Published in final edited form as:

Hum Mutat. 2018 February ; 39(2): 237–254. doi:10.1002/humu.23366.

A comprehensive approach to identification of pathogenic FANCA variants in Fanconi anemia patients and their families

Danielle C. Kimble¹, Francis P. Lach², Siobhan Q. Gregg², Frank X. Donovan¹, Elizabeth K. Flynn¹, Aparna Kamat¹, Alice Young³, Meghana Vemulapalli³, James W. Thomas³, James C. Mullikin³, Arleen D. Auerbach⁴, Agata Smogorzewska^{2,*}, and Settara C. Chandrasekharappa^{1,*}

¹Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, NIH, Bethesda, MD 20892, USA

²Laboratory of Genome Maintenance, The Rockefeller University, New York, NY 10065, USA

³NIH Intramural Sequencing Center, National Human Genome Research Institute, NIH, Bethesda, MD 20892, USA

⁴Human Genetics and Hematology Program, The Rockefeller University, New York, NY 10065, USA

Abstract

Fanconi anemia (FA) is a rare recessive DNA repair deficiency resulting from mutations in one of at least 22 genes. Two-thirds of FA families harbor mutations in *FANCA*. To genotype patients in the International Fanconi Anemia Registry (IFAR) we employed multiple methodologies, screening 216 families for *FANCA* mutations. We describe identification of 57 large deletions and 261 sequence variants, in 159 families. All but seven families harbored distinct combinations of two mutations demonstrating high heterogeneity. Pathogenicity of the 18 novel missense variants was analyzed functionally by determining the ability of the mutant cDNA to improve the survival of a *FANCA*-null cell line when treated with MMC. Overexpressed pathogenic missense variants were found to reside in the cytoplasm, and nonpathogenic in the nucleus. RNA analysis demonstrated that two variants (c.522G>C and c.1565A>G), predicted to encode missense variants, which were determined to be nonpathogenic by a functional assay, caused skipping of exons 5 and 16 respectively, and are most likely pathogenic. We report 48 novel *FANCA* sequence variants. Defining both variants in a large patient cohort is a major step towards cataloging all *FANCA* variants, and permitting studies of genotype-phenotype correlations.

Keywords

Fanconi anemia; *FANCA*; recessive disorder; functional assay; pathogenic mutations

*correspondence to Settara C. Chandrasekharappa: Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, National Institutes of Health, 50 South Drive, Building 50, Room 5232, Bethesda, MD. chandra@mail.nih.gov or Agata Smogorzewska: Laboratory of Genome Maintenance, The Rockefeller University, 1230 York Avenue, New York, NY. asmogorzewska@rockefeller.edu.

Disclosure Statement: The authors declare no conflict of interest.

INTRODUCTION

Fanconi anemia (FA) is a DNA repair deficiency syndrome associated with congenital abnormalities, hematological manifestations, and predisposition to cancer (Alter, et al., 2010; Kutler, et al., 2003a; Rosenberg, et al., 2003). Two-thirds of FA patients exhibit congenital defects involving multiple organs including the skeletal and gastrointestinal systems as well as heart and kidneys (Alter and Giri, 2016; Skagen, et al., 1978). Hematological manifestations presenting at a median age of seven years include bone marrow failure (BMF), myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML). FA patients are at a 700-fold increased risk for developing squamous cell carcinoma, primarily of the head and neck, as well as the esophagus and anogenital region (Kutler, et al., 2003b; Rosenberg, et al., 2003). The cellular phenotype of increased chromosomal breakage upon exposure to DNA crosslinking agents, diepoxybutane (DEB) or mitomycin C (MMC) has been employed in the clinical diagnosis of FA (Auerbach, 2009). FA patients have been found to carry biallelic pathogenic variants in any of 20 genes, *FANCA* (MIM #607139), *FANCC* (MIM #613899), *FANCD1/BRCA2* (MIM #600185), *FANCD2* (MIM #613984), *FANCE* (MIM #613976), *FANCF* (MIM #613897), *FANCG* (MIM #602956), *FANCI* (MIM #611360), *FANCI/BRIP1* (MIM #605882), *FANCL* (MIM #608111), *FANCM* (MIM #609644), *FANCN* (MIM #610832)/*PALB2* (MIM #610355), *FANCO* (MIM #613390)/*RAD51C* (MIM #602774), *FANCP* (MIM #613951)/*SLX4* (MIM #613278), *FANCQ* (MIM #615272)/*ERCC4* (MIM #133520)/*XPF* (MIM #278760), *FANCS/BRCA1* (MIM #113705), *FANCT* (MIM #616435)/*UBE2T* (MIM #610538), *FANCU* (MIM #617247)/*XRCC2* (MIM #600375), *FANCV* (MIM #617243)/*MAD2L2* (MIM #604094)/*REV7*, and *FANCW/RFWD3* (MIM #614151). Variants in *FANCB* (MIM #300515), an X-linked gene and an autosomal dominant variant in *FANCR* (MIM #617244)/*RAD51* (MIM #179617), can also cause the disease, resulting in a total of 22 FA genes reported thus far (Knies, et al., 2017; Kottemann and Smogorzewska, 2013; Mamrak, et al., 2016; Rickman, et al., 2015; Wang and Smogorzewska, 2015). The underlying defect in FA patients is the inability to repair DNA interstrand crosslinks (ICLs), although many proteins encoded by the FA genes have functions outside of ICL repair (Boersma, et al., 2015; Schlacher, et al., 2011; Schlacher, et al., 2012; West, et al., 2015; Xu, et al., 2015).

Two-thirds of FA patients harbor pathogenic variants in *FANCA*, which makes it the most frequently mutated FA gene, while 20% of patients carry mutations in *FANCC* or *FANCG*, and causative mutations in the other 19 FA genes account for 0.1%–4% of cases per gene (Wang and Smogorzewska, 2015). *FANCA* variants include single nucleotide variations, small insertions and deletions (indels), and large deletions. *FANCA* variants are largely private, except for three founder mutations: exon12-31del, c.295C>T (p.Q99*), and exon15del in South Africaner (Tipping, et al., 2001), Spanish Gypsy (Callen, et al., 2005; Castella, et al., 2011), and Tunisian (Amouri, et al., 2014) families, respectively. In addition there are two frequently observed mutations across ethnic groups: c.3788_3790delTCT in exon 38 and c.1115_1118delTTGG in exon 13 (Levrn, et al., 1997) (FA MutDB: <http://www2.rockefeller.edu/fanconi/>). Twenty to forty percent of *FANCA* mutations are large deletions which were found, primarily, by multiplex ligation probe analysis (MLPA) (Castella, et al., 2011; Centra, et al., 1998; Flynn, et al., 2014; Levrn, et al., 1998;

Moghrabi, et al., 2009; Morgan, et al., 1999). Our recent study scanning for large deletions in the entire gene and 200kb 5' and 3' to the gene regions by high-resolution aCGH (Comparative Genomic Hybridization) indicates that nearly half of the *FANCA* large deletions extend beyond boundaries of the gene (Flynn, et al., 2014). In addition, because the *FANCA* gene is large, with 43 exons, it harbors complex heterogeneity of mostly private mutations. As a result, many patients are left with uncharacterized disease-causing *FANCA* mutations.

Complete molecular diagnosis of the disease-causing gene and specific pathogenic variants is required for confirmation of an FA diagnosis and understanding of genotype-phenotype correlations. Molecular diagnosis is also critical for detection of mutation carriers, identification of appropriate donors for bone marrow transplantation, counseling of families, and exploring novel therapeutic interventions. We present here a summary of screening results from 216 FA families, and characterization of both variants in 159 *FANCA* families. We employed high throughput sequencing strategies for detection of variants, and MLPA, aCGH and SNP arrays for large deletions. In situations where pathogenicity was uncertain, we performed functional evaluation to obtain a more accurate characterization of potentially pathogenic variants. This work highlights the importance of a multipronged approach for identifying all causative mutations in FA patients.

METHODS

Further details on the sequencing methods and analysis, and additional methods are presented online *as* Supplemental Material

Study Subjects

Genomic DNA samples, fibroblast cultures, and EBV immortalized lymphoblastoid cell lines were from individuals diagnosed with FA and registered in the International Fanconi Anemia Registry (IFAR), following written informed consent. The Institutional Review Board of the Rockefeller University, New York, USA, has approved these studies. The Office of Human Subjects Research at the National Institutes of Health and Institutional Review Board of the National Human Genome Research Institute (NHGRI) approved the reception of de-identified cell lines and DNA samples from The Rockefeller University and analysis of the underlying molecular variants. Genomic DNA was isolated from peripheral blood (PB), fibroblast, or EBV immortalized lymphoblastoid cell lines (LCL). Genomic DNA was extracted using heat lysis/proteinase K/RNaseA/SDS digestion of the MNCs and cultured cells and purified by 2-propanol precipitation/70% ethanol wash or using the DNeasy blood & tissue DNA extraction kit (Qiagen, Inc). Phenol/chloroform extraction and ethanol precipitation was included as a final step in the preparation of DNA.

High throughput Sanger sequencing

Forty-nine amplicons ranging in size from 222bp–740bp were designed to cover the exons and the flanking regions of *FANCA* (NG_011706.1, NM_000135.3). We adopted published ClinSeq methodologies (Biesecker, et al., 2009) for high throughput Sanger sequencing of genomic DNA samples from 179 *FANCA* families. PCR products were subjected to

bidirectional Sanger sequencing. Sequence data from ~19,000 sequence reads was analyzed to identify sequence variants and to find rare, likely pathogenic variants. Common polymorphic variants were removed based on their frequency in the 1000 Genomes Project (1000G), NHLBI Exome Sequencing Project (ESP6500), ExAC, and dbSNP databases. Further details are reported in the online supplementary material.

