels of exon 22 skipping and retention (Figure 4A

and D). Patients 3, 4, 5 and 13 showed nearly complete exon 22 skipping, but we consistently observed a small amount of correctly spliced mRNA retaining exon 22 (Figure 4A and E). Genomic sequencing identified a common underlying mutation, the base substitution c.1948-16T>G in IVS21. Homozygosity for this mutation was observed in patients 3, 4, 5 and 13 with nearly complete skipping of exon 22 and also in the deceased patient 25. All of these patients were products of consanguineous matings. Patients 9 and 10 with balanced levels of exon 22 skipping and retention were compound heterozygous carriers of the mutation.

A different base substitution preceding exon 22, c.1948-6C>A, was present on one allele of the compound-heterozygous patients 2, 8, 14, 15 and 20, likewise resulting in similar levels of exon 22 skipping and retention. Both mutations, c.1948-16C>T and 1948-6C>A, are predicted to disrupt the splice acceptor recognition in intron 21 suggested by impaired scores of the 3´splice site relative to wildtype (cf. Supplementary Table S6A).

Three apparently unrelated patients (patients 6, 12 and 30) showed balanced levels of skipping and retention of exon 5 due to heterozygous insertional mutagenesis by an *Alu* element between positions c.274-57 and c.274-56 into an AT-rich target sequence in IVS4. This *Alu* was identical to the evolutionary young subfamily Yb8[40,41]. It was lacking its annotated nucleotides 1-35, had integrated in reverse orientation (with its poly-A tail towards the 5´ end of *FANCD2*) and had duplicated the 13-nt sequence c.274-69 to c.274-57 of *FANCD2* IVS4 such that this duplicated sequence flanked the *Alu* repeat on either side. Altogether the insertion was 298 bp long. Integration site, type, length and orientation of the *Alu* and the duplicated FANCD2 intron sequence were identical in all three patients.

Aberrant splicing of exons 4, 5, 10, 13, 15-17, 28 and 37 was observed also in other patients. Patients 28 and 29 showed skipping of exon 4 due to a base

substitution in the preceding canonical splice acceptor site (c.206-2A>T). Patients 26 and 27 had a base substitution in exon 5 (c.376A>G) abrogating the downstream splice donor. This change led to the inclusion of 13 bp of IVS5 into the transcript by activating a cryptic 5'-splice site in intron 5 (r.377_378ins13; also previously reported[24]). Patient 18 showed skipping of exon 10 due a base substitution in the upstream splice acceptor (c.696-2A>T). Exon 10 skipping was observed in patient 31, who had a substitution of the last minus one base of exon 10 (c.782A>T). In patient 8, we detected a splice acceptor mutation upstream of exon 13 (c.990-1G>A). This change results in the activation of a cryptic splice acceptor 8 bp downstream and exclusion of the corresponding sequence from the mature mRNA. A 2-bp deletion in exon 16 (c.1321_1322delAG) in patient 18 causes skipping of exons 15-17. In this case, aberrant splicing occurs in the same position as low-grade alternative splicing in normal controls, but at heterozygous levels. Patients 10 and 22 showed inclusion of a 27-bp sequence of intron 28 into mRNA due to a splice donor mutation (c.2715+1G>A) and the usage of a cryptic splice donor downstream. Patient 11 had a base substitution in exon 37 (c.3707G>A, previously reported[24]) that abrogates the normal splice acceptor 25 bp upstream and activates a cryptic site 19 bp downstream of the mutation, resulting in skipping of 44 bp. Interestingly, an adjacent base substitution (c.3706C>A) in patient 32 generates a new splice acceptor that is used instead of the normal one 23 bp upstream, leading to skipping of the 24 nt in between. All of these splicing aberrations were due to heterozygous mutations whereas patient 1 showed homozygous exonization of an IVS9 fragment due to a mutation in intron 9 (c.696-121C>G), which activates cryptic splice sites. Predicted scores and consequences of some of these splice mutations are computed in Supplementary Tables S6. Apart from 1321_1322delAG causing skipping of exons 15-17, all mutations affecting splicing in the patients of this study result in frameshifts

and subsequent premature termination of translation. More than half, i.e., 30/58 mutation alleles of the 29 fully informative FA-D2 patients, or 34/66 of all, were splicing mutations. Thus, the most prevalent effect of *FANCD2* mutations involves abnormal splicing patterns.

Other mutations

There were five different heterozygous nonsense mutations in nine patients from six families (c.757C>T, siblings 23 and 24; c.1092G>A, patient 7; c.2404C>T, patient 21; c.2775_2776CC>TT, siblings 14 and 15; c.3803G>A, patient 6, siblings 26 and 27; Table 1 and Supplementary Table S4). In addition, we detected five different missense mutations in eleven patients from nine families (c.692T>G, patient 19; c.904C>T, patient 7, identical to a previously reported mutation[24]; c.1367T>G, siblings 23 and 24; c.1370T>C, patient 31; c.2444G>A, siblings 16 and 17, patients 19, 21, 22 and 30). These amino acid substitutions were classified as missense mutations because of their absence from normal controls, their absence from FA-D2 patients of our cohort with other biallelic mutations and their occurrence at evolutionary conserved residues. Missense mutations were either compound heterozygous in combination with other types of *FANCD2* mutations or homozygous in consanguineous families. Three unrelated patients had small deletions (c.2660delA, patient 20; c.3453_3456delCAAA, patient 12; c.3599delT, patient 2) resulting in frameshifts. Another small deletion was in frame and affected a single codon (c.810_812delGTC, patient 9). There was only a single small frameshift duplication (c.2835dupC, patient 11). A large genomic deletion (g.22875_23333del459) spanning the entire exon 17 (similar to a mutation previously reported without defined breakpoints[24]) adjacent 71 bp of intron 16 and 256 bp of intron 17 was found in sibling pair 28 and 29. This deletion resulted in a net loss of 41

aa. A large genomic duplication in patient 33 included exons 11-14 and resulted in the insertion of 132 aa. Both gross gene rearrangements retained the reading frame. In all of our patients, nonsense mutations, deletions and insertions were exclusively affecting single alleles in combination with splice or missense mutations.

A unique case was a compound heterozygous start codon mutation (c.2T>C) in patient 32.

Figure 5 illustrates the distribution of *FANCD2* mutations that were identified in this study, including those of three FA-D2 patients previously reported[24].

Ethnic associations and shared alleles

Relatively severe birth defects and early hematological onset were observed in three patients (4, 5 and 13) homozygous for the splice mutation c.1948-16T>G with exon 22 skipping. These three patients and two other homozygotes with reverse mosaicism in the hematopoietic system (patients 3 and 25) were all from four consanguineous Turkish families. Of two FA-D2 patients compound heterozygous for this mutation, one was also of Turkish origin; the other came from eastern Czech Republic. The splice mutation c.1948-6C>A, likewise leading to exon 22 skipping, was detected in five patients (patients 2, 8, 14, 15 and 20), including two sisters (patients 14 and 15). These patients came from three families in Northern Germany and a German immigrant family in the US (patient 20). They presented with intermediate phenotypic and hematological severity. Relatively mild birth defects and a protracted hematological course into adulthood was observed in two siblings from a consanguineous Spanish family (patients 16 and 17) with the homozygous missense substitution c.2444G>A. Of four compound heterozygotes for this mutation with mild disease manifestations, one had mixed ethnicity (patient 19), one was Hispanic American (patient 21), one had Sicilian (patient 22) and another Spanish and

Portuguese ancestry (patient 30). The insertion of an *Alu*Yb8 element was found compound-heterozygous in a patient each of German (patient 6), Danish (patient 12), and Spanish/Portuguese FA-D2 (patient 30) descent. We therefore considered the latter mutation as recurrent rather than ethnically associated. All other mutations did not occur in more than two families.

On haplotype analysis, all patients homozygous for the mutation detected in the Turkish population (c.1948-16T>G; patients 3, 4, 5, 13 and 25) were homozygous for markers D3S1597, D3S1938, D3S3611 and D3S1675. The resulting haplotype was shared, in heterozygous state, with the non-consanguineous compound heterozygous Turkish patient (no. 10). The Czech patient (no. 9) with this mutation had a different haplotype. Lack of homozygotes for the intron 21 mutation prevalent in the German population (c.1948-6C>A; patients 2, 8, 14, 15 and 20) and unavailability of patients' parents precluded construction of a mutation-associated haplotype. However, all patients with this mutation had one or two identical marker(s) at least on one side of their mutated *FANCD2* gene. This finding suggests that c.1948-6C>A is an old mutation with erosion of an ancient haplotype. The consanguineous siblings (patients 16 and 17) homozygous for the mutation prevalent in Spanish or Southern European populations (c.2444G>A) were also homozygous for the set of markers used. Of their common haplotype, the microsatellite markers adjacent to *FANCD2* were shared with a Hispanic patient (no. 21), a patient with Sicilian ancestry (no. 22) and a patient of Spanish/Portuguese descent (no. 30), all compound heterozygotes for this mutation. Additional support for a conserved haplotype came from linkage disequilibrium. All of the patients homo- or heterozygous for the mutation c.2444G>A were also homo- or heterozygous for the polymorphism c.2702G>T (p.G901V). Sequence analysis of the parents indicated that both substitutions were on the same allele. A single patient (no. 19) with the

mutation c.2444G>A neither shared the haplotype nor the polymorphism c.2702G>T. Apart from c.2702G>T that was also observed without association with the mutation c.2444G>A, the only new *FANCD2* polymorphisms detected in our study were c.3978C>T and c.4478A>G in the 3'-UTR, all others have been previously reported[24]. Despite clear ethnical association of the patients with the insertion of an *Alu*Yb8 element in intron 4, it nevertheless seems unlikely that an identical event would have occurred three times independently. Two of these patients (6 and 12) shared all of the four markers studied. Patient 30 with the same mutation had retained a single identical marker adjacent to *FANCD2*. A base substitution in the *Alu* sequence, 260G>A, present in all three cases but in less than 10% of complete *Alu*Yb8 elements in the human genome (BLAT) further suggests that the *Alu* insertion goes back to a single event and is an ancient rather than a recurrent mutation.

Reverse mosaicism

Among the 28 fully informative FA-D2 patients in this study (excluding the fetal case no. 19), five (no. 3, 14, 15, 25 and 26) developed reverse mosaicism in the hematopoietic system. Mosaic patients were recognized by the facts that they had levels of both FANCD2-S and -L in protein from LCLs, comparable to normal controls (Figure 6A), that they had low chromosome breakage rates in blood and blood-derived LCLs (Supplementary Table S4) and that they had lost the typical G2 phase arrest of their lymphocytes after exposure to MMC (Figure 6B). Nonetheless, these patients had the characteristic clinical FA phenotype and their cultured fibroblasts had preserved MMC sensitivity, indicated by elevated chromosome breakage and G2 phase arrest (Figure 6B). Molecular studies confirmed these findings. Two patients with heterozygous base substitutions in the coding sequence, resulting in a nonsense (patient 14) and a splice mutation (patient 26), showed reversion to the respective

wildtype bases in primary blood cells and LCLs. The mechanism of these reversions is not clear and could involve back mutation, recombination with LOH or recombination with gene conversion. Intragenic mitotic crossover is the likely but not proven mechanism of mosaicism in the sibling of patient 14 (no. 15) who had retained her dinucleotide substitution in her peripheral blood cells. Two patients (3 and 25) with the c.1948-16T>G splice mutation had different second site compensatory mutations nearby. Clinically, 3/5 mosaic patients (3, 14 and 15) in the present cohort experienced a mild or protracted hematological course. The other 2/5 patients (25 and 26) had no apparent benefit from their mosaicism; one of them required relatively early HSCT and the other died of intracranial hemorrhage (Supplementary Table S5). The rate of 17% mosaic FA-D2 patients in our study is within the 15%[42] to 20%[43] or 25%[44] range reported for other complementation groups. With a rate comparable to *FANCA*, *FANCD2* appears to be another FA gene particularly prone to reverse mosaicism.

Residual FANCD2 protein

A surprising finding was the presence of residual FANCD2 protein in PBLs and LCLs of every FA-D2 patient tested. Detection of residual protein required overexposure of FANCD2 immunoblots (Figure 7A). Unlike standard exposure that showed no FANCD2 bands in most of the FA-D2 cell lines (cf. Figure 1), both FANCD2-S and FANCD2-L bands were detected when films were exposed overnight. As the study progressed, it became evident that the cell lines initially detected with residual protein were those with the highest levels. When we systematically re-examined all of our FA-D2 lines, all 21 LCLs available from our 29 fully informative FA-D2 patients had minute but unequivocal amounts of residual protein (Supplementary Table S4). This was also true for CD3/CD28/IL-2 stimulated PBL cultures from patient 13. Primary

fibroblasts normally have lower levels of FANCD2 relative to total protein than LCLs, and this might be the reason why detection of residual FANCD2 remained ambiguous in a prenatal case (patient 19) with only fibroblasts available. Because of lack of LCLs, two affected siblings (patients 4 and 17) could not be tested. Finally, five of our patients were mosaic leaving 8/29 patients unconfirmed for residual protein. Given the normal amounts of FANCD2 protein in the mosaic patients and the fact that the non-mosaic patients had high chromosome breakage rates and G2 phase arrest, we consider it unlikely that undetected mosaicism accounts for the presence of residual protein in the remainder of our patients. Densitometry suggested reductions of residual FANCD2 protein in the order of 1/100 to 1/1000 relative to wildtype, with the expression differing greatly amongst individual LCLs (Figure 7A). FA-D2 LCLs with the highest levels of residual FANCD2 were used to examine its characteristics on overexposed blots. The intensity of the FANCD2-L band increased as a function of the concentration of the DNA crosslinking agent (Figure 7B) and the period of treatment (not shown). This time and concentration dependency suggests genuine biochemical activity of the residual FANCD2 protein, implying that most, if not all, cases of FA-D2 result from functionally hypomorphic mutations.

