Whole exome sequencing reveals uncommon mutations in the recently identified Fanconi anemia gene \textit{SLX4/FANCP}


* These authors contributed equally to this work.

\textit{Published in Human Mutation, 2013. 34(1):93-96.}
Fanconi anemia (FA) is a rare genetic disorder characterized by congenital malformations, progressive bone marrow failure (BMF), and susceptibility to malignancies. FA is caused by biallelic or hemizygous mutations in one of 15 known FA genes, whose products are involved in the FA/BRCA DNA damage response pathway. Here, we report on a patient with previously unknown mutations of the most recently identified FA gene, \textit{SLX4/FANCP}. Whole exome sequencing (WES) revealed a nonsense mutation and an unusual splice site mutation resulting in the partial replacement of exonic with intronic bases, thereby removing a nuclear localization signal. Immunoblotting detected no residual \textit{SLX4} protein, which was consistent with abrogated interactions with XPF/ERCC1 and MUS81/EME1. This cellular finding did not result in a more severe clinical phenotype than that of previously reported FA-P patients. Our study additionally exemplifies the versatility of WES for the detection of mutations in heterogenic disorders such as FA.

Fanconi anemia (FA; MIM\# 227650) is an autosomal or X-chromosomal recessive disorder first described in 1927 by the Swiss pediatrician Guido Fanconi. Recently, the carrier frequency in the United States was estimated to be 1:181, corresponding to an incidence of FA of less than 1:100,000 (ref 2). Higher rates have been reported for certain ethnicities or due to isolation or founder effects. The clinical manifestations of FA are variable yet characteristic. Typical congenital malformations include short stature, skin hyper- or hypopigmentations, radial ray defects, and malformations of ears, eyes, and inner organs. Most FA patients develop progressive bone marrow failure in childhood. Furthermore, they have an increased risk of myelodysplastic syndrome (MDS) and hematological malignancies, in particular acute myelogenous leukemia. In addition, they are predisposed for solid tumors occurring in young adulthood. FA patients have an up to 700-fold increased risk for squamous cell carcinomas, which arise most frequently in the mucosa of the head and neck or genital regions. The reasons for the increased susceptibility of FA patients to neoplasms are not fully understood. Most likely, this is due to a DNA repair defect and genomic instability that characterize the cellular phenotype. FA cells show highly increased rates of chromosomal breakage especially after exposure to DNA-crosslinking agents, accumulate in the G2 phase of the cell cycle, and encounter diminished survival. Like the clinical phenotype, the genetic background of FA is very heterogeneous. To date, 15 complementation groups (FA-A, -B, -C, -D1, -D2, -E, -F, -G, -I, -J, -L, -M, -N, -O, and -P) have been delineated. The first identified FA gene, \textit{FANCC} (MIM\# 613899), was reported in 1992 (ref 12). Since 2000, nearly every year a new FA gene has been added, most recently \textit{FANCP} (MIM\# 613951), encoded by the gene \textit{SLX4/BTBD12} (approved symbol \textit{SLX4}; MIM\# 613278).
Whole exome sequencing reveals uncommon mutations in \textit{SLX4}

Its product interacts with different structure-specific endonucleases such as XPF/ERCC1 (MIM\#s 133520, 126380), MUS81/EME1 (MIM\#s 606591, 610885) and the Holliday junction resolvase SLX1, by coordinating their activity in DNA repair and recombination\textsuperscript{15}. FANCP is involved in the FA/BRCA pathway and the network of DNA interstrand crosslink (ICL) repair. A key step in this pathway is FANCD2 (MIM\# 613984) and FANCI (MIM\# 611360) monoubiquitination through the FA core complex following DNA damage and replication fork stalling. FANCP acts downstream of these protein modifications, similar to FANCD1 (MIM\# 600185), FANCJ (MIM\# 605882), FANCN (MIM\# 610355), and FANCO (MIM\# 602774). FANCP-mutated cells are proficient of RAD51 foci formation, unlike FA-D1- or FA-O cells. Given these facts, a role in the coordination of DNA incision for ICL unhooking seems more likely than one in Holliday junction resolution, even though the precise function of FANCP in the FA/BRCA pathway remains elusive\textsuperscript{14,16}. Four families with a total of six affected children have been assigned to complementation group FA-P. Their underlying \textit{SLX4} mutations result in protein truncation and degradation. The presence of residual protein, retained function, and other factors may explain the variable severity of clinical FA manifestations\textsuperscript{13-15}.