TruSeq Custom Amplicon capturing and sequencing and analysis

TruSeq Custom Amplicon design ([Illumina.com](http://illumina.com)) consisted of 67 amplicons targeting 14642 bp, and generating a 250bp paired-end sequence for each amplicon. The uniquely tagged amplicon library was used for cluster generation and sequencing. The sequence data was processed with RTA (v1.17.28) and analyzed using the TruSeq Amplicon application within the MiSeq Reporter software (v2.2.0). These alignments were fed in BAM format into bam2mpg (<http://research.nhgri.nih.gov/software/bam2mpg/index.shtml>), which calls genotypes at all covered positions using a probabilistic Bayesian algorithm (Most Probable Genotype, or MPG). Genotypes with MPG score ≥ 10 were considered reliable (Teer, et al., 2010). The variants were annotated using Annovar (http://www.openbioinformatics.org/annovar/annovar_gene.html). Variants detected in dbSNP (version 137), 1000 Genomes, NHLBI 6500ESP, ExAC, and HGMD are annotated. All variants, genotypes, and annotations are viewed using VarSifter (<http://research.nhgri.nih.gov/software/VarSifter/>). To be considered highly reliable, genotypes should have an MPG score of at least 10 and a score to coverage ratio greater than 0.5. Further details are reported in the online supplementary data.

arrayCGH

A custom CGH 12×135K Array was designed using NimbleDesign (NimbleGen, Madison, WI). The design, hybridization, and data analysis were as described previously (Flynn, et al., 2014).

SNP array

Genotyping was performed using the HumanOmniExpress (750,000 SNPs), HumanOmniExpressExome (1M SNPs), or HumanOmni2.5 (2.5M SNPs) beadchips. DNA analysis using these BeadChip kits (Illumina, Inc.) was as described earlier (Sood, et al., 2016).

MLPA

The SALSA MLPA probe mixes for *FANCA*, P031-B1 mix 1 and P032-B1 mix 2, were used to detect large deletions in *FANCA* (MRC-Holland). Seventy-five to 100 ng of genomic DNA in a volume of 5 μ l was used for each probe mix reaction. The denaturation, hybridization, ligation, and PCR reactions were performed using the protocol provided by MRC-Holland.

Cell lines

Cell lines were derived from individuals who had been diagnosed with Fanconi anemia and were registered in IFAR. BJ cells are normal foreskin fibroblasts obtained from ATCC.

RA3087 cells are fibroblasts derived from a Fanconi anemia patient with biallelic deletions in *FANCA* (c.793-?_4368+? del; delEx9-43), resulting in no expression of FANCA protein (Kim, et al., 2013).

Vectors and mutagenesis

The *FANCA* gene was amplified by polymerase chain reaction using LFAPEG (gift from Dr. Detlev Shindler) and cloned into pDONR223. *FANCA* mutants were generated with the QuikChange II XL Site-Directed Mutagenesis kit (Agilent Technologies) using mutagenesis primers. Mutants were recombined into pHAGE vector using LR clonase (Invitrogen). HPV E6 and E7 and hTERT expression plasmids were described previously (Kim, et al., 2011).

Antibodies

Anti-HA antibody (Covance, MMS-101R) was used for immunofluorescence and western blotting. Anti-GAPDH (Santa Cruz, 25778), and anti-FANCA (Bethyl, A301-980A) antibodies were used for western blotting.

Immunofluorescence

Immunofluorescence was performed on human fibroblasts as described previously (Smogorzewska, et al., 2007). Cells were grown on coverslips and washed twice with PBS. Cells were pre-extracted with 0.1% Triton X-100 in PBS, fixed in 3.7% formaldehyde and permeabilized with NP-40. Cells were washed with PBS twice and blocked with PBG (0.2% [w/v] cold fish gelatin, 0.5% [w/v] BSA in PBS) for 20 minutes. Coverslips were incubated with relevant antibodies overnight at 4°C. Cells were washed three times with PBG and incubated with the appropriate secondary antibody. After three additional washes in PBG the coverslips were embedded in Vectashield (Vector Laboratories) supplemented with DAPI.

MMC sensitivity assay

2.5×10^4 cells were plated in each well of a six-well plate in triplicate. Twenty-four hours later, MMC (Sigma, M4287) was added at final concentrations from 0–75 nM. After 3–4 days in culture, cells were split at appropriate dilutions and after total of 6–8 days in culture, cell numbers were counted with a Z2 Coulter counter (Beckman Coulter). The cell numbers at each dose of drug were divided by the cell number in the untreated sample to calculate percent survival.

RESULTS

High throughput FANCA screening of 216 FA families

We screened 216 families diagnosed with FA, but whose associated pathogenic variants were unknown. The cohort included 87 families assigned to the FA-A group based on prior complementation tests and 129 families that had not been previously complementation tested (Figure 1). Using a variety of approaches, as described in Methods and Table 1, we identified two likely pathogenic variants in 79 of the 87 families known to be FA-A (Figure 1). In seven of the 87 families, only one putative pathogenic variant was identified. In one family, no *FANCA* mutations were found, but, rather, two putative pathogenic *FANCE*

variants were observed. In 129 families without complementation testing, we found two mutations in 80 families and one mutation in 17 families. No *FANCA* mutations were identified in 32 families and they most likely represent non-FA-A complementation groups. In total, we were able to identify two *FANCA* pathogenic variants in 159 FA families (Figure 1).

Identification and characterization of two variants in each of 159 *FANCA* families

Characteristics of *FANCA* variants identified in the 159 FA families are presented in Table 2 and Supp. Table S2. Of the 318 variant alleles detected, 57 were large deletions and 261 were sequence variants that included 130 distinct variants, illustrating the extensive heterogeneity observed among *FANCA* pathogenic variants. Of these 130 sequence variants, 82 were present in the Fanconi Anemia Mutation Database (FAMutDB) (<http://www.rockefeller.edu/fanconi/>). These known variants accounted for 198 alleles, while 48 were novel, contributing to the remaining 63 alleles (Table 3). The 82 known variants included 24 indel, 22 missense, 18 nonsense, 16 splice, and two predicted to be synonymous but affect splicing. The 48 novel variants included 12 indel, 10 missense, 9 nonsense, and 17 splice variants. We report more novel than known splice mutations (17 vs 16), and 37% of the identified sequence variants are novel. Strikingly, 46 known and 35 novel mutations appear only once, and the highest number of alleles seen for any novel variant is three, demonstrating a high prevalence of private *FANCA* variants in Fanconi anemia patients.

Indel, nonsense, or splice variants affecting the canonical splice donor GT or acceptor AG, were considered pathogenic based on the type of variant and their presumed effect on the expression of *FANCA* RNA/protein (Table 3). Among the previously identified missense variants, pathogenicity of some have been shown through functional analysis of engineered missense expression constructs or from RNA analysis of patient cell lines, and all by prediction programs (Supp. Table S3).

Evaluation of novel missense variants by functional assay

We identified 18 novel rare missense variants that were absent in dbSNP and FAMutDB. To confirm their pathogenicity, we performed functional analysis of all 18 variants by expressing a mutagenized *FANCA*-expression construct in a *FANCA* null cell line (RA3087), and assessing expression, cellular localization, and ability to complement survival of the cells after MMC treatment (Figure 2). Ten of the 18 variants did not complement the *FANCA* null cells in this assay and are most likely pathogenic: p.R435H, p.R435L, p.Y448C, p.D944A, p.Q436R, p.P497S, p.L946P, p.L1181P, p.W1224S, p.M1360I (Figure 2A). All of the variants that did not complement MMC sensitivity were well-expressed but were excluded from the nucleus (Figure 2B and C). Sensitivity assay revealed that 6/18 candidates tested were most likely non-pathogenic as missense variants: p.Q174H, p.V230I, p.K522R, p.V697I, p.Q1099H, and p.V1418M (Figure 2A). They were expressed well and were localized to the nucleus (Figure 2B and C). The p.D953E and p.R756C variants did not fully complement *FANCA*-null cells. The p.D953E was expressed well and localized to the nucleus so we consider it to be most likely non-pathogenic. The p.R756C variant was expressed very poorly and although it was localized to the nucleus, staining was weak. Considering that the patient was homozygous for this variant and the patient's cells

were clinically complemented by the FANCA cDNA, we consider this variant to be most likely pathogenic.

Aberrant splicing caused by two nonsynonymous variants and an intronic indel

FANCA variants, c.522G>C (predicted p.Q174H) and c.1565A>G (predicted p.K522R), were identified as second alleles in two patients (FAM117 and FAM89 respectively) who were assigned to the FA-A group by complementation. As described above, the p.Q174H and p.K522R variants could complement MMC sensitivity in *FANCA*^{-/-} cells. Our exhaustive search had not produced any other variants in *FANCA*, which prompted us to evaluate the patients' RNA for potential splicing effects. RNA analysis from the patient cell line with the novel variant c.522G>C (exon 5) reveals that the variant allele causes exon 5 skipping (Figure 3A). RT-PCR products spanning the variant site in exon 5 were cloned and sequenced: 9/12 colonies contained the presence of exon 5, while the other three lacked it. The exon 5 retaining transcripts carried the WT p.Q174 allele, so we conclude that the c.522G>C variant results in exon skipping. The transcripts with exon 5 skipping are expected to result in an in-frame deletion of 32 amino acids. Similarly, RNA analysis from the patient cell line with the novel variant c.1565A>G (exon 16), which is expected to cause a nonsynonymous change (p.K522R), also revealed skipping, in this case, of exon 16 (Figure 3B). The transcripts with exon 16 skipped are expected to result in an in-frame deletion of 32 amino acids in the FANCA protein.

