haematologica
Journal of The Ferrata Storti Foundation

Pathogenic mutations identified by a multimodality approach in 117 Japanese Fanconi anemia patients

by Minako Mori, Asuka Hira, Kenichi Yoshida, Hideki Muramatsu, Yusuke Okuno, Yuichi Shiraishi, Michiko Anmae, Jun Yasuda, Shu Tadaka, Kengo Kinoshita, Tomoo Osumi, Yasushi Noguchi, Souichi Adachi, Ryoji Kobayashi, Hiroshi Kawabata, Kohsuke Imai, Tomohiro Morio, Kazuo Tamura, Akifumi Takaori-Kondo, Masayuki Yamamoto, Satoru Miyano, Seiji Kojima, Etsuro Ito, Seishi Ogawa, Keitaro Matsuo, Hiromasa Yabe, Miharu Yabe, and Minoru Takata

Haematologica 2019 [Epub ahead of print]

Citation: Minako Mori, Asuka Hira, Kenichi Yoshida, Hideki Muramatsu, Yusuke Okuno, Yuichi Shiraishi, Michiko Anmae, Jun Yasuda, Shu Tadaka, Kengo Kinoshita, Tomoo Osumi, Yasushi Noguchi, Souichi Adachi, Ryoji Kobayashi, Hiroshi Kawabata, Kohsuke Imai, Tomohiro Morio, Kazuo Tamura, Akifumi Takaori-Kondo, Masayuki Yamamoto, Satoru Miyano, Seiji Kojima, Etsuro Ito, Seishi Ogawa, Keitaro Matsuo, Hiromasa Yabe, Miharu Yabe, and Minoru Takata. Pathogenic mutations identified by a multimodality approach in 117 Japanese Fanconi anemia patients. Haematologica. 2019; 104:xxx doi:10.3324/haematol.2018.207241

Original Article

Pathogenic mutations identified by a multimodality approach in 117 Japanese Fanconi anemia patients

Minako Mori[1,2], Asuka Hira[1], Kenichi Yoshida[3], Hideki Muramatsu[4], Yusuke Okuno[4], Yuichi Shiraishi[5], Michiko Anmae[6*], Jun Yasuda[7**], Shu Tadaka[7], Kengo Kinoshita[7, 8, 9], Tomoo Osumi[10], Yasushi Noguchi[11], Souichi Adachi[12], Ryoji Kobayashi[13], Hiroshi Kawabata[14], Kohsuke Imai[15], Tomohiro Morio[16], Kazuo Tamura[6], Akifumi Takaori-Kondo[2], Masayuki Yamamoto[7,17], Satoru Miyano[5], Seiji Kojima[4], Etsuro Ito[18], Seishi Ogawa[3, 19], Keitaro Matsuo[20], Hiromasa Yabe[21], Miharu Yabe[21], Minoru Takata[1]

[1]Laboratory of DNA Damage Signaling, Department of Late Effects Studies, Radiation Biology Center, Graduate School of Biostudies, Kyoto University, Kyoto, Japan.
[2]Department of Hematology and Oncology, Graduate School of Medicine, Kyoto University, Kyoto, Japan.
[3]Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan
[4]Department of Pediatrics, Nagoya University Graduate School of Medicine, Nagoya, Japan
[5]Laboratory of DNA Information Analysis, Human Genome Center, The Institute of Medical Science, University of Tokyo, Tokyo Japan
[6]Medical Genetics Laboratory, Graduate School of Science and Engineering, Kindai University, Osaka, Japan
[7]Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan
[8]Department of Applied Information Sciences, Graduate School of Information Sciences, Tohoku University, Sendai, Japan
[9]Institute of Development, Aging, and Cancer, Tohoku University, Sendai, Japan.

[10] Children's Cancer Center, National Center for Child Health and Development, Tokyo, Japan
[11] Department of Pediatrics, Japanese Red Cross Narita Hospital, Chiba, Japan
[12] Department of Pediatrics, Kyoto University Graduate School of Medicine, Kyoto, Japan
[13] Department of Pediatrics and Adolescence, Sapporo Hokuyu Hospital, Sapporo, Japan
[14] Department of Hematology and Immunology, Kanazawa Medical University, Uchinada-machi, Japan
[15] Department of Community Pediatrics, Perinatal and Maternal Medicine, Tokyo Medical and Dental University, Tokyo, Japan
[16] Department of Pediatrics and Developmental Biology, Tokyo Medical and Dental University, Tokyo, Japan
[17] Department of Medical Biochemistry, Graduate School of Medicine, Tohoku University, Sendai, Japan
[18] Department of Pediatrics, Hirosaki University Graduate School of Medicine, Hirosaki, Japan.
[19] Department of Medicine, Center for Hematology and Regenerative Medicine, Karolinska Institute, Sweden
[20] Division of Molecular and Clinical Epidemiology, Aichi Cancer Center Research Institute, Nagoya, Japan
[21] Department of Cell Transplantation and Regenerative Medicine, Tokai University, Isehara, Japan

*Current address. Sunkaky Medical Corporation IVF Namba Clinic, Osaka, Japan
**Current address. Division of Molecular and Cellular Oncology, Miyagi Cancer Center Research Institute. 47-1, Nodayama, Megushima-Siote, Natori, Miyagi 981-1293 Japan

Correspondence

Minoru Takata, Laboratory of DNA Damage Signaling, Department of Late Effects Studies, Radiation Biology Center, Graduate School of Biostudies, Kyoto University, Kyoto, Japan.
Tel +81-75-753-7563
E-mail mtakata@house.rbc.kyoto-u.ac.jp
Miharu Yabe, Department of Cell Transplantation and Regenerative Medicine, Tokai University, Isehara, Japan
Tel +81-463-93-1121
E-mail miharu@is.icc.u-tokai.ac.jp

Abstract

Fanconi anemia is a rare recessive disease characterized by multiple congenital abnormalities, progressive bone marrow failure, and a predisposition to malignancies, resulting from mutations in one of the 22 known *FANC* genes. The number of Japanese Fanconi anemia patients with a defined genetic diagnosis was relatively limited. In this study, we reveal the genetic subtyping and the characteristics of mutated *FANC* genes in Japan and clarify the genotype-phenotype correlations. We studied 117 Japanese patients and successfully subtyped 97% of the cases. *FANCA* and *FANCG* pathogenic variants accounted for the disease in 58% and 25% of Fanconi anemia patients, respectively. We found one *FANCA* and two *FANCG* hot spot mutations, which exist at low percentages (0.04-0.1%) in the whole-genome reference panel of 3554 Japanese individuals (Tohoku Medical Megabank). *FANCB* was the third most common complementation group and only one *FANCC* case was identified in our series. Based on the data from Tohoku Medical Megabank, we estimate that ~2.6% of Japanese are carriers of disease-causing *FANC* gene variants, excluding missense mutations. This is the largest series of subtyped Japanese Fanconi anemia patients to date and the results will be useful for future clinical management.

Introduction

Fanconi anemia (FA) is a rare recessive disease characterized by multiple congenital abnormalities, progressive bone marrow failure, and predisposition to malignancies, resulting from mutations in one of the 22 known FA genes[1]. These genes are summarized in Supplemental Table 1. The proteins encoded by these genes participate in a DNA interstrand cross-link repair pathway that deals with DNA damage due to endogenous aldehydes, which are particularly deleterious to hematopoietic stem cells[2]. However, newer studies showed that biallelic mutations in *FANCM* cause infertility and early onset cancer but not a typical FA phenotype, and some of the FA genes are actually FA-like since the patients with mutations in these genes do not display hematological defects (Supplemental Table 1 and references therein). Molecular subtyping is critical for the accurate diagnosis and clinical management of the FA patients. However, finding causative mutations for a FA patient is not an easy task[3,4].

In this study, we successfully subtyped 113 of the 117 Japanese FA patients and identified 215 mutant alleles through a comprehensive strategy starting from a simple genome PCR-direct sequencing approach, then progressing to next generation sequencing. The coordinated strategies included whole-exome sequencing (WES) and targeted exome sequencing (targeted-seq). In some cases in which we could not reach a conclusive diagnosis, additional methods, such as array-CGH (aCGH) or RNA-sequencing (RNA-seq) and whole-genome sequencing (WGS) analysis, were highly useful to detect deletions or splicing abnormalities, respectively. Similar to the other ethnic groups, we found that the FA-A and FA-G groups are the most prevalent in Japan. The *FANCC* mutation is rare and, a bit surprisingly, FA-B is the third most prevalent subtype in Japan. The patients with the rare complementation groups, such as FA-D1, E, F, I, N, P, and T, were detected in less than 5% of the cases. We noted striking genotype-phenotype correlation in Japanese FA-B, D1, I, and N cases. In addition, we report the allele frequency of FA-associated deleterious genetic variations in the general Japanese population using the 3.5KJPNv2 database from the Tohoku Medical Megabank Organization (ToMMo).

Methods

Patients and samples

We studied 117 Japanese FA patients from 104 families in total. They overlap with previously reported cases (Table 1)[5-10] and additional 13 new FA patients were recruited. The diagnosis of FA was confirmed on the basis of chromosomal breakage tests and clinical features. Informed consent was obtained from the family for all subjects involved in this study, and the study was approved by the Research ethics committees of all participating hospitals and universities including Tokai University, Kyoto University, and Nagoya University. Genomic DNA or total RNA was isolated from peripheral blood or cultured fibroblasts using Puregene (Qiagen) or RNAeasy (Qiagen) kit, respectively. cDNA was synthesized with a PrimeScript RT reagent Kit (Takara).

Mutation screening for *FANCA and FANCG,* and *ALDH2* genotyping

Mutation analyses by PCR of *FANCA or FANCG* genes, Multiplex Ligation-mediated Probe Amplification (MLPA) tests for *FANCA* (Falco Biosystems), and *ALDH2* genotyping were performed as previously described[11,12]

Targeted-seq and WES

Ten and 67 patients were examined by targeted-seq and WES, respectively, as previously described[8]. In targeted-seq, 184 genes, including 15 FA genes (*FANCA, B, C, D1, D2, E, F, G, I, J, L, M, N, O* and *P*), were covered. All the mutation variants identified by targeted-seq or WES were verified by PCR and Sanger sequencing.

The aCGH analysis

For 10 patients, aCGH was performed as previously described[6]. The probes covered 19 FA genes (*FANCA, B, C, D1, D2, E, F, G, I, J, L, M, N, O, P, Q, S, T, U*) as well as FA-related genes, including *NBS1*, three *RAD51* paralogs *(XRCC3, RAD51B, and RAD51D)*, *FAAP20*, *FAAP24*, and *FAAP100*.

RNA-seq

We performed RNA-seq for three patients (Cases 62, 98, and 104). Libraries for RNA-seq were prepared using the TruSeq RNA Sample Prep Kit (Illumina) at Macrogen, and sequenced using the Illumina HiSeq 2500 platform with a standard 126-bp paired-end read protocol. Exon skipping events were identified using Genomon-fusion[13] in which patients' specific spliced junctions were identified compared with those identified in a control sample.

WGS

We performed WGS of DNA samples from one patient (Case 64) and his parents. The TruSeq DNA PCR-Free Library Preparation Kit (Illumina, San Diego, CA, USA) was used for library preparation. The prepared libraries were subjected to next-generation sequencing using a HiSeq X platform. We detected mutation variants as previously described[8]

Estimating allele frequencies of the FA-associated deleterious genetic variations in the general Japanese population

We analyzed the 3.5KJPNv2 database, which was created with data generated by WGS of 3554 individuals of the resident cohort of the ToMMo Project. The ToMMo project was established to develop a biobank that combines medical and genome information in the Tohoku area[14]. As of November 5th, 2018, the allele frequencies, including indel variations, were released in the publicly accessible 3.5KJPNv2 database (https://jmorp.megabank.tohoku.ac.jp/201811/). Our analysis focused on nonsense mutations, frameshift mutations (indels) and splicing donor or acceptor site mutations with less than 1% allele frequencies.

Results

The genetic subtyping of 117 Japanese FA patients through a comprehensive mutation screening

We started mutation analysis of FA patients by direct sequencing of *FANCA* and *FANCG*, and MLPA analysis for *FANCA* in 2009. WES and targeted-seq analyses were initiated in 2012, and molecular diagnosis was successfully achieved in 107 (91.5%) of the 117 patients (Figure 1A). We also examined the *ALDH2* genotype which has been reported to affect FA phenotypes (see below in Discussion) (Table 1)[5,10]. Unfortunately, mutations were found in only one allele in seven (six FA-A and one FA-G) of the 107 patients. Since the mutations in these patients were clearly pathogenic and rare, we assumed this was diagnostic, and did not perform further analysis.

For the remaining 10 unclassified cases, we screened large deletions in FA and related genes using our custom-designed aCGH in 2014. It revealed large deletions in two FA-B cases and one FA-T case (Figure 1B). The *FANCB* deletions spanned the entire genic area of *FANCB* (complete loss), and the defects extended into neighboring genes *MOSPD2* and/or *GLRA2*. Reanalysis of the WES data suggested putative junctions, where were amplified and sequenced. While the junction in Case 60 had a 3 bp overlapping microhomology, implying microhomology-mediated end joining as the mechanism (explained in Supplemental Fig.1), there was no such homologous sequence in the break point in Case 61, suggesting that the re-ligation was mediated by non-homologous end joining (Figure 1B)[15]. Two cases of entire *FANCB* deletion have been described in the literature[16,17] without elucidation of the junctional sequence. All of these *FANCB* large deletions seem to be distinct, but uniformly accompany severe phenotypic malformations (see below). The FA-T case with a large deletion was previously described[6].

After aCGH, seven FA cases remained unclassified. We performed WGS for Case 64, in which parents' genome was available, and RNA-seq analysis was carried out for three cases (Case 62, 98, and 104), in which patient's fibroblast cell lines were available. Interestingly, these analyses identified three cases with aberrant splice site mutations. WGS revealed

that Case 64 harbored a homozygous mutation (c.1154+5G>A) in intron 12 of the *FANCC* gene. RT-PCR confirmed that the mutation caused a splicing abnormality, resulting in retention of 120bp of intron 12 and a subsequent in-frame nonsense codon (Figure 1C). In Case 62, RNA-seq analysis revealed skipping of *FANCB* exon 7 (Figure 1D). This was likely to be caused by a mutation in the first nucleotide of exon 7, which did not alter the encoded amino acid (p.Leu499Leu). This mutation was considered non-pathogenic when the WES results were originally evaluated. However, it has been increasingly recognized that similar synonymous mutations affect splicing and cause genetic disorders and cancer[18,19]. RNA-seq and WES also revealed that Case 98 had a homozygous mutation (c.3350+5G>A) in intron 12 of *PALB2/FANCN* gene, resulting in skipping of exon 12 (Figure 1D).