Discussion

Our results suggest that FA-D2 is a more frequent FA complementation group than previously reported[2,22,23]. The relatively large proportion of Turkish FA-D2 patients in the present study–about 10% of the patients studied in Germany–appears to be due to a founder effect for the *FANCD2* mutation c.1948-16T>G among individuals of Turkish origin. This is similar to the disparity in the frequency of FA-C patients in the IFAR database compared to the European FA population. The proportion of FA-C patients in the IFAR is 15%[45], compared to only 10% in the European dataset[2]. This is due to the relatively high frequency of Ashkenazi Jewish FA patients in the IFAR with the prevalent *FANCC* mutation c.456+4A>G (formerly IVS4+4A>G)[30], comprising 7.5% of all IFAR patients and 50% of the FA-C patients therein. We calculate roughly that about 6% of FA patients belong to complementation group D2, which is supported by independent studies with a figure of 4/53 ≈ 7.5%[42], 3/73 ≈ 4.1%[46] and an estimate of 5%[45], as recently reported.

The D2 patients in our cohort displayed anomalies and malformations typical of FA such that there were no exceptional clinical features that had not previously been observed[47]. However, it is remarkable that not a single D2 patient lacked phenotypical manifestations, whereas the proportion of FA patients without anomalies and malformations is generally estimated as high as 30%[22]. Growth retardation was present in 86% of the present cohort, substantially higher than the 58%[48] and 63%[22] reported. Microcephaly was present in 89% of the FA-D2 cases; in contrast, Faivre et al.[48] found anomalies of the head in only 56%. Anomalies of skin pigmentation were present in 75% of our FA-D2 cohort compared to 71% and 64%[22,48]. 72% of our FA-D2 patients had radial ray defects in contrast to only 47%[48] or 49.1% [47] of all FA patients. 61% of the patients in the present study had microphtalmia, whereas 38% have been reported in other FA patients[22]. As with

these rather common phenotypic alterations, FA-D2 patients showed also higher rates of rare FA features such as psychomotor retardation and hyperactivity attention deficit disorder. Psychomotor retardation was present in 29% of our FA-D2 cohort versus 12% or 10% mentally retarded individuals in other studies[22,48]. As many as 31% of our FA-D2 patients had anomalies of the brain, whereas other studies report such alterations in the order of 4.5%[48], 7.7%[47] and 8%[22] of their FA patients. 17% of our D2 patients with brain anomalies had hydrocephalus, in contrast to 4.6% reported[47]. Since several labs contributed to the present study, and since all of our D2 patients came from previously unassigned FA patients, it is unlikely that our rates reflect major biases. A more severe D2 phenotype has also been observed in drosophila comparing *Fancd2* and *Fancl* knock-down[21]. Given the high frequency of phenotypic alterations, it is not surprising that in 30% of our FA-D2 patients the diagnosis of FA was made by the age of 2 y, and the median of age at diagnosis was 4.5 y which is considerably younger than in other FA patients where only in 30% the diagnosis is made before onset of hematological manifestations at the median age of 7.6 y[49]. In addition to an earlier median age of hematological onset (BMF) in our FA patients, there was a shorter median period between BMF and HSCT, earlier HSCT and a tendency towards shorter median survival than all FA in the IFAR[39]. However, due to relatively small numbers and the relative deficit of older patients in our cohort, statistical significance was not reached for all of these end points. HSCT appears to be a therapeutic option also in group FA-D2 as nine transplanted non-mosaic patients and one mosaic patient of our cohort suggest, although deficient ATM/ATR-dependent phosphorylation of FANCD2[50,51,52] could theoretically involve additional toxicity of conditioning. Again, our data suggests that FA-D2 patients represent a group with frequent but typical congenital anomalies and malformations, and with

relatively early hematological manifestations, compared to most other FA complementation groups.

Among the FA proteins, FANCD2 is unique since the presence of residual protein and the demonstration of its activation can be accomplished in a single assay. In our cohort, LCLs and PBLs from 21 fully informative, non-mosaic FA-D2 patients studied showed traces of residual FANCD2 protein. Importantly, the residual protein always consisted of both FANCD2 isoforms, and the typical time- and dose-dependent induction of FANCD2-L was maintained, suggesting a preserved function. Differences in expression levels of residual FANCD2 between individual LCLs might result from variations of conserved splice site recognition, in mRNA and protein stability, and, very clearly, from differences in cell growth. FANCD2 is highly expressed and monoubiquitinated in the S-phase of the cell cycle [8,53]. The proportion of S-phase cells is a function of cell growth such that differences in cell proliferation between individual cell lines account for the wide variation of FANCD2 protein levels and render any quantitative mutation-specific comparisons of residual FANCD2 protein levels close to impossible. The existence of residual protein has previously been described for other FA-D2 patients[7,24,42] but our study confirms residual protein as a consistent and in all likelihood essential feature of FA-D2 patient cells. Somatic reversion as a cause of residual protein levels could be excluded because the diagnosis of FA in these of our D2 patients was based on hypersensitivity towards crosslinking agents.

FANCD2 is targeted to chromatin following DNA damage-dependent monoubiquitination where it interacts with the highly conserved C-terminal region of BRCA2[54]. FANCD2-L promotes BRCA2 loading onto a chromatin complex that is required for effective, but RAD51-independent DNA repair[13,55]. The examination of DT40 cell lines revealed that components of the FA core complex have additional

functions in DNA repair pathways which seem to be independent of monoubiquitination and chromatin targeting of Fancd2[56]. However, the common pathway of FANCD2 and FANCD1/BRCA2 appears to be crucial for functional resolution of ICL-induced stalled replication forks and, in order for humans to be viable, may require residual protein activity.

Despite a rather severe phenotype in most of the FA-D2 patients, the vast majority of our FA-D2 patients were found to carry leaky mutations, merely affecting splicing, and displayed residual FANCD2 protein of both isotypes in their cell lines. Splicing mutations have become an increasingly successful target for experimental therapeutic approaches. Modified and antisense oligonucleotides have been used to inhibit cryptic exons or to activate regular exons weakened by mutations via targeting of the oligonucleotides to the desired transcript. This approach could eventually lead to effective therapies for the correction of erroneous splicing (reviewed in[56,57]). The tight regulation of FANCD2 expression and activation, and the presence of low-abundant wildtype gene products associated with *FANCD2* mutations should render *FANCD2* an ideal candidate for RNA-reprogramming strategies such as spliceosome-mediated RNA trans-splicing (SMaRT; reviewed in[57,58]).

Acknowledgments

We thank Richard Friedl, Wurzburg, for expert flow cytometry; Birgit Gottwald, Wurzburg, for dedicated cell culture work; Daniela Endt, Wurzburg, for sequencing; Dr. Sabine Herterich, Wurzburg, for microsatellite analyses; and Kerstin Goettsche and Silke Furlan, Dusseldorf, for assistance with retroviral vectors. We are grateful to Dr. Heidemarie Neitzel, Berlin, for providing patient DNAs and Ralf Dietrich, Unna, for facilitating personal contacts with FA families and arranging for insight into their

medical histories. The GR plasmid for construction of diagnostic retroviral vectors was kindly provided by Dr. Christopher Baum, Hamburg. We thank Dr. Birgit Pils, Oxford, UK, for help with database searches, and Drs. Adrian Krainer; Cold Spring Harbor, NY, and Chris Smith, Cambridge, UK, for advice with the characterization of some of the splice site mutations. Dr. Markus Schmugge, Zurich, Switzerland, and Dr. Eva Seemanova, Prague, Czech Republic, are gratefully acknowledged for providing clinical information, as are Drs. John Wagner, Minneapolis, David Williams, Cincinnati, Farid Boulad, New York, and many other physicians who provided clinical data for the IFAR. We are deeply obliged to all of the participating patients and families, and to the many clinicians who contributed to the present work through their patient care.

Web resources

Genomic *FANCD2* sequences were compared by BLAT homology searches (http://genome.ucsc.edu/cgi-bin/hgBlat) and the Ensembl genome browser (http://www.ensembl.org/). Polypeptide sequences were compared using the Windows interface Clustal X (ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX), version 1.81, for the Clustal W multiple sequence alignment program. Promoter analyses were done using the CpG island explorer[59] (http://bioinfo.hku.hk/cpgieintro.html), version 1.9, at the settings GC 60%, CpG O/E ratio 0.7 and minimum length 500 nt. Analysis of repetitive elements was done using the Repeatmasker software (http://www.repeatmasker.org). Predicted splice donor performance was calculated using the Splicefinder algorithm (http://www.splicefinder.org). Deduced splice acceptor function was estimated using a maximum entropy model (http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq_acc.html). Regulatory

splice sequences were analyzed using the ESE finder (http://rulai.cshl.edu/tools/ESE) and the Rescue-ESE (http://genes.mit.edu/burgelab/rescue-ese).

The NCBI (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi) nucleotide sequences NM 033084 (43 exons) and AF 340183 (44 exons) were used as the human *FANCD2* cDNA reference. The genomic reference sequence was ENSG00000144554. *Fancd2* sequence information of other species is available at the same website. Fancd2 protein sequences of different species including Homo sapiens were from the Swiss-Prot database (http://www.expasy.org/sprot/).

References

1. Joenje H, Patel KJ. The emerging genetic and molecular basis of Fanconi anaemia. Nat Rev Genet. 2001;2:446-457.
2. Levitus M, Rooimans MA, Steltenpool J, et al. Heterogeneity in Fanconi anemia: evidence for 2 new genetic subtypes. Blood. 2004;103:2498-2503.
3. Levran O, Attwooll C, Henry RT, et al. The BRCA1-interacting helicase BRIP1 is deficient in Fanconi anemia. Nat Genet. 2005;37:931-933.
4. Meetei AR, Medhurst AL, Ling C, et al. A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. Nat Genet. 2005;37:958-963.
5. Levitus M, Waisfisz Q, Godthelp BC, et al. The DNA helicase BRIP1 is defective in Fanconi anemia complementation group J. Nat Genet. 2005;37:934-935.
6. Meetei AR, Sechi S, Wallisch M, et al. A multiprotein nuclear complex connects Fanconi anemia and Bloom syndrome. Mol Cell Biol. 2003;23:3417-3426.
7. Garcia-Higuera I, Taniguchi T, Ganesan S, et al. Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. Mol Cell. 2001;7:249-262.
8. Taniguchi T, Garcia-Higuera I, Andreassen PR, Gregory RC, Grompe M, D'Andrea AD. S-phase-specific interaction of the Fanconi anemia protein, FANCD2, with BRCA1 and RAD51. Blood. 2002;100:2414-2420.
9. Thompson LH, Hinz JM, Yamada NA, Jones NJ. How Fanconi anemia proteins promote the four Rs: Replication, recombination, repair, and recovery. Environ Mol Mutagen. 2005;45:128-142.
10. Nakanishi K, Yang YG, Pierce AJ, et al. Human Fanconi anemia monoubiquitination pathway promotes homologous DNA repair. Proc Natl Acad Sci U S A. 2005;102:1110-1115.

11. Niedzwiedz W, Mosedale G, Johnson M, Ong CY, Pace P, Patel KJ. The Fanconi anaemia gene FANCC promotes homologous recombination and error-prone DNA repair. Mol Cell. 2004;15:607-620.
12. Digweed M, Rothe S, Demuth I, et al. Attenuation of the formation of DNA-repair foci containing RAD51 in Fanconi anaemia. Carcinogenesis. 2002;23:1121-1126.
13. Yamamoto K, Hirano S, Ishiai M, et al. Fanconi anemia protein FANCD2 promotes immunoglobulin gene conversion and DNA repair through a mechanism related to homologous recombination. Mol Cell Biol. 2005;25:34-43.
14. Mirchandani KD, D'Andrea AD. The Fanconi anemia/BRCA pathway: A coordinator of cross-link repair. Exp Cell Res. 2006.
15. Levitus M, Joenje H, de Winter JP. The Fanconi anemia pathway of genomic maintenance. Cell Oncol. 2006;28:3-29.
16. Blom E, van de Vrugt HJ, de Winter JP, Arwert F, Joenje H. Evolutionary clues to the molecular function of fanconi anemia genes. Acta Haematol. 2002;108:231-236.
17. Houghtaling S, Timmers C, Noll M, et al. Epithelial cancer in Fanconi anemia complementation group D2 (Fancd2) knockout mice. Genes Dev. 2003;17:2021-2035.
18. Carreau M. Not-so-novel phenotypes in the Fanconi anemia group D2 mouse model. Blood. 2004;103:2430.
19. Dequen F, St-Laurent JF, Gagnon SN, Carreau M, Desnoyers S. The Caenorhabditis elegans FancD2 ortholog is required for survival following DNA damage. Comp Biochem Physiol B Biochem Mol Biol. 2005;141:453-460.
20. Liu TX, Howlett NG, Deng M, et al. Knockdown of zebrafish Fancd2 causes developmental abnormalities via p53-dependent apoptosis. Dev Cell. 2003;5:903-914.
21. Marek LR, Bale AE. Drosophila homologs of FANCD2 and FANCL function in DNA repair. DNA Repair (Amst). 2006.