In this study, we report on an additional FA-P patient who was assigned to that complementation group due to \textit{SLX4} mutations identified by whole exome sequencing (WES). \textit{SLX4} proved to be the only FA gene carrying compound heterozygous pathogenic sequence changes that were confirmed by Sanger sequencing. A nonsense and a splice site mutation followed Mendelian segregation in the family of the patient.

The 21-year-old girl of German descent was diagnosed with FA at the age of 5 years, showing FA-typical features including prenatal dystrophy, short stature, hypoplasia of the right thumb, microcephaly, speckles of skin hyperpigmentation at the arms and legs, minor café-au-lait and vitiligo spots, trivial mitral valve prolapse, and hypothyroidism. Apart from the platelets (reduced since age 5, lowest number about 20,000/μl) her blood counts were relatively stable until she developed MDS at 19 years of age. She was successfully transplanted with hematopoietic stem cells from a 10/10 matched unrelated donor. The patient has not developed malignancies up to her current age of 21; nor does she have a strong family history of cancer.

Initially, the clinical suspicion of FA was confirmed by elevated spontaneous and mitomycin C-induced chromosome breakage rates (data not shown) and G2-phase accumulation in lymphocyte, lymphoblastoid (\textbf{Fig. 1A}), and fibroblast cultures, which was shown by flow cytometric cell cycle analysis as described in Vaz \textit{et al.} (ref 17).

We isolated genomic DNA from fibroblasts using the GeneJet\textsuperscript{TM} Genomic DNA Purification Kit (Fermentas, St. Leon-Rot, Germany). For isolation of RNA, we employed the Quick-RNA\textsuperscript{TM} MiniPrep Kit (Zymo Research, Freiburg, Germany).
Translation into cDNA was performed by SuperScript® II Reverse Transcriptase (Invitrogen, Darmstadt, Germany).

Because the patient was among those who remained without detected mutation or assignment to a distinct subtype, we got interested in the significance of WES for molecular diagnostics of FA. We commissioned enrichment and sequencing of the patient’s exome to a service provider. Target enrichment was achieved by means...
Whole exome sequencing reveals uncommon mutations in SLX4

of the SureSelect Human All Exon 50 Mb Kit (Agilent, Boeblingen, Germany) and was followed by next-generation sequencing on a SOLiD5500xl instrument (Applied Biosystems, Darmstadt, Germany). Afterward we performed in house analysis of the WES data using NextGENe™ v2.18 software (Softgenetics, State College, PA, USA). The data revealed a total of 103,222,641 reads (Supplementary Table 1). Sixty-one percent of these mapped on target and resulted in an 87-fold average coverage of the exome. Altogether we detected 32,013 variants, including novel mutations as well as listed SNPs. Because of the patient’s nonconsanguineous descent, we restricted our search to compound heterozygous changes and detected 14,715 unknown (excluding reported polymorphisms) heterozygous variants in coding sequences and adjacent intron portions. In particular, 15 base substitutions were detected in 21 FA and FA-associated genes (91% of exons covered by ≥5 reads). SLX4 (NM_032444.2) carried two bona fide pathogenic variants. Even though WES could have missed pathogenic mutations in other FA genes, the compound heterozygous finding in SLX4 makes this most unlikely. The mutated positions were covered by 6 and 15 reads, respectively. We observed the nonsense mutation c.1538G>A in exon 7 resulting in a premature stop codon with the predicted effect of protein truncation, p.W513X, and the splice acceptor mutation c.1367-2A>G in intron 6 (Fig. 1B). Mutation nomenclature is based on cDNA sequence of SLX4 transcript ENST00000294008 and nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of ATG initiation codon.

Mutation validation was performed by Sanger technique on a 3130xl instrument (Applied Biosystems, Darmstadt, Germany). The primers for gDNA sequencing included FANCP_exon7_for 5’-CAGAAGCAGGTGTTGTGA-‘3 and FANCP_exon7_rev 5’-CCTTCCTGGACTTCCATCA-‘3. We resequenced the corresponding regions of SLX4 in the patient and additionally confirmed the biallelic mutation status and Mendelian segregation of the mutations by sequencing genomic DNA from both parents. The results showed that the splice site mutation was paternally inherited, whereas the nonsense mutation was transmitted maternally (Supplementary Fig. 1A).