Collectively, 113 (97%) of 117 Japanese FA patients were subtyped, and a total of 215 mutant alleles were identified (Table 1, Figure 2A and 2B). As shown in Figure 2A, FA-A and FA-G accounted for 58% and 25% of FA patients, respectively. Interestingly, *FANCB* was the third most common complementation group in our series (~3%). In a notable contradistinction to a previous report from the Rockefeller University Fanconi Anemia Mutation Database[20], FA-C was very rare complementation group in Japan (Table S1). In keeping this, there was not a single record with an IVS4+4 mutation in the 3.5KJPN or the East Asian population represented in the Exome Aggregation Consortium (ExAC) database. In Europeans, the allele frequency of the mutation was relatively high (0.04%) in the ExAC database, which reflects a high frequency of the IVS4+4A>T mutation in Ashkenazi-Jewish FA-C cases[21].

Characteristics of Japanese *FANCA* mutation variants

In 68 FA-A patients (from 59 unrelated families), 130 mutant alleles were identified that consisted of 55 different *FANCA* mutational variants (listed in Supplemental Table 2 and Supplemental Fig 2A). The mutant alleles included 9 missense mutations, 8 nonsense mutations, 16 small insertions/deletions (indels), 12 large deletions, 1 large duplication, and 9 splicing mutations. All of the 9 missense mutations were rated as

"damaging" by both SIFT and PolyPhen-2 prediction programs, including two novel variants (c.2723_2725TCT>GCC, p.LS908_909RP; c.3965T>G, p.V1322G). Three of the 8 nonsense mutations, 6 of the 16 small indels, and 4 of the 9 splicing mutations were novel (Supplemental Table 2). We consider that these 13 novel mutations are all pathogenic. The large duplication and all of the large deletions except one (c.3765+827_3814del) were detected by the MLPA assay. We did not identify the precise breakpoints of these *FANCA* deletions; therefore, it was unclear whether they were novel or not.

Similar to the past reports from Western countries[20,22,23], the mutational spectrum in Japanese FA patients was broad (Figure 2B). However, some mutations were recurrently detected. The *FANCA* c.2546delC mutation was the most frequent (41/130 alleles; 31.5%), and other mutations such as c.978_c.979delGA, c.2602-2A>T, and c.2602-1G>A were detected in at least three unrelated families. c.1303C>T, c.2170A>C, c.2840C>G, c.3720_3724del, c.4168-2A>G were each detected in two unrelated families. The 45 remaining mutation variants were unique and were detected in single patients. *FANCA* c.2546delC existed at 0.08% frequency among 3554 individuals from 3.5KJPNv2 in the ToMMo (Table 2), but not in the ExAC database (0%). This mutation was also commonly identified in Korean FA-A patients[24] and therefore seems to be a hotspot in the East Asian population.

Characteristics of Japanese *FANCG* mutation variants

In 29 FA-G patients (from 27 unrelated families), 57 mutant alleles were identified, and seven different *FANCG* variants were detected (Supplemental Table 3 and Supplemental Fig 2B). The number of unique mutation variants in FA-G was fewer, compared with FA-A (Figure 2B). Three of the 7 mutation variants were novel. Of the three novel variants, two (c.907_908del and c.1386delC) were clearly pathogenic, whereas one mutation in intron 12 (c.1637-15G>A) was of uncertain significance. As previously reported, c.307+1G>C and 1066C>T accounted for most of the *FANCG* mutant alleles (49/57; 86%) in the Japanese FA-G patients[25,26]. 13

of the 29 FA-G patients were homozygous for c.307+1G>C, and 8 were compound heterozygous with one c.307+G>C allele. 5 of the 8 remaining FA-G patients had homozygous c.1066C>T mutations. Four cases were compound heterozygous for the c.307+G>C and c.1066C>T mutations. In the 3.5KJPNv2 data, *FANCG* c.307+1G>C and c.1066C>T mutation variants were present with frequencies of 0.1% and 0.04%, respectively (Table 2). These mutations were similarly detected in Korean FA-G patients[24] but hardly observed in the other ethnic populations according to the ExAC database.

VACTERL-H phenotype caused by *FANCB*, *FANCI*, and other FA gene variants

We identified *FANCB* mutations in four affected males. The *FANCB* gene maps to the X-chromosome. Two of the four FA-B patients had a complete loss of the *FANCB* gene, as detected by aCGH (Figure 1B). In the remaining two patients, one harbored a nonsense mutation (c.516G>A/p.W172X) and one had a synonymous mutation (c.1497G>T/p.L499L) resulting in exon 7 skipping (Figure 1D and Supplemental Fig 4A). All four mutations were unique. The two FA-B cases with complete loss of *FANCB* displayed severe somatic abnormalities, consistent with VACTERL-H association (Table 3). The VACTERL-H association is defined as having three or more of the following defects: vertebral anomalies, anal atresia, cardiac anomalies, tracheal-esophageal fistula, esophageal atresia, renal structural abnormalities, limb anomalies, and hydrocephalus[27]. This set of anomalies has been reported in rare cases of FA, and is particularly associated with FA-B, I, J, N, or O cases[28]. The most frequent combination patterns in these patients with VACTERL-H association were CRL (cardiac-renal-limb anomalies), ARL (anal-renal-limb anomalies), and VRL (vertebral-renal-limb anomalies), which accounted for more than half of the patients. Case 60 and 61 had 5 and 7 features of the VACTERL-H anomalies, respectively.

Compared with these two FA-B cases, Case 62 with C-terminally truncated FANCB protein showed a less severe phenotype and experienced later onset of bone marrow failure (Supplemental Fig 4A). A recent

biochemical study revealed that FANCB together with FAAP100 and FANCL are the central subcomplex components of the FA core complex, which is essential for ID2 complex monoubiquitination, a key activation event in the FA pathway. The FANCB:FAAP100 subunits form a scaffold that drives dimer formation of FANCL[29], which is the E3 ligase component in the FA core complex. The truncated FANCB protein in Case 62 might maintain the ability to interact with FAAP100 or FANCL protein to some extent[30]. We could not get clinical information from another FA-B patient (Case 63).

Two FA-I cases were identified, and both had compound heterozygous mutations (Supplemental Fig. 4B). Case 96, with N-terminal premature termination codons, had the five features of the VACTERL-H anomalies and died within 2 months after hematopoietic stem cell transplantation (HSCT) (Table 3). On the other hand, Case 97, with C-terminal mutations, had only two features of the VACTERL-H and survived for more than 17 years after HSCT. In Case 96, a c.158-2A>G mutation in intron 3 and a c.288G>A mutation in the last codon of exon 4 caused splicing defects that resulted in a single nucleotide (guanine) insertion after exon 3 and skipping of exon 4, respectively (Figure 3). For Case 97, cells were not available and we could not verify the actual splicing defect caused by the c.3006+3A>G mutation. The patient's mother had only the c.3346_3347 insT mutation, while the father's genome was unavailable. The mutation at the +3 splice donor position was indicative of a potential splice defect [31] and we therefore considered that c.3006+3A>G would be a pathogenic mutation. This mutation was very rare and not reported as an SNV in the 3.5KJNv2 and ExAC database.

We also revisited available clinical data from 103 additional FA patients, and identified seven more cases with VACTERL-H (Table 3). These include three FA-A, one FA-C case, two FA-G cases, and one FA-P case. All these seven cases met with VACTERL-H criteria with only three features. Four of the seven cases showed the CRL defect combination pattern. Compared with these cases, FA-B and FA-I cases with VACTERL-H association appeared to have higher number of malformations (five to

seven). We could not get detailed clinical information from the remaining 9 patients. Thus, altogether there were 10 VACTERL-H cases out of 108 cases with clinical data in our series, which seems slightly high compared to the prior report by Alter and Rosenberg (108 cases out of 2245)[28].

Early-onset malignancies associated with the *FANCD1* (*BRCA2*) or *FANCN* (*PALB2*) complementation group

We identified two FA-D1 patients and one FA-N patient in our series. To the best of our knowledge, no FA-N cases and only one FA-D1 case (AP37P in Table 4) have been previously reported from Japan[32,33]. The two FA-D1 cases in our study had compound heterozygous mutations, of which one was an N-terminal splice site mutation and the other was a nonsense or missense mutation (Supplemental Fig. 4C). Both of the two *FANCD1* (*BRCA2*) splice site mutations (c.475+1G>A, c.517-2A>G) were regarded as deleterious. The one missense mutation (*FANCD1* c.7847C>T/p.S2616F) was rated as "damaging" by both SIFT and PolyPhen-2 prediction programs. It is notable that this missense mutation falls into the region termed "FA cluster" (amino acid position 2336-2729) where all of the five FA-D1 associated *BRCA2* missense mutations are found[34]. One FA-N patient had a homozygous splice mutation (c.3350+5C>T), resulting in skipping of exon 12 and C-terminal truncation (Figure 1D, Supplemental Fig. 4D). This truncation may affect PALB2 interaction with RNF168 or BRCA2 which is mediated by the PALB2 C-terminal WD40 domain[35,36].

As shown in Table 4, the three FA-D1 patients, including the previous Japanese case, as well as the one FA-N patient developed early-onset malignancies, in line with previous reports from Western countries[34,37,38]. Although it is important to note that the first clinical manifestation in such cases could be onset of malignancy without prior clinical problems, Cases 65 and 98 had severe physical anomalies as well. Their malformations did not fully meet VACTERL-H criteria (Table 4). Alter *et al.* previously reported that FA-D1 and FA-N patients were characterized by frequent VACTERL-H association and early-onset tumors, such as Wilms tumor, or acute myeloid leukemia (AML), with a cumulative

incidence of malignancy as high as 97% by the age of 5.2[34]. Thus, Case 66 was highly unusual as a FA-D1 patient. He developed T lymphoblastic lymphoma at 23 years of age, with a mild short stature and severe microcephaly (see Supplemental Note). He received standard chemotherapy for the lymphoma, which caused prolonged pancytopenia. Then a mitomycin C induced chromosome breakage test was performed, and he was diagnosed as FA. We list Case 66 as FA-D1, since he had biallelic, likely deleterious, *BRCA2* variants but no other FA gene mutations. This case may expand the clinical spectrum of FA-D1. Alternatively, it is currently impossible to exclude hidden FA gene variants caused his FA phenotype.

Allele frequency of pathogenic variations in 22 FA genes in the Japanese population

To estimate the frequency of pathogenic FA gene variations in the Japanese population, we analyzed WGS data for 22 FA genes from the 3.5KJPNv2 database. We identified 66 deleterious genetic variations (nonsense, frameshifts, and splicing site mutations) in 19 FA genes (Table 2). In addition to the three common *FANCA* (c.2546delC (0.08%)) and *FANCG* mutations (c.307+1G>C (0.1%); c.1066C>T (0.04%)), carriers with *FANCA* c.2602-2A>T, *FANCD1* c.6952C>T, *FANCG* c.194delC, or *FANCI* c.157-2A>G mutations were detected at low percentages (0.01-0.08%), and these variants were identified as causative mutations in Japanese FA patients. Allele frequencies of *FANCL* c.170G>A (p.W57X) variants were relatively high (0.08%); however, no patients with these variants were identified in our FA collection.

Monoallelic mutations in some FA genes, such as *BRCA1, BRCA2, BRIP1, PALB2 and RAD51C*, cause adult-onset cancer predisposition[39-41] and we identified 25 deleterious variants in these genes (five in *BRCA1,* ten in *BRCA2*, three in *BRIP1*, six in *PALB2,* and one in *RAD51C*). *BRCA1* c.188T>A (p.L63X) and *BRCA2* c.6952C>T (p.R2318X) are well-known mutations in hereditary breast and ovarian cancer (HBOC) in Japan[42]. The *BRCA2* c.10150C>T (p.R3384X) was more prevalent than p.R2318X, but it has been classified as non-pathogenic because of its location near the 3'-

end[43]. The *PALB2* c.2834+2T>C was recently identified in a Japanese female with bilateral breast cancer[44].

From these analyses of allele frequency of FA-associated deleterious mutation variants in 3554 individuals, we estimated that ~2.6% of the Japanese could be considered to be carriers of pathogenic variations in FA genes.

Discussion

In this study, we report the largest series of subtyped Japanese FA patients to date by compiling our previously reported cases with additional 13 new cases (Table 1). We employed various methods including PCR-direct sequencing and next generation sequencing. WES and targeted exome sequencing were highly useful to identify mutations, as reported previously[8]. However, about half of the cases were undiagnosed even after these procedures[8]. When combined with the data generated by *FANCA*-MLPA, the diagnosis rate was much enhanced since *FANCA* deletion was frequent, and WES/target-seq is not necessarily effective in identifying deletions. We also noted that mutations affecting splicing, such as intronic or synonymous variants, were difficult to detect by WES or targeted-seq. The former weak point was complemented by the use of aCGH, while RNA-seq was useful for detecting splicing abnormalities. We think the identification of two synonymous mutations affecting splicing is of high significance, since this type of mutations could have been easily overlooked. Thus our approach ultimately achieved molecular diagnosis in most of the cases, and many private and novel mutations were identified in 11 of the 22 known FA genes.

Given the present results, we suggest that a molecular work-up of Japanese FA patients should start with screening for the three most-common mutations (*FANCA* c.2546delC, *FANCG* c.307+1G>C, and *FANCG* c.1066C>T) along with an MLPA assay for *FANCA*. As a next step, targeted-seq or WES analysis should be considered. For the remaining unclassified cases, aCGH, WGS, and RNA-seq analysis may be useful to identify large indels or splicing defects. Through these combined and comprehensive

efforts, correct genetic diagnosis may be obtained in more than 90% of the Japanese FA patients.

Aldehyde dehydrogenase 2 (*ALDH2)* converts acetaldehyde to acetate, and potentially catalyzes other aldehydes as well. In East Asian countries including Japan, a significant fraction (~50%) of the population carries *ALDH2* variant *ALDH2*504Lys* which is encoded by the so-called A allele, and affects alcohol tolerance and some aspects of human health[45]. We have previously described a subset of severe FA cases that were homozygous for the *ALDH2*504Lys* variant (the AA genotype), and who experienced bone marrow failure and/or MDS immediately after birth[5,7]. We also found several FA-B and FA-I cases that were accompanied by severe physical abnormalities, termed VACTERL-H. Two (Case 18-1, 99-1) of the six previously reported FA cases carrying a homozygous *ALDH2* AA genotype also displayed these severe malformations[7] (Table 3), but interestingly, their siblings (Case 18-2, 99-2) with *ALDH2* GG genotype displayed milder physical abnormalities (Table 1). We note here that the FA-B or FA-I patients with VACTERL-H anomalies were carriers of the *ALDH2* GG or the GA genotype. The impact of endogenous aldehyde catabolism on bone marrow stem cells is very clear, and this effect also extends to the role of the FA proteins in preventing severe malformations. It has been suggested that the extent of physical abnormalities and severity of hematological defects tend to be correlated[46]. In any event, FA-B and FA-I groups often exhibit severe malformations, as described previously[47,48] and confirmed here in Japanese cases. Since many of our cases were referred to us in order to carry out HSCT, our data could be biased toward a severer fraction of patients and may not reflect the entirety of individuals carrying FA gene variants. The relatively high incidence of VACTERL-H anomalies in our series could reflect this[46] and/or may be due to the impact of the *ALDH2* genotype.