22. Tischkowitz M, Dokal I. Fanconi anaemia and leukaemia - clinical and molecular aspects. Br J Haematol. 2004;126:176-191.

23. Taniguchi T, D'Andrea AD. The molecular pathogenesis of fanconi anemia: recent progress. Blood. 2006.

24. Timmers C, Taniguchi T, Hejna J, et al. Positional cloning of a novel Fanconi anemia gene, FANCD2. Mol Cell. 2001;7:241-248.

25. Joenje H. Fanconi anemia: cytogenetic diagnosis. Protocol Free University of Amsterdam. 1997.

26. Auerbach AD. Diagnosis of Fanconi anemia by diepoxybutane analysis.: John Wiley & Sons; 2003.

27. Berger R, Le Coniat M, Gendron MC. Fanconi anemia. Chromosome breakage and cell cycle studies. Cancer Genet Cytogenet. 1993;69:13-16.

28. Seyschab H, Friedl R, Sun Y, et al. Comparative evaluation of diepoxybutane sensitivity and cell cycle blockage in the diagnosis of Fanconi anemia. Blood. 1995;85:2233-2237.

29. Heinrich MC, Hoatlin ME, Zigler AJ, et al. DNA cross-linker-induced G2/M arrest in group C Fanconi anemia lymphoblasts reflects normal checkpoint function. Blood. 1998;91:275-287.

30. Kutler DI, Auerbach AD. Fanconi anemia in Ashkenazi Jews. Fam Cancer. 2004;3:241-248.

31. Neitzel H. A routine method for the establishment of permanent growing lymphoblastoid cell lines. Hum Genet. 1986;73:320-326.

32. Schindler D, Hoehn H. Fanconi anemia mutation causes cellular susceptibility to ambient oxygen. Am J Hum Genet. 1988;43:429-435.

33. Hildinger M, Abel KL, Ostertag W, Baum C. Design of 5' untranslated sequences in retroviral vectors developed for medical use. J Virol. 1999;73:4083-4089.

34. Hanenberg H, Xiao XL, Dilloo D, Hashino K, Kato I, Williams DA. Colocalization of retrovirus and target cells on specific fibronectin fragments increases genetic transduction of mammalian cells. Nat Med. 1996;2:876-882.

35. Hanenberg H, Hashino K, Konishi H, Hock RA, Kato I, Williams DA. Optimization of fibronectin-assisted retroviral gene transfer into human CD34+ hematopoietic cells. Hum Gene Ther. 1997;8:2193-2206.

36. Hanenberg H, Batish SD, Pollok KE, et al. Phenotypic correction of primary Fanconi anemia T cells with retroviral vectors as a diagnostic tool. Exp Hematol. 2002;30:410-420.

37. Chandra S, Levran O, Jurickova I, et al. A rapid method for retrovirus-mediated identification of complementation groups in Fanconi anemia patients. Mol Ther. 2005;12:976-984.

38. Shimamura A, de Oca RM, Svenson JL, et al. A novel diagnostic screen for defects in the Fanconi anemia pathway. Blood. 2002;100:4649-4654.

39. Kutler DI, Singh B, Satagopan J, et al. A 20-year perspective on the International Fanconi Anemia Registry (IFAR). Blood. 2003;101:1249-1256.

40. Carter AB, Salem AH, Hedges DJ, et al. Genome-wide analysis of the human Alu Yb-lineage. Hum Genomics. 2004;1:167-178.

41. Roy-Engel AM, Carroll ML, Vogel E, et al. Alu insertion polymorphisms for the study of human genomic diversity. Genetics. 2001;159:279-290.

42. Soulier J, Leblanc T, Larghero J, et al. Detection of somatic mosaicism and classification of Fanconi anemia patients by analysis of the FA/BRCA pathway. Blood. 2005;105:1329-1336.

43. Callen E, Casado JA, Tischkowitz MD, et al. A common founder mutation in FANCA underlies the world's highest prevalence of Fanconi anemia in Gypsy families from Spain. Blood. 2005;105:1946-1949.

44. Lo Ten Foe JR, Kwee ML, Rooimans MA, et al. Somatic mosaicism in Fanconi anemia: molecular basis and clinical significance. Eur J Hum Genet. 1997;5:137-148.

45. Kennedy RD, D'Andrea AD. The Fanconi Anemia/BRCA pathway: new faces in the crowd. Genes Dev. 2005;19:2925-2940.

46. Casado JA, Callen E, Jacome A, et al. A comprehensive strategy for the subtyping of Fanconi Anemia patients: conclusions from the Spanish Fanconi Anemia research network. J Med Genet. 2006.

47. Auerbach AD, Buchwald M, Joenje H. Fanconi anemia. Vol. 1 (ed 8.). New York: McGraw-Hill; 2001.

48. Faivre L, Guardiola P, Lewis C, et al. Association of complementation group and mutation type with clinical outcome in fanconi anemia. European Fanconi Anemia Research Group. Blood. 2000;96:4064-4070.

49. Huret JL. Fanconi anaemia. Atlas Genet Cytogentic Oncol Haematol. 2002.

50. Andreassen PR, D'Andrea AD, Taniguchi T. ATR couples FANCD2 monoubiquitination to the DNA-damage response. Genes Dev. 2004;18:1958-1963.

51. Pichierri P, Rosselli F. The DNA crosslink-induced S-phase checkpoint depends on ATR-CHK1 and ATR-NBS1-FANCD2 pathways. Embo J. 2004;23:1178-1187.

52. Ho GP, Margossian S, Taniguchi T, D'Andrea AD. Phosphorylation of FANCD2 on two novel sites is required for mitomycin C resistance. Mol Cell Biol. 2006;26:7005-7015.

53. Holzel M, van Diest PJ, Bier P, et al. FANCD2 protein is expressed in proliferating cells of human tissues that are cancer-prone in Fanconi anaemia. J Pathol. 2003;201:198-203.

54. Hussain S, Wilson JB, Medhurst AL, et al. Direct interaction of FANCD2 with BRCA2 in DNA damage response pathways. Hum Mol Genet. 2004;13:1241-1248.

55. Ohashi A, Zdzienicka MZ, Chen J, Couch FJ. Fanconi Anemia Complementation Group D2 (FANCD2) Functions Independently of BRCA2- and RAD51-associated Homologous Recombination in Response to DNA Damage. J Biol Chem. 2005;280:14877-14883.

56. Matsushita N, Kitao H, Ishiai M, et al. A FancD2-Monoubiquitin Fusion Reveals Hidden Functions of Fanconi Anemia Core Complex in DNA Repair. Mol Cell. 2005;19:841-847.

57. Garcia-Blanco MA, Baraniak AP, Lasda EL. Alternative splicing in disease and therapy. Nat Biotechnol. 2004;22:535-546.

58. Mansfield SG, Chao H, Walsh CE. RNA repair using spliceosome-mediated RNA trans-splicing. Trends Mol Med. 2004;10:263-268.

59. Wang Y, Leung FC. An evaluation of new criteria for CpG islands in the human genome as gene markers. Bioinformatics. 2004;20:1170-1177.

Table 1. Identified *FANCD2* mutations and their effects

Location	Mutation*		Consequence*	Patient
Exon/Intron	gDNA	RNA	Protein	No.
Exon 2	c.2T>C	r.2T>C	*Failure of normal translation initiation*	32
Intron 3	c.206–2A>T *(IVS3–2A>T)*	r.206_273del68 *(exon 4 skipping)*	p.A69DfsX7	28, 29
Intron 4	c.274–57_–56insinv*AluYb8* nt36_319 +dup c.274–69_–57	r.274_377del104 *(exon 5 skipping)*	p.I92YfsX7	6, 12, 30
Exon 5	c.376A>G	r.376A>G+r.377_378ins13 *(aberrant splicing)*	p.S126RfsX12	26, 27
Exon 9	c.692T>G	r.692T>G	p.L231R	19
Intron 9	c.696–121C>G *(IVS9–121C>G)*	r.695+1619_696-126ins34 *(exonization)*	p.S232insQNNFX	1
	c.696–2A>T *(IVS9–2A>T)*	r.696_783del88 *(exon 10 skipping)*	p.S232RfsX6	18
Exon 10	c.757C>T	r.757C>T	p.R253X	23, 24
	c.782A>T	r.696_783del88 *(exon 10 skipping)*	p.S232RfsX6	31, 33
Exon 11	c.810_812delGTC	r.810_812delGTC	p.S271del	9
Exon 12	c.904C>T	r.904C>T	p.R302W	7
Intron 12	c.990–1G>A *(IVS12–1G>A)*	r.990del8 *(aberrant splicing)*	p.S330RfsX16	8
Exon 13	c.1092G>A	r.1092G>A	p.W364X	7
Intron 14	g.13377_17458dup4082 *(Duplication including exons 11_14)*	r.784_1134dup *(duplication of 351 nt in frame)*	p.262_378dup *(duplication of 117 aa)*	33
Exon 16	c.1321_1322delAG	r.1135_1545del411 *(exon 15-17 skipping)*	p.V379_K515del	18
	c.1367T>G	r.1367T>G	p.L456R	23, 24
	c.1370T>C	r.1370T>C	p.L457P	31
Exon 17	g.22875_23333del459 (c.1414–71_c.1545+256del459)	r.1414_1545del132	p.E472_K515del	28, 29
Intron 21	c.1948–16T>G *(IVS21–16T>G)*	r.1948_2021del74 *(exon 22 skipping)*	p.E650X	3, 4, 5, 9, 10, 13, 25
	c.1948–6C>A *(IVS21–6C>A)*	r.1948_2021del74 *(exon 22 skipping)*	p.E650X	2, 8, 14, 15, 20
Exon 26	c.2404C>T	r.2404C>T	p.Q802X	21
	c.2444G>A	r.2444G>A	p.R815Q	16, 17, 19, 21, 22, 30
Exon 28	c.2660delA	r.2660delA	p.E888RfsX16	20
Intron 28	c.2715+1G>A *(IVS28+1G>A)*	r.2715_2716ins27 *(aberrant splicing)*	p.E906LfsX4	10, 22
Exon 29	c. 2775_2776CC>TT	r. 2775_2776CC>TT	p.R926X	14, 15
	c.2835dupC	r.2835dupC	p.D947RfsX3	11
Exon 34	c.3453_3456delCAAA	r.3453_3456delCAAA	p.N1151KfsX46	12
Exon 36	c.3599delT	r.3599delT	p.I1200KfsX12	2
Exon 37	c.3706C>A	r.3684_3707del24 *(aberrant splicing)*	p.R1228S_F1235del	32
	c.3707G>A	r.3684_3727del44 *(aberrant splicing)*	p.H1229EfsX7	11
Exon 38	c.3803G>A	r.3803G>A	p.W1268X	6, 26, 27

* Nomenclature according to the Human Genome Variation Society (http://hgvs.org/mutnomen/recs)

Figure Legends

Figure 1. Delineation of FA-D2. (A) Schematic representation of the retroviral vector construct S11FD2IN expressing *FANCD2* cDNA. Used for cloning were the 5' EcoR I and the 3' Sal I (insert) and BamH I (vector) sites; the two latter were destroyed by blunting. The target vector S11IN without FANCD2 is shown underneath. Abbreviations: L, long terminal repeat; I, internal ribosomal entry site; N, neomycin resistance gene. (B) Assignment to group FA-D2 based on the absence of either FANCD2 band on immunoblots after exposure of the patients' cells to MMC, here shown for a LCL from patient 6 (lane 2). Transduction with *FANCD2* cDNA using S11FD2IN restores both isoforms of FANCD2, S and L (lane 3), similar to a non-transduced normal control (lane 1). Transduction with *FANCA* cDNA in the same vector fails to show such restoration (lane 4). (C) Assignment to group FA-D2 based on cell cycle analysis: After exposure to MMC, the LCL of the same patient shows pronounced G2 phase arrest (56.6%, lane 2, Hoechst 33342 staining). Transduction with *FANCD2* cDNA using S11FD2IN reduces the G2 phase to normal (14.9%, lane 3, arrow), similar to the non-transduced normal control (16.6%, lane 1). Transduction with *FANCA* cDNA in the same vector fails to reverse the G2 phase arrest (53.1%, lane 4). (D) and (E) are analogous to (B) and (C) and show complemention with cultured fibroblasts from patient 10; staining in (E) was with DAPI. G2 phase proportions in (E) are 20.3% (lane 1, control), 61.3% (lane 2, non-transduced FA), 19.9% (lane 3, *FANCD2*-transduced FA) and 58.5% (lane 4, *FANCA*-transduced FA).

Figure 2. Clinical course of 23 fully informative, non-mosaic FA-D2 patients in this study. (A) The cumulative incidence of bone marrow failure (BMF) of the FA-D2 patients in the present study (FA-D2) precedes that of all FA patients in the IFAR (IFAR)[39] (p=0.001). (B) The period from BMF to hematological stem cell

transplantation (HSCT) was shorter in the patients of the present study than in those of the IFAR[39] (trending, $p<0.08$. (C) Cumulative incidence of HSCT of the FA-D2 patients in our study likewise antedates that of all FA patients in the IFAR[39] ($p<0.01$). (D) Kaplan-Meier curves of survival suggest higher death rates of the FA-D2 patients than of all FA patients in the IFAR after 10 years of age

Figure 3. Topography of *FANCD2*, its pseudogenes and the superamplicons. (A) The two pseudogenes, *FANCD2-P1* and *FANCD2-P2,* are located upstream and downstream of the functional *FANCD2* gene, respectively. All three have the same orientation. The scale denotes Mbp on chromosome 11. (B) *FANCD2* exons and their pseudogene equivalents and are connected by dashed lines containing percentages of nucleotide identity. Homology also extends into many introns nearby as indicated by the boxes beyond and below the active gene. (C) Graphical presentation of the positions and sizes of 15 superamplicons relative to the active gene in B. These amplicons represent *FANCD2* exon-exon or exon-intron regions. Unique primer binding sites ensure specific amplification.

