A founder variant in the South Asian population leads to a high prevalence of FANCL Fanconi anemia cases in India

Frank X. Donovan1* | Avani Solanki2* | Minako Mori3,4* | Niranjan Chavan2 | Merin George2 | Selvaa Kumar C5 | Yusuke Okuno6 | Hideki Muramatsu7 | Kenichi Yoshida8 | Akira Shimamoto9 | Akifumi Takaori-Kondo4 | Hiromasa Yabe10 | Seishi Ogawa8 | Seiji Kojima7 | Miharu Yabe10 | Ramanagouda Ramanagoudr-Bhojappa1 | Agata Smogorzewska11 | Sheila Mohan12 | Aruna Rajendran13 | Arleen D. Auerbach14 | Minoru Takata3 | Settara C. Chandrasekharappa1 | Babu Rao Vundinti2

1 Cancer Genomics Unit, Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland
2 Department of Cytogenetics, National Institute of Immunohaematology (ICMR), Mumbai, Maharashtra, India
3 Laboratory of DNA Damage Signaling, Department of Late Effects Studies, Radiation Biology Centre, Graduate School of Biostudies, Kyoto University, Kyoto, Japan
4 Department of Hematology and Oncology, Graduate School of Medicine, Kyoto University, Kyoto, Japan
5 School of Biotechnology and Bioinformatics, Level 6, D.Y. Patil Deemed to be University, Navi Mumbai, Maharashtra, India
6 Center for Advanced Medicine and Clinical Research, Nagoya University Hospital, Nagoya, Japan
7 Department of Pediatrics, Nagoya University Graduate School of Medicine, Nagoya, Japan
8 Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan
9 Department of Regenerative Medicine Research, Faculty of Pharmaceutical Sciences, Sanyo-Onoda City University, Sanyo Onoda, Yamaguchi, Japan
10 Department of Innovative Medical Science, Tokai University School of Medicine, Isehara, Kanagawa, Japan
11 Laboratory of Genome Maintenance, The Rockefeller University, New York, New York
12 Apollo Speciality Hospital, Chennai, India
13 Department of Pediatric Hematology, Institute of Child Health and Hospital for Children, Chennai, India
14 Human Genetics and Hematology Program, The Rockefeller University, New York, New York

Correspondence
Minoru Takata, Laboratory of DNA Damage Signaling, Department of Late Effects Studies, Radiation Biology Centre, Graduate School of Biostudies, Kyoto University, Yoshida-Konoe, Sakyo-ku, Kyoto 606-8501, Japan.
Email: mtakata@house.rcb.kyoto-u.ac.jp

Settara C. Chandrasekharappa, Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bldg 50, Room 5232, 50 South Drive, Bethesda, MD 20892, USA.
Email: chandra@mail.nih.gov

Babu Rao Vundinti, Department of Cytogenetics, National Institute of Immunohaematology, 13th Floor.

Abstract
Fanconi anemia (FA) is a rare genetic disorder characterized by bone marrow failure, predisposition to cancer, and congenital abnormalities. FA is caused by pathogenic variants in any of 22 genes involved in the DNA repair pathway responsible for removing interstrand crosslinks. FANCL, an E3 ubiquitin ligase, is an integral component of the pathway, but patients affected by disease-causing FANCL variants are rare, with only nine cases reported worldwide. We report here a FANCL founder variant, anticipated to be synonymous, c.1092G>A;p.K364=, but demonstrated to induce aberrant splicing, c.1021_1092del;p.W341_K364del, that accounts for the onset of FA in 13 cases from South Asia, 12 from India and one from Pakistan. We comprehensively illustrate the pathogenic nature of the variant, provide evidence for a founder effect, and propose including this variant in genetic screening of suspected
Fanconi anemia (FA) is a rare, mostly recessive, genetic disorder caused by a DNA repair deficiency which fails to coordinate the removal of interstrand crosslinks, leading to increased genomic instability. FA is characterized by the development of progressive bone marrow failure resulting in aplastic anemia, congenital skeletal malformations, developmental delays, and an increased predisposition to solid tumors and hematological malignancies (Faivre et al., 2000; Kottemann & Smogorzewska, 2013). The onset of hematologic disease occurs at an average age of 7 years (Butturini et al., 1994; Faivre et al., 2000; Kutler, 2003), with varying hematologic severity between FA patients as rather rare in Japan (Mori et al., 2019); FA patients in India and Pakistan, as well as those with ancestry from these regions of South Asia.

**KEYWORDS**

FA patients in India and Pakistan, as well as those with ancestry from these regions of South Asia.

** Funding information**

National Institutes of Health, Grant/Award Number: RO1HL120922; Japan Society for the Promotion of Science, Grant/Award Number: JP15H01738; Indo-JSPS, Grant/Award Number: DST/INT/JSPS/P-206/2015; Science and Engineering Research Board, Grant/Award Number: EEQ/2016/000510; National Center for Advancing Translational Sciences, Grant/Award Number: UL1 TR001866; intramural research program of the National Human Genome Research Institute.

**Technological Impact**

The technological impact of this study is significant in understanding the genetic diversity of FA patients in India and Pakistan, particularly for those originating from South Asian ancestry. The presence of founder variants in India and Pakistan suggests a likely founder effect in these populations. This information is crucial for genetic counseling, clinical management, and future research in FA.

**Funding information**

The research was supported by the National Institutes of Health, Grant/Award Number: RO1HL120922, Japan Society for the Promotion of Science, Grant/Award Number: JP15H01738, Indo-JSPS, Grant/Award Number: DST/INT/JSPS/P-206/2015, Science and Engineering Research Board, Grant/Award Number: EEQ/2016/000510, National Center for Advancing Translational Sciences, Grant/Award Number: UL1 TR001866. The intramural research program of the National Human Genome Research Institute was also pivotal.