A third patient (FAM8) known to belong to the FANCA group by complementation, but for whom the second pathogenic variant was unknown, was found to carry a novel variant in intron 27, c.2602-9_c.2602-8delCT. RNA from this patient's cell line was tested and found to affect splicing, resulting in aberrant transcripts skipping either exons 28–30 or exon 29 alone (Figure 3C). A combination of DNA and RNA sequencing may be necessary to identify all FANCA alleles.

Consequence of FANCA mutations on FANCA protein expression

We studied FANCA protein expression in a subset of patients from whom cell lines had been created (Figure 4). As anticipated, the carriers whose two mutations belong to nonsense, indel, or large deletions showed none or low expression of FANCA protein (RA2227/FAM95, RA2357/FAM100, RA2448/FAM102, RA1535/FAM33, and RA1759/FAM46). In addition, RA2250/FAM96, with a large deletion and a codon 1 mutation, also lacks FANCA protein expression. Among the patient cell lines with FANCA expression similar to WT, RA1038 is likely to be a non-FANCA group family as the only known FANCA variant, p.Q1099H, is non-pathogenic as determined by a functional assay. RA2572/FAM89 was shown to belong to FA-A group by complementation testing, however the robust FANCA signal is reflective of the two missense variants, with one of the variants [c.1565A>G (p.K522R)] resulting in the in-frame deletion of 32 aa as described above. The patient cell line, RA2689/FAM117, which carries the exon skipping variant in exon 5 [c.522G>C (p.Q174H)] shows a weaker expression, as the second variant results in a stop codon. The other cell lines expressing a low level of FANCA, RA1760/FAM25, RA2408/FAM1, and RA2259/FAM97, each carry a missense variant and a nonsense variant or a large deletion. RA2531 and RA475 carry one *FANCA* pathogenic mutation while the other pathogenic

variant is yet to be identified. There remains, however, a possibility that these two cell lines may represent non-FANCA complementation groups.

DISCUSSION

Several studies have reported mutations in the FANCA gene since the discovery of *FANCA* in 1996 (Apostolou, et al., 1996; Lo Ten Foe, et al., 1996). This report includes the largest number of FANCA patients, and identifies both mutations in 159 patients. As the precise characterization of both disease-causing mutations in an FA patient is a prerequisite for genetic counseling and reliable correlation with the patient phenotype, multipronged approaches are required for molecular diagnosis.

A distinct set of *FANCA* mutations in nearly all FANCA families studied

A distinct set of mutations accounted for the disease in 152 out of 159 FA families, illustrating the exceptional genetic heterogeneity of *FANCA* mutations in FA patients (Table 2). Only two families had an identical pair of compound heterozygous mutations (FAM10 and FAM8) and in fact, we subsequently learned that they were related. Also, two distinct homozygous mutations were present in two and three families each: c.2851C>T in FAM141 and FAM129, and c.1827-1G>A in FAM23, FAM138 and FAM107. The complex heterogeneity is also apparent from the fact that we identified 48 distinct and novel variants accounting for 63 mutant alleles, of which none were homozygous, and no more than three families carried the same novel variant (Table 3). However, our study adds seven and eleven more patients respectively to the list of two most often encountered *FANCA* variants in the FAMutDB, c.3788_3790delTCT (present 261 times), c.1115_1118delTTGG (present 67 times). We also add eleven more patients to the 19 known patients who carry the p.T1131A mutation, and four more to the known c.3624C>T list. The latter is predicted to be a synonymous variant, (p.S1208S) but causes aberrant splicing (Ameziane, et al., 2008). Eleven of the novel pathogenic variants were present in the Exome Aggregate Consortium database (ExAC), but none were observed at a frequency greater than 0.009 (Table 3). Figure 5 illustrates the location of the 130 novel and known sequence variants, and their frequencies, accounting for 261 most likely pathogenic alleles in this analysis.

Our large-scale effort to screen *FANCA* variants identified two *FANCA* mutations in 159 FA families. Efforts are underway to identify the second pathogenic variant for 24 families for whom only one mutation has been identified. However, for many of these, we may not identify additional *FANCA* variants as we may not have the appropriate DNA/cell line, the variant could have been lost from the cell line due to hematopoietic stem cell mosaicism, which is common in FA patients, or causative mutations could be present in a different FA gene despite the patient carrying a *FANCA* variant. Thirty-two families with no *FANCA* mutations (Table 1) are likely to be non-FANCA families and their group can now be determined by methods described earlier (Chandrasekharappa, et al., 2013).

Large deletions account for nearly a quarter of all pathogenic variants in *FANCA*

Of the 318 mutant alleles, we observed 57 deletions in this study. We have reported previously ten FANCA families with biallelic deletions (Flynn, et al., 2014), and ten other

families with a total of four deletions (Chandrasekharappa, et al., 2013; Donovan, et al., 2016), accounting for 24 deletions among 20 *FANCA* families. Thus, large deletions (81/358 alleles) appear to account for 23% of all *FANCA* mutations in our cohort. The frequency of large deletions compared to other types of *FANCA* pathogenic variants from this study is similar to three of the four previous reports 15% (Castella, et al., 2011), 20% (Moghrabi, et al., 2009), 25% (Ameziane, et al., 2008) and 40% (Morgan, et al., 1999). The lower end of the range is based on 20 large deletions in 65 individuals (130 mutations) that included 12 patients carrying a founder nonsense mutation. The upper end of 40% is the earliest study of large deletions and is based on 17 deletions in 26 individuals but only 43 mutations identified and thus may be an overestimate. We provide the precise breakpoint coordinates as determined by cloning and sequencing for 2/3 of the deletions (Table 2). Precise deletion coordinates allow recognition of shared deletions among FA families. Of the 37 deletions with known precise deletion intervals, four deletions are shared by a total of 11 families, and the rest are distinct. Of the four shared deletions, one is found in a set of two related families, the only two families with the same two compound heterozygous *FANCA* mutations. Our extensive screening for deletion variants allowed us to clearly distinguish genuine homozygous variants from those sequence variants that appeared homozygous due to an overlapping deletion.

High throughput and combinations of technologies ensure comprehensive mutation analysis

We employed high throughput sequencing methodologies for variant detection. Although the depth and the extent of coverage may have varied to some extent among methods, each method was competent to recognize the variants in the regions covered, with a few exceptions. One issue we encountered is the overlap of an individual SNP or a pathogenic intronic variant with a PCR primer, and subsequent lack of PCR product. The pathogenic variants, p.R1117G (c.3349A>G) in exon 34 in two patients, and c.2778+83C>G in intron 28 of another patient were, thus, initially missed by Sanger sequencing. Similarly, a SNP (rs7190823, 89866043T>C hg19) and an indel variant (c.1734_1739delCTACGT) in two families prevented binding of MLPA probes resulting in calling the respective exon nine and 19 deletions incorrectly. Application of multiple methodologies, next-gen sequencing, and aCGH helped sort out the respective sequence and deletion variations. Our experience highlights that extra caution is warranted to recognize technical issues that provide false positive or false negative results.

Evaluation of pathogenicity of nonsynonymous and intronic indel variants

The ability of a missense variant to complement the growth deficiency of a null cell is a standard and reliable assay to determine pathogenicity (or lack thereof). This functional evaluation helped identify eleven new pathogenic variants. However, two of the seven variants (c.522G>C and c.1565A>G) that behaved as wild type in functional assays, turned out to instead affect splicing. In both cases, the consensus AG exonic sequences at the splice junction where the U1 snRNA binds (AG|gtragt) are altered, CAG|gt>CAC|gt in exon 5 and AAG|gt>AGG|gt in exon 16, and thus, may prevent U1snRNP from performing its normal function (Cartegni, et al., 2002). We had observed earlier that altering the G in the wobble position (last) of the same codon in exon 16 in another patient (c.1566G>A; pK522K)

produced exon 16 skipping as well (Chandrasekharappa, et al., 2013). We also report an intronic variant, c.2602-9_c.2602-8delCT, located eight nucleotides from the splice junction which causes splicing errors. Recognition of pathogenic synonymous and nonsynonymous *FANCA* variants, and the resulting in-frame deletions in the transcript, will allow the effect of these variations on *FANCA* protein function and disease phenotype to be studied.

Pathogenic nonsynonymous variants are retained in the cytoplasm

All nonsynonymous variants that were shown to be pathogenic by functional analysis resided in the cytoplasm and, conversely, those categorized as nonpathogenic localized to the nucleus. Nuclear localization of *FANCA* protein is critical for normal biological function (Naf, et al., 1998). Impairment of nuclear localization by pathogenic missense variants has been recognized previously by several investigators (Adachi, et al., 2002; Castella, et al., 2011; Garcia-Higuera, et al., 2000). The mechanism causing so many deleterious variants to localize erroneously in the cytoplasm is unexplored; however, a likely explanation is that the resultant proteins are inappropriately folded. Exploring ways to correct the impaired subcellular localization may lead to novel therapeutic approaches. A recent report indicates that relative levels of interaction between the *FANCA* mutant proteins with chaperonins HSP90 and HSP70 can directly contribute to the clinical course of the disease (Karras, et al., 2017). It remains to be tested if *FANCA* variants can localize appropriately to the nucleus following upregulation of HSP90 activity.

Final Remarks

Identification of the precise molecular nature of variants at the genomic level is a prerequisite for efficient clinical management of FA patients, genetic counseling for families, understanding the disease process to explore novel therapeutic options, and discovery of genotype-phenotype correlations with prognostic implications. The enrollment of these patients in registries, such as the IFAR, enables us to obtain the patient-derived cell lines and RNA necessary for functional analysis of FA variants, thereby confirming their pathogenicity, or lack thereof, which is especially important for genes like *FANCA* which exhibit complex heterogeneity. This ensures the correct diagnosis is given, and allows for accurate carrier and prenatal testing.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the many physicians who have registered their FA patients in the IFAR, without whom this research could not have been possible, and all the FA family members for their participation in the study. We thank Dr. Elaine Ostrander for critical reading of the manuscript. We thank Uma Veturi for technical assistance. We acknowledge the support from the intramural research program of the National Human Genome Research Institute, National Institutes of Health. This work was also supported by NIH RO1 HL120922 (AS) and # UL1TR001866 from the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health (NIH) Clinical and Translational Science Award (CTSA) program. AS is a Howard Hughes Faculty Scholar.