Figure 4. Exon 22 splicing. (A) Schematical depiction of the splicing patterns resulting from exon 22 retention or skipping. (B) cDNA sequencing in normal controls shows abundance of exon 22 sequence following that of exon 21 but also low level underlying sequence readable as exon 23. (C) Treatments of normal control cells with CHX for 4 h prior to cDNA synthesis increase the relative level of sequence with exon 22 skipping. (D) Heterozygotes for splice acceptor mutations in intron 21 show comparable levels of inclusion and exclusion of exon 22 sequence following that of exon 21. (E) Homozygotes for splice acceptor mutations in intron 21 reveal

abundance of exon 23 sequence following that of exon 21 but also low level underlying sequence readable as exon 22.

Figure 5. Positions and identity of mutations detected in *FANCD2*. Mutations identified in the present study are shown above, previously reported mutations[24] underneath the schematical display of *FANCD2* cDNA. Solid squares (■) represent mutations resulting in aberrant splicing patterns, solid circles (●) nonsense mutations, open circles (○) missense mutations, solid triangles (▲) frameshift deletions or duplications and open triangles (Δ) *in frame* deletions or duplications. [1] denotes homozygous occurrence (2 alleles), [2] affected sibling (relationship bias).

Figure 6. Reverse mosaicism. (A) Blood-derived cells from FA-D2 patients with reverse mosaicism of the hematopoietic system (patients 3 and 26, LCLs; patient 14, stimulated PBL; lanes 2, 3 and 4) reveal both FANCD2 bands at levels similar to a random normal control (lane 1) after exposure to MMC. RAD50 was used as loading control. (B) In addition, these LCLs and PBL fail to show G2 phase arrest on flow cytometric cell cycle distributions in response to MMC (black histograms, DAPI stain; CON, 8.0% G2; patient 3, 8.8% G2; patient 14, 8.8% G2; patient 26, 10.4% G2), whereas corresponding cultured FA-D2 fibroblasts retain high G2 phase accumulations, which is in contrast to the non-FA control (superimposed grey histograms; CON, 22.6% G2; patient 3, 53.2% G2; patient 14, 56.0% G2; patient 26, 54.8% G2).

Figure 7. Residual FANCD2 protein. (A) Blood-derived cells from non-mosaic FA-D2 patients (examplified 13, 5, 1, 21, 2, 6, 11 and 28) show faint, but conspicuous FANCD2 bands of both species in response to MMC exclusively on overexposed

immunoblots as indicated by the very intense FANCD2 signals of the normal controls (patient 13, stimulated PBL; patients 5, 1, 21, 2, 6, 11 and 28, LCLs; loading control RAD50). The individual abundance of residual protein varies considerably at low levels. (B) LCLs were subjected to the indicated concentrations of hydroxyurea (HU) for 16 h. On an overexposed blot, the FANCD2-L band of the residual protein in the LCL from patient 21 increases with the HU concentration in a dose-dependent response. This reaction is similar to that of a normal control LCL distinctive by its prominent FANCD2 signals.

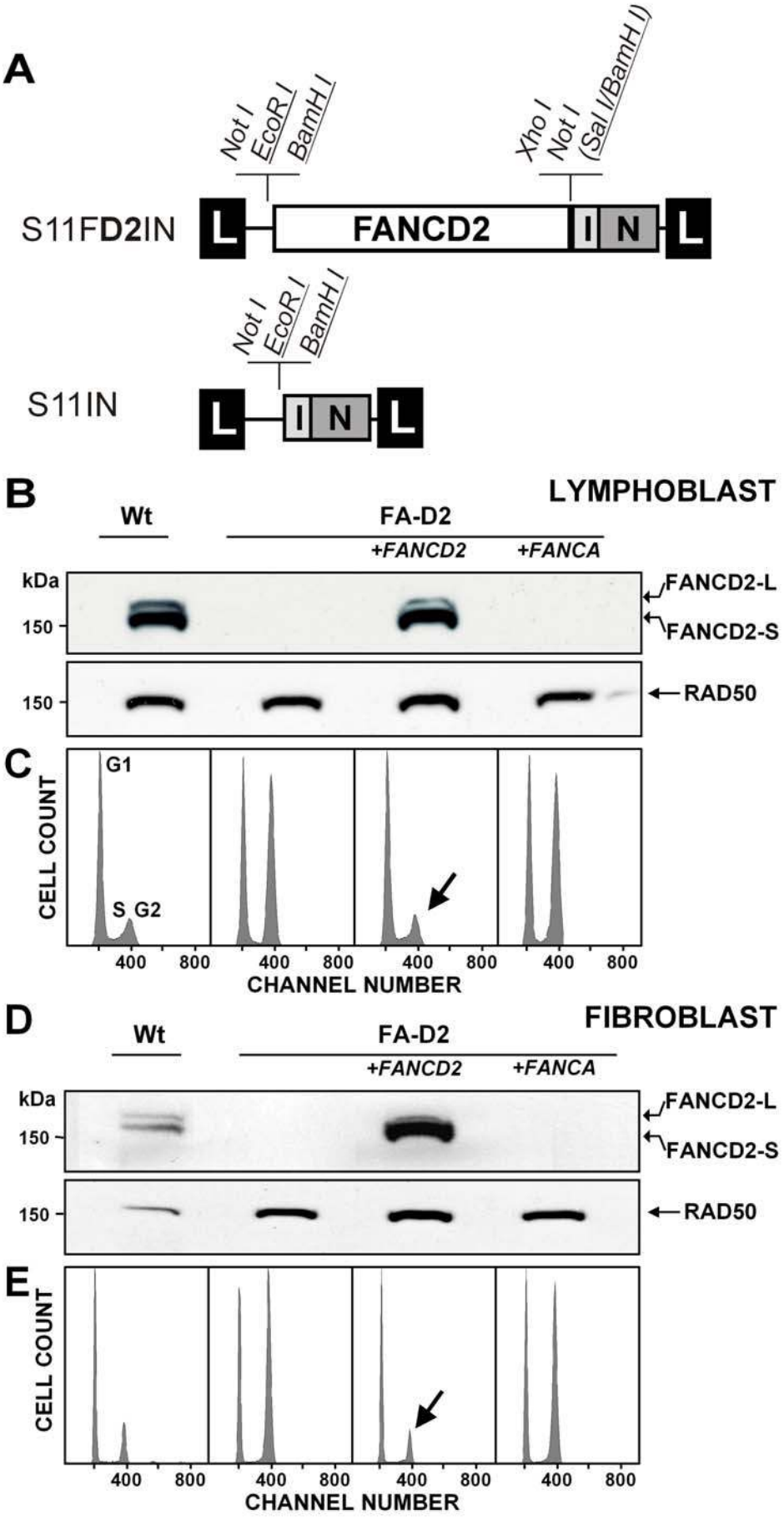
A
Not I
EcoR I
BamH I
Xho I
Not I
(Sal I/BamH I)
S11FD2IN
L
FANCD2
I
N
L
Not I
EcoR I
BamH I
S11IN
L
I
N
L
B
LYMPHOBLAST
Wt
FA-D2
+FANCD2
+FANCA
kDa
150
FANCD2-L
FANCD2-S
150
RAD50
C
CELL COUNT
G1
S G2
400
800
CHANNEL NUMBER
D
FIBROBLAST
Wt
FA-D2
+FANCD2
+FANCA
kDa
150
FANCD2-L
FANCD2-S
150
RAD50
E
CELL COUNT
400
800
CHANNEL NUMBER