An important issue is how prevalent the FA-causing variants in the Japanese population are. We estimate that at least ~2.6% of the Japanese population might carry pathogenic variants in FA genes, using the 3.5KJPNv2 database. In Japan, ~10 individuals with FA are born per ~1

million births each year according to the report from the Japanese Society of Pediatric Hematology/Oncology[49]. FA-G accounted for 25% of Japanese FA patients according to our study and ~2 FA-G patients are estimated to be born each year in Japan. Our estimated allele frequency for *FANCG* (0.16%) from the 3.5KJPNv2 database is reasonable to the birth rate of the FA-G patients. Rogers *et al.* reported that at least one FA disease-causing variant among 16 FA genes (nonsense, splice altering, frame shifts, and a subset of missense variants that are judged to be highly deleterious) was identified in 4.3% of individuals from the ESP and 1KGP studies[50]. This estimate was substantially higher than ours, but our number may increase if we can include deleterious missense mutation data in the future.

In conclusion, the molecular diagnostic strategy and data described in this study provide a basis for future molecular work-ups and clinical management for Japanese FA patients. In four cases, we failed to achieve a definitive subtyping (this could be due to technical problems or due to novel FA genes awaiting discovery). These remain as "unclassified", and could be of particular interest in further elucidation of the etiology of FA.

Acknowledgments

The authors would like to thank patients and families for participating in the study, Dr. James Hejna (Kyoto University) for critical reading of the manuscript, and Akiko Watanabe and Fan Peng for technical and secretarial help. This work was supported by JSPS KAKENHI Grant Number JP15H01738 [to M.T.], grants from the Ministry of Health, Labor, and Welfare [to S.K. and to E.I.], and grants from Uehara Memorial Foundation [to M.T.], and Astellas Foundation for Research on Metabolic Disorders [to M.T.].

Author contributions

M.Yabe, H.Y. and M.T. initiated the study with help from K.T., A.T.-K., S.K., E.I., S.M., S.O., and K.M.; M.M., A.H., K.Y., H.M., Y.O., Y.S., and M.A. performed subtyping experiments; J.Y., S.T., K.K., and M.Yamamoto provided the data from the ToMMo; H.Y., M.Y., T.O., Y.N., R.K., S.A., H.K.

K.I., and T.M. provided clinical samples and analyzed clinical data; M.M. compiled the subtyping data and wrote the paper with M.T. and M.Yabe.

Conflict of Interest Disclosures

The authors declare that there is no conflict of interest.

References

1. Nalepa G, Clapp DW. Fanconi anaemia and cancer: an intricate relationship. Nat Rev Cancer. 2018;18(3):168–185.

2. Garaycoechea JI, Crossan GP, Langevin F, et al. Alcohol and endogenous aldehydes damage chromosomes and mutate stem cells. Nature. 2018;553(7687):171–177.

3. Chandrasekharappa SC, Lach FP, Kimble DC, et al. Massively parallel sequencing, aCGH, and RNA-Seq technologies provide a comprehensive molecular diagnosis of Fanconi anemia. Blood. 2013;121(22):e138–148.

4. De Rocco D, Bottega R, Cappelli E, et al. Molecular analysis of Fanconi anemia: the experience of the Bone Marrow Failure Study Group of the Italian Association of Pediatric Onco-Hematology. Haematologica. 2014;99(6):1022–1031.

5. Hira A, Yabe H, Yoshida K, et al. Variant ALDH2 is associated with accelerated progression of bone marrow failure in Japanese Fanconi anemia patients. Blood. 2013;122(18):3206–3209.

6. Hira A, Yoshida K, Sato K, et al. Mutations in the Gene Encoding the E2 Conjugating Enzyme UBE2T Cause Fanconi Anemia. Am J Hum Genet. 2015;96(6):1001–1007.

7. Yabe M, Yabe H, Morimoto T, et al. The phenotype and clinical course of Japanese Fanconi Anaemia infants is influenced by patient, but not maternal ALDH2 genotype. Br J Haematol. 2016;175(3):457–461.

8. Muramatsu H, Okuno Y, Yoshida K, et al. Clinical utility of next-generation sequencing for inherited bone marrow failure syndromes. Genet Med. 2017;19(7):796–802.

9. Sekinaka Y, Mitsuiki N, Imai K, et al. Common Variable Immunodeficiency Caused by FANC Mutations. J Clin Immunol. 2017;37(5):434–444.

10. Yabe M, Koike T, Ohtsubo K, et al. Associations of complementation group, ALDH2 genotype, and clonal abnormalities with hematological outcome in Japanese patients with Fanconi anemia. Ann Hematol. 2019;98(2):271-280.

11. Tachibana A, Kato T, Ejima Y, et al. The FANCA gene in Japanese Fanconi anemia: reports of eight novel mutations and analysis of sequence variability. Hum Mutat. 1999;13(3):237–244.

12. Matsuo K, Wakai K, Hirose K, Ito H, Saito T, Tajima K. Alcohol dehydrogenase 2 His47Arg polymorphism influences drinking habit independently of aldehyde dehydrogenase 2 Glu487Lys polymorphism: analysis of 2,299 Japanese subjects. Cancer Epidemiol Biomarkers Prev. 2006;15(5):1009–1013.

13. Shiraishi Y, Fujimoto A, Furuta M, et al. Integrated analysis of whole genome and transcriptome sequencing reveals diverse transcriptomic aberrations driven by somatic genomic changes in liver cancers. Creighton C, editor. PLoS One. 2014;9(12):e114263.

14. Yasuda J, Kinoshita K, Katsuoka F, et al. Genome analyses for the Tohoku Medical Megabank Project toward establishment of personalized healthcare. J Biochem. 2019;165(2):139-158.

15. Ceccaldi R, Rondinelli B, D'Andrea AD. Repair Pathway Choices and Consequences at the Double-Strand Break. Trends Cell Biol. 2016;26(1):52–64.

16. Umaña LA, Magoulas P, Bi W, Bacino CA. A male newborn with VACTERL association and Fanconi anemia with a FANCB deletion

detected by array comparative genomic hybridization (aCGH). Am J Med Genet A. 2011;155A(12):3071–3074.

17. Flynn EK, Kamat A, Lach FP, et al. Comprehensive analysis of pathogenic deletion variants in Fanconi anemia genes. Hum Mutat. 2014;35(11):1342–1353.

18. Chamary JV, Parmley JL, Hurst LD. Hearing silence: non-neutral evolution at synonymous sites in mammals. Nat Rev Genet. 2006;7(2):98–108.

19. Supek F, Miñana B, Valcárcel J, Gabaldón T, Lehner B. Synonymous Mutations Frequently Act as Driver Mutations in Human Cancers. Cell. 2014;156(6):1324–1335.

20. Neveling K, Endt D, Hoehn H, Schindler D. Genotype-phenotype correlations in Fanconi anemia. Mutat Res. 2009;668(1-2):73–91.

21. Whitney MA, Saito H, Jakobs PM, Gibson RA, Moses RE, Grompe M. A common mutation in the FACC gene causes Fanconi anaemia in Ashkenazi Jews. Nat Genet. 1993;4(2):202–205.

22. Kimble DC, Lach FP, Gregg SQ, et al. A comprehensive approach to identification of pathogenic FANCA variants in Fanconi anemia patients and their families. Hum Mutat. 2018;39(2):237–254.

23. Castellà M, Pujol R, Callen E, et al. Origin, functional role, and clinical impact of Fanconi anemia FANCA mutations. Blood. 2011;117(14):3759–3769.

24. Park J, Chung N-G, Chae H, et al. FANCA and FANCG are the major Fanconi anemia genes in the Korean population. Clin Genet. 2013;84(3):271–275.

25. Yamada T, Tachibana A, Shimizu T, Mugishima H, Okubo M, Sasaki MS. Novel mutations of the FANCG gene causing alternative splicing in Japanese Fanconi anemia. J Hum Genet. 2000;45(3):159–166.

26. Yagasaki H, Oda T, Adachi D, et al. Two common founder mutations of the fanconi anemia group G gene FANCG/XRCC9 in the Japanese population. Hum Mutat. 2003;21(5):555.

27. Solomon BD, Pineda-Alvarez DE, Raam MS, et al. Analysis of component findings in 79 patients diagnosed with VACTERL association. Am J Med Genet A. 2010;152A(9):2236–2244.

28. Alter BP, Rosenberg PS. VACTERL-H Association and Fanconi Anemia. Mol Syndromol. 2013;4(1-2):87–93.

29. van Twest S, Murphy VJ, Hodson C, et al. Mechanism of Ubiquitination and Deubiquitination in the Fanconi Anemia Pathway. Mol Cell. 2017;65(2):247–259.

30. McCauley J, Masand N, McGowan R, et al. X-linked VACTERL with hydrocephalus syndrome: further delineation of the phenotype caused by FANCB mutations. Am J Med Genet A. 2011;155A(10):2370–2380.

31. Krawczak M, Thomas NST, Hundrieser B, et al. Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. Hum Mutat. 2007;28(2):150–158.

32. Ikeda H, Matsushita M, Waisfisz Q, et al. Genetic reversion in an acute myelogenous leukemia cell line from a Fanconi anemia patient with biallelic mutations in BRCA2. Cancer Res. 2003;63(10):2688–2694.

33. Bakker JL, Thirthagiri E, van Mil SE, et al. A novel splice site mutation in the noncoding region of BRCA2: implications for Fanconi

anemia and familial breast cancer diagnostics. Hum Mutat. 2014;35(4):442–446.

34. Alter BP, Rosenberg PS, Brody LC. Clinical and molecular features associated with biallelic mutations in FANCD1/BRCA2. J Med Genet. 2007;44(1):1–9.

35. Xia B, Sheng Q, Nakanishi K, et al. Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. Mol Cell. 2006;22(6):719–729.

36. Luijsterburg MS, Typas D, Caron M-C, et al. A PALB2-interacting domain in RNF168 couples homologous recombination to DNA break-induced chromatin ubiquitylation. Elife. 2017;6.

37. Reid S, Schindler D, Hanenberg H, et al. Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer. Nat Genet. 2007;39(2):162–164.

38. Wagner JE, Tolar J, Levran O, et al. Germline mutations in BRCA2: shared genetic susceptibility to breast cancer, early onset leukemia, and Fanconi anemia. Blood. 2004;103(8):3226–3229.

39. Prakash R, Zhang Y, Feng W, Jasin M. Homologous Recombination and Human Health: The Roles of BRCA1, BRCA2, and Associated Proteins. Cold Spring Harb Perspect Biol. 2015;7(4):a016600.

40. Katsuki Y, Takata M. Defects in homologous recombination repair behind the human diseases: FA and HBOC. Endocr Relat Cancer. 2016;23(10):T19–37.

41. Paulo P, Maia S, Pinto C, et al. Targeted next generation sequencing identifies functionally deleterious germline mutations in novel genes in early-onset/familial prostate cancer. PLoS Genet. 2018;14(4):e1007355.

42. Hirotsu Y, Nakagomi H, Sakamoto I, Amemiya K, Mochizuki H, Omata M. Detection of BRCA1 and BRCA2 germline mutations in Japanese population using next-generation sequencing. Mol Genet Genomic Med. 2015;3(2):121–129.

43. Borg A, Haile RW, Malone KE, et al. Characterization of BRCA1 and BRCA2 deleterious mutations and variants of unknown clinical significance in unilateral and bilateral breast cancer: the WECARE study. Hum Mutat. 2010;31(3):E1200–1240.

44. Nakagomi H, Hirotsu Y, Okimoto K, et al. PALB2 mutation in a woman with bilateral breast cancer: A case report. Mol Clin Oncol. 2017;6(4):556–560.

45. Gross ER, Zambelli VO, Small BA, Ferreira JCB, Chen C-H, Mochly-Rosen D. A personalized medicine approach for Asian Americans with the aldehyde dehydrogenase 2*2 variant. Annu Rev Pharmacol Toxicol. 2015;55:107–127.

46. Rosenberg PS, Huang Y, Alter BP. Individualized risks of first adverse events in patients with Fanconi anemia. Blood. 2004;104(2):350–355.

47. Holden ST, Cox JJ, Kesterton I, Thomas NS, Carr C, Woods CG. Fanconi anaemia complementation group B presenting as X linked VACTERL with hydrocephalus syndrome. J Med Genet. 2006;43(9):750–754.

48. Savage SA, Ballew BJ, Giri N, et al. Novel FANCI mutations in Fanconi anemia with VACTERL association. Am J Med Genet A. 2016;170A(2):386–391.

49. Yabe M, Yabe H. [Diagnosis and management of inherited bone marrow failure syndrome]. Rinsho Ketsueki. 2015;56(10):1914–1921.

50. Rogers KJ, Fu W, Akey JM, Monnat RJ. Global and disease-associated genetic variation in the human Fanconi anemia gene family. Hum Mol Genet. 2014;23(25):6815–6825.

Table 1. The list of *FANC* genes mutation variants, *ALDH2* genotype, and clinical information in 117 Japanese patients with Fanconi anemia