References

- Adachi D, Oda T, Yagasaki H, Nakasato K, Taniguchi T, D'Andrea AD, Asano S, Yamashita T. Heterogeneous activation of the Fanconi anemia pathway by patient-derived FANCA mutants. *Hum Mol Genet.* 2002; 11(25):3125–34. [PubMed: 12444097]
- Alter BP, Giri N. Thinking of VACTERL-H? Rule out Fanconi Anemia according to PHENOS. *Am J Med Genet A.* 2016; 170(6):1520–4. [PubMed: 27028275]
- Alter BP, Giri N, Savage SA, Peters JA, Loud JT, Leathwood L, Carr AG, Greene MH, Rosenberg PS. Malignancies and survival patterns in the National Cancer Institute inherited bone marrow failure syndromes cohort study. *Br J Haematol.* 2010; 150(2):179–88. [PubMed: 20507306]
- Ameziane N, Errami A, Leveille F, Fontaine C, de Vries Y, van Spaendonck RM, de Winter JP, Pals G, Joenje H. Genetic subtyping of Fanconi anemia by comprehensive mutation screening. *Hum Mutat.* 2008; 29(1):159–66. [PubMed: 17924555]
- Amouri A, Talmoudi F, Messaoud O, d'Enghien CD, Rekaya MB, Allegui I, Azaiez H, Kefi R, Abdelhak A, Meseddi SH, Torjemane L, Ouederni M, Mellouli F, Abid HB, Aissaoui L, Bejaoui M, Othmen TB, Lyonnet DS, Soulier J, Hachicha M, Dellagi K, Abdelhak S, Fanconi T. High frequency of exon 15 deletion in the FANCA gene in Tunisian patients affected with Fanconi anemia disease: implication for diagnosis. *Mol Genet Genomic Med.* 2014; 2(2):160–5. [PubMed: 24689079]
- Apostolou S, Whitmore SA, Crawford J, Lennon G, Sutherland GR, Callen DF, Ianzano L, Savino M, D'Apolito M, Notarangelo A, Memeo E, Piemontese MR, Zelante L, Savoia A, Gibson RA, Tipping AJ, Morgan NV, Hassock S, Jansen S, de Ravel TJ, Van Berkell S, Pronk JC, Easton DF, Mathew CG, Levran O, Verlander PC, Batish SD, Erlich T, Auerbach AD, Cleton-Jansen AM, Moerland EW, Cornelisse CJ, Doggett NA, Deaven LL, Moyzis RK. Positional cloning of the Fanconi anaemia group A gene. *Nat Genet.* 1996; 14(3):324–8. [PubMed: 8896564]
- Auerbach AD. Fanconi anemia and its diagnosis. *Mutation research.* 2009; 668(1–2):4–10. [PubMed: 19622403]
- Biesecker LG, Mullikin JC, Facio FM, Turner C, Cherukuri PF, Blakesley RW, Bouffard GG, Chines PS, Cruz P, Hansen NF, Teer JK, Maskeri B, Young AC, NISC Comparative Sequencing Program. Manolio TA, Wilson AF, Finkel T, Hwang P, Arai A, Remaley AT, Sachdev V, Shamburek R, Cannon RO, Green ED. The ClinSeq Project: piloting large-scale genome sequencing for research in genomic medicine. *Genome Res.* 2009; 19(9):1665–74. [PubMed: 19602640]
- Boersma V, Moatti N, Segura-Bayona S, Peuscher MH, van der Torre J, Wevers BA, Orthwein A, Durocher D, Jacobs JJ. MAD2L2 controls DNA repair at telomeres and DNA breaks by inhibiting 5' end resection. *Nature.* 2015; 521(7553):537–40. [PubMed: 25799990]
- Callen E, Casado JA, Tischkowitz MD, Bueren JA, Creus A, Marcos R, Dasi A, Estella JM, Munoz A, Ortega JJ, de Winter J, Joenje H, Schindler D, Hanenberg H, Hodgson SV, Mathew CG, Surrallés J. A common founder mutation in FANCA underlies the world's highest prevalence of Fanconi anemia in Gypsy families from Spain. *Blood.* 2005; 105(5):1946–9. [PubMed: 15522956]
- Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet.* 2002; 3(4):285–98. [PubMed: 11967553]
- Castella M, Pujol R, Callen E, Trujillo JP, Casado JA, Gille H, Lach FP, Auerbach AD, Schindler D, Benítez J, Porto B, Ferro T, Muñoz A, Sevilla J, Madero L, Cela E, Beléndez C, de Heredia CD, Olivé T, de Toledo JS, Badell I, Torrent M, Estella J, Dasí A, Rodríguez-Villa A, Gómez P, Barbot J, Tapia M, Molinés A, Figuera A, Bueren JA, Surrallés J. Origin, functional role, and clinical impact of Fanconi anemia FANCA mutations. *Blood.* 2011; 117(14):3759–69. [PubMed: 21273304]
- Centra M, Memeo E, d'Apolito M, Savino M, Ianzano L, Notarangelo A, Liu J, Doggett NA, Zelante L, Savoia A. Fine exon-intron structure of the Fanconi anemia group A (FAA) gene and characterization of two genomic deletions. *Genomics.* 1998; 51(3):463–7. [PubMed: 9721219]
- Chandrasekharappa SC, Lach FP, Kimble DC, Kamat A, Teer JK, Donovan FX, Flynn E, Sen SK, Thongthip S, Sanborn E, Smogorzewska A, Auerbach AD, Ostrander EA. NISC Comparative Sequencing Program. Massively parallel sequencing, aCGH, and RNA-Seq technologies provide a comprehensive molecular diagnosis of Fanconi anemia. *Blood.* 2013; 121(22):e138–48. [PubMed: 23613520]

- Donovan FX, Kimble DC, Kim Y, Lach FP, Harper U, Kamat A, Jones M, Sanborn EM, Tryon R, Wagner JE, MacMillan ML, Ostrander EA, Auerbach AD, Smogorzewska A, Chandrasekharappa SC. Paternal or Maternal Uniparental Disomy of Chromosome 16 Resulting in Homozygosity of a Mutant Allele Causes Fanconi Anemia. *Hum Mutat.* 2016; 37(5):465–8. [PubMed: 26841305]
- Flynn EK, Kamat A, Lach FP, Donovan FX, Kimble DC, Narisu N, Sanborn E, Boulad F, Davies SM, Gillio AP 3rd, Harris RE, MacMillan ML, Wagner JE, Smogorzewska A, Auerbach AD, Ostrander EA, Chandrasekharappa SC. Comprehensive analysis of pathogenic deletion variants in Fanconi anemia genes. *Hum Mutat.* 2014; 35(11):1342–53. [PubMed: 25168418]
- Garcia-Higuera I, Kuang Y, Denham J, D'Andrea AD. The fanconi anemia proteins FANCA and FANCG stabilize each other and promote the nuclear accumulation of the Fanconi anemia complex. *Blood.* 2000; 96(9):3224–30. [PubMed: 11050007]
- Karras GI, Yi S, Sahni N, Fischer M, Xie J, Vidal M, D'Andrea AD, Whitesell L, Lindquist S. HSP90 Shapes the Consequences of Human Genetic Variation. *Cell.* 2017; 168(5):856–866 e12. [PubMed: 28215707]
- Kim Y, Lach FP, Desetty R, Hanenberg H, Auerbach AD, Smogorzewska A. Mutations of the SLX4 gene in Fanconi anemia. *Nat Genet.* 2011; 43(2):142–6. [PubMed: 21240275]
- Kim Y, Spitz GS, Veturi U, Lach FP, Auerbach AD, Smogorzewska A. Regulation of multiple DNA repair pathways by the Fanconi anemia protein SLX4. *Blood.* 2013; 121(1):54–63. [PubMed: 23093618]
- Knies K, Inano S, Ramirez MJ, Ishiai M, Surrallés J, Takata M, Schindler D. Biallelic mutations in the ubiquitin ligase RFWF3 cause Fanconi anemia. *J Clin Invest.* 2017; 127(8):3013–3027. [PubMed: 28691929]
- Kottemann MC, Smogorzewska A. Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. *Nature.* 2013; 493(7432):356–63. [PubMed: 23325218]
- Kutler DI, Auerbach AD, Satagopan J, Giampietro PF, Batish SD, Huvos AG, Goberdhan A, Shah JP, Singh B. High incidence of head and neck squamous cell carcinoma in patients with Fanconi anemia. *Archives of otolaryngology--head & neck surgery.* 2003a; 129(1):106–12. [PubMed: 12525204]
- Kutler DI, Singh B, Satagopan J, Batish SD, Berwick M, Giampietro PF, Hanenberg H, Auerbach AD. A 20-year perspective on the International Fanconi Anemia Registry (IFAR). *Blood.* 2003b; 101(4):1249–56. [PubMed: 12393516]
- Levrán O, Doggett NA, Auerbach AD. Identification of Alu-mediated deletions in the Fanconi anemia gene FAA. *Human mutation.* 1998; 12(3):145–52. [PubMed: 9711872]
- Levrán O, Erlich T, Magdalena N, Gregory JJ, Batish SD, Verlander PC, Auerbach AD. Sequence variation in the Fanconi anemia gene FAA. *Proc Natl Acad Sci U S A.* 1997; 94(24):13051–6. [PubMed: 9371798]
- Lo Ten Foe JR, Rooimans MA, Bosnoyan-Collins L, Alon N, Wijker M, Parker L, Lightfoot J, Carreau M, Callen DF, Savoia A, Cheng NC, van Berkel CG, Strunk MH, Gille JJ, Pals G, Kruyt FA, Pronk JC, Arwert F, Buchwald M, Joenje H. Expression cloning of a cDNA for the major Fanconi anaemia gene, FAA. *Nat Genet.* 1996; 14(3):320–3. [PubMed: 8896563]
- Mamrak NE, Shimamura A, Howlett NG. Recent discoveries in the molecular pathogenesis of the inherited bone marrow failure syndrome Fanconi anemia. *Blood Rev.* 2016
- Moghrabi NN, Johnson MA, Yoshitomi MJ, Zhu X, Al-Dhalimy MJ, Olson SB, Grompe M, Richards CS. Validation of Fanconi anemia complementation Group A assignment using molecular analysis. *Genetics in medicine : official journal of the American College of Medical Genetics.* 2009; 11(3):183–92. [PubMed: 19367192]
- Morgan NV, Tipping AJ, Joenje H, Mathew CG. High frequency of large intragenic deletions in the Fanconi anemia group A gene. *American journal of human genetics.* 1999; 65(5):1330–41. [PubMed: 10521298]
- Naf D, Kupfer GM, Suliman A, Lambert K, D'Andrea AD. Functional activity of the fanconi anemia protein FAA requires FAC binding and nuclear localization. *Mol Cell Biol.* 1998; 18(10):5952–60. [PubMed: 9742112]
- Rickman KA, Lach FP, Abhyankar A, Donovan FX, Sanborn EM, Kennedy JA, Sougnez C, Gabriel SB, Elemento O, Chandrasekharappa SC, Schindler D, Auerbach AD, Smogorzewska A.