Figure 1

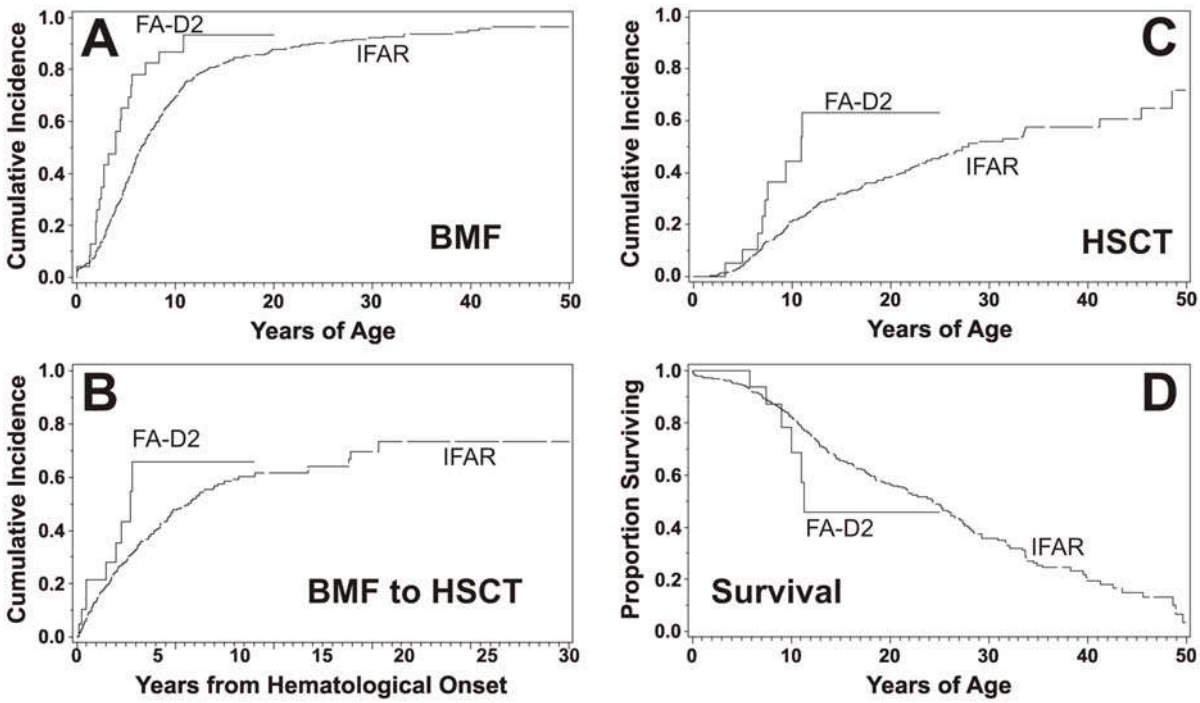

Figure 2

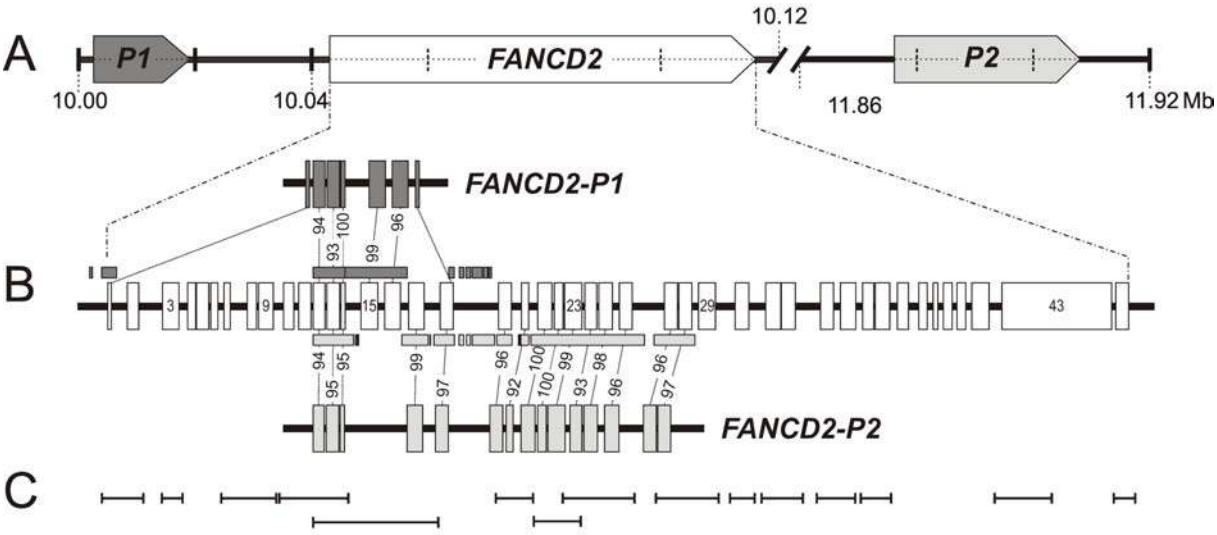

Figure 3

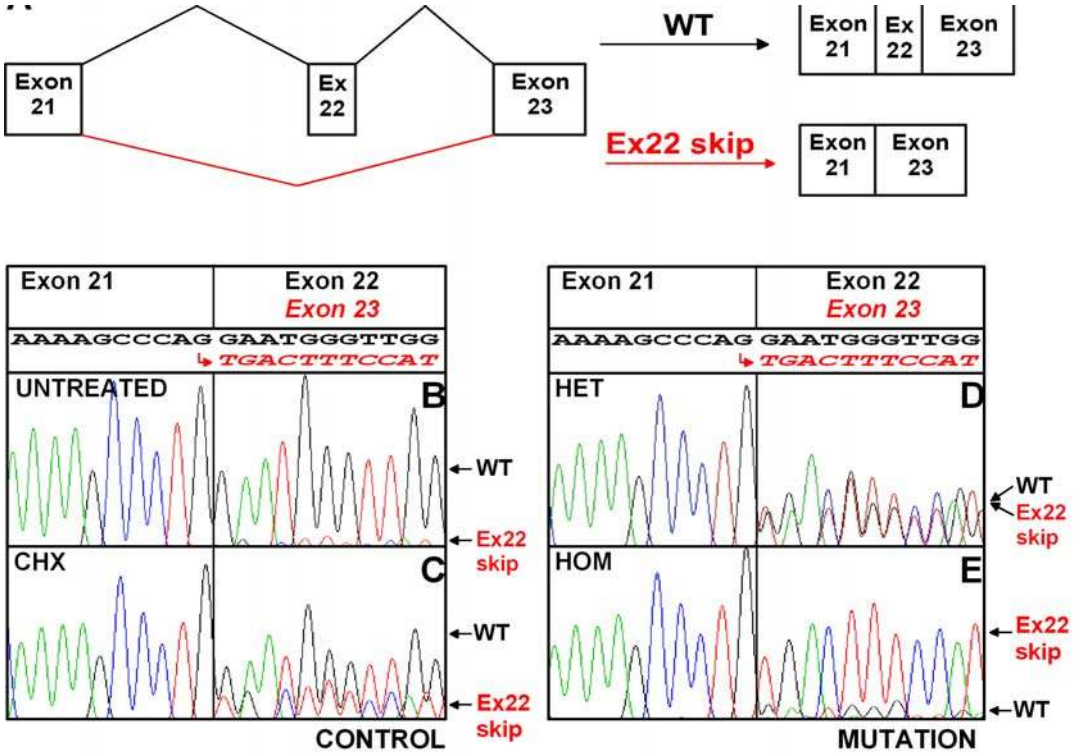

Figure 4

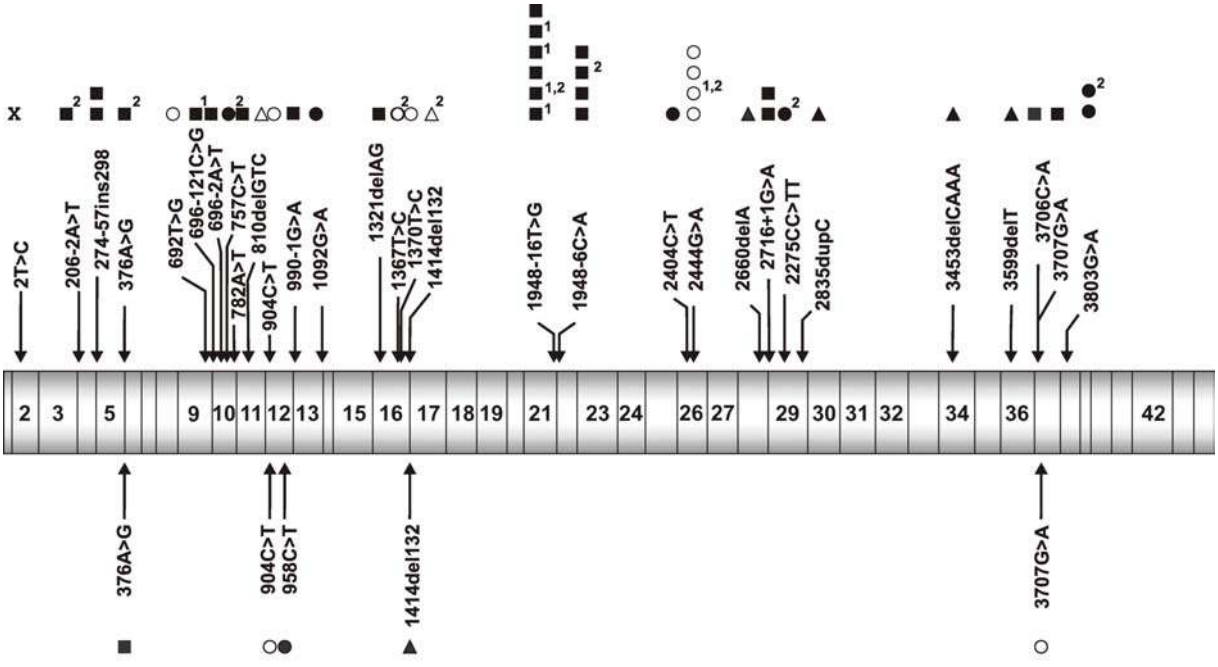

Figure 5

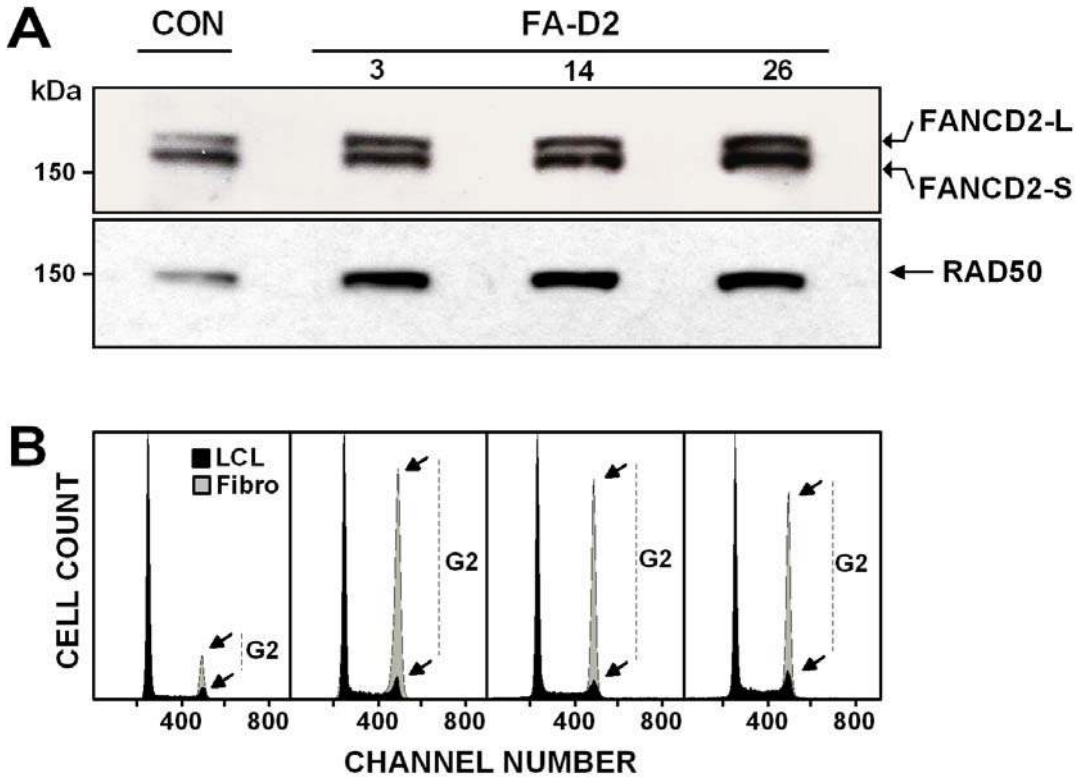

Figure 6

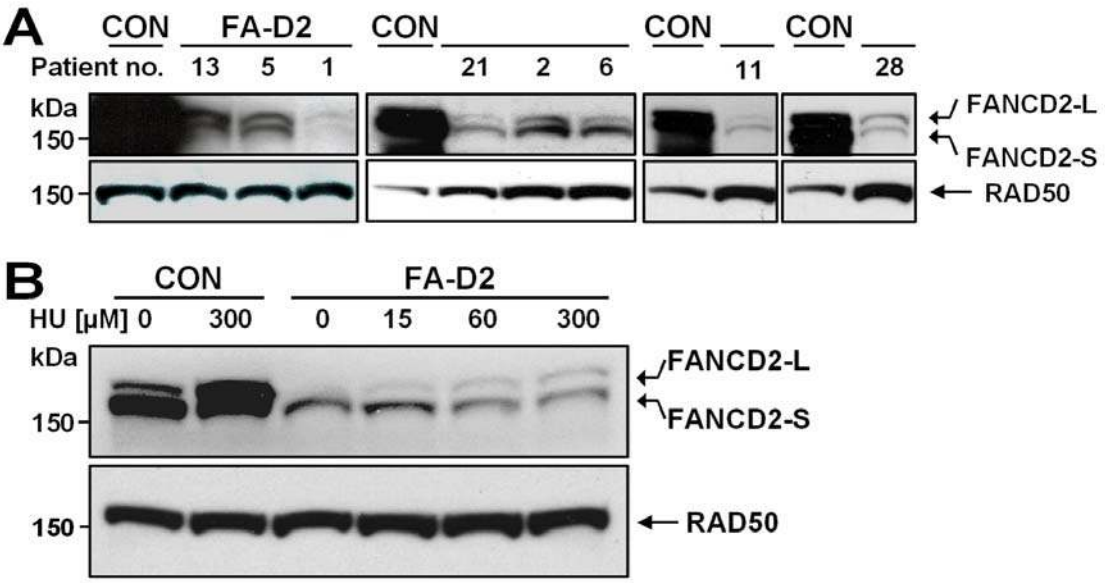

Figure 7

Supplementary Table S1A. *FANCD2* cDNA amplification primers

PCR Fragment	Designation	Binding Position	Sequence (5´→3´)	Designation	Binding position	Sequence (5´→3´)	PCR Product Size (bp)
1	FA-D2, Fr.1 F	-47 to -27	GCGACGGCTTCTCGGAAGTAA	FA-D2, Fr.1 R	998 to 976	CTGTAACCGTGATGGCAAAACAC	998
2	FA-D2, Fr.2 F	763 to 787	GACCCAAACTTCCTATTGAAGGTTC	FA-D2, Fr.2 R	1996 to 1975	CTACGAAGGCATCCTGGAAATC	1234
3	FA-D2, Fr.3 F	1757 to 1777	CGGCAGACAGAAGTGAATCAC	FA-D2, Fr.3 R	2979 to 2958	GTTCTTGAGAAAGGGGACTCTC	1223
4	FA-D2, Fr.4 F	2804 to 2829	TTCTACATTGTGGACTTGTGACGAAG	FA-D2, Fr.4 R	3942 to 3922	GTCTAGGAGCGGCATACATTG	1139
5	FA-D2, Fr.5(L) F	3761 to 3781	CAGCAGACTCGCAGCAGATTC	FA-D2, Fr.5(L) R	4700 to 4679	GACTCTGTGCTTTGGCTTTCAC	940

Supplementary Table S1B. *FANCD2* cDNA sequencing primers

Designation	Binding Position	Sequence (5´→3´)	Designation	Binding position	Sequence (5´→3´)
sFA-D2, 244 F	244 to 263	ACCCTGAGGAGACACCCTTC	sFA-D2, 367 R	367 to 347	CATCCTGCAGACGCTCACAAG
sFA-D2, 545 F	545 to 566	GGCTTGACAGAGTTGTGGATGG	sFA-D2, 621 R	621 to 600	CAGGTTCTCTGGAGCAATACTG
sFA-D2, 1011 F	1011 to 1033	CAGCGGTCAGAGCTGTATTATTC	sFA-D2, 951 R	951 to 929	CTGTAACCGTGATGGCAAAACAC
sFA-D2, 1308 F	1308 to 1327	GTCGCTGGCTCAGAGTTTGC	sFA-D2, 1158 R	1183 to 1158	TCTGAGTATTGGTGCTATAGATGATG
sFA-D2, 1574 F	1574 to 1596	CCCCTCAGCAAATACGAAAACTC	sFA-D2, 1414 R	1414 to 1396	CCTGCTGGCAGTACGTGTC
sFA-D2, 2142 F	2142 to 2162	GGTGACCTCACAGGAATCAGG	sFA-D2, 1704 R	1704 to 1684	GAATACGGTGCTAGAGAGCTG
sFA-D2, 2381 F	2381 to 2404	GAGAGATTGTAAATGCCTTCTGCC	sFA-D2, 2253 R	2253 to 2232	CTCCTCCAAGTTTCCGTTATGC
sFA-D2, 2679 F	2679 to 2699	TGACCCTACGCCATCTCATAG	sFA-D2, 2526 R	2526 to 2505	GTTTCCAAGAGGAGGGACATAG
sFA-D2, 3268 F	3268 to 3288	GCCCTCCATGTCCTTAGTAGC	sFA-D2, 3346 R	3346 to 3328	GGACGCTCTGGCTGAGTAG
sFA-D2, 3573 F	3573 to 3594	GCACACAGAGAGCATTCTGAAG	sFA-D2, 3674 R	3674 to 3653	GTAGGGAATGTGGAGGAAGATG
sFA-D2, 4049 F	4049 to 4069	ACACGAGACTCACCCAACATG	sFA-D2, 4159 R	4159 to 4139	CCAGCCAGAAAGCCTCTCTAC
sFA-D2, 4303 F	4303 to 4323	GAGTCTGGCACTGATGGTTGC	sFA-D2, 4409 R	4409 to 4387	GGGAATGGAAATGGGCATAGAAG