Family No.	Case No.	Sex	Affected gene	Methods for identifying the mutations	Mutation 1		Mutation 2		*ALDH2* genotype*	Hematological/Oncologic phenotype	Age at BMF/Malignancy diagnosis (months)	FA-features**	VACTERL-H	Reference/comments
					cDNA	Protein	cDNA	Protein						
1	1	F	*FANCA*	WES	c.2870G>A	p.W957X	**c.2723_2725TCT>GCC**	p.LS908_909RP	GG	MDS	121/335	Yes	No	5, 10
2	2	M	*FANCA*	WES	c.1303C>T	p.R435C	c.1303C>T	p.R435C	GA	AML	unknown/289	Yes	No	8, 10
3	3	F	*FANCA*	WES	c.2170A>C	p.T724P	c.505G>T	p.E169X	GA	MDS	unknown/143	Yes	No	8, 10
4	4	M	*FANCA*	WES, MLPA	c.2546delC	p.S849FfsX40	ex30del	·	GA	AA	37	Yes	No	5, 8, 10
5	5	F	*FANCA*	WES	c.1303C>T	p.R435C	c.4168-1G>C	aberrant splicing	GA	AA	26	Yes	No	5, 8, 10
6	6-1	M	*FANCA*	WES, MLPA	c.3765+1G>T	aberrant splicing	ex30del	·	GG	AA	96	Yes	No	5, 10
	6-2	F	*FANCA*	MLPA, Sanger	c.3765+1G>T	aberrant splicing	ex30del	·	GG	AA	51	Yes	No	5, 10
7	7	M	*FANCA*	WES	c.4240_4241delAG	p.S1414LfsX10	c.2602-1G>A	aberrant splicing	GG	AML	41/115	Yes	No	5, 7, 8, 10
8	8	M	*FANCA*	Sanger	c.2546delC	p.S849FfsX40	c.2546delC	p.S849FfsX40	GA	AA	38	Yes	No	5, 10
9	9-1	M	*FANCA*	MLPA, Sanger	c.978_979delGA	p.Q326HfsX12	ex30del	·	GA	MDS	60/192	Yes	No	5, 7, 10
	9-2	F	*FANCA*	MLPA, Sanger	c.978_979delGA	p.Q326HfsX12	ex30del	·	GG	AA	92	Yes	No	5, 7, 10
	9-3	F	*FANCA*	WES, MLPA	c.978_979delGA	p.Q326HfsX12	ex30del	·	GG	AA	45	Yes	No	5, 7, 10
10	10-1	F	*FANCA*	WES	c.2602-2A>T	aberrant splicing	c.4198C>T	p.R1400C	GG	AA	120	Yes	No	5, 8, 10
	10-2	F	*FANCA*	WES	c.2602-2A>T	aberrant splicing	c.4198C>T	p.R1400C	GA	AA	48	Yes	No	5, 10
11	11	M	*FANCA*	WES, MLPA	c.3568C>T	p.Q1190X	**ex11-15dupli**	·	GG	AA	297	Yes	No	5, 10
12	12	M	*FANCA*	WES	c.3919_3920insT	p.Q1307LfsX6	c.2546delC	p.S849FfsX40	GG	MDS	144/145	Yes	No	5, 8, 10
13	13	F	*FANCA*	WES, MLPA	c.2546delC	p.S849FfsX40	ex1-28del	·	GG	MDS	72/72	Yes	No	5, 7, 8, 10
14	14	M	*FANCA*	WES	c.2602-1G>A	aberrant splicing	c.2602-2A>T	aberrant splicing	GG	AA	134	Yes	No	5, 8, 10
15	15	F	*FANCA*	WES	**c.1007-2A>G**	aberrant splicing#	c.4168-2A>G	aberrant splicing	GA	MDS	48/60	No	No	10
16	16	F	*FANCA*	WES	c.2546delC	p.S849FfsX40	**c.3965T>G**	p.V1322G	GA	AA	24	Yes	No	5, 7, 8, 10
17	17	F	*FANCA*	WES	**c.190_191insT**	p.E65RfsX6	**c.190_191insT**	p.E65RfsX6	not examined	CVID	No	Yes	No	9
18	18-1	M	*FANCA*	Sanger	c.2546delC	p.S849FfsX40	**c.4042_4043insC**	p.I1348TfsX77	AA	MDS	0/12	Yes	Yes	5, 7, 10
	18-2	F	*FANCA*	Targeted-seq	c.2546delC	p.S849FfsX40	**c.4042_4043insC**	p.I1348TfsX77	GG	AML	69/69	Yes	No	7, 8, 10
19	19-1	M	*FANCA*	Targeted-seq	**c.283+2T>C**	aberrant splicing#	c.2730_2731delCT	p.W911DfsX31	GA	AA	30	Yes	No	7, 8, 10
	19-2	M	*FANCA*	Targeted-seq	**c.283+2T>C**	aberrant splicing#	c.2730_2731delCT	p.W911DfsX31	GA	AA	16	Yes	No	7, 8, 10
20	20	F	*FANCA*	Sanger	c.2546delC	p.S849FfsX40	**c.3781_3785delTTCTT**	p.F1261LfsX15	AA	MDS	7/7	Yes	No	5, 10
21	21	F	*FANCA*	Sanger	c.2546delC	p.S849FfsX40	**c.3931_3932delAG**	p.S1311X	GA	AA	21	Yes	No	5, 7, 10
22	22-1	F	*FANCA*	Sanger	c.2546delC	p.S849FfsX40	c.4168-2A>G	aberrant splicing	GG	AA	106	Yes	No	5, 10
	22-2	M	*FANCA*	Sanger	c.2546delC	p.S849FfsX40	c.4168-2A>G	aberrant splicing	GA	MDS	28/168	Yes	No	5, 10
23	23	F	*FANCA*	Sanger	c.2593delA	p.I879LfsX24	c.2840C>G	p.S947X	GA	AA/HNSCC	53/457	No	No	5, 10
24	24	M	*FANCA*	Sanger, MLPA	c.2546delC	p.S849FfsX40	ex1-3del	·	GA	AA	22	Yes	No	5, 7, 10
25	25	M	*FANCA*	Sanger	c.2602-2A>T	aberrant splicing	c.2527T>G	p.Y843D	GG	AA	78	Yes	No	5, 7, 10
26	26	M	*FANCA*	Sanger	c.2546delC	p.S849FfsX40	c.2546delC	p.S849FfsX40	GG	AA	114	Yes	No	5, 10
27	27	F	*FANCA*	Sanger	c.2602-2A>T	aberrant splicing	c.2602-2A>T	aberrant splicing	GG	AML	62/311	Yes	No	5, 7, 10

28	28	F	*FANCA*	Sanger	c.4124·4125delCA	p.T1375SfsX49	c.2290C>T	p.R764W	GG	AML	156/156	Yes	No	5, 10
29	29	F	*FANCA*	Sanger	c.2546delC	p.S849FfsX40	**c.3765+827_3814del**	·	GG	AA	72	Yes	No	5, 10
30	30	F	*FANCA*	Sanger	c.2546delC	p.S849FfsX40	c.2546delC	p.S849FfsX40	GG	AA	70	Yes	Yes	5, 7, 10
31	31	F	*FANCA*	Sanger	c.2546delC	p.S849FfsX40	**c.1567·1G>A**	aberrant splicing#	GG	MDS	82/82	Yes	No	5, 10
32	32	F	*FANCA*	Sanger	c.2546delC	p.S849FfsX40	c.3720_3724del	p.E1240DfsX36	GG	AA	88	Yes	No	5, 10
33	33	M	*FANCA*	Sanger	c.2546delC	p.S849FfsX40	c.3720_3724del	p.E1240DfsX36	GG	MDS	68/105	Yes	No	10
34	34	F	*FANCA*	Sanger	c.2546delC	p.S849FfsX40	c.2602·1G>A	aberrant splicing	GA	AML	60/282	Yes	No	10
35	35·1	M	*FANCA*	Sanger	c.2546delC	p.S849FfsX40	c.2546delC	p.S849FfsX40	AA	MDS	0/4	Yes	No	7, 10
	35·2	M	*FANCA*	Sanger	c.2546delC	p.S849FfsX40	c.2546delC	p.S849FfsX40	GA	AA	21	Yes	No	7, 10
36	36	M	*FANCA*	Sanger	c.44_69del	p.P15RfsX13	c.2170A>C	p.T724P	GG	MDS/HNSCC	108/348/348	Yes	No	5, 10
37	37	F	*FANCA*	Sanger	c.2546delC	p.S849FfsX40	**c.3295C>T**	p.Q1099X	GG	MDS	49/189	Yes	Yes	5, 10
38	38	M	*FANCA*	WES, MLPA	c.2840C>G	p.S947X	ex24·28del	·	GG	AA	60	Yes	No	5, 10
39	39	F	*FANCA*	Targeted·seq, MLPA	**c.462T>G**	p.Y154X	ex6del	·	GG	AA	unknown	No	No	8
40	40	F	*FANCA*	Sanger	c.2602·1G>A	aberrant splicing	not detected	·	GG	AML	108/384	Yes	No	5, 10
41	41	M	*FANCA*	Sanger, MLPA	c.2546delC	p.S849FfsX40	ex37del	·	GG	AML	136/176	Yes	No	5, 10
42	42	F	*FANCA*	WES, MLPA	c.4199G>C	p.R1400P	ex16_17del	·	GG	AML	61/61	Yes	No	5, 10
43	43	M	*FANCA*	Targeted·seq	c.2T>C	p.M1T	c.15G>A	p.W5X	GA	AA	37	No	No	7, 10
44	44	M	*FANCA*	Targeted·seq	c.2546delC	p.S849FfsX40	c.2972delT	p.F991SfsX35	GA	MDS	50/73	Yes	No	7, 10
45	45·1	F	*FANCA*	Targeted·seq, MLPA	ex1_43del	·	ex19_29 del	·	GG	AA	108	Yes	No	10
	45·2	F	*FANCA*	Targeted·seq, MLPA	ex1_43del	·	ex19_29 del	·	GA	AA	12	Yes	No	10
46	46	F	*FANCA*	WES, MLPA	c.2546delC	p.S849FfsX40	ex1_5del	·	GG	AA	unknown	Yes	No	New case
47	47	M	*FANCA*	WES	c.2546delC	p.S849FfsX40	c.2546delC	p.S849FfsX40	GA	unknown	unknown	unknown	unknown	8
48	48	F	*FANCA*	WES	c.4015_4017delCTC	p.L1339del	c.3638_3639delCT	p.P1213RfsX64	GA	unknown	unknown	unknown	unknown	8
49	49	F	*FANCA*	WES	c.2546delC	p.S849FfsX40	c.2546delC	p.S849FfsX40	GG	AA	71	Yes	No	8
50	50	F	*FANCA*	WES	c.2546delC	p.S849FfsX40	c.2546delC	p.S849FfsX40	GG	AA	71	Yes	No	8
51	51	F	*FANCA*	WES	**c.1464C>A**	p.Y488X	**c.1464C>A**	p.Y488X	GG	AA	157	Yes	No	8
52	52	F	*FANCA*	WES, MLPA	c.978_979delGA	p.Q326HfsX12	ex30del	·	GG	unknown	unknown	unknown	unknown	New case
53	53	M	*FANCA*	WES, MLPA	c.978_979delGA	p.Q326HfsX12	ex30del	·	GG	unknown	unknown	unknown	unknown	New case
54	54	F	*FANCA*	WES	c.2546delC	p.S849FfsX40	not detected	·	not examined	AA	85	No	No	8
55	55	F	*FANCA*	WES, MLPA	c.2546delC	p.S849FfsX40	ex30_31del	·	GA	AA	80	Yes	No	8
56	56	F	*FANCA*	WES	**c.2316+2T>A**	aberrant splicing#	not detected	·	GG	AA	59	Yes	No	8
57	57	F	*FANCA*	Sanger	c.2546delC	p.S849FfsX40	not detected	·	GA	MDS	unknown/234	No	No	New case
58	58	F	*FANCA*	Sanger	c.2546delC	p.S849FfsX40	not detected	·	GG	AA	82	Yes	No	New case
59	59	F	*FANCA*	Sanger	c.2546delC	p.S849FfsX40	not detected	·	GG	AA	80	Yes	No	New case
60	60	M	*FANCB*	aCGH, Sanger	**chrX: g.14730104·14904216del**	complete loss	·	·	GG	AA	58	Yes	Yes	5 Current study identified causative FA gene mutation.
61	61	M	*FANCB*	aCGH, Sanger	**chrX: g.14810970·14932973del**	complete loss	·	·	GA	MDS	24/51	Yes	Yes	5, 10 Current study identified causative FA gene mutation.

62	62	M	*FANCB*	WES, RNA-seq	**c.1497G>T**	aberrant splicing (p.S500AfsX14)	·	·	GG	AA	96	Yes	No	5 Current study identified causative FA gene mutation.
63	63	M	*FANCB*	WES	**c.516G>A**	p.W172X	·	·	not examined	unknown	unknown	unknown	unknown	8
64	64	F	*FANCC*	WGS	**c.1154+5G>A**	aberrant splicing (p.S386X)	**c.1154+5G>A**	aberrant splicing (p.S386X)	GG	AA	40	Yes	Yes	5, 7 Current study identified causative FA gene mutation.
65	65	F	*FANCD1*	WES	**c.517-2A>G**	aberrant splicing#	**c.6952C>T**	p.R2318X	GG	immmature teratoma	No/9	Yes	No	8
66	66	M	*FANCD1*	Targeted-seq	**c.475+1G>A**	aberrant splicing#	**c.7847C>T**	p.S2616F	GA	T-LBL, Lung cancer	No/508	No	No	8
67	67	F	*FANCE*	WES	**c.419T>C**	p.L140P	**c.648delC**	p.R219DfsX77	GA	CVID	No	Yes	No	9
68	68	F	*FANCF*	WES	**c.484_485delCT**	p.L162DfsX103	**c.66C>A**	p.Y22X	GG	AA	43	Yes	No	8
69	69	M	*FANCG*	WES	c.1066C>T	p.Q356X	c.307+1G>C	aberrant splicing	GA	MDS	12/61	Yes	Yes	5, 8, 10
70	70	F	*FANCG*	WES	c.1066C>T	p.Q356X	c.1066C>T	p.Q356X	GG	AA	66	Yes	No	5, 8, 10
71	71	M	*FANCG*	WES	c.1066C>T	p.Q356X	c.1066C>T	p.Q356X	GG	AA	72	Yes	No	5, 8, 10
72	72-1	M	*FANCG*	WES	c.91C>T	p.Q31X	c.307+1G>C	aberrant splicing	GG	AA	27	Yes	No	5, 8, 10
	72-2	M	*FANCG*	WES	c.91C>T	p.Q31X	c.307+1G>C	aberrant splicing	GG	AA	60	Yes	No	5, 10
73	73-1	M	*FANCG*	WES	c.307+1G>C	aberrant splicing	c.307+1G>C	aberrant splicing	GA	AA	48	Yes	Yes	5, 8, 10
	73-2	M	*FANCG*	WES	c.307+1G>C	aberrant splicing	c.307+1G>C	aberrant splicing	GA	AA	39	Yes	No	5, 10
74	74	F	*FANCG*	WES	c.1066C>T	p.Q356X	c.1066C>T	p.Q356X	GA	AA	24	Yes	No	5, 7, 8, 10
75	75	M	*FANCG*	WES	**c.907_908del**	p.L303GfsX5	c.307+1G>C	aberrant splicing	GA	AA	21	Yes	No	5, 8, 10
76	76	F	*FANCG*	WES	c.307+1G>C	aberrant splicing	c.307+1G>C	aberrant splicing	GA	AA	69	Yes	No	5, 8, 10
77	77	F	*FANCG*	WES	c.307+1G>C	aberrant splicing	c.307+1G>C	aberrant splicing	GA	AA	18	Yes	No	5, 7, 8, 10
78	78	F	*FANCG*	WES	c.307+1G>C	aberrant splicing	c.307+1G>C	aberrant splicing	GG	MDS	78/78	Yes	No	7, 8, 10
79	79	F	*FANCG*	WES	c.307+1G>C	aberrant splicing	c.307+1G>C	aberrant splicing	GG	AA	37	Yes	No	10
80	80	M	*FANCG*	WES	c.1066C>T	p.Q356X	c.1066C>T	p.Q356X	GA	AA	24	Yes	No	10
81	81	F	*FANCG*	WES	c.307+1G>C	aberrant splicing	c.307+1G>C	aberrant splicing	AA	MDS	0/18	Yes	No	7, 8, 10
82	82	M	*FANCG*	WES	c.307+1G>C	aberrant splicing	c.307+1G>C	aberrant splicing	AA	AA	0	Yes	No	7, 8, 10
83	83	F	*FANCG*	WES	c.1066C>T	p.Q356X	c.307+1G>C	aberrant splicing	GA	MDS	24/38	Yes	No	7, 8, 10
84	84	M	*FANCG*	Sanger	c.307+1G>C	aberrant splicing	c.307+1G>C	aberrant splicing	GA	AA	38	Yes	No	5, 10
85	85	M	*FANCG*	Sanger	c.307+1G>C	aberrant splicing	c.307+1G>C	aberrant splicing	GA	AA	36	Yes	No	5, 10
86	86	F	*FANCG*	Sanger	c.307+1G>C	aberrant splicing	c.1066C>T	p.Q356X	GG	AA	50	Yes	No	5, 7, 10
87	87	M	*FANCG*	Sanger	c.307+1G>C	aberrant splicing	c.307+1G>C	aberrant splicing	GG	AA	28	No	No	5, 7, 10
88	88	M	*FANCG*	WES	**c.1386delC**	p.W463GfsX55	**c.1637-15G>A**	VUS	GG	MDS	69/120	Yes	No	5, 7, 10
89	89	F	*FANCG*	Sanger	c.307+1G>C	aberrant splicing	c.307+1G>C	aberrant splicing	GG	AA	46	Yes	No	7, 10
90	90	M	*FANCG*	WES	c.194delC	p.P65LfsX7	c.307+1G>C	aberrant splicing	GG	AA	29	Yes	No	8
91	91	F	*FANCG*	WES	c.1066C>T	p.Q356X	c.307+1G>C	aberrant splicing	GG	AA/MDS	30/66	unknown	unknown	8
92	92	M	*FANCG*	Sanger	c.194delC	p.P65LfsX7	c.194delC	p.P65LfsX7	GA	AA	unknown	Yes	No	New case
93	93	M	*FANCG*	Sanger	c.307+1G>C	aberrant splicing	c.307+1G>C	aberrant splicing	GG	AA	67	Yes	No	New case