Deficiency of UBE2T, the E2 Ubiquitin Ligase Necessary for FANCD2 and FANCI Ubiquitination, Causes FA-T Subtype of Fanconi Anemia. *Cell Rep.* 2015; 12(1):35–41. [PubMed: 26119737]

Rosenberg PS, Greene MH, Alter BP. Cancer incidence in persons with Fanconi anemia. *Blood.* 2003; 101(3):822–6. [PubMed: 12393424]

Schlacher K, Christ N, Siaud N, Egashira A, Wu H, Jasin M. Double-strand break repair-independent role for BRCA2 in blocking stalled replication fork degradation by MRE11. *Cell.* 2011; 145(4): 529–42. [PubMed: 21565612]

Schlacher K, Wu H, Jasin M. A distinct replication fork protection pathway connects Fanconi anemia tumor suppressors to RAD51-BRCA1/2. *Cancer Cell.* 2012; 22(1):106–16. [PubMed: 22789542]

Skagen K, Hansen JF, Olesen KH. Closed mitral valvulotomy after the age of fifty. *Scand J Thorac Cardiovasc Surg.* 1978; 12(2):85–9. [PubMed: 715403]

Smogorzewska A, Matsuoka S, Vinciguerra P, McDonald ER 3rd, Hurov KE, Luo J, Ballif BA, Gygi SP, Hofmann K, D'Andrea AD, Elledge SJ. Identification of the FANCI protein, a monoubiquitinated FANCD2 paralog required for DNA repair. *Cell.* 2007; 129(2):289–301. [PubMed: 17412408]

Sood R, Hansen NF, Donovan FX, Carrington B, Bucci D, Maskeri B, Young A, Trivedi NS, Kohlschmidt J, Stone RM, Caligiuri MA, Chandrasekharappa SC, Marcucci G, Mullikin JC, Bloomfield CD, Liu P. Somatic mutational landscape of AML with inv(16) or t(8;21) identifies patterns of clonal evolution in relapse leukemia. *Leukemia.* 2016; 30(2):501–4. [PubMed: 26139325]

Teer JK, Bonnycastle LL, Chines PS, Hansen NF, Aoyama N, Swift AJ, Abaan HO, Albert TJ, Program NCS, Margulies EH, Green ED, Collins FS, Mullikin JC, Biesecker LG. Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing. *Genome Res.* 2010; 20(10):1420–31. [PubMed: 20810667]

Tipping AJ, Pearson T, Morgan NV, Gibson RA, Kuyt LP, Havenga C, Gluckman E, Joenje H, de Ravel T, Jansen S, Mathew CG. Molecular and genealogical evidence for a founder effect in Fanconi anemia families of the Afrikaner population of South Africa. *Proc Natl Acad Sci U S A.* 2001; 98(10):5734–9. [PubMed: 11344308]

Wang AT, Smogorzewska A. SnapShot: Fanconi anemia and associated proteins. *Cell.* 2015; 160(1–2): 354–354 e1. [PubMed: 25594185]

West SC, Blanco MG, Chan YW, Matos J, Sarbajna S, Wyatt HD. Resolution of Recombination Intermediates: Mechanisms and Regulation. *Cold Spring Harb Symp Quant Biol.* 2015; 80:103–9. [PubMed: 26370409]

Xu G, Chapman JR, Brandsma I, Yuan J, Mistrik M, Bouwman P, Bartkova J, Gogola E, Warmerdam D, Barazas M, Jaspers JE, Watanabe K, Pieterse M, Kersbergen A, Sol W, Celie PHN, Schouten PC, van den Broek B, Salman A, Nieuwland M, de Rink I, de Ronde J, Jalink K, Boulton SJ, Chen J, van Gent DC, Bartek J, Jonkers J, Borst P, Rottenberg S. REV7 counteracts DNA double-strand break resection and affects PARP inhibition. *Nature.* 2015; 521(7553):541–4. [PubMed: 25799992]

Figure 1.

A schematic presentation of 216 FA families screened for *FANCA* mutations with their complementation and mutation status after completion of the study. The 40 patients with one *FANCA* mutation known is based on previously performed screening for a couple of very common FA mutations or deletions by MLPA.

B. Immunoblot with anti-HA antibody to test expression levels of HA-tagged FANCA variants which were transduced into FANCA-null RA3087/E6E7/hTERT fibroblasts. The FANCA-null RA3087 cell line was derived from a FA patient with biallelic FANCA deletions (delEx9-43).

C. Immunofluorescence to detect HA-tagged FANCA expressed in FANCA-null RA3087/E6E7/hTERT fibroblasts. Localization is categorized as being nuclear or cytoplasmic.

Figure 3. Splicing errors from analysis of RNA from patient cell lines harboring variants predicted to cause nonsynonymous and intronic indel variants

A. cDNA was prepared from the patient (FAM117) LCL cell line harboring the c.522G>C (predicted p.Q174H) mutation affecting the last base of exon 5. The cDNA region spanning exons 1–6 was amplified using primers located in exons 1 and 7 and sequenced. Image shows the RT-PCR products including a WT (686bp) and a fainter smaller product indicated with an asterisk (590bp). Sequence of the total cDNA (top right) indicates presence of the WT product and the product resulting from skipping of exon 5. A representative cloned cDNA with exon 5 skipping is shown (bottom right).

B. cDNA from the patient (FAM89) fibroblast cell line harboring the c.1565A>G, (predicted p.K522R) mutation affecting the second to last base of exon 16. The cDNA region spanning exons 15–22 was amplified using primers located on the flanking exons 14 and 23 and sequenced. Image shows RT-PCR products including a WT (777bp) and a smaller product indicated with an asterisk (681bp). Sequence of the total cDNA (top right) indicates presence of the WT product and the product resulting from skipping of exon 16. A representative cloned cDNA with exon 16 skipping is shown (bottom right).

C. cDNA prepared from patient (FAM8) LCL cell line carrying the c.2602-9_2602-8delCT mutation in the pyrimidine track of intron 27 was tested for splicing errors. RT-PCR products spanning exons 27–31 were generated using primers from the flanking exons 26 and 32. Sequencing the RT-PCR product revealed skipping of exons 28–30, which appears to be predominant (band indicated with an asterisk on the image and in the top sequence tracing; reverse sequence). RT-PCR and sequencing of the region spanning exons 23–29 region revealed exon 29 skipping (bottom sequence tracing; reverse sequence); the exon 28–30 skipping is not seen in this sequence due to reverse primer binding within exon 30.

Figure 4. Analysis of FANCA expression in FANCA patient cell lines

Expression of the endogenous FANCA was analyzed in patient-derived LCL (A) and fibroblasts (B). Cells were grown in the presence or absence of 1 μ M mitomycin C (MMC) for 24 hours. A cell line from a normal individual (WT) is a positive control. RA3087 fibroblasts derived from a patient with known biallelic null mutations in *FANCA* is a negative control. Cell line IDs (RA numbers), and family IDs (FAM numbers) are on top and the mutations present in the cell lines are indicated below the image. For the three patients with no FAM ID, only one *FANCA* variant is known and the presence of a second mutation has not been determined (ND). The variants denoted with * cause in-frame exon skipping.

Figure 5. Mapping of all the coding and splicing variants identified in this study of *FANCA* and its encoded protein

A. *FANCA* protein diagram displaying the distribution of mutations. Novel mutations extend upward and known mutations extend downward. Nuclear localization signal (NLS) at aa 18–34; FANCG binding domain (FANCG) at aa 18–29; Region of reported interaction w/ BRCA1 at aa 740–1083; FAAP20 binding domain (FAAP20) at aa 1095–1200; phosphorylation sites, serine residues at aa 849, 850, 858, and 1449 (<http://atlasgeneticsoncology.org/Genes/FA1ID102.html>).

B. *FANCA* diagram displaying the distribution of intronic splicing mutations. Exons are represented by vertical lines and not drawn to scale. Novel splicing mutations extend upwards and the known splicing mutations extend downwards.