Supplementary Table S2A. *FANCD2* superamplicon primers

Super-amplicon	Containing Exons	Designation	Sequence (5´→3´)	Designation	Sequence (5´→3´)	PCR Product Size (bp)
I	1, 2	hFANCD2_exon1_F	TATGCCCGGCTAGCACAGAA	hFANCD2_super_1_2_R	GGCCCACAGTTTCCGTTTCT	4346
II	3	hFANCD2_super_3_3_F	GTGTCACGTGTCTGTAATCTC	hFANCD2_super_3_3_R	CTGGGACTACAGACACGTTTT	2323
III	7,8,9	hFANCD2_super_7_14_F	TGGGTTTGGTAGGGTAATGTC	hFANCD2_exon9_R	TACTCATGAAGGGGGGTATCA	4595
IV	10,11,12,13,14	hFANCD2_exon10_F	GCCCAGCTCTGTTCAAACCA	hFANCD2_super_7_14_R	TTAAGACCCAGCGAGGTATTC	5635
V	13,14,15,16,17	FA-D2, sup13-I17 F	CATGGCAGGAACTCCGATCTTG	FA-D2, sup13-I17 F	CTCCCTTAAAAGCTCAAAGCTCAAGTTC	8858
VI	19, 20	hFANCD2_super_19_22_F	ACGTAATCACCCCTGTAATCC	hFANCD2_exon20_R	TGACAGAGCGAGACTCTCTAA	2749
VII	21,22,23	FA-D2, 21_23, F	GCTTCTAGTCACTGTCAGTTCACCAG	FA-D2, 21_23, R	ACGTTGGCCAGAAAGTAATCTCAG	2518
VIII	23,24,25,26	hFANCD2_super_23_29_F	GGCCTTGTGCTAAGTGCTTTT	hFANCD2_exon26_R	TCAGGGATATTGGCCTGAGAT	3252
IX	27,28,29	hFANCD2_exon27_F	GCATTCAGCCATGCTTGGTAA	hFANCD2_super_23_29_R	CACTGCAAACTGCTCACTCAA	3371
X	30	hFANCD2_super_30_32_F	CCAAAGTACTGGGAGTTTGAG	hFANCD2_exon30_R	TACCCAGTGACCCAAACACAA	2186
XI	31,32	hFANCD2_exon31_F	CCATTGCGAACCCTTAGTTTC	hFANCD2_super_30_32_R	ACCCTGGTGGACATACCTTTT	299
XII	33,34	hFANCD2_super_33_36_F	GAGCAATTTAGCCTGTGGTTTT	hFANCD2_exon34_R	TATAGCAAGAGGGCCTATCCA	3457
XIII	35,36	hFANCD2_exon35_F	TTAGACCGGGAACGTCTTAGT	hFANCD2_super_33_36_R	TCTGGGCAACAGAACAAGCAA	2040
XIV	43a	hFANCD2_super_43_44_F	AGGGTCCTGAGACTATATACC	hFANCD2_exon43a_R	AGCATGATCTCGGCTCACCA	2040
XV	44	hFANCD2_exon44_F	CACCCAGAGCAGTAACCTAAA	hFANCD2_super_43_44_R	ACCATCTGGCCGACATGGTA	464

Supplementary Table S2B. *FANCD2* exon primers

Exon	Designation	Sequence (5´→3´)	Designation	Sequence (5´→3´)	PCR Product Size (bp)
1	hFANCD2_exon1_F	TATGCCCGGCTAGCACAGAA	hFANCD2_exon1_R	TCCCATCTCAGGGCAGATGA	324
2	hFANCD2_exon2_F	CCCCTCTGATTTTGGATAGAG	hFANCD2_exon2_R	TCTCTCACATGCCTCACACAT	258
3	hFANCD2_exon3_F	GACACATCAGTTTTCCTCTCAT	hFANCD2_exon3_R	AAGATGGATGGCCCTCTGATT	354
4	hFANCD2_exon4_F	TGGTTTCATCAGGCAAGAAACT	hFANCD2_exon4_R	AATCATTCTAGCCCACTCAACT	253

4/5	FA-D2, exon 4 II F	GAGAAGGAAAACTATGGTAGGAAAC	FA-D2, exon 5 II R	GTGTAAGCTCTGTTTTCCTCAGAG	509
5	hFANCD2_exon5_F	GCTTGTGCCAGCATAACTCTA	hFANCD2_exon5_R	AGCCCCATGAAGTTGGCAAAA	298
6	hFANCD2_exon6_F	GAGCCATCTGCTCATTTCTGT	hFANCD2_exon6_R	GCTGTGCTAAAGCTGCTACAA	341
7	hFANCD2_exon7_F	AATCTCGGCTCACTGCAATCT	hFANCD2_exon7_R	CAGAGAAACCAATAGTTTTCAG	280
8	hFANCD2_exon8_F	TAGTGCAGTGCCGAATGCATA	hFANCD2_exon8_R	AGCTAATGGATGGATGGAAAAG	333
9	hFANCD2_exon9_F	TTCACACGTAGGTAGTCTTTCT	hFANCD2_exon9_R	TACTCATGAAGGGGGGTATCA	323
10	hFANCD2_exon10_F	GCCCAGCTCTGTTCAAACCA	hFANCD2_exon10_R	CATTACTCCCAAGGCAATGAC	229
	FA-D2, exon10, F	GTCTGCCCAGCTCTGTTCAAAC	FA-D2, exon10, R	ATTACTCCCAAGGCAATGACTGACTG	232
11	hFANCD2_exon11_F	GTGGGAAGATGGAGTAAGAGA	hFANCD2_exon11_R	AGCTCCATTCTCTCCTCTGAA	341
	FA-D2, exon11, F	CAGTTCAGTACAAAGTTGAGGTAGTG	FA-D2, exon11, R	CCGGATTAGTCAGTATTCTCAGTTAG	267
12	hFANCD2_exon12_F	TGCCTACCCACTATGAATGAG	hFANCD2_exon12_R	TCTGACAGTGGGATGTCAGAA	211
13	hFANCD2_exon13_F	CAGGAACTCCGATCTTGTAAG	hFANCD2_exon13_R	ATGTGTCCATCTGGCAACCAT	321
	FA-D2, exon 13 F P1+2	CCGATCTTGTAAGTTCTTTTCTGGTACG	FA-D2, exon 13 R P1+2	TGGCAACCATCAGCTATCATTTCCAC	302
14	hFANCD2_exon14_F	CGTGTTTCGCTGATGTGTCAT	hFANCD2_exon14_R	TGGAGGGGGGAGAAAGAAAG	186
15	hFANCD2_exon15a_F	GTGTTTGACCTGGTGATGCTT	hFANCD2_exon15a_R	GGAAGGCCAGTTTGTCAAAGT	325
	hFANCD2_exon15b_F	GTGGAACAAATGAGCATTATCC	hFANCD2_exon15b_R	CTTATTTCTTAGCACCCTGTCAA	204
	FA-D2, exon 15 F uniq	GGAACAAATGAGCATTATCCATTCTGTG	FA-D2, exon 15 R/ P1	CTCAATGGGTTTGAACAATGGACTG	363
16	hFANCD2_exon16_F	AGGGAGGAGAAGTCTGACATT	hFANCD2_exon16_R	TTCCCCTTCAGTGAGTTCCAA	332
	FA-D2, exon 16 F P1	GTCTGACATTCCAAAAGGATAAGCAAC	FA-D2, exon 16 R	CTTGAGACCCAGGTCAGAGTTC	344
17	hFANCD2_exon17_F	GATGGGTTTGGGTTGATTGTG	hFANCD2_exon17_R	GATTAGCCTGTAGGTTAGGTAT	422
	FA-D2, exon 17 F P1+2	CTGGCATATTCCTAAATCTCCTGAAG	FA-D2, exon 17 R	GCCTGTAGGTTAGGTATAAAGAAGTG	472
18	hFANCD2_exon18_F	GGCTATCTATGTGTGTCTCTTT	hFANCD2_exon18_R	CCAGTCTAGGAGACAGAGCT	282
19	hFANCD2_exon19_F	CGATATCCATACCTTCTTTTGC	hFANCD2_exon19_R	ACGATTAGAAGGGAACATGGAA	328
20	hFANCD2_exon20_F	CACACCAACATGGCACATGTA	hFANCD2_exon20_R	TGACAGAGCGAGACTCTCTAA	239
21	hFANCD2_exon21_F	AAAGGGGCGAGTGGAGTTTG	hFANCD2_exon21_R	GAGACAGGGTAGGGCAGAAA	339
22	hFANCD2_exon22_F	ATGCACTCTCTCTTTTCTACTT	hFANCD2_exon22_R	GTAACTTCACCAGTGCAACCAA	279
23	hFANCD2_exon23_F	TTCCCTGTAGCCTTGCGTATT	hFANCD2_exon23_R	ACAAGGAATCTGCCCCATTCT	356
24	hFANCD2_exon24_F	CTCCCTATGTACGTGGAGTAA	hFANCD2_exon24_R	CCCCACATACACCATGTATTG	258
25	hFANCD2_exon25_F	AGGGGAAAGTAAATAGCAAGGA	hFANCD2_exon25_R	GTGGGACATAACAGCTAGAGA	350
26	hFANCD2_exon26_F	GACATCTCTCAGCTCTGGATA	hFANCD2_exon26_R	TCAGGGATATTGGCCTGAGAT	324
27	hFANCD2_exon27_F	GCATTCAGCCATGCTTGGTAA	hFANCD2_exon27_R	CCAATTACTGATGCCATGATAC	324
28	hFANCD2_exon28_F	TCTACCTCTAGGCAGTTTCCA	hFANCD2_exon28_R	GATTACTCCAACGCCTAAGAG	354

	FA-D2, exon 28 F	TCTACCTCTAGGCAGTTTCCA	FA-D2, exon 28 R	GATTACTCCAACGCCTAAGAG	354
29	hFANCD2_exon29_F	CTTGGGCTAGAGGAAGTTGTT	hFANCD2_exon29_R	TCTCCTCAGTGTCACAGTGTT	384
30	hFANCD2_exon30_F	GAGTTCAAGGCTGGAATAGCT	hFANCD2_exon30_R	TACCCAGTGACCCAAACACAA	348
	FA-D2, exon 30 F	CATGAAATGACTAGGACATTCCTG	FA-D2, exon 30 F	GCAAGATGAATATTGTCTGGCAATACG	319
31	hFANCD2_exon31_F	CCATTGCGAACCCTTAGTTTC	hFANCD2_exon31_R	ACCGTGATTCTCAGCAGCTAA	341
32	hFANCD2_exon32_F	CCACCTGGAGAACATTCACAA	hFANCD2_exon32_R	AGTGCCTTGGTGACTGTCAAA	336
33	hFANCD2_exon33_F	CACGCCCGACCTCTCAATTC	hFANCD2_exon33_R	TACTGAAAGACACCCAGGTTAT	340
34	hFANCD2_exon34_F	TTGGGCACGTCATGTGGATTT	hFANCD2_exon34_R	TATAGCAAGAGGGCCTATCCA	349
	FA-D2, exon 34 II F	GGCAATCTTCTTGGGCTTATTACTGAG	FA-D2, exon 34 II R	CAACTTCCAAGTAATCCAAAGTCCACTTC	327
35	hFANCD2_exon35_F	TTAGACCGGGAACGTCTTAGT	hFANCD2_exon35_R	GTCCAGTCTCTGACAAACAAC	300
36	hFANCD2_exon36_F	CCTCTGGTTCTGTTTTATACTG	hFANCD2_exon36_R	GGCCAAGTGGGTCTCAAAAC	398
37	hFANCD2_exon37_F	CTTCCCAGGTAGTTCTAAGCA	hFANCD2_exon37_R	TCTGGGCAACAGAACAAGCAA	277
	FA-D2, exon 37 II F	CATCCTCTTACTAAGGACCCTAGTGAAAG	FA-D2, exon 37 II R	CAGCAACTTCCAAGTAATCCAAAGTCCAC	288
38	hFANCD2_exon38_F	GCACTGGTTGCTACATCTAAG	hFANCD2_exon38_R	AAGCCAGGACACTTGGTTTCT	274
39	hFANCD2_exon39_F	TGCTCAAAGGAGCAGATCTCA	hFANCD2_exon39_R	GCATCCATTGCCTTCCCTAAA	236
40	hFANCD2_exon40_F	CCTTGGGCTGGATGAGACTA	hFANCD2_exon40_R	CAGTCCAATTTGGGGATCTCT	309
41	hFANCD2_exon41_F	GATTGCAAGGGTATCTTGAATC	hFANCD2_exon41_R	CCCCAATAGCAACTGCAGATT	214
42	hFANCD2_exon42_F	AACATACCGTTGGCCCATACT	hFANCD2_exon42_R	GCTTAGGTGACCTTCCTTACA	356
43	hFANCD2_exon43a_F	GTGGCTCATGCTTGTAATCCT	hFANCD2_exon43a_R	AGCATGATCTCGGCTCACCA	366
	hFANCD2_exon43b_F	CTGCCACCTTAGAGAACTGAA	hFANCD2_exon43b_R	TCAGTAGAGATGGGGTTTCAC	358
	hFANCD2_exon43c_F	TAGAATCACTCCTGAGTATCTC	hFANCD2_exon43c_R	CTCAAGCAATCCTCCTACCTT	405
	hFANCD2_exon43d_F	AGTTGGTGGAGCAGAACTTTG	hFANCD2_exon43d_R	CAGCTTCTGACTCTGTGCTTT	367
	hFANCD2_exon43e_F	TCAACCTTCTCCCCTATTACC	hFANCD2_exon43e_R	CTCGAGATACTCAGGAGTGAT	381
	hFANCD2_exon43f_F	GGTATCCATGTTTGCTGTGTTT	hFANCD2_exon43f_R	AGTTCTGCTCCACCAACTTAG	306
44	hFANCD2_exon44_F	CACCCAGAGCAGTAACCTAAA	hFANCD2_exon44_R	GAAAGGCAAACAGCGGATTTC	213
	FA-D2, exon 44 II F	CTAGGAGCTGTATTCCAGAGGTCAC	FA-D2, exon 44 II R	GGATCCTACCAGTAAGAAAGGCAAAC	250