94	94	M	*FANCG*	Sanger	c.1066C>T	p.Q356X	not detected	·	GG	MDS	unknown/396	Yes	No	New case
95	95	M	*FANCG*	Sanger	c.1066C>T	p.Q356X	c.1066C>T	p.Q356X	GA	AA	75	Yes	No	New case
96	96	M	*FANCI*	WES	**c.158·2A>G**	aberrant splicing (p.S54FfsX5)	**c.288G>A**	aberrant splicing (p.C56FfsX8)	GA	AA	7	Yes	Yes	5 Current study identified causative FA gene mutations.
97	97	M	*FANCI*	WES	**c.3346_3347insT**	p.S1116FfsX16	**c.3006+3A>G**	aberrant splicing#	GA	AA	15	Yes	No	5, 7 Current study identified causative FA gene mutations.
98	98	M	*FANCN*	WES, RNA-seq	**c.3350+5C>T**	aberrant splicing (p.G1068VfsX5)	**c.3350+5C>T**	aberrant splicing (p.G1068VfsX5)	GA	Wilms tumor	No/12	Yes	No	10 Current study identified causative FA gene mutation.
99	99·1	M	*FANCP*	WES	**c.343delA**	p.S115AfsX11	**c.343delA**	p.S115AfsX11	AA	MDS	0/4	Yes	Yes	5, 7, 8, 10
	99·2	F	*FANCP*	Sanger	**c.343delA**	p.S115AfsX11	**c.343delA**	p.S115AfsX11	GG	Normal	No at 8	Yes	No	7
100	100	F	*FANCT*	WES, aCGH	c.4C>G	p.Q2E	chr1:g202288583_202309772del	complete loss	GA	AA	87	Yes	No	5, 6, 7
101	101	M	*FANCT*	WES	c.4C>G	p.Q2E	c.180+5G>A	aberrant splicing (p.E37RfsX49)	GA	AML	41/41	Yes	No	5, 6, 10
102	102	F	unclassified	·	·	·	·	·	GA	MDS	12/108	Yes	No	5, 10
103	103·1	F	unclassified	·	·	·	·	·	not examined	unknown	unknown	unknown	unknown	New case
	103·2	M	unclassified	·	·	·	·	·	not examined	unknown	unknown	unknown	unknown	New case
104	104	M	unclassified	·	·	·	·	·	not examined	unknown	unknown	unknown	unknown	New case

Novel mutations (not included in the Rockefeller University FA mutation data base) are indicated in boldface type.
#Effects of these splicing mutations are unverified.
*The *ALDH2* wild type and the inactivating mutation (p.Glu504Lys) allele is referred to as G and A, respectively[5]. **FA features include physical abnormalities such as short stature, malformations or skin pigmentation.

AA, aplastic anemia; *ALDH2*, aldehyde dehydrogenase·2; AML, acute myeloid leukemia; aCGH, array·CGH; BMF, bone marrow failure; CVID, common variable immunodeficiency; FA, Fanconi anemia; HNSCC, head and neck squamous cell carcinoma; MDS, myelodysplastic syndrome; MLPA, multiplex ligation·dependent probe amplification; T·LBL, T cell lymphoblastic lymphoma; VACTERL·H, VACTERL·H, vertebral anomalies, anal atresia, cardiac anomalies, tracheal·esophageal fistula, esophageal atresia, renal structural abnormalities, limb anomalies, and hypocephalus; VUS, variation of unknown significant; WES, whole·exome sequencing; WGS, whole·genome sequencing

Table 2. Allele frequency of FA-associated deleterious mutation variants* in Japanese population

Gene	Genomic location (hg19)	Reference allele /Alternative allele	cDNA	Protein	Frequency
FANCA	chr16:89882954	CGGCCCAGGCCCTCCGGCGGCCCCCTG/C	c.77_102del	p.P15fs	0.0001
FANCA	chr16:89833603	AG/A	c.2546delC	p.S849fs	0.0008
FANCA	chr16:89831476	T/A	c.2602-2A>T	aberrant splicing	0.0001
FANCA	chr16:89805357	TTTG/T	c.4189_4191del	p.T1397del	0.0001
FANCC	chr9:97897635	G/C	c.836C>G	p.S279X	0.0001
FANCC	chr9:97864024	G/A	c.1642C>T	p.R548X	0.0003
FANCD1 (BRCA2)	chr13:32903604	CTG/C	c.657_658del	p.Val220fs	0.0001
FANCD1 (BRCA2)	chr13:32911557	ATAACAT/A	c.3067_3072del	p.N1023_I1024del	0.0001
FANCD1 (BRCA2)	chr13:32911577	A/AT	c.3085_3086insT	p.M1029fs	0.0001
FANCD1 (BRCA2)	chr13:32913261	AGT/A	c.4770_4771del	p.C1591fs	0.0001
FANCD1 (BRCA2)	chr13:32914065	CAATT/C	c.5574_5577del	p.I1859fs	0.0003
FANCD1 (BRCA2)	chr13:32914209	ACT/A	c.5718_5719del	p.L1908fs	0.0001
FANCD1 (BRCA2)	chr13:32914893	ATAACT/A	c.6402_6406del	p.N2135fs	0.0001
FANCD1 (BRCA2)	chr13:32920978	C/T	c.6952C>T	p.R2318X	0.0003
FANCD1 (BRCA2)	chr13:32930713	AG/A	c.7585delG	p.G2529fs	0.0001
FANCD1 (BRCA2)	chr13:32972800	C/T	c.10150C>T	p.R3384X	0.0004
FANCD2	chr3:10122879	T/TA	c.3072_3073insA	p.N1025fs	0.0001
FANCD2	chr3:10130510	A/C	c.3561-2A>C	aberrant splicing	0.0001
FANCE	chr6:35425734	GCTT/G	c.943_945del	p.L316del	0.0001
FANCG	chr9:35078714	AG/A	c.194delC	p.P65fs	0.0001
FANCG	chr9:35078601	C/G	c.307+1G>C	aberrant splicing	0.001
FANCG	chr9:35076439	G/A	c.1066C>T	p.Q356X	0.0004
FANCI	chr15:89801943	TCTC/T	c.94_96del	p.L33del	0.0001
FANCI	chr15:89803942	A/G	c.157-2A>G	aberrant splicing	0.0008
FANCI	chr15:89833476	G/GC	c.1854_1855insC	p.L619fs	0.0001
FANCI	chr15:89843085	GAA/G	c.2692_2693del	p.K898fs	0.0001
FANCI	chr15:89843605	C/CGGCAAT	c.2878_2879insGGCAAT	p.Q961_F962insWE	0.0004
FANCI	chr15:89850868	A/AC	c.3616_3617insC	p.L1208fs	0.0003
FANCJ (BRIP1)	chr17:59763487	G/C	c.2615C>G	p.S872X	0.0001
FANCJ (BRIP1)	chr17:59761334	AC/A	c.3072delG	p.S1025fs	0.0001
FANCJ (BRIP1)	chr17:59761166	C/CA	c.3240_3241insT	p.A1081fs	0.0003
FANCL	chr2:58456995	C/T	c.170G>A	p.W57X	0.0008
FANCL	chr2:58453870	ATCT/A	c.263_265del	p.K88del	0.0003
FANCL	chr2:58453867	AG/A	c.268delC	p.L90fs	0.0001
FANCL	chr2:58387305	C/CT	c.1044_1045insA	p.G349fs	0.0001
FANCM	chr14:45642287	A/ACT	c.2190_2191insCT	p.E735fs	0.0001
FANCM	chr14:45644477	TAAAC/T	c.2521_2522insAAAC	p.Q842fs	0.0001
FANCM	chr14:45650888	CGCAGA/C	c.4367_4371del	p.R1456fs	0.0001

FANCM	chr14:45658082	TGAA/T	c.4858_4860del	p.E1620del	0.0001
FANCM	chr14:45668139	G/A	c.6008+1G>A	aberrant splicing	0.0003
FANCN (PALB2)	chr16:23647568	AG/A	c.298delC	p.D101fs	0.0001
FANCN (PALB2)	chr16:23647395	G/A	c.472C>T	p.Q158X	0.0001
FANCN (PALB2)	chr16:23646369	AC/A	c.1497delG	p.L499fs	0.0001
FANCN (PALB2)	chr16:23646192	G/A	c.1675C>T	p.Q559X	0.0003
FANCN (PALB2)	chr16:23641004	CA/C	c.2470delT	p.C824fs	0.0001
FANCN (PALB2)	chr16:23635328	A/G	c.2834+2T>C	aberrant splicing	0.0001
FANCO(RAD51C)	chr17:56787352	G/C	c.837+1G>C	aberrant splicing	0.0001
FANCP (SLX4)	chr16:3651155	CAGA/C	c.985_987del	p.Ser329del	0.0001
FANCP (SLX4)	chr16:3647443	C/T	c.1620G>A	p.W540X	0.0001
FANCP (SLX4)	chr16:3644451	TA/T	c.2160+2delT	aberrant splicing	0.0003
FANCP (SLX4)	chr16:3640407	CAGCTGG/C	c.3226_3231del	p.P1076_A1077del	0.0001
FANCP (SLX4)	chr16:3639742	CCT/C	c.3895_3896del	p.R1299fs	0.0001
FANCP (SLX4)	chr16:3639379	T/TG	c.4259_4260insC	p.I1421fs	0.0001
FANCQ (ERCC4)	chr16:14042182	C/G	c.2729C>G	p.S910X	0.0001
FANCR(RAD51)	chr15:40994106	C/T	c.328C>T	p.R110X	0.0003
FANCS (BRCA1)	chr17:41258497	A/T	c.188T>A	p.L63X	0.0003
FANCS (BRCA1)	chr17:41245553	G/GAAA	c.1995_1997insTTT	p.N665_L666insF	0.0001
FANCS (BRCA1)	chr17:41244748	G/A	c.2659C>T	p.Q887X	0.0001
FANCS (BRCA1)	chr17:41244333	AG/A	c.3214delC	p.L1072fs	0.0001
FANCS (BRCA1)	chr17:41226421	C/CA	c.4664_4665insT	p.E1556fs	0.0001
FANCT (UBE2T)	chr1:202304773	C/T	c.109+1G>A	aberrant splicing	0.0004
FANCU (XRCC2)	chr7:152346394	TA/T	c.175delT	p.T59fs	0.0001
FANCW (RFWD3)	chr16:74695317	G/A	c.31C>T	p.Q11X	0.0001
FANCW (RFWD3)	chr16:74685992	G/GA	c.546_547insT	p.Q183fs	0.0001
FANCW (RFWD3)	chr16:74678352	C/T	c.988-1G>A	aberrant splicing	0.0001
FANCW (RFWD3)	chr16:74660405	G/A	c.2017C>T	p.R673X	0.0001

These data were obtained from 3.5KJPNv2 database (https://jmorp.megabank.tohoku.ac.jp/201811/).

*We focused on nonsense mutations, frameshift mutations, and splicing site mutations. Missense mutation variants were not included.

Table 3. Clinical phenotype of 10 Japanese FA patients with VACTERL-H association

Individual	Affected gene	Mutations	VACTERL-H features	FA features	Family history of FA*	Birth weight SD score	DEB induced chromosome breakage (breaks /cell)	*ALDH2* genotype	Hematological findings				
									Onset of BMF (months)	BM status at HSCT	Karyotype of BM	Age at HSCT (months)	Outcome after HSCT (months)
Case 18-1	*FANCA*	c.2546delC: p.S849FfsX40 c.4042_4043insC: p.I1348TfsX77	C: PDA R: Left renal agenesis L: Bilateral absent thumbs/Bilateral radial hypoplasia	Short stature	+	-1.9	0.44	AA	0	RCMD	46,XY,add(2)(q33)	13	Alive (105)
Case 30	*FANCA*	c.2546delC: p.S849FfsX40 c.2546delC: p.S849FfsX40	V: scoliosis C: ASD/Persistent left superior vena E: Esophageal atresia	Skin pigmentation Deafness Right inguinal hernia Bicornuate uterus Short stature (-1.8SD)	-	-2.1	2.06	GG	70	SAA	46,XX	153	Dead/Esophageal cancer (165)
Case 37	*FANCA*	c.2546delC: p.S849FfsX40 c.3295C>T: p.Q1099X	E: Esophageal atresia R: Right pelvic kidney with malrotation L: Bilateral thumb hypoplasia	Jejunal atresia Strabismus Short stature (-4SD)	-	-2.3	0.12	GG	49	RAEB1	46,XX, complex	192	Alive (66)
Case 60	*FANCB*	complete loss of *FANCB* gene (chrX g.14730104-14904216 del) .	V: Spina bifida occulta/Abnormal ribs A: Anal atresia C: PDA R: Right renal agenesis L: Right absent thumb/Partial loss of left thumb	Skin pigmentation Microphthalmus/Stenocephaly/Ptosis Duodenal stenosis Annular pancreas/Intestinal malrotation Hypospadias/Undescended testis Short stature (-6SD)	-	-4.8	3.8	GG	58	SAA	46,XY	72	Alive (167)
Case 61	*FANCB*	complete loss of FANCB gene (chrX g.14810970-14932973 del)	V: Abnormal ribs/Scoliosis A: Anal atresia C: VSD/PS E: Duodenal atresia** R: Left renal agenesis	Skin pigmentation Microphthalmus/Deafness/Ear canal stenosis Undescended testis (Short stature (-1SD))	-	-2.8	4.2	GA	24	RCMD	46,XY,add(5)(p15)	51	Alive (160)