Number of patients carrying two pathogenic *FANCA* mutations and the methods employed to discover each mutation

Table 1

		Mutation 2						Total # of patients
		Sanger	NGS (TSCA)	NGS (TC)	aCGH	SNP array	MLPA	
Mutation 1	Sanger	80	2	2	47	6	3	140
	NGS (TSCA)		18			1		19
								159

NGS = NextGen Sequencing; TSCA = TruSeq Custom Amplicon; TC = Targeted Capture; MLPA = Multiplex Ligation-dependent Probe Amplification

Table 2

FA families with two pathogenic *FANCA* mutations

FAM ID	Mutation 1			Mutation 2		
	Complementation	cDNA#	Protein	cDNA#	Protein	
1	No	c.4198C>T	p.R1400C	c.2749C>T	p.R917*	
2	Yes	c.862G>T	p.E288*	c.597-I>G>C		
3	Yes	c.3338A>T	p.N1113I	large deletion		
4	Yes	c.1827-I>G>A		large deletion		
5	No	c.862G>T	p.E288*	c.1303C>T	p.R435C	
6	Yes	c.154C>T	p.R52*	c.190-2A>T		
7	No	c.1827-I>G>A		large deletion		
8	Yes	c.2602-9_2602-8delCT		large deletion		
9	No	c.14G>A	p.W5*	large deletion		
10	No	c.2602-9_2602-8delCT		large deletion		
11	No	c.1285_1288dupACTG	p.A430Dfs*56	c.1285_1288dupACTG	p.A430Dfs*56	
12	No	c.2535_2536delCT	p.C846Qfs*20	c.2535_2536delCT	p.C846Qfs*20	
13	No	c.2638C>T	p.R880*	c.710-10G>A		
14	No	c.1827-I>G>A		c.2840C>G	p.S947*	
15	No	c.3391A>G	p.T1131A	large deletion		
16	Yes	c.862G>T	p.E288*	c.862G>T	p.E288*	
17	No	c.1734_1739delCTACGT	p.Y578*	large deletion		
18	No	c.2837T>C	p.L946P	large deletion		
19	No	c.3971C>T	p.P1324L	large deletion		
20	No	c.2708G>A	p.W903*	c.3931_3932delAG	p.S1311*	
21	No	c.2641C>T	p.Q881*	large deletion		
22	No	c.1827-I>G>A		c.1115_1118delTTGG	p.V372Afs*42	
23	No	c.1827-I>G>A		c.1827-I>G>A		
24	No	c.1303C>T	p.R435C	large deletion		
25	No	c.1304G>A	p.R435H	c.2854C>T	p.Q952*	
26	No	c.2738A>C	p.H913P	c.4017_4021delCTCCT		
27	No	c.3391A>G	p.T1131A	c.522+1G>T		

Mutation 1				Mutation 2	
FAM ID	Complementation	cDNA#	Protein	cDNA#	Protein
28	No	c.3490C>T	p.P1164S	large deletion	
29	No	c.2T>C	p.MIT	large deletion	
30	No	c.1827-1G>A		c.2778+83C>G	
31	No	c.3788_3790delTCT	p.F1263del	large deletion	
32	No	c.1159C>T	p.Q387*	large deletion	
33	No	c.3494T>G	p.L1165*	c.3558dupG	p.R1187Efs*28
34	Yes	c.2639G>A	p.R880Q	c.597-1G>C	
35	Yes	c.2606A>C	p.Q869P	large deletion	
36	No	c.710-2A>C		c.2730_2731delCT	p.W911Dfs*31
37	Yes	c.4015delC	p.L1339Sfs*24	c.2839dupT	p.S947Ffs*4
38	No	c.3349-1G>A		large deletion	
39	No	c.3788_3790delTCT	p.F1263del	c.4261-2A>C	
40	Yes	c.3634dupT	p.S1212Ffs*3	large deletion	
41	Yes	c.3391A>G	p.T1131A	c.811C>T	p.Q271*
42	Yes	c.3391A>G	p.T1131A	c.2639G>A	p.R880Q
43	No	c.597-1G>C		c.597-1G>C	
44	No	c.1489C>T	p.P497S	c.3762_3763insAG	p.E1255Rfs*12
45	No	c.1084-1G>A		large deletion	
46	No	c.1645C>T	p.Q549*	c.2812_2830dup	p.D944Gfs*5
47	No	c.1777-1G>C		large deletion	
48	No	c.2T>C	p.MIT	c.2606A>C	p.Q869P
49	No	c.752C>G	p.S251*	large deletion	
50	Yes	c.2266C>T	p.R756C	c.2266C>T	p.R756C
51	No	c.3791_3793delCCT	p.S1264del	c.3791_3793delCCT	p.S1264del
52	Yes	c.154C>T	p.R52*	c.2534T>C	p.L845P
53	No	c.2535_2536delCT	p.C846Qfs*20	c.3791_3793delCCT	p.S1264del
54	No	c.3828+1G>C		large deletion	
55	No	c.4198C>T	p.R1400C	c.3828+2dupT	
56	No	c.3490C>T	p.P1164S	large deletion	
57	No	c.3130C>T	p.Q1044*	c.1470+2T>C	

Mutation 1				Mutation 2		
FAM ID	Complementation	cDNA#	Protein	cDNA#	Protein	Protein
58	No	c.1073_1074delTTG	p.Y359Pfs*49	c.3624C>T	p.S1208S (affects splicing)	p.S1208S (affects splicing)
59	No	c.2233dupT	p.W745Lfs*49	large deletion		
60	No	c.1827-1G>A		c.3920delA		p.Q1307Rfs*2
61	No	c.3408+1G>C		large deletion		
62	Yes	c.862G>T	p.E288*	c.3934+2T>C		
63	No	c.3391A>G	p.T1131A	large deletion		p.E45*
64	Yes	c.1226-6_1226-2delAACTA		c.133G>T		
65	No	c.2638C>T	p.R880*	c.190-2A>T		
66	No	c.1115_1118delTTGG	p.V372Afs*42	c.65G>A	p.W22*	
67	No	c.2678G>A	p.W893*	c.2678G>A	p.W893*	
68	No	c.1777-2A>G		c.1115_1118delTTGG	p.V372Afs*42	
69	No	c.3813dupA	p.H1272Tfs*6	c.3788_3790delTCT	p.F1263del	
70	No	c.426+1G>A		large deletion		
71	Yes	c.1115_1118delTTGG	p.V372Afs*42	large deletion		
72	Yes	c.2021C>A	p.S674*	large deletion		
73	No	c.3520_3522delTTGG	p.W1174del	large deletion		
74	No	c.14G>A	p.W5*	large deletion		p.F954I
75	Yes	c.2708G>A	p.W903*	c.2860T>A	p.S1212Ffs*3	
76	No	c.1115_1118delTTGG	p.V372Afs*42	c.3634dupT	p.G972Afs*17	
77	Yes	c.1827-1G>A		c.2910delA	p.S947Ffs*4	
78	No	c.2852G>A	p.R951Q	large deletion		
79	No	c.3624C>T	p.S1208S (affects splicing)	large deletion		
80	Yes	c.2839dupT	p.S947Ffs*4	large deletion		
81	No	c.4261-19_4261-12delACCTGCTC		large deletion		
82	Yes	c.2812_2830dup	p.D944Gfs*5	large deletion		
83	No	c.862G>T	p.E288*	c.3788_3790delTCT	p.F1263del	
84	Yes	c.2524delT	p.S842Lfs*47	c.1115_1118delTTGG	p.V372Afs*42	
85	Yes	c.154C>T	p.R52*	c.2535_2536delCT	p.C846Qfs*20	
86	Yes	c.2639G>A	p.R880Q	c.1776+1G>A		
87	Yes	c.4198C>T	p.R1400C	c.2557C>T	p.R853*	

Mutation 1				Mutation 2		
FAM ID	Complementation	cDNA#	Protein	cDNA#	Protein	Protein
88	Yes	c.1A>C	p.MIP	c.2T>C	p.MIP	p.MIP
89	Yes	c.3391A>G	p.T1131A	c.1565A>G	p.K522R (affects splicing)	p.K522R (affects splicing)
90	Yes	c.1827-1G>A		c.3520_3522delTTGG	p.W1174del	p.W1174del
91	Yes	c.3391A>G	p.T1131A	large deletion		
92	No	c.3971C>T	p.P1324L	large deletion		
93	Yes	c.3762_3763insAG	p.E1255Rfs*12	c.3915_3916delCT	p.F1306Sfs*6	p.F1306Sfs*6
94	No	c.3391A>G	p.T1131A	c.2606A>C	p.Q869P	p.Q869P
95	Yes	c.3064C>T	p.Q1022*	c.3558dupG	p.R1187Efs*28	p.R1187Efs*28
96	Yes	c.1A>C	p.MIL	large deletion		
97	Yes	c.1307A>G	p.Q436R	large deletion		
98	Yes	c.2831A>C	p.D944A	c.3762_3763insAG	p.E1255Rfs*12	p.E1255Rfs*12
99	No	c.3788_3790delTTCT	p.F1263del	c.3913C>T	p.L1305F	p.L1305F
100	Yes	c.3601C>T	p.Q1201*	c.3601C>T	p.Q1201*	p.Q1201*
101	Yes	c.2852G>A	p.R951Q	c.3605_3607delAAG	p.E1202del	p.E1202del
102	No	c.97G>T	p.E33*	c.862G>T	p.E288*	p.E288*
103	Yes	c.4010+1_4010+18del		c.2T>C	p.MIT	p.MIT
104	Yes	c.1A>C	p.MIL	large deletion		
105	Yes	c.1027C>T	p.Q343*	large deletion		
106	Yes	c.597-1G>C		large deletion		
107	Yes	c.1827-1G>A		c.1827-1G>A		
108	No	c.4261-2A>C		c.3558dupG		
109	No	c.3624C>T	p.S1208S (affects splicing)	c.1027C>T	p.Q343*	p.Q343*
110	No	c.190-2A>T		c.1115_1118delTTGG	p.V372Afs*42	p.V372Afs*42
111	No	c.1776+2T>C		large deletion		
112	Yes	c.2533_2536delCTCT	p.L845Afs*43	c.1979T>C	p.L660P	p.L660P
113	Yes	c.1226-2A>G		c.3558dupG	p.R1187Efs*28	p.R1187Efs*28
114	Yes	c.1A>G	p.MIV	c.3520_3522delTTGG	p.W1174del	p.W1174del
115	No	c.2852G>A	p.R951Q	c.1827-1G>A		
116	Yes	c.1115_1118delTTGG	p.V372Afs*42	large deletion		
117	Yes	c.2021C>A	p.S674*	c.522G>C	p.Q174H (affects splicing)	p.Q174H (affects splicing)