Supplementary Table S2C. *FANCD2* mutation-specific primers

PCR/ Sequencing	Designation	Sequence (5´→3´)
PCR/Seq	FA-D2, exon4-6 F	GAAGGAAAACTATGGTAGGAAACTGGTG
	FA-D2, exon 6 R	CAGATGTATTAGGCTAATAAGCACAG
	FA-D2, exon4-i6 R	CCAGAAGCAGTTTGATGAGACTCTTAG
	FA-D2, exon i4F	GCTTTCCAAAAGAAGCTCTTTCAGAC
	FA-D2, exon4-IVS F	GGAGACACCCTTCCTATCCCAAAG
	FA-D2, exon 5F	GAGTGGGCTAGAATGATTTTTAACAGC
	FA-D2, exon 5 R	CTCTGAGGAAAACAGAGCTTACAC
	D2_AluYb9 F	GCAATCTCGGCTCACTGCAAGCTC
	D2_IVS4/AluYb9, R	GCTGTTAAAAATCATTCTACTTTGGGAGG
PCR/Seq	FA-D2, ex 10 F	GACTTGACCCAAACTTCCTATTGAA*
	FA-D2, ex 14 F	TCGTGTTTCGCTGATGTGT
	FA-D2, ex 11 R	CCGGATTAGTCAGTATTCTCAGTTAG
	IVS14+2411 R	CGAGACCATCCTGACTAACACG
	IVS14+2512 R	GATACCCCTTAAGAATACAGAGC
PCR/Seq	FANCD2_16S	AGAGCTAGGGAGGAGAAGTCTGA
	FANCD2_18A	GAGCTGAGATCGTGCCAACT
	FANCD2_17S	TGGTCAAGTTACACTGGCATATT
	FANCD2_17A	CCATCCTTCAGCAATCACTC
PCR/Seq	D2_P2_21_23 F	GTTTTCTGATACTTGGAAACTACTGGCTTG
	D2_P1_21_23 R	GACACAGAGGTAGCAAAGGATGTTC
Seq	FA-D2, ex21_23, int1	CTATGATGAATTTGCCAACCTGATCC
	FA-D2, ex21_23, int2	GAGGGCTCCTTCACTTAATAACAATC
	FA-D2, ex21_23, int3	GTATTGTTTACCTGCTGGCTGGTTG
PCR	FA-D2_sup_exon26 II F uniq	TAGGGTCACAAGCCTAATCTCCTTT
	FA-D2_sup_exon26 II R uniq	GGCCATGATGAATAATCTTTCTTTTGTTTG

Supplementary Table S3. Microsatellite primers

STR	Genomic Position [Mb]	Sense Primer Sequence (5´→3´)	Antisense Primer Sequence (5´→3´)
D3S1597	9,34	AGTACAAATACACACAAATGTCTC	CAATTCGCAAATCGTTCATTGCT
D3S1038	10,49	AAAGGGGTTCAGGAAACCTG	CCCTCCAGTAAGAGGCTTCCTAG
D3S3611	10,53	GCTACCTCTGCTGAGCATATTC	CACATAGCAAGACTGTTGGGGGC
D3S1675	10,64	GGATAGATGGATGAATGGATGGC	CCTCTCTAACTACCAATTCATCCA

Supplementary Table S4. Laboratory diagnostic data of the 29 cohort FA-D2 patients

Patient number	Kindred Sibling	Cell type of lab diagnosis	G2-phase arrest, G2/GF		Breaks/cell		Technique of complementation group assignment	*FANCD2* mutation		Somatic mosaicism
			Spon	MMC/DEB	Spon	MMC/DEB		Allele 1	Allele 2	
1	1 / I	Lymphocyte	65.7%	n.d.	0.07	4.5 (M) [6.6 (M)]	IB of LCL	c.696–121C>G (exonization)	c.696–121C>G (exonization)	None (residual protein)
2	2 / I	Lymphocyte	54.3%	70.1% (M)	0.09	1.4 (M), 1.5 (D)	IB and RC of LCL	c.1948–6C>A (exon 22 skipping)	c.3599delT	None (residual protein)
3	3 / I	Lymphocyte (prior to mosaicism)	38.6% (prior to mosaicism)	46.6% (M) (prior to mosaicism)	0.04	0.06 (M)	RC of fibroblasts	c.1948–16T>G (exon 22 skipping)	c.1948–16T>G (exon 22 skipping)	1954G>A (exon 22), V652I, reconstitutes exon 22 recognition (blood, BM, LCL)
4	4 / I	Lymphocyte	45.7%	63.6% (M)	n.d.	n.d.	RC of fibroblasts	c.1948–16T>G (exon 22 skipping)	c.1948–16T>G (exon 22 skipping)	None (no LCL)
5	4 / II	Lymphocyte	44.5%	58.9% (M)	n.d.	n.d.	RC of fibroblasts	c.1948–16T>G (exon 22 skipping)	c.1948–16T>G (exon 22 skipping)	None (residual protein)
6	5 / I	Lymphocyte	34.5%	64.7% (M)	0.05	4.7 (M) 5.6 (D)	IB and RC of LCL	c.274–57_–56insinvAluYb8nt36_319 +dup c.274–69_–57 (exon 5 skipping)	c.3803G>A	None (residual protein)
7	6 / I	Lymphocyte	34.8%	51.5% (M)	n.d.	n.d.	IB of LCL	c.904C>T	c.1092G>A	None (residual protein)
8	7 / I	Lymphocyte	45.6%	58.4 (M)	0.06	1.3 (M)	IB of LCL	c.990–1G>A (aberrant splicing)	c.1948–6C>A (exon 22 skipping)	None (residual protein)
9	8 / I	Lymphocyte	65.3%	70.9% (M)	0.12	n.d.	IB of LCL	c.810_812delGTC	c.1948–16T>G (exon 22 skipping)	None (residual protein)

Patient number	Kindred Sibling	Cell type of lab diagnosis	G2-phase arrest, G2/GF		Breaks/cell		Technique of complementation group assignment	*FANCD2* mutation		Somatic mosaicism
			Spon	MMC/DEB	Spon	MMC/DEB		Allele 1	Allele 2	
10	9 / I	Lymphocyte	38.4%	58.4% (M)	n.d.	n.d.	RC of fibroblasts	c.1948−16T>G (exon 22 skipping)	c.2715+1G>A (aberrant splicing)	None (residual protein)
11	10 / I	Lymphocyte	55.4%	65.8% (M)	? (Wien)	? (Wien)	IB of LCL	c.3707G>A (aberrant splicing)	c.2835dupC	None (residual protein)
12	11 / I	Lymphocyte	40.2%	61.1% (M)	n.d.	n.d.	IB of LCL	c.274−57_−56insinvAluYb8nt36_319 +dup c.274−69_−57	c.3453_3456delCAAA	None (residual protein)
13	12 / I	Lymphocyte	27.6%	57.8%	0.11	1.08 (M) 2.9 (D)	IB of LCL	c.1948−16T>G (exon 22 skipping)	c.1948−16T>G (exon 22 skipping)	None (residual protein in T cells and LCL)
14	13 / I	Lymphocyte Fibroblast	20.9%	32.1% (M)	n.d.	n.d.	Mutation analysis (by sibling)	c.1948−6C>A (exon 22 skipping)	2775_2776CC>TT	2775_2776 CC (blood, LCL)
15	13 / II	Lymphocyte Fibroblast	25.0%	35.4% (M) 69.2% (M)	0	0.11 (M) 0.02 (D)	RC of fibroblasts	c.1948−6C>A (exon 22 skipping)	2775_2776CC>TT	Recombination
16	14 / I	Lymphocyte	n.d.	n.d.	0.04	1.78 (D)	IB of LCL	c.2444G>A	c.2444G>A	None (residual protein)
17	14 / II	Lymphocyte	n.d.	n.d.	0.12	3.1 (D)	Mutation analysis (by sibling)	c.2444G>A	c.2444G>A	None (no LCL)
18	15 / I	Lymphocyte	n.d.	n.d.	0.12	1.5 (D)	IB of LCL	c.696−2A>T (exon 10 skipping)	c.1321_1322delAG (aberrant splicing)	None (residual protein)
19	16 / I	Fetal blood	n.d.	n.d.	n.d.	3.7 (D)	RC of fetal fibroblasts	c.692T>Gpat	c.2444G>Amat	not done
20	17 / I	Lymphocyte	n.d.	n.d.	0.02	8.4 (D)	IB and RC of LCL	c.1948−6C>Amat, (exon 22 skipping)	2660delApat	None (residual protein)
21	18 / I	Lymphocyte	n.d.	n.d.	0.02	5.4 (D) 10.3 (D)	IB of LCL	c.2404C>T	c.2444G>A	None (residual protein)

Patient number	Kindred Sibling	Cell type of lab diagnosis	G2-phase arrest, G2/GF		Breaks/cell		Technique of complementation group assignment	*FANCD2* mutation		Somatic mosaicism
			Spon	MMC/DEB	Spon	MMC/DEB		Allele 1	Allele 2	
22	19 / I	Lymphocyte	n.d.	n.d.	0.04	3.7 (D)	RC of fetal fibroblasts from 880/2 (early spontaneous abortion)	c.2444G>Apat	c.2715+1G>Amat (aberrant splicing)	None (residual protein)
23	20 / I	Lymphocyte	n.d.	n.d.	0.08	7.4 (D)	IB of LCL	c.757C>T	c.1367T>G	None (residual protein)
24	20 / II	Lymphocyte	n.d.	n.d.	0.20	8.9 (D)	IB of LCL	c.757C>T	c.1367T>G	None (residual protein)
25	21	Lymphocyte	n.d.	n.d.	Data missing	Data missing	Mutation analysis	c.1948−16T>G (exon 22 skipping)	c.1948−16T>G (exon 22 skipping)	1953G>T (W651C) (blood, LCL)
26	22 / I	Fibroblast	22.2% (fibroblast)	54% (fibroblast, 300 nM ≈ 100 ng/ml MMC)	0.04	0.16 (300 nM ≈ 100 ng/ml MMC)	Mutation analysis in fibroblasts (by sibling)	c.376A>G (aberrant splicing)	c.3803G>A	376A (blood,)
27	22 / II	Lymphocyte	n.d.	n.d.	0.12	>10 (M)	IB and IP of LCL	c.376A>G (aberrant splicing)	c.3803G>A	None (residual protein)
28	23 / I	Lymphocyte	n.d.	n.d.	0.10	6.0 (M)	IB, IP and RC of LCL	c.206−2A>T (exon 4 skipping)	g.22875_23333del459 (c.1414−71_c.1545+256del459)	None (residual protein)
29	23 / II	Lymphocyte	n.d.	n.d.	0.12	8.1 (M)	Mutation analysis (by sibling)	c.206−2A>T (exon 4 skipping)	g.22875_23333del459 (c.1414−71_c.1545+256del459)	None (residual protein)

MMC, M, mitomycin C; DEB, D, diepoxybutane; RC, retroviral complementation; IB, immunoblotting; IP, immunoprecipitation; LCL lymphoblast cell line; n.d., not determined; G2, G2 phase fraction of the cell cycle; GF, growth fraction; G2/GF, ration G2 phase fraction over GF

Supplementary Table S5. Clinical diagnostic data of the 29 cohort FA-D2 patients

Patient number	Kindred/ Sibling	Consanguinity Gender	Ethnicity Nationality	Age at diagnosis	Clinical presentation	Hematologic manifestations	Survival at last follow-up	Family history
1	1 / I	unkown f	Asian Indian	6 mo	IUGR, patent ductus arteriosus, pigmentation anomalies, microcephaly, low-set ears, hypoplastic thumb with duplicate nail (R), radial ray aplasia with cutaneous thumb (L), pelvic kidney (R), congenital hip dislocation (L), aplasia of the corpus callosum	BMF as of 2 y 4mo, transfusions from 3 y 2 mo, AML at 7.0 y	† 7 y 6 mo (AML, pneumonia)	No SABs; no known cancer
2	2 / I	absent f	Caucasian German	5 y 7 mo	GR, pigmentation anomalies, microcephaly, microphtalmia, low-set thumbs, duplicate kidney (R), dysplastic hips	BMF as of 5 y 7 mo, cortisol from 8 y, transfusions from 8 y 4 mo, androgen from 9 y 2 mo	† 11 y 4 mo (subarachnoidic hemorrhage)	1 SAB; MGM:Cervix ca, 40 y
3	3 / I	cousins of 1st° m	Caucasian Turkish	1 y 11 mo	IUGR, pigmentation anomalies, microcephaly, hypoplastic thumbs (L>R), syndactyly II/III toes, hypogenitalism, glomerulosclerosis	Stable partial mosaïcism, BMF as of 11 y, cortisol and androgen from 12 y, transfusions from 18 y 9 mo, BMT at 19 y 7 mo	† 20 y 7 mo (viral encephalitis following BMT)	No SABs, no known cancer
4	4 / I	cousins of 2nd° f	Caucasian Turkish	5 y 10 mo	IUGR, pigmentation anomalies, microcephaly, microphtalmia, hypoplastic thumb (R), hydocephalus internus, hypoplastic corpus callosum, mental retardation, hyperactivity attention deficit disorder	BMF as of 2 y 6 mo, transfusions from 2 y 6 mo, subdural hemorrhage 6 y, BMT at 7 y	8 y 3 mo	No SABs; no known cancer
5	4 / II	cousins of 2nd° m	Caucasian Turkish	4 y 5 mo	IUGR, microcephaly, microphtalmia, strabism, mental retardation, hyperactivity attention deficit disorder	BMF as of 3 y 3 mo, transfusions from 3 y 3 mo, oxymetholon from 5y 9 mo	6 y 11 mo	No SABs; no known cancer