		·	L: Bilateral absent thumbs H: Hydrocephalus										
Case 64	*FANCC*	c.1154+5G>A: p.S386X c.1154+5G>A: p.S386X	A: Anal atresia C: VSD, PDA E: Esophageal atresia	Skin pigmentation Deafness/Left aural stenosis/Right aural atresia Cleft palate Short stature (<2SD)	·	<2.53	7.8	GG	40	SAA	46,XX	61	Alive (73)
Case 69	*FANCG*	c.307+1G>C c.1066C>T: p.Q356X	C: Coarctation complex R: Right renal agenesis/Left renal cyst L: Bilateral absent thumbs/Right radial hypoplasia	Skin pigmentation Short stature (<8SD)	·	<1.7	8.54	GA	12	RCMD	46,XY	62	Alive (144)/Tongue SCC at 14 years old
Case 73-1	*FANCG*	c.307+1G>C c.307+1G>C	C: PDA R: Left renal agenesis L: Right absent thumb/Bilateral radial hypoplasia	Skin pigmentation Bilateral aural atresia Short stature (<2.7SD)	+	<0.9	3.49	GA	48	SAA	46,XY	88	Dead/Oral SCC (111)
Case 96	*FANCI*	c.158-2A>G:p.S54FfsX5 c.288G>A:p.C56FfsX8	A: Anal atresia C: VSD/PDA R: Right renal agenesis/ Left renal hypoplasia L: Bilateral absent thumb/Bilateral absent radius H: Hydrocephalus	Skin pigmentation Microphthalmus Hypogenitalia Short stature (<8SD)	·	<3.9	0.52	GA	7	SAA	46,XY	45	Dead (2)
Case 99-1	*FANCP*	c.343delA: p.S115AfsX11 c.343delA: p.S115AfsX11	C: ASD/VSD/PS R: horseshoe kidney L: Bilateral floating thumbs/bilateral radial hypoplasia	Intestinal malrotation Duodenal stenosis Short stature (<5.8SD)	+	<2.3	0.91	AA	0	RCMD	46,XY,+del(3)(q12)	13	Alive (59)

* Case 18-1, 73-1, and 99-1 had a sibling with FA.
** Duodenal atresia is considered to be a part of the VACTERL association by some reports[28].
ALDH2, aldehyde dehydrogenase-2; ASD, atrial septal defect; BM, bone marrow; BMF, bone marrow failure; DEB, diepoxybutane; FA, Fanconi anemia; HSCT, hematopoietic stem cell transplantation; PDA, patent ductus arteriosus; PS, pulmonary stenosis; RAEB, refractory anemia with excess of blasts; RCMD, refractory cytopenia with multilineage dysplasia; SAA, severe aplastic anemia; SCC; squamous cell carcinoma; SD, standard deviation; VACTERL-H, vertebral anomalies, anal atresia, cardiac anomalies, tracheal-esophageal fistula, esophageal atresia, renal structural abnormalities, limb anomalies, and hypocephalus; VSD, ventricular septal defect

Table 4. Clinical features of Japanese FA-D1 and FA-N cases

Individual	Case 65	Case 66	Case 98	AP37P*
Sex	Female	Male	Male	Male
FA mutations	*FANCD1*	*FANCD1*	*FANCN*	*FANCD1*
	c.517-2A>G,	c.475+1G>A	c.3350+5C>T	c.-40+1G>A,
	c.6952C>T: p.R2318X	c.7847C>T:p.S2616F	c.3350+5C>T	c.8504C>A: p.S2835X
FA-features	Short stature	Short stature	Short stature	Short stature
	Left thumb polydactyly	Microcephaly	ASD, PDA	Mid-face hypoplasia
	Right renal agenesis		Congenital absence of inferior vena cava,	Sprengel's deformity
	Microphthalmus		Congenital tracheal stenosis	Multiple café-au-lait spots
	Microcephaly		Microcephaly	
Chromosome breakage test	Positive (MMC)	Positive (MMC)	Positive (DEB)	Positive (MMC)
ALDH2 genotype	GG	GA	GA	GG
Hematological abnormality (onset)	None	None	None	Acute myeloid leukemia (2-year-old)
Solid tumors (onset)	Immature teratoma (9-month-old)	T-lymphoblastic lymphoma, Adenosquamous lung carcinoma (23-year-old)	Wilms tumor (1-year-old)	None
Outcome	Alive with progressive teratoma at 1.7 years old	Died of lymphoma at 25.5 years old	Died of Wilms tumor at 1.5 years old	Died of leukemia at 2 years old

* a previously reported case[32]

MMC, mitomycin C. Other abbreviations are explained in Table 3.

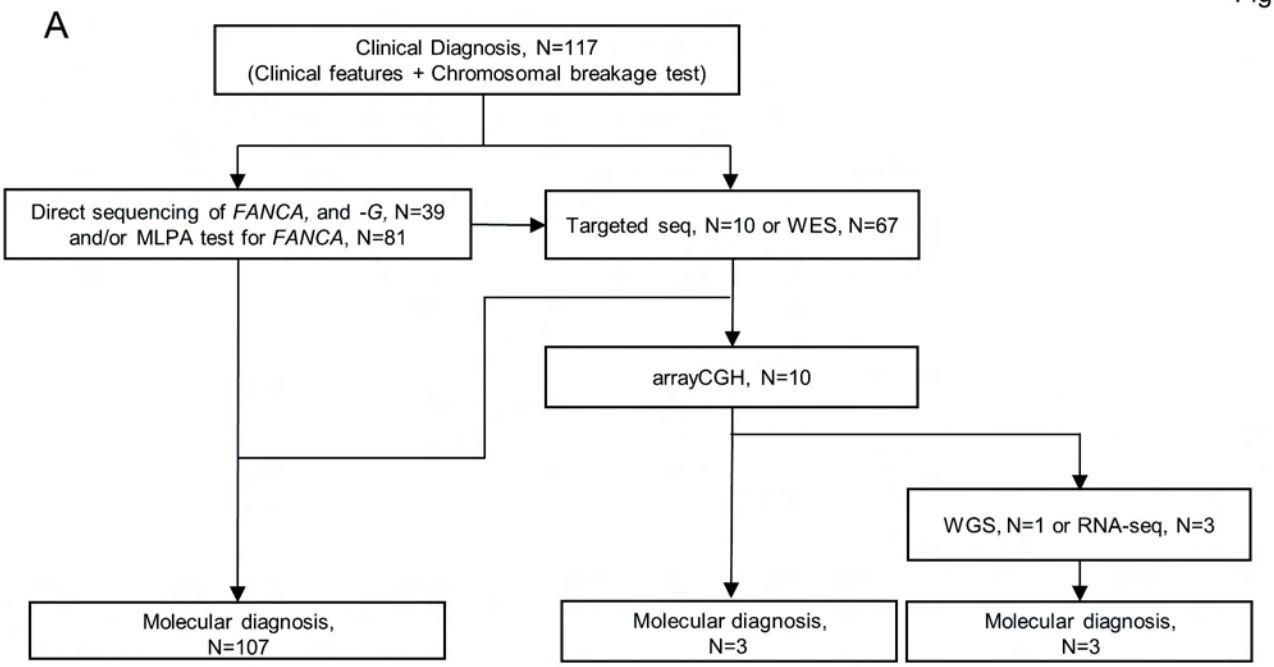
Figure legends

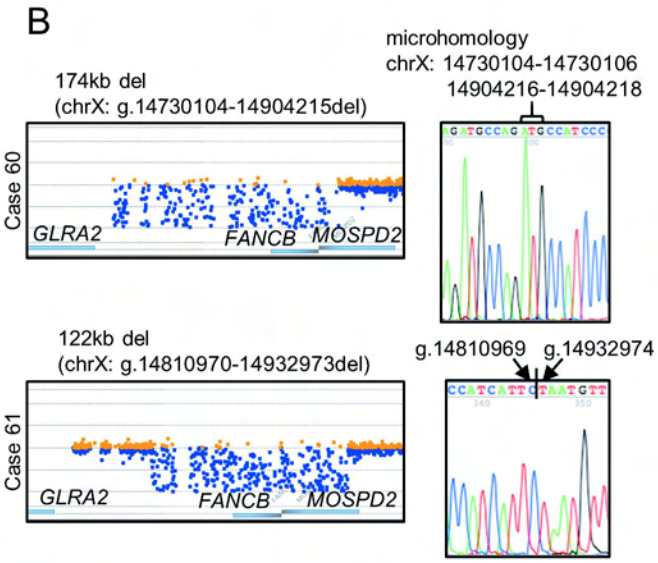
Figure 1. A comprehensive analysis successfully subtyped most of the Japanese FA patients (A) A schematic presentation of the diagnostic strategy for the 117 FA patients (B) The aCGH data displayed complete loss of the *FANCB* gene in Case 60 and Case 61. Sanger sequencing data identified the precise junctions in the two cases. (C) The WGS analysis detected homozygous *FANCC* mutations in intron 12, resulting in a splicing defect. The Sanger sequencing data (left side) identified the homozygous mutations in the patient (Case 64) and the heterozygous mutation in the patient's mother. The RT-PCR analysis showed a larger product (arrowhead) than the wild-type product, and sequencing analysis of the RT-PCR product (right side) revealed the 120bp intron retention (*) after exon 12, resulting in a stop codon. (D) The RNA sequence reads of exon 7 in *FANCB* and exon 12 in *FANCN* were absent for Case 62 and Case 98, respectively. Corresponding WES read alignments for Case 62 and Case 98 were diagnostic for the *FANCB* or *FANCN* mutations as shown in Supplemental Fig 2A and 2B.

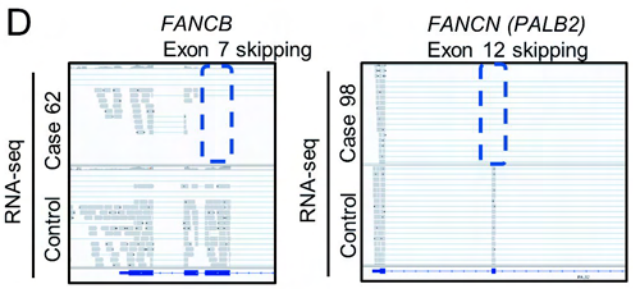
Figure 2. Frequency distribution of total (A) versus unique (B) FA gene mutations in the 117 Japanese FA patients. The frequency of the total FA gene mutation was based on subtyping of 117 FA cases, while frequency of unique FA gene mutations was derived from 84 mutation variants detected in the 117 FA patients

Figure 3. The two *FANCI* mutation variants in Case 96 caused two types of splicing defects. RT-PCR analysis was carried out using a forward flanking primer on exon 3 and a reverse flanking primer on exon 5 as indicated. Two types of products were obtained, and the sequencing analyses revealed a single nucleotide insertion (top) and exon 4 skipping (bottom)

Figure 1


A
Clinical Diagnosis, N=117
(Clinical features + Chromosomal breakage test)
Direct sequencing of FANCA, and -G, N=39
and/or MLPA test for FANCA, N=81
Targeted seq, N=10 or WES, N=67
arrayCGH, N=10
WGS, N=1 or RNA-seq, N=3
Molecular diagnosis, N=107
Molecular diagnosis, N=3
Molecular diagnosis, N=3


B
174kb del
(chrX: g.14730104-14904215del)
microhomology
chrX: 14730104-14730106
14904216-14904218
Case 60
GLRA2
FANCB
MOSPD2
122kb del
(chrX: g.14810970-14932973del)
g.14810969
g.14932974
Case 61
GLRA2
FANCB
MOSPD2


D
FANCB
Exon 7 skipping
FANCN (PALB2)
Exon 12 skipping
RNA-seq
Case 62
Control
RNA-seq
Case 98
Control

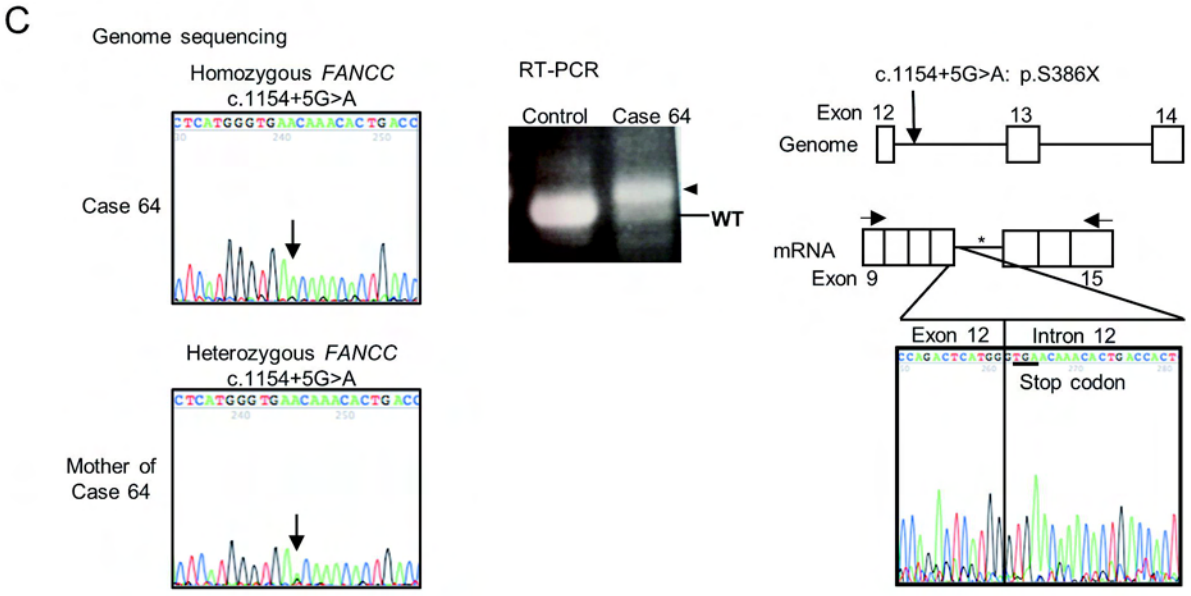

C
Genome sequencing
Homozygous FANCC
c.1154+5G>A
Case 64
Heterozygous FANCC
c.1154+5G>A
Mother of
Case 64
RT-PCR
Control
Case 64
WT
c.1154+5G>A: p.S386X
Exon 12
13
14
Genome
mRNA
Exon 9
15
Exon 12
Intron 12
Stop codon