We analyzed the consequences of c.1367-2A>G by Sanger sequencing of patient’s cDNA using the primers FANCP_c.1-65_for 5’-CAGTACTTTTGTTCAATTGTGCAAACTC-‘3 and FANCP_c. 1570_rev 5’-CACAGAAAGCTCTGCTTGCGTTC-‘3. This analysis demonstrated that a cryptic splice acceptor in exon 7 at position c.1417_1418 is used instead of the mutated in intron 6, as it was predicted by in silico analysis. Splice site score calculation using the Web tool http://rulai.cshl.edu/new_alt_exon_db2/HTML/score.html revealed a score of 1.2 of the cryptic splice acceptor at positions c.1417_1418 and a score of −1.7 of the mutated one. Of note, this change of the splice acceptor altered the usage of the unaffected wild-type splice donor of intron 6 (score 5.4) to that of a cryptic
splice donor at position c.1366+52_1366+53 (score 10). This fact is indicated by the substitution of the 5'-terminal 51 bases of exon 7 with the 5'-terminal 51 bases of intron 6. The electropherogram of cDNA sequencing demonstrates this finding by the superposition of exactly 51 bases starting at cDNA position 1,367 and ending at 1,417 (Fig. 1C and Supplementary Fig. 1B), designated as r.1367_1417delins gtttggtgatcagagaagtgacctggagaggccatcagaggtcccgg. The length of the open reading frame does not differ as a result of this aberrant splicing pattern. Other in silico analyses revealed that the deduced wild-type amino acids at positions p.456_472, ENKSRKKPPVSPPLL, are predicted to include at positions p.460_464 one of five potential SLX4/FANCP nuclear localization signals (NLS) (http://psort.hgc.jp/form2.html) (Fig. 2A). The mutation, denoted p.E456_L4772delinsGLCDQKSDPGRGHQQVP, results in the loss of that potential NLS.

Further experiments similar to those described by Stoepker et al. (ref 14) showed that no residual protein is present. Neither by immunoprecipitation (Fig. 2B), nor in a cell fractionation assay (Fig. 2) SLX4/FANCP was detected on Western blots. We conclude that the allele carrying the stop mutation gives rise to a truncated protein that is unstable and rapidly degraded. Similarly, the allele with the splice mutation does not express a stable protein, which could locate to the nucleus. Therefore, it is not surprising that interactions with the structure-specific nucleases XPF/ERCC1 and EME1/MUS81 are disrupted (Fig. 2B) and that ERCC1 is not able to form nuclear foci (Fig. 2D), as described for other FA-P patients14.

In summary, our study adds a seventh patient to the most recently described FA subtype, FA-P. Neither of her compound heterozygous mutations has previously been reported. They extend the mutation spectrum of the latest member of the FA gene family, FANCP, and have been added to the Fanconi Anemia Mutation Database (http://www.rockefeller.edu/fanconi/). In contrast to the FA-P patients reported so far, cells derived from the present patient do not seem to be able to express any SLX4/FANCP protein13,14. The failure of coordination of structure-specific nucleases in ICL unhooking due to the absence of SLX4/FANCP does not result in a more severe phenotype as that of other FA-P patients previously reported, which is not comparable to the cancer-prone phenotype of subtypes FA-D1 or -N, but falls into the clinical spectrum of the other FA groups. These insights were facilitated by WES that proved a valuable tool for molecular diagnostics of FA, as of other heterogeneous diseases, by the identification of disease-causing genes so that it may increasingly replace classical genetic approaches.
Whole exome sequencing reveals uncommon mutations in SLX4