Patient number	Kindred/ Sibling	Consanguinity Gender	Ethnicity Nationality	Age at diagnosis	Clinical presentation	Hematologic manifestations	Survival at last follow-up	Family history
6	5 / I	absent m	Caucasian German	2 y 6 mo	GR, microcephaly, micro-phtalmia, absent anthelix (R), radial ray hypoplasia, preaxial hexadactyly (R), duplicate pelvic kidney (R), maldescensus of the testes, micropenis, dysplastic hips, hypoplastic corpus callosum, misshaped brain ventricles, psychomotor retardation	BMF as of 2 y 9 mo, BMT at 3 y 3 mo.	4 y 4 mo	No SABs; PGM cancer 70 y, otherwise no cancer history
7	6 / I	absent f	Caucasian Italian	2 y	IUGR, pigmentation anomalies, microcephaly, microphtalmia, absent thumbs, short radii, absent anthelix (R), closed auditory canals	BMF as of 4.5 y	12 y	No SABs; no cancer history
8	7 / I	absent m	Caucasian German	3 y 9 mo	Pigmentation anomalies, microcephaly, 'flat' auricles-absent anthelix?, ptosis, short thumbs, hyperactivity attention deficit disorder	BMF as of 4 y	4 y 4 mo	No SABs; no cancer history
9	8 / I	absent f	Caucasian Czech	2 y 11 mo	IUGR, microcephaly, brain atrophy, patent ductus arteriosus, esophagus atresy, tracheoesophageal fistula (IIIb), hypoplastic kidneys, polycystic ovary (L), triphangeal digitalized thumbs, pedes equinovari, rib anomaly (VACTERL-like association)	BMF as of 2 y 10 mo, transfusions from 2 y 11mo	† 5 y 10 mo (hemorrhage)	1 SAB (first trimester); no cancer in the family

Patient number	Kindred/ Sibling	Consanguinity Gender	Ethnicity Nationality	Age at diagnosis	Clinical presentation	Hematologic manifestations	Survival at last follow-up	Family history
10	9 / I	absent f	Caucasian Turkish	7 mo	IUGR, pigmentation anomalies, microcephaly, hydrocephalus internus, absent corpus callosum, microphtalmia, small mouth, low-set ears, hypoplastic thumbs, unilateral triphalangeal (R), pelvic kidney (L), hip luxation, psychomotor retardation	BMF as of 2 y	2 y 3 mo	1 SAB (first trimester); 1 pregnancy terminated because of hydrocephalus and renal agenesy; PGF bronchus ca
11	10 / I	absent f	Caucasian Austrian	10 y 10 mo	IUGR, pigmentation anomalies, microcephaly, hypoplastic thumbs, ectopic kidney (R)	BMF as of 10 y 10 mo, transfusions from 10 y 10 mo, MDS (RAEB-t) with del(7)(q32) at 10 y 10 mo, BMT at 11y 1 mo	11y 11 mo	No SABs; no cancer history
12	11 / I	absent m	Caucasian Danish	3 mo	IUGR, atresy of the duodenum, microcephaly, dilated lateral ventricles and stenosis of the aquaeduct (hydrocephalus), hypoplasia of the carpous callosum, microphtalmia, closed auditory channels, hypoplastic thumbs, micropenis	BMF as of 2 wks	5 mo	No SABs; PGM and MPGM breast ca., MGGF prostate ca.
13	12 / I	cousins of 1st° m	Caucasian Turkish	5 y 5 mo	IUGR, pigmentation anomalies, microcephaly, microphalmia, psychomotor retardation, Michelin tire baby syndrome	BMF as of 1 y 5 mo	5 y 8 mo	No SABs; no cancer history
14	13 / I	absent f	Caucasian German	34 y 2 mo	IUGR, microcephaly, mild radial ray hypoplasia	None	34 y	No SABs; no cancer history
15	13 / II	absent f	Caucasian German	21 y 11 mo	IUGR, microcephaly, radial ray hypoplasia, dysplasia of mandibula, anomalies of the teeth, dysplasia of hip (R), mental retardation	Transfusions from 17 y 6 mo, MDS(RARS-RAEB) at 17 y 6 mo	† 23 y 5 mo (pneumonia, invasive aspergillosis, hemorrhage)	No SABs; no cancer history

Patient number	Kindred/ Sibling	Consanguinity Gender	Ethnicity Nationality	Age at diagnosis	Clinical presentation	Hematologic manifestations	Survival at last follow-up	Family history
16	14 / I	cousins of 3° m	Caucasian Spanish	6 y	Patent ductus arteriosus, pigmentation anomalies, bifid thumb (R), hypogonadism	BMF as of 7 y (very mild hypoplasia of the myeloid series)	25 y	No SABs; MGF lung, PGFstomach ca.
17	14 / II	cousins of 3° m	Caucasian Spanish	8 mo	Pigmentation anomalies, microphtalmia, hypoplastic thumb (R), absent os metacarpale I (L), glandular hypospadia	Blood cell counts at low-range normal levels	20 y	No SABs; MGF lung, PGF stomach ca.
18	15 / I	absent f	Caucasian Spanish	5 y 3 mo	IUGR, pigmentation anomalies, microcephaly, microphtalmia, hypotelorism, annular pancreas	BMF as of 5 y 3 mo, androgen, G-SCF, EPO and transfusions from 5 y 3 mo, BMT at 10 y 11 mo	† 11 y 1 mo (graft failure / no take)	No SABs; MMGM colon ca.
19	16 / I	absent m	Caucasian, maternal Irish and English, paternal Irish and Italian	22 wk of gestation	IUGR, absent thumb and radial aplasia (R), lateral cerebral ventricular dilation (hydrocephalus)	N/A	N/A, terminated with diagnosis of FA	3 first trimester SABs, 4th fetus with IUGR, radial aplasia, cystic hygromas, encephalocele, probably heart defects, terminated; MGM pancreas, MMGM breast, MGF melanoma & basal cell ca.
20	17 / I	absent m	Caucasian maternal German, paternal Dutch	4 y 5 mo	IUGR, pigmentation anomalies, microcephaly, microphtalmia	BMF as of 2 y, BMT at 5 y	9 y (4y post BMT)	1 SAB, M.3x basal cell, MMMGM melanoma, MMGF breast, PGM bowel ca.

Patient number	Kindred/ Sibling	Consanguinity Gender	Ethnicity Nationality	Age at diagnosis	Clinical presentation	Hematologic manifestations	Survival at last follow-up	Family history
21	18 / I	absent m	Caucasian Hispanic (Mexican)	7 mo	IUGR, café au lait spots, microcephaly, micro-phtalmia, hearing loss (auditory canals? sensory hearing impairment?), absent thumbs and radii, intestinal atresia, renal defects, genital anormalities (undescended testes), learning disabilities	None yet	10 y 3 mo	No SAB; one cancer;
22	19 / I	absent m	Caucasian maternal Irish, Dutch Yugoslavian French and Native American, paternal Irish and Sicilian	Newborn	Patent ductus arteriosus, pigmentation anomalies, low-set ears, malformed auricle (R), constriction bands of mid forearms (Michelin tire baby syndrome?), preaxial hexadactyly (R), hypoplastic thumb with ponce flottant (L)	BMF as of 1 y 4 mo	2 y 6 mo	5 miscarriages with one positive for FA; PGM breast, MPGM cervix and lung ca., MMGM brain tumor
23	20 / I	absent m	maternal African American / Caucasian, paternal African American	(1 mo)	IUGR; microcephaly; microphtalmia; hypoplastic thumb (L); hypoplastic metacarpal I (R); horseshoe kidney	none	1 y 4 mo	GM: 2 miscarriages. Cancer only in GreatGP generation
24	20 / II	absent f	maternal African American / Caucasian, paternal African American	4 y 6 mo	IUGR; café-au-lait spots; microcephaly; microphtalmia	BMF starting from 4 y 6 mo	5 y 9 mo	GM: 2 miscarriages Cancer only in GreatGP generation
25	21 / I	cousins of 1st° m	Caucasian Turkish	5 y 5 mo	IUGR, pigmentation anomalies, microcephaly, microphtalmia, pelvic kidney	BMF starting from 4 y; oxymethalone and prednisone from 5 y 5 mo	† 9 y (intracranial hemorrhage)	No SAB; no cancer history

Patient number	Kindred/ Sibling	Consanguinity Gender	Ethnicity Nationality	Age at diagnosis	Clinical presentation	Hematologic manifestations	Survival at last follow-up	Family history
26	22 / I	absent f	Caucasian Dutch	5 y	IUGR (asymmetrical); pigmentation anomalies; microcephaly; ventriculo-megaly (hydrocephalus) and multiple developmental anomalies of the brain, possibly holopros-encephaly; hypotelorism; microphtalmia; narrow auditory canals; hypoplastic os metacarpale I, renal aplasia (R); dysplasia of the hip (L); growth hormon deficiency	BMF as of 5 y; transfusions from 6 y 9 mo + GCSF; BMT at 7 y 10 mo	9 y	No SAB; no cancer history
27	22 / II	absent m	Caucasian Dutch	3 y	IUGR; pigmentation anomalies; microcephaly; hypoplastic corpus callosum; hypertelorism; blepharophimosis; preaxial hexadactyly (L)	BMF as of 2 y 1 mo; transfusions from 5 y 8 mo; BMT at 7 y 7 mo	7 y 9 mo	No SAB; no cancer history
28	23 / I	absent m	Caucasian Dutch	8 y 6 mo	IUGR, pigmentation anomalies, microcephaly, Kartagener syndrome with situs inversus, mild mental retardation	BMF as of 8 y 5 mo; no transfusions; BMT at 9 y 5 mo	† 10 y 1 mo (gastro-intestinal hemorrhage due to necrotizing enterocolitis post BMT)	1 SAB; no cancer history
29	23 / II	absent m	Caucasian Dutch	5 y 8 mo	IUGR, pigmentation anomalies, microcephaly, microphtalmia	BMF as of 5 y 8 mo; BMT at 6 y 7 mo	7 y 10 mo	1 SAB; no cancer history

f, female; m, male; L, left; R, right; (IU)GR, (intrauterine) growth retardation; BMF, bone marrow failure; BMT, bone marrow transplantation; MDS, myelodysplastic syndrome; AML, acute myelogenous leukemia; SAB, spontaneous abortion

Supplementary Table S6A. *FANCD2* splice acceptor calculations

3´Splice site (acceptor)			
Exon	**Wild type/ variant**	**Sequence**	**Score (Maxent*)**
4	Consensus	ctcttctttttctgcatagCTG	9.12
	IVS3-2A>T	ctcttctttttctgcat**t**gCTG	0.76
10	Consensus	tctttttctaccattcacagTGA	7.39
	IVS9-2A>T	tctttttctaccattcac**t**gTGA	-0.97
13	Consensus	ttcctctctgctacttgtagTTC	6.19
	IVS12-1G>A	ttcctctctgctacttgta**t**TTC	-2.56
22	Consensus	tgtttgtttgcttcctgaagGAA	6.43
	IVS21-16T>G	tgtt**g**gtttgcttcctgaagGAA	5.58
	IVS21-6C>A	tgtttgtttgcttc**a**tgaagGAA	4.51
37a	Ex37 (consensus)	ACTTTTGTTGTTTTCTTCCGTGT	2.10
	3706C>A	ACTTTTGTTGTTTTCTTC**A**GTGT	10.14

* MaxEntScan::score3ss for human 3' splice sites (http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq_acc.html)

Supplementary Table S6B. *FANCD2* splice donor calculations

5´Splice site (donor)						
Exon	Wild type/ variant	Sequence	Score (splicefinder*)		Difference	Result
5	Consensus	CAGgtgtggag	LC4	12||3|2|		
	376A>G	CGGgtgtggag	LC2	2|8||3|2	large	malfunction
9a	IVS (consensus)	acggtaactta	LC4	||12|2||		
	IVS9-121C>G	ACGgtaagtta	HC3	||17||	large	gain of function
10	Consensus	AAGgtagaaaa	LC4	|12|||2|		
	782A>T	ATGgtagaaaa	LC3	||10|||2|	small	malfunction

* Splicefinder (http://www.uni-duesseldorf.de/rna/html/5__ss_mutation_assessment.php)

Supplementary Figure S1. Circular map of the vector S11FD2IN. The retroviral expression vector S11FD2IN contains a bicistronic construct of the full-length *FANCD2* cDNA (FANCD2) and the neomycin resistance gene (NEO). Translation of the latter is ensured by an internal ribosomal entry site (IRES). Shown are also the LTRs, the restriction sites and their positions and the bacterial resistance (*AmpR*).