Figure 2

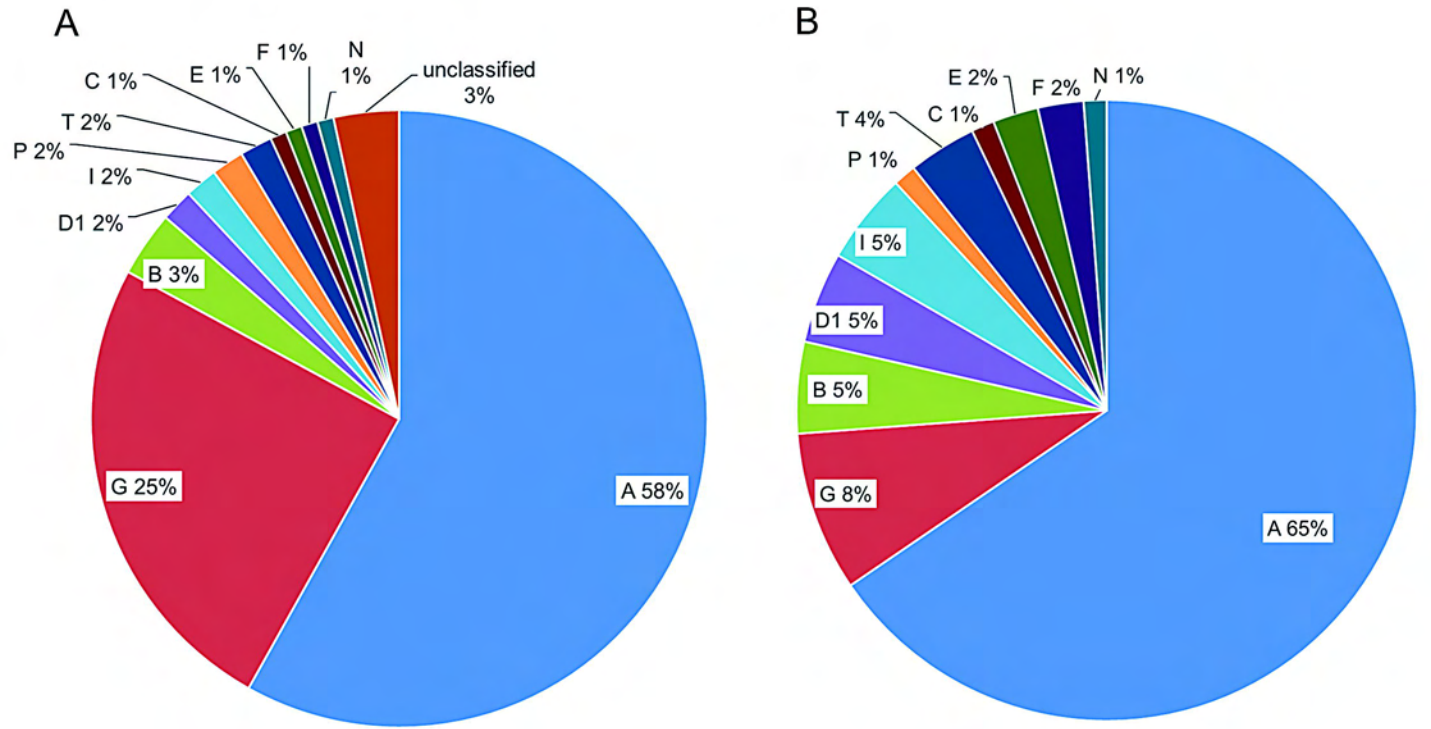
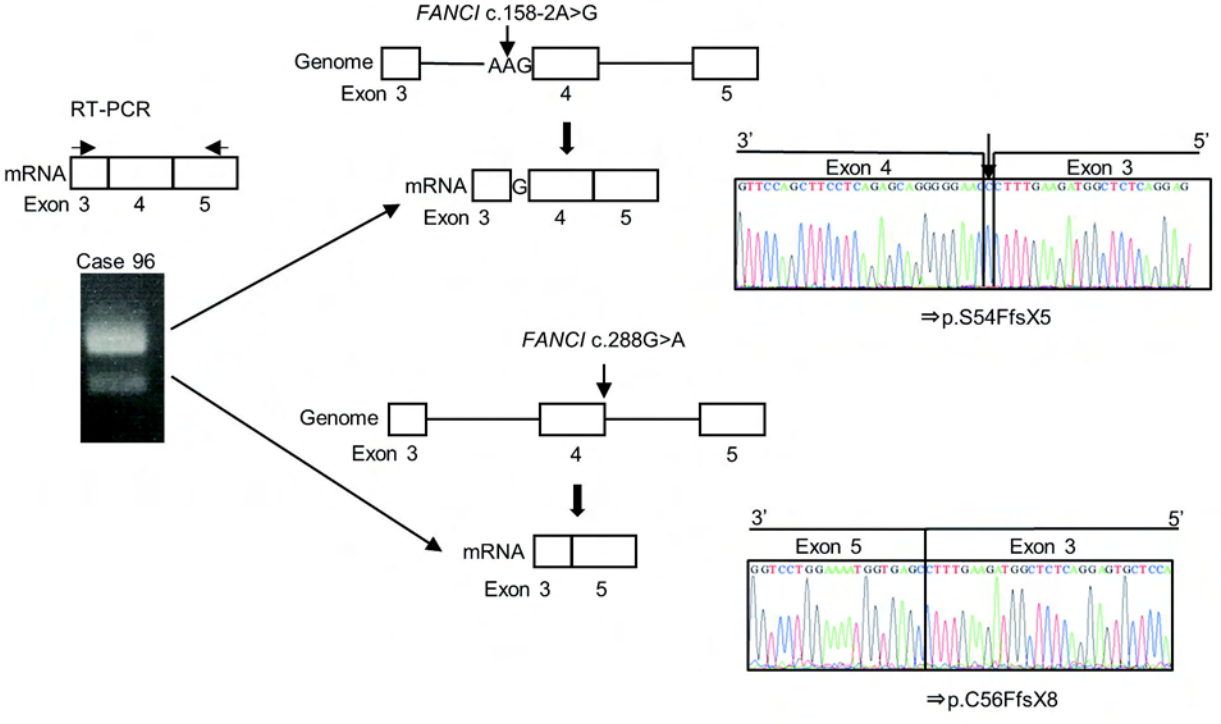

A
A 58%
G 25%
B 3%
D1 2%
I 2%
P 2%
T 2%
C 1%
E 1%
F 1%
N 1%
unclassified 3%
B
A 65%
G 8%
B 5%
D1 5%
I 5%
P 1%
T 4%
C 1%
E 2%
F 2%
N 1%

Figure 3


FANCI c.158-2A>G
Genome
AAG
Exon 3
4
5
RT-PCR
mRNA
Exon 3 4 5
mRNA
G
Exon 3 4 5
3'
5'
Exon 4
Exon 3
⇒p.S54FfsX5
Case 96
FANCI c.288G>A
Genome
Exon 3
4
5
mRNA
Exon 3 5
3'
5'
Exon 5
Exon 3
⇒p.C56FfsX8

Supplemental Data

Supplemental Note: Case 66 Presentation

A 23-year-old man was admitted to the hospital because of a giant mediastinal tumor. He was born to unrelated healthy parents and had no significant past medical history. Physically, he presented with short stature (155cm, -2.7SD) and severe microcephaly (49.4cm, -5SD). Hematological and bone marrow examination were normal (neutrophils, 3.88×10^9/L; hemoglobin, 14.1 g/dl; hematocrit, 42.2%; reticulocytes, 1.5%; platelets 244×10^9/L). On the basis of a CT scan and mediastinal tumor biopsy, he was diagnosed with mediastinal T-cell lymphoblastic lymphoma (T-LBL). Induction chemotherapy consisting of cyclophosphamide, vincristine, daunorubicin, prednisone, and l-asparaginase was performed and he suffered from severe sepsis due to prolonged pancytopenia. After hematological recovery, a mitomycin C induced chromosomal breakage test showed an increased rate of chromosomal breakage. Targeted-exome sequencing identified the splice site mutation c.475+1G>A and the missense mutation c.7847C>T in *FANCD1 (BRCA2)*. He was diagnosed as Fanconi anemia.

After induction chemotherapy, the T-LBL achieved a partial remission but he was found to have adenosquamous lung cancer and bilateral renal tubule-papillary adenoma. He underwent focal radiation therapy to the mediastinal lesion and reduced-intensity chemotherapy. However, he relapsed and died of T-LBL 15 months after the initial chemotherapy.

Supplemental Table 1. Summary of 22 FA-related genes

FA gene (Alternative name)	OMIM No.	FA /FBOC/FA-like*	Functions	Distribution of genetic subtyping		
				this study (2018)	Rockefeller Fanconi Anemia Mutation Database (2008)[1]	National Network of the Italian Association of Pediatric Hematology and Oncology (2014)[2]
FANCA	607139	FA	Component of the FA core complex	58%	57%	85%
FANCB	300515	FA	Component of the FA core complex	3%	0.9%	1%
FANCC	613899	FA	Component of the FA core complex	1%	15%	3%
FANCD1 (BRCA2)	600185	FBOC, FA	HR repair, mediator fucntion for RAD51, Protects stalled replication fork	2%	2.9%	0%
FANCD2	613984	FA	Monoubiquitylated by the FA core complex, Forms ID2 complex, Regulates the DNA damage response	0%	3.9%	2%
FANCE	613976	FA	Component of the FA core complex	1%	2.3%	0%
FANCF	613897	FA	Component of the FA core complex	1%	2%	0%
FANCG	602956	FA	Component of the FA core complex	25%	11.0%	9%
FANCI	611360	FA	Monoubiquitylated by the FA core complex, Forms ID2 complex, Regulates the DNA damage response	2%	1.7%	0%
FANCJ (BRIP1)	605882	FBOC,FA	HR repair, DNA helicase	0%	2.4%	0%
FANCL	608111	FA	Component of the FA core complex, E3 ubiquitin ligase	0%	0.1%	0%
FANCM	609644	FBOC, FA-associated**	Component of the FA core complex, DNA translocase	0%	0%	0%
FANCN (PALB2)	610355	FBOC, FA	HR repair, Facilitaties BRCA2 function	1%	0.8%	0%
FANCO (RAD51C)	602774	FBOC, FA-like	RAD51 paralog, HR repair, Stabilizes RAD51 nucleoprotein filament	0%	—***	0%
FANCP (SLX4)	613278	FA	Resolutes Holliday junctions, Nuclease regulation, Incises DNA-ICL damage	2%	—***	0%
FANCQ (ERCC4)	133520	FA	DNA repair nuclease	0%	—***	0%
FANCR (RAD51)	179617	FA-like	HR repair, Protects stalled replication fork	0%	—***	—***
FANCS (BRCA1)	113705	FBOC, FA-like	HR repair, Promotes RAD51 recruitment	0%	—***	—***
FANCT (UBE2T)	610538	FA	E2 ubiquitin-conjugating enzyme	2%	—***	—***
FANCU (XRCC2)	600375	FA-like	RAD51 paralog, HR repair, Stabilizes RAD51 nucleoprotein filament	0%	—***	—***
FANCV (MAD2L2/REV7)	604094	FA	Translesion DNA synthesis	0%	—***	—***
FANCW (RFWD3)	614151	FA	HR repair, E3 ligase	0%	—***	—***

* FA-like genes cause a chromosome fragility syndrome with FA-related malformations but without bone marrow failure[3].

** FANCM was originally thought to be FA gene but it turned out that biallelic FANCM mutations do not cause any overt FA phenotype[4] but early onset cancer[5, 6].

*** These genes were not identified at the time of the publication[1, 2].

FA, Fanconi anemia; FBOC, familial breast and ovarian cancer; HR, homologous recombination; ID2 complex, FANCD2-FANCI heterodimer; ICL, interstrand crosslink

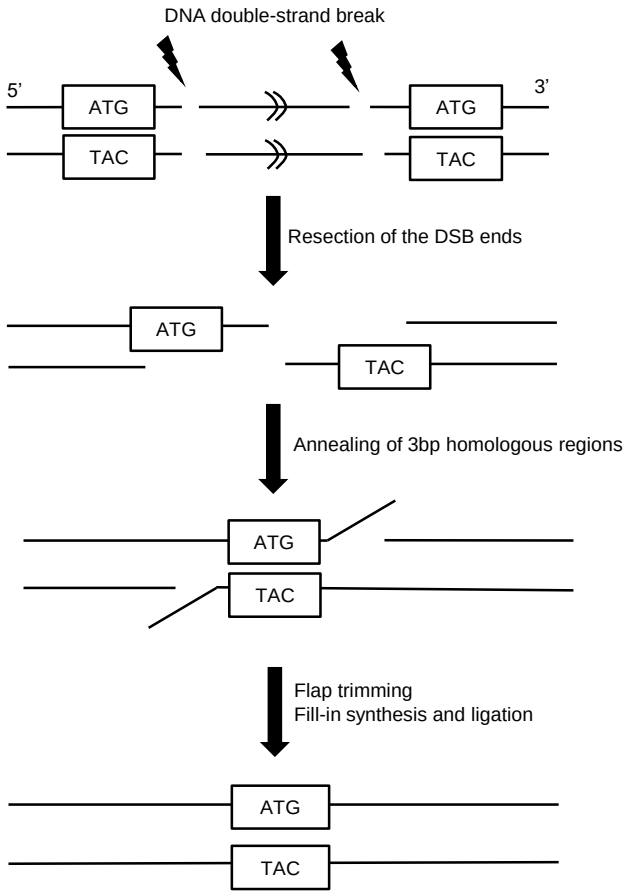
Supplemental Table 2. 55 different *FANCA* mutational variants detected in Japanese FA-A patients

DNA change	Location	Effect	Comments	No. of alleles	No. of patients	No.of unrelated families
missense mutations						
c.2T>C	exon 1	p.M1T	known mutation	1	1	1
c.1303C>T	exon 14	p.R435C	known mutation	3	2	2
c.2170A>C	exon 24	p.T724P	known mutation	2	2	2
c.2290C>T	exon 25	p.R764W	known mutation	1	1	1
c.2527T>G	exon 27	p.Y843D	known mutation	1	1	1
c.2723_2725TCT>GCC	exon 28	p.LS908_909RP	novel mutation	1	1	1
c.3965T>G	exon 40	p.V1322G	novel mutation	1	1	1
c.4198C>T	exon 42	p.R1400C	known mutation	2	2	1
c.4199G>C	exon 42	p.R1400P	known mutation	1	1	1
nonsense mutations						
c.15G>A	exon 1	p.W5X	known mutation	1	1	1
c.462T>G	exon 5	p.Y154X	novel mutation	1	1	1
c.505G>T	exon 5	p.E169X	known mutation	1	1	1
c.1464C>A	exon 15	p.Y488X	novel mutation	2	1	1
c.2840C>G	exon 29	p.S947X	known mutation	2	2	2
c.2870G>A	exon 30	p.W957X	known mutation	1	1	1
c.3295C>T	exon 33	p.Q1099X	novel mutation	1	1	1
c.3568C>T	exon 36	p.Q1190X	known mutation	1	1	1
small insertions/deletions						
c.44-69del	exon 1	p.P15RfsX40	known mutation	1	1	1
c.190_191insT	exon 3	p.E65RfsX6	novel mutation	2	1	1
c.978_979delGA	exon 11	p.Q326HfsX12	known mutation	5	5	3
c.2546delC	exon 27	p.S849FfsX40	known mutation	41	33	30
c.2593delA	exon 27	p.I879LfsX24	novel mutation	1	1	1
c.2730_2731delCT	exon 28	p.W911DfsX31	known mutation	2	2	1
c.2972delT	exon 30	p.F991SfsX35	known mutation	1	1	1
c.3638_3639delCT	exon 37	p.P1213RfsX64	known mutation	1	1	1
c.3720_3724 del	exon 37	p.E1240DfsX36	known mutation	2	2	2
c.3781_3785delTTCTT	exon 38	p.F1261LfsX15	novel mutation	1	1	1
c.3919_3920insT	exon 39	p.Q1307LfsX6	novel mutation	1	1	1
c.3931-3932delAG	exon 39	p.S1311X	novel mutation	1	1	1
c.4015_4017delCTC	exon 41	p.L1339del	known mutation	1	1	1
c.4042_4043insC	exon 41	p.I1348TfsX77	novel mutation	2	2	1
c.4124-4125delCA	exon 41	p.T1375SfsX49	known mutation	1	1	1
c.4240_4241delAG	exon 42	p.S1414LfsX10	known mutation	1	1	1
splicing mutations						
c.283+2T>C	intron 3	aberrant splicing	novel mutation	2	2	1
c.1007-2A>G	intron 11	aberrant splicing	novel mutation	1	1	1
c.1567-1G>A	intron 16	aberrant splicing	novel mutation	1	1	1
c.2316+2T>A	intron 25	aberrant splicing	novel mutation	1	1	1
c.2602-2A>T	intron 27	aberrant splicing	known mutation	6	5	4
c.2602-1G>A	intron 27	aberrant splicing	known mutation	4	4	4
c.3765+1G>T	intron 37	aberrant splicing	known mutation	2	2	1
c.4168-1G>C	intron 41	aberrant splicing	known mutation	1	1	1
c.4168-2A>G	intron 41	aberrant splicing	known mutation	3	3	2
large deletions						
ex1-3 del	—	—		1	1	1
ex1-5 del	—	—		1	1	1
ex1-28 del	—	—		1	1	1
ex1-43 del	—	—		2	2	1
ex6 del	—	—		1	1	1
ex16-17 del	—	—		1	1	1
ex19-29 del	—	—		2	2	1
ex24-28 del	—	—		1	1	1
ex30 del	—	—		8	8	5
ex30-31 del	—	—		1	1	1
ex37 del	—	—		1	1	1
c.3765+827_3814 del	intron 37-exon 38	—	novel mutation	1	1	1
large duplication						
ex11-15 dupi	—	—		1	1	1
Total				130		