Figure 2. Mutant SLX4/FANCP expression and function.
(A) Ideograms of the SLX4/FANCP domain structure (modified after Svendsen et al. (ref 18)). Wild-type (WT) protein (top) contains in addition to reported domains (UBZ, yellow; BTB, blue; SAP, purple; and SBD, green) five potential nuclear localization signals (NLS, red) predicted by the Web tool http://psort.hgc.jp/form2.html, spanning amino acid (aa) positions 109_124 (NLS 1), 397_412 (NLS 2), 460_464 (NLS 3), 1079_1085 (NLS 4), and 1814_1830 (NLS 5). The ideograms below (middle and bottom) show the predicted protein effect of the SLX4/FANCP mutations in the present patient. c.1538G>A leads to protein truncation at p.W513X, whereas c.1367-2A>G leads to p.E456_472Ldel/ins17 and the loss of NLS 3. (B) Immunoprecipitation and Western blot analysis shows SLX4/FANCP deficiency in fibroblasts of the patient compared to WT and FA-P controls. There is no co-precipitation of XPF, MUS81 and ERCC1 with the mutant protein. SE and LE indicate short and long exposure of the blot, respectively. (C) Subcellular fractionation of patient’s fibroblasts fails to demonstrate SLX4/FANCP protein in any fraction. Chromatin loading of XPF and MUS81 were not detected. Tubulin, p300 and HDAC1 served as loading controls for the cytoplasmic fraction (CE), nuclear extract (NE), and chromatin fraction (CB). The faint band observed in NE slightly below SLX4/FANCP is unspecific. (D) Formation of nuclear ERCC1 foci is abolished in patient’s fibroblasts as in other FA-P cells (EUF1354) in contrast to the wild-type control line (LN9SV). The ERCC1 antibody FL297 was used for immunofluorescence and TOPRO3 as a nuclear counterstain.
**Acknowledgements**
We are grateful to Helmut Hanenberg (Indianapolis) and Kornelia Neveling (Nijmegen) for earlier pre-classifications of patient’s cells; Ralf Dietrich, executive director of the German FA support association “Deutsche Fanconi-Anamie-Hilfe e. V.,” for facilitating contact to the family. The patient and her parents generously provided information on the disease course.

**Author contributions**
B.S. and K.K. designed and performed experiments, generated data, and wrote the manuscript; C.S. performed experiments (Immunoprecipitation and western blot analysis, Immunofluorescence ERCC1 foci) and contributed data; E.V. contributed vital materials; R.F. and B.M.-G. performed experiments; J.P.deW. and D.S. contributed data, directed experiments, and revised the manuscript.

**Disclosure statement**
The authors declare no financial conflict of interest.
Supplementary data

Supplementary Figure 1. Validation of Whole Exome Sequencing results by Sanger resequencing.

(A) Mendelian segregation of the FANCP mutations in the patient’s family. The splice site change c.1367-2A>G was inherited from the father, while the nonsense mutation c.1538G>A is of maternal origin. (B) Consequence of c.1367-2A>G on transcript level. Sequencing the patient’s cDNA revealed the replacement of exactly 51 exonic bases starting at cDNA position 1367 and ending at 1417 by the first 51 bases of the following intron.

Supplementary Table 1. Statistical summary of WES results including read counts and detected variants

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total read number</td>
<td>109,637,890</td>
</tr>
<tr>
<td>Reads passing QC</td>
<td>103,222,641</td>
</tr>
<tr>
<td>Reads on target</td>
<td>62,868,844 (61%)</td>
</tr>
<tr>
<td>Average exome coverage</td>
<td>87x</td>
</tr>
<tr>
<td>Total number of variants</td>
<td>32,013</td>
</tr>
<tr>
<td>Known SNPs/MNPs</td>
<td>17,194</td>
</tr>
<tr>
<td>Unknown variants in coding sequence</td>
<td>13,519</td>
</tr>
<tr>
<td>Unknown variants at canonical splice sites</td>
<td>377</td>
</tr>
<tr>
<td>Unknown homozygous variants</td>
<td>91</td>
</tr>
<tr>
<td>Unknown heterozygous variants</td>
<td>14,715</td>
</tr>
<tr>
<td>Unknown silent variants</td>
<td>2,983</td>
</tr>
<tr>
<td>Unknown missense variants</td>
<td>9,629</td>
</tr>
<tr>
<td>Unknown nonsense variants</td>
<td>510</td>
</tr>
<tr>
<td>Unknown insertions/deletions</td>
<td>468</td>
</tr>
<tr>
<td>Genes with heterozygous variants (≥2 altered bases)</td>
<td>131</td>
</tr>
<tr>
<td>Heterozygous variants in FA and FA associated genes</td>
<td>15</td>
</tr>
</tbody>
</table>

QC: quality control; SNP: single nucleotide polymorphism; MNP: multiple nucleotide polymorphism
Chapter 3

References

Whole exome sequencing reveals uncommon mutations in SLX4