Mutation 1				Mutation 2		
FAM ID	Complementation	cDNA#	Protein	cDNA#	Protein	Protein
118	No	c.426+1G>A		large deletion		
119	Yes	c.154C>T	p.R52*	c.2812_2830dup		p.D944Gfs*5
120	Yes	c.1A>T	p.MIL	c.1566G>A		p.K522K (affects splicing)
121	Yes	c.2851C>T	p.R951W	c.3931_3932delAG		p.S1311*
122	Yes	c.2557C>T	p.R853*	c.1083+1G>T		
123	Yes	c.2639G>A	p.R880Q	c.894-3C>G		
124	Yes	c.1303C>T	p.R435C	c.3558dupG		p.R1187Efs*28
125	No	c.4199G>A	p.R1400H	c.1734_1739delCTACGT		p.Y578*
126	Yes	c.3141_3146delCCTCTT	p.L1048_F1049del	c.709+5G>A		
127	Yes	c.1771C>T	p.R591*	large deletion		
128	No	c.2639G>A	p.R880Q	c.2T>C		p.MIT
129	Yes	c.2851C>T	p.R951W	c.2851C>T		p.R951W
130	Yes	c.916_917delAC	p.T306Afs*32	c.3349A>G		p.R1117G
131	Yes	c.1115_1118delTTGG	p.V372Afs*42	large deletion		
132	Yes	c.3004G>T	p.E1002*	c.3091C>T		p.Q1031*
133	Yes	c.1771C>T	p.R591*	large deletion		
134	Yes	c.3391A>G	p.T1131A	c.4267_4368+36del		p.A1423*
135	No	c.2638C>T	p.R880*	large deletion		
136	Yes	c.1115_1118delTTGG	p.V372Afs*42	c.3349A>G		p.R1117G
137	Yes	c.4010+1G>A		large deletion		
138	Yes	c.1827-1G>A		c.1827-1G>A		
139	Yes	c.1343A>G	p.Y448C	c.987_990delTCAC		p.H330Afs*4
140	Yes	c.1007-2A>G		c.3520_3522delTTGG		p.W1174del
141	Yes	c.2851C>T	p.R951W	c.2851C>T		p.R951W
142	Yes	c.3542T>C	p.L1181P	c.1delA		p.MICfs*43
143	Yes	c.2021C>A	p.S674*	c.3624C>T		p.S1208S (affects splicing)
144	Yes	c.2812_2830dup	p.D944Gfs*5	large deletion		
145	Yes	c.827-1G>T		c.3934+2T>C		
146	Yes	c.3240-2A>G		c.1683_1689delGGGAAG		p.G562Sfs*41
147	Yes	c.4168-2A>G		large deletion		

FAM ID	Mutation 1			Mutation 2		
	Complementation	cDNA #	Protein	cDNA #	Protein	Protein
148	Yes	c.3391A>G	p.T1131A	large deletion		
149	No	c.1303C>T	p.R435C	c.1A>G		p.M1V
150	Yes	c.4080G>A	p.M1360I	c.80-1G>C		
151	Yes	c.1771C>T	p.R591*	c.4284_4287dupCGAC		p.P1430Rfs*17
152	Yes	c.991delA	p.S331Afs*4	large deletion		
153	No	c.710-10G>A		c.710-10G>A		
154	No	c.2812_2830dup	p.D944Gfs*5	c.3788_3790delTCT		p.F1263del
155	No	c.3520_3522delTTGG	p.W1174del	c.3391A>G		p.T1131A
156	No	c.1115_1118delTTGG	p.V372Afs*42	large deletion		
157	Yes	c.284-2A>C		c.284-2A>C		
158	Yes	c.1307A>G	p.Q436R	c.2778+1G>A		
159	No	c.3788_3790delTCT	p.F1263del	large deletion		

bold typeface = novel mutation, plain typeface = known mutation; "-"- = large deletion;

§ cloned deletion coordinates; *FANCA* (NG_011706.1, NC_000016.9, NM_000135.3); cDNA numbering is based on c.1 being the A of the ATG translation initiation codon

Table 3

Known and novel pathogenic *FANCA* mutations

FANCA Mutations			Occurrences in Analyzed Cohort		Frequency in Control Databases		# of Occurrences in the FA Mutation Database		
Exon	cDNA #	Nucleotide (hg19)	Protein	Mutation Type	# of alleles	# of patients	NHLBI ESP6500 ^a	ExAC ^b	FA Mutation Database ^c
1	c.14G>A	89883010	p.W5*	Nonsense	2	2	--	--	1×
1	c.1A>C	89883023	p.MIL	Missense	3	3	--	0.00008422	1×
1	c.1A>G	89883023	p.MIV	Missense	2	2	--	--	4×
1	c.1A>T	89883023	p.MIL	Missense	1	1	--	--	4×
1	c.1delA	89883023	p.MICfs*43	Indel	1	1	--	--	--
1	c.2T>C	89883022	p.MIT	Missense	5	5	--	0.00008389	10×
1	c.65G>A	89882959	p.W22*	Nonsense	1	1	--	0.0002	9×
2	c.133G>T	89882341	p.E45*	Nonsense	1	1	7.70E-05	0.000009416	--
2	c.154C>T	89882320	p.R52*	Nonsense	4	4	--	0.000009416	1×
2	c.97G>T	89882377	p.E33*	Nonsense	1	1	--	--	--
5	c.522G>C	89877115	p.Q174H	Missense**	1	1	--	--	--
8	c.752C>G	89869707	p.S251*	Nonsense	1	1	--	--	--
9	c.811C>T	89866028	p.Q271*	Nonsense	1	1	7.70E-05	--	5×
10	c.862G>T	89865605	p.E288*	Nonsense	7	6	7.70E-05	0.00003771	13×
11	c.916_917delAC	89862403–89862404	p.T306Afs*32	Indel	1	1	--	0.000009419	1×
11	c.987_990delTCAC	89862330–89862333	p.H330Afs*4	Indel	1	1	--	0.00003766	13×
11	c.991delA	89862329	p.S331Afs*4	Indel	1	1	--	0.000009416	--
12	c.1027C>T	89858935	p.Q343*	Nonsense	2	2	--	0.000009439	1×
12	c.1073_1074delITG	89858888–89858889	p.Y359Pfs*49	Indel	1	1	--	--	--
13	c.1115_1118delTTGG	89858442–89858445	p.V372Afs*42	Indel	11	11	0.0003	0.00005649	67×
13	c.1159C>T	89858401	p.Q387*	Nonsense	1	1	--	--	--
14	c.1285_1288dupACTG	89857882–89857885	p.A430Dfs*56	Indel	2	1	--	--	--
14	c.1303C>T	89857867	p.R435C	Missense	4	4	7.70E-05	0.00001915	11×
14	c.1304G>A	89857866	p.R435H	Missense	1	1	--	--	1×
14	c.1307A>G	89857863	p.Q436R	Missense	2	2	--	--	--
14	c.1343A>G	89857827	p.Y448C	Missense	1	1	--	0.000009837	--

FANCA Mutations			Occurrences in Analyzed Cohort			Frequency in Control Databases		# of Occurrences in the FA Mutation Database	
Exon	cDNA #	Nucleotide (hg19)	Protein	Mutation Type	# of alleles	# of patients	NHLBI ESP6500 ^a	ExAC ^b	FA Mutation Database ^c
16	c.1489C>T	89849492	p.P497S	Missense	1	1	--	--	--
16	c.1565A>G	89849416	p.K522R	Missense ^{**}	1	1	--	--	--
16	c.1566G>A	89849415	p.K522K	Synonymous ^{**}	1	1	--	--	1×
18	c.1645C>T	89846347	p.Q549*	Nonsense	1	1	--	0.000009416	1×
18	c.1683_1689delGGGGAAC	89846303–89846309	p.G562Sfs*41	Indel	1	1	--	--	1×
19	c.1734_1739delCTACGT	89845388–89845393	p.Y578*	Indel	2	2	8.00E-05	0.000009461	--
19	c.1771C>T	89845356	p.R591*	Nonsense	3	3	--	--	12×
22	c.1979T>C	89839714	p.L660P	Missense	1	1	--	--	1×
23	c.2021C>A	89838216	p.S674*	Nonsense	3	3	--	--	2×
25	c.2233dupT	89836657	p.W745Lfs*49	Indel	1	1	--	--	2×
25	c.2266C>T	89836624	p.R756C	Missense	2	1	--	0.0003	--
27	c.2524delT	89833626	p.S842Lfs*47	Indel	1	1	--	--	4×
27	c.2533_2536delCTCT	89833614–89833617	p.L845Afs*43	Indel	1	1	--	--	1×
27	c.2534T>C	89833616	p.L845P	Missense	1	1	--	--	5×
27	c.2535_2536delCT	89833614–89833615	p.C846Qfs*20	Indel	4	3	--	0.000009449	10×
27	c.2557C>T	89833593	p.R853*	Nonsense	2	2	--	--	2×
28	c.2606A>C	89831470	p.Q869P	Missense	3	3	--	0.00001887	4×
28	c.2638C>T	89831438	p.R880*	Nonsense	3	3	--	--	2×
28	c.2639G>A	89831437	p.R880Q	Missense	5	5	7.70E-05	0.00003767	8×
28	c.2641C>T	89831435	p.Q881*	Nonsense	1	1	--	--	1×
28	c.2678G>A	89831398	p.W893*	Nonsense	2	1	--	--	1×
28	c.2708G>A	89831368	p.W903*	Nonsense	2	2	--	--	1×
28	c.2730_2731delCT	89831345–89831346	p.W911Dfs*31	Indel	1	1	--	--	2×
28	c.2738A>C	89831338	p.H913P	Missense	1	1	--	--	4×
28	c.2749C>T	89831327	p.R917*	Nonsense	1	1	--	--	--
29	c.2812_2830dup	89828379–89828397	p.D944Gfs*5	Indel	5	5	--	--	7×
29	c.2831A>C	89828378	p.D944A	Missense	1	1	--	--	--
29	c.2837T>C	89828372	p.L946P	Missense	1	1	--	--	--
29	c.2839dupT	89828370	p.S947Tfs*4	Indel	3	3	--	--	--