Supplemental Table 3. Seven different *FANCG* mutational variants found in Japanese FA-G patients

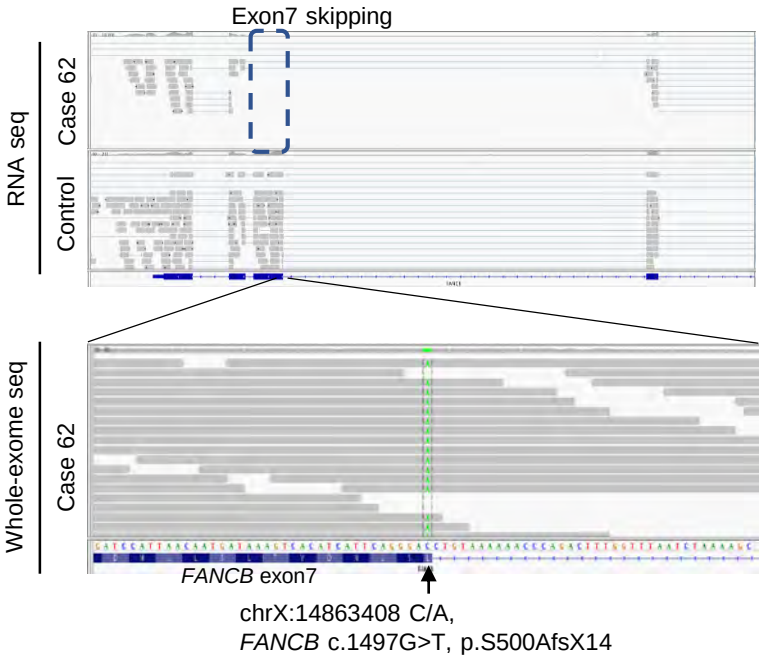
DNA change	Location	Effect	Comments	No. of alleles	No. of patients
nonsense mutations					
c.91C>T	exon 2	p.Q31X	known mutation	2	2
c.1066C>T	exon 8	p.Q356X	known mutation	15	10
small deletions					
c.194delC	exon 3	p.P65LfsX7	known mutation	3	2
c.907_908del	exon 7	p.L303GfsX5	novel muation	1	1
c.1386delC	exon 10	p.W463GfsX55	novel muation	1	1
splicing mutations					
c.307+1G>C	intron 3	aberrant splicing	known mutation	34	21
c.1637-15G>A	intron 12	VUS	novel mutation	1	1
Total				57	

variant of unknown significance

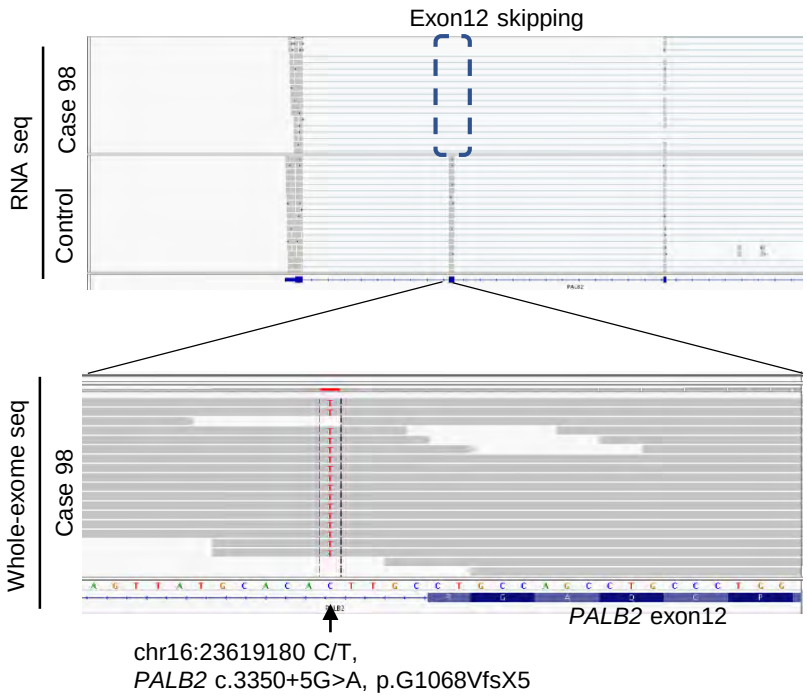


Supplemental Figure 1. Proposed model for mechanism of microhomology-mediated end joining (MMEJ) to repair DNA double-strand break (DSB). This repair model consists of at least five steps: resection of the DSB ends by nuclease digestion, annealing of 3bp homologous regions, removal of heterologous flaps, and fill-in synthesis and ligation[7]. The mutation is speculated to be created by two DSBs and subsequent religation of the two distant ends by MMEJ repair.

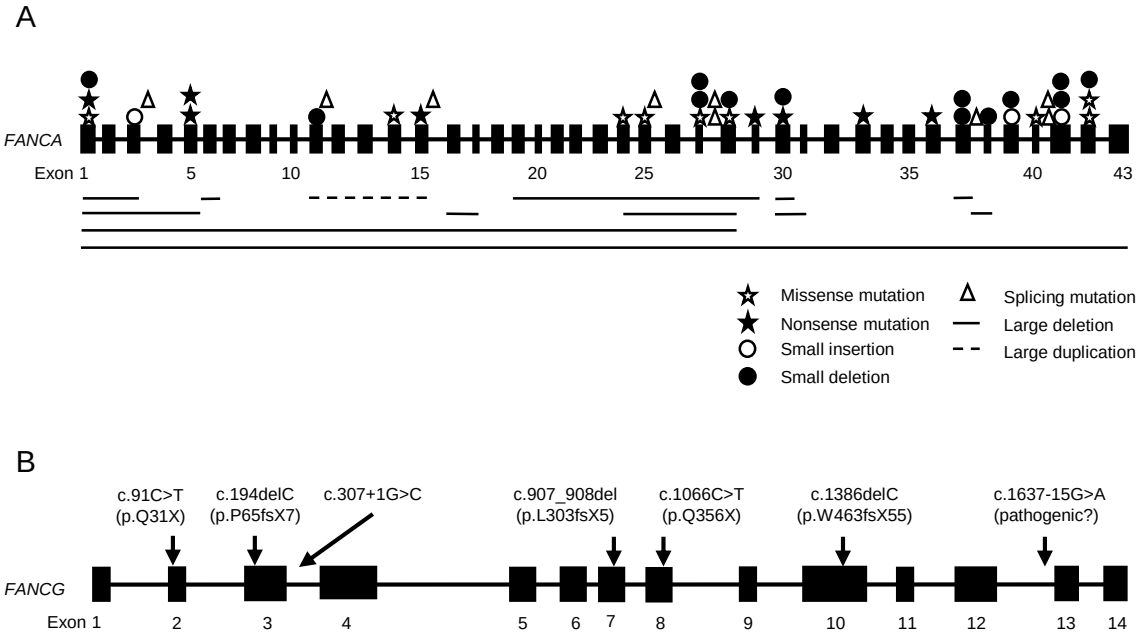
A



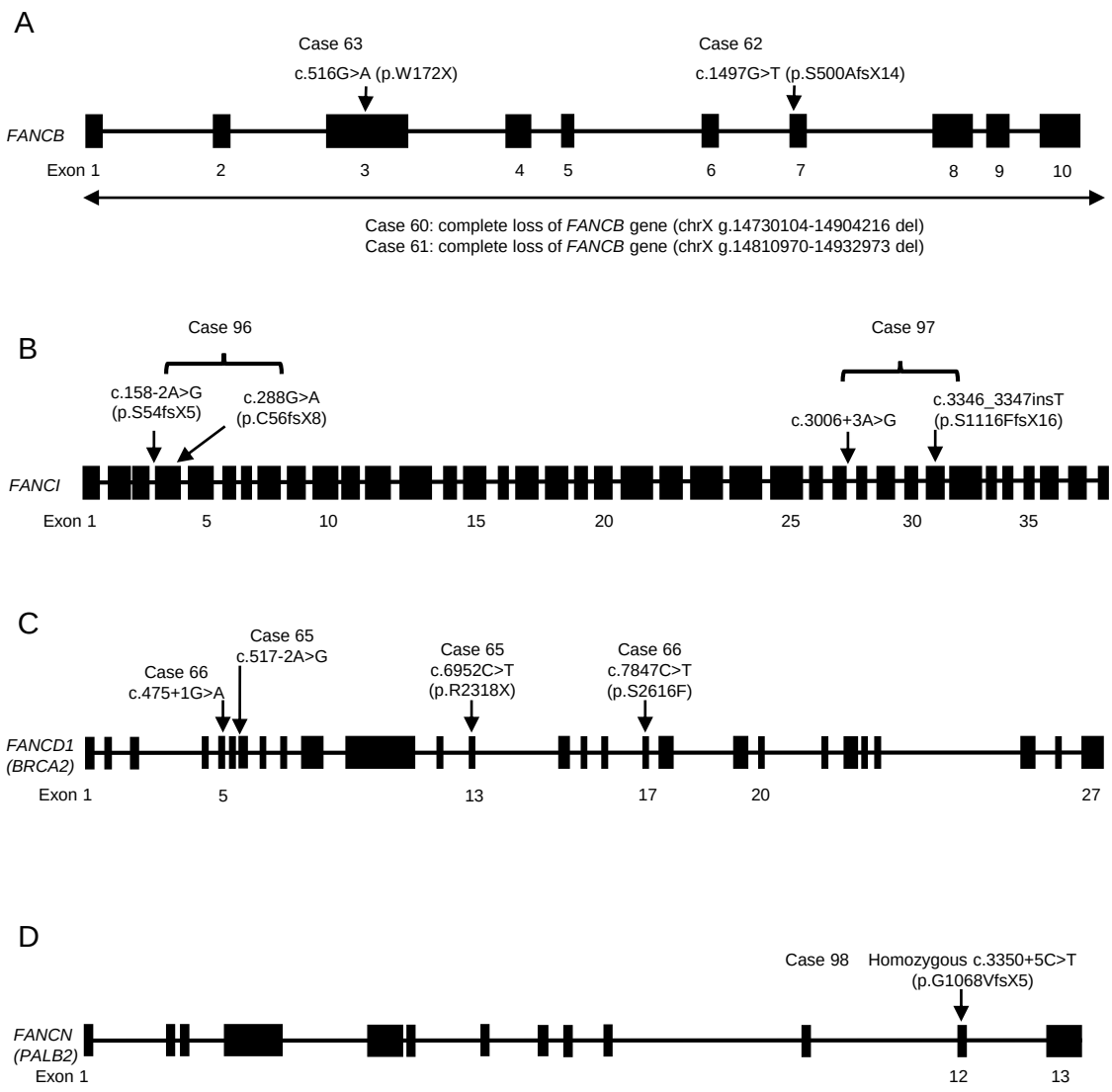
B



Supplemental Figure 2. Display of a cross section of RNA-sequencing (top) and Whole-exome sequencing (bottom). RNA sequence reads of exon 7 in *FANCB* and exon 12 in *FANCN* were absent for Case 62 (A) and Case 98 (B), respectively, which enabled us to identify exon skipping. WES analysis revealed a synonymous mutation (*FANCB* c.1497G>T) in Case 62, resulting in skipping of exon7, and a homozygous mutation (*PALB2* c.3350+5G>A) in intron 12 in Case 98, resulting in skipping of exon 12. These mutation variants were also verified by PCR and Sanger sequencing.



Supplemental Figure 3. Localization of mutation variants found in *FANCA* (A) or *FANCG* (B).



Supplemental Figure 4. Localization of mutation variants in *FANCB* (A), *FANCI* (B), *FANCD1 (BRCA2)* (C), and *FANCN (PALB2)* (D).

Supplemental References

1. Neveling K, Endt D, Hoehn H, Schindler D. Genotype-phenotype correlations in Fanconi anemia. *Mutat Res*. 2009;668(1-2):73–91.
2. De Rocco D, Bottega R, Cappelli E, et al. Molecular analysis of Fanconi anemia: the experience of the Bone Marrow Failure Study Group of the Italian Association of Pediatric Onco-Hematology. *Haematologica*. 2014;99(6):1022–1031.
3. Bogliolo M, Surrallés J. Fanconi anemia: a model disease for studies on human genetics and advanced therapeutics. *Curr Opin Genet Dev*. 2015;33:32-40.
4. Singh TR, Bakker ST, Agarwal S, Jansen M, Grassman E, Godthelp BC, et al. Impaired FANCD2 monoubiquitination and hypersensitivity to camptothecin uniquely characterize Fanconi anemia complementation group M. *Blood*. 2009;114(1):174-80.
5. Bogliolo M, Bluteau D, Lespinasse J, Pujol R, Vasquez N, d'Enghien CD, et al. Biallelic truncating FANCM mutations cause early-onset cancer but not Fanconi anemia. *Genet Med*. 2018;20(4):458-463.
6. Catucci I, Osorio A, Arver B, Neidhardt G, Bogliolo M, Zanardi F, et al. Individuals with FANCM biallelic mutations do not develop Fanconi anemia, but show risk for breast cancer, chemotherapy toxicity and may display chromosome fragility. *Genet Med*. 2018;20(4):452-457.
7. Ceccaldi R, Rondinelli B, D'Andrea AD. Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends Cell Biol*. 2016;26(1):52–64.