FANCA Mutations			Occurrences in Analyzed Cohort		Frequency in Control Databases		# of Occurrences in the FA Mutation Database		
Exon	cDNA#	Nucleotide (hg19)	Protein	Mutation Type	# of alleles	# of patients	NHLBI ESP6500 ^a	ExAC ^b	FA Mutation Database ^c
29	c.2840C>G	89828369	p.S947*	Nonsense	1	1	--	0.000009416	9×
29	c.2851C>T	89828358	p.R951W	Missense	5	3	--	0.000009416	15×
29	c.2852G>A	89828357	p.R951Q	Missense	3	3	--	0.00003767	9×
30	c.2854C>T	89825112	p.Q952*	Nonsense	1	1	--	--	1×
30	c.2860T>A	89825106	p.F954I	Missense	1	1	--	--	2×
30	c.2910delA	89825056	p.G972Afs*17	Indel	1	1	--	--	--
31	c.3004G>T	89818608	p.E1002*	Nonsense	1	1	--	--	--
31	c.3064C>T	89818548	p.Q1022*	Nonsense	1	1	--	--	--
32	c.3091C>T	89816286	p.Q1031*	Nonsense	1	1	--	--	1×
32	c.3130C>T	89816247	p.Q1044*	Nonsense	1	1	--	--	4×
32	c.3141_3146delCCTCTT	89816231–89816236	p.L1048_F1049del	Indel	1	1	--	--	1×
33	c.3338A>T	89815077	p.N1113I	Missense	1	1	--	--	1×
34	c.3349A>G	89813298	p.R1117G	Missense	2	2	7.70E-05	--	5×
34	c.3391A>G	89813256	p.T1131A	Missense	11	11	--	0.00005029	19×
35	c.3490C>T	89813015	p.P1164S	Missense	2	2	--	--	1×
35	c.3494T>G	89813011	p.L1165*	Nonsense	1	1	--	--	--
36	c.3520_3522delTTGG	89811471–89811473	p.W1174del	Indel	5	5	--	--	31×
36	c.3542T>C	89811451	p.L1181P	Missense	1	1	--	--	--
36	c.3558dupG	89811435	p.R1187Efs*28	Indel	5	5	--	--	23×
36	c.3601C>T	89811392	p.Q1201*	Nonsense	2	1	--	--	--
36	c.3605_3607delAAG	89811386–89811388	p.E1202del	Indel	1	1	--	--	1×
36	c.3624C>T	89811369	p.S1208S	Synonymous**	4	4	0.0002	0.0001	4×
37	c.3634dupT	89809339	p.S1212Ffs*3	Indel	2	2	--	--	--
37	c.3761_3762dupAG	89809211–89809212	p.E1255Rfs*12	Indel	3	3	--	--	1×
38	c.3788_3790delTTCT	89807250–89807252	p.F1263del	Indel	7	7	--	0.0001	261×
38	c.3791_3793delCCT	89807247–89807249	p.S1264del	Indel	3	2	--	--	--
38	c.3813dupA	89807227	p.H1272Tfs*6	Indel	1	1	--	--	2×
39	c.3913C>T	89806423	p.L1305F	Missense	1	1	--	0.00001883	5×
39	c.3915_3916delCT	89806420–89806421	p.F1306Sfs*6	Indel	1	1	--	0.000009418	--

FANCA Mutations			Occurrences in Analyzed Cohort			Frequency in Control Databases		# of Occurrences in the FA Mutation Database	
Exon	cDNA #	Nucleotide (hg19)	Protein	Mutation Type	# of alleles	# of patients	NHLBI ESP6500 ^a	ExAC ^b	FA Mutation Database ^c
39	c.3920delA	89806416	p.Q1307Rfs*2	Indel	1	1	--	--	4×
39	c.3931_3932delAG	89806404–89806405	p.S1311*	Indel	2	2	--	--	--
40	c.3971C>T	89805925	p.P1324L	Missense	2	2	--	0.0006489	4×
41	c.4015delC	89805693	p.L1339Sfs*24	Indel	1	1	8.00E-05	0.000009454	2×
41	c.4017_4021delCTCCT	89805687–89805691		Indel	1	1	--	0.00002834	1×
41	c.4080G>A	89805628	p.M1360I	Missense	1	1	--	--	--
42	c.4198C>T	89805352	p.R1400C	Missense	3	3	--	0.00000942	4×
42	c.4199G>A	89805351	p.R1400H	Missense	1	1	0.0002	0.00002826	2×
43	c.4267_4368+36del	89804973–89805110		Indel	1	1	--	--	3×
43	c.4284_4287dupCGAC	89805090–89805093	p.P1430Rfs*17	Indel	1	1	--	--	--
10i	c.894–3C>G	89862429		Splicing	1	1	--	--	1×
11i	c.1007–2A>G	89858957		Splicing	1	1	--	--	--
12i	c.1083+1G>T	89858878		Splicing	1	1	--	--	--
12i	c.1084–1G>A	89858477		Splicing	1	1	--	--	--
13i	c.1226–2A>G	89857946		Splicing	1	1	--	0.00003795	1×
13i	c.1226–6_1226–2delAACTA	89857946–89857950		Indel	1	1	--	--	1×
15i	c.1470+2T>C	89851260		Splicing	1	1	--	--	--
19i	c.1776+1G>A	89845350		Splicing	1	1	--	0.000009427	--
19i	c.1776+2T>C	89845349		Splicing	1	1	--	--	--
19i	c.1777–1G>C	89845259		Splicing	1	1	--	0.00002825	--
19i	c.1777–2A>G	89845260		Splicing	1	1	--	--	--
1i	c.80–1G>C	89882395		Splicing	1	1	--	--	--
20i	c.1827–1G>A	89842224		Splicing	1	1	--	0.000009435	--
27i	c.2602–9_2602–8delCT	89831482–89831487		Splicing	15	12	--	0.00001263	13×
28i	c.2778+1G>A	89831297		Splicing	2	2	--	0.0009	--
28i	c.2778+83C>G	89831215		Splicing	1	1	--	--	1×
2i	c.190–2A>T	89881023		Splicing	1	1	--	0.000009624	3×
32i	c.3240–2A>G	89815177		Splicing	3	3	--	0.000009442	2×
33i	c.3349–1G>A	89813299		Splicing	1	1	--	--	1×
				Splicing	1	1	--	--	--

FANCA Mutations			Occurrences in Analyzed Cohort			Frequency in Control Databases		# of Occurrences in the FA Mutation Database	
Exon	cDNA [#]	Nucleotide (hg19)	Protein	Mutation Type	# of alleles	# of patients	NHLBI ESP6500 ^a	ExAC ^b	FA Mutation Database ^c
34i	c.3408+1G>C	89813238		Splicing	1	1	--	--	1×
38i	c.3828+1G>C	89807211		Splicing	1	1	--	--	--
38i	c.3828+2dupT	89807210		Splicing	1	1	--	--	2×
39i	c.3934+2T>C	89806400		Splicing	2	2	--	0.00000942	5×
3i	c.284-2A>C	89877481		Splicing	2	1	--	0.000009562	--
40i	c.4010+1_4010+18del	89805868-89805885		Indel	1	1	--	--	4×
40i	c.4010+1G>A	89805885		Splicing	1	1	--	--	1×
41i	c.4168-2A>G	89805384		Splicing	1	1	--	--	2×
42i	c.4261-19_4261-12delACCTG	89805128-89805135		Indel	1	1	--	--	4×
42i	c.4261-2A>C	89805118		Splicing	2	2	--	--	--
4i	c.426+1G>A	89877336		Splicing	2	2	--	--	--
5i	c.522+1G>T	89877114		Splicing	1	1	--	--	--
6i	c.597-1G>C	89871801		Splicing	5	4	7.70E-05	0.000009939	1×
7i	c.709+5G>A	89871683		Splicing	1	1	--	0.00001989	4×
7i	c.710-10G>A	89869759		Splicing	3	2	--	--	2×
7i	c.710-2A>C	89869751		Splicing	1	1	--	--	--
9i	c.827-1G>T	89865641		Splicing	1	1	--	--	2×

[#] **bold typeface** = novel mutation, plain typeface = known mutation; FANCA (NG_011706.1, NC_000016.9, NM_000135.3); cDNA numbering is based on c.1 being the A of the ATG translation initiation codon

^{**} affects splicing;

^a as of December 2014;

^b as of April 01, 2016