**Ethics Committee**

The study was carried out with the consent of FA patients or parents, as appropriate. The study protocols were approved by the Institutional Ethics Committee (IEC) for human subjects in NIIH (Mumbai), Tokai University Hospital in Kanagawa, Japan, and Rockefeller University in New York, USA.

**Data Access**

The dataset is available upon request from the corresponding author (vbaburao@hotmail.com).
breaks/metaphase, with a mean of 2 breaks/metaphase (Figure 1a and Table S1), which is typical of an FA patient (Auerbach, 2009). The age at diagnosis and clinical presentations are presented in Table 1. Patients’ ages at diagnosis ranged from 2.8 to 10 years, with a mean of 5.3 years and median of 5 years. The clinical phenotype of the 13 patients is generally milder with regard to congenital malformations, with skin pigmentation abnormalities being the most common manifestation, observed in 11/13 (85%) patients. Five (38%) patients exhibited short stature. Skeletal anomalies were also observed in five (38%) patients, including limb and facial malformations and microcephaly. Cardiac, limb, and renal phenotypes were, individually, observed in three patients, but no patients had a diagnosis of VACTERL-H association, as it requires the presentation of at least three of the phenotypes from its name (Alter & Rosenberg, 2013). At the time of diagnosis, bone marrow failure was recorded in all 13 patients, with an average age of onset at 5.3 years, and included diagnoses of aplastic anemia, pancytopenia, hypoplastic hypocellular marrow, and thrombocytopenia. Seven patients were born to consanguineous parents, whereas parental consanguinity was not known for four patients.

Evaluation of FANCD2 ubiquitination informs whether the affected protein is upstream or downstream of the ubiquitination event in the FA DNA repair pathway. In each patient only the non-ubiquitinated form of FANCD2 was present (S-form, 155 kDa),
TABLE 1  Clinical observations and hematologic presentation of patients with the FANCL c.1092G>A variant

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>FANCL mutations</th>
<th>Ethnic/regional origin</th>
<th>Parental consanguinity</th>
<th>Age at diagnosis</th>
<th>Hematologic phenotype(^a)</th>
<th>Congenital abnormalities</th>
<th>Other clinical observations</th>
<th>VACTERL-H</th>
</tr>
</thead>
<tbody>
<tr>
<td>IN01(^b)</td>
<td>c.1092G&gt;A; c.1092G&gt;A</td>
<td>Karnataka, India</td>
<td>3rd degree</td>
<td>10</td>
<td>Severe AA</td>
<td>Hyper- and hypopigmentation; bilateral hypoplastic testes</td>
<td>Failure to thrive; bronze skin tone</td>
<td></td>
</tr>
<tr>
<td>IN02(^b)</td>
<td>c.1092G&gt;A; c.1092G&gt;A</td>
<td>Maharashtra, India</td>
<td>1st degree</td>
<td>10</td>
<td>Pancytopenia</td>
<td>short stature</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IN03(^b)</td>
<td>c.1092G&gt;A; c.1092G&gt;A</td>
<td>Maharashtra, India</td>
<td>2nd degree</td>
<td>2.8</td>
<td>Pancytopenia w/ macrocytosis</td>
<td>Café-au lait spots; epistaxis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IN04(^b)</td>
<td>c.1092G&gt;A; c.1092G&gt;A</td>
<td>Karnataka, India</td>
<td>2nd degree</td>
<td>4</td>
<td>Severe AA</td>
<td>Hyperpigmentation; short stature</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IN05(^b)</td>
<td>c.1092G&gt;A; c.1092G&gt;A</td>
<td>Maharashtra, India</td>
<td>2nd degree</td>
<td>5</td>
<td>Hypoplastic hypocellular marrow</td>
<td>White spots on palm; increased folic acid levels</td>
<td>Increased prothrombin time</td>
<td></td>
</tr>
<tr>
<td>IN06(^b)</td>
<td>c.1092G&gt;A; c.1092G&gt;A</td>
<td>Karnataka, India</td>
<td>Not known; estimated to be 2nd-degree from genotype data</td>
<td>3</td>
<td>Thrombocytopenia</td>
<td>Bilateral index finger; short stature</td>
<td>Brother diagnosed with FA @ age 6 (died of infection/bleeding)</td>
<td>L</td>
</tr>
<tr>
<td>IN07</td>
<td>c.1092G&gt;A; c.1092G&gt;A</td>
<td>Maharashtra, India</td>
<td>3rd degree</td>
<td>3</td>
<td>Pancytopenia</td>
<td>Hyperpigmentation; short stature; triangular facies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IN08</td>
<td>c.1092G&gt;A; c.1092G&gt;A</td>
<td>Tamil Nadu, India</td>
<td>Not known</td>
<td>7</td>
<td>Pancytopenia</td>
<td>Hyperpigmentation; café-au lait spots</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IN09</td>
<td>c.1092G&gt;A; c.1092G&gt;A</td>
<td>Tamil Nadu, India</td>
<td>Not known</td>
<td>3</td>
<td>Pancytopenia</td>
<td>Hyperpigmentation; short stature; microcephaly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IN10</td>
<td>c.1092G&gt;A; c.592delA</td>
<td>Madhya Pradesh, India</td>
<td>None</td>
<td>4</td>
<td>Severe AA</td>
<td>Hyperpigmentation; white spots—face and mouth; deep eyes; high arched palate</td>
<td>Epistaxis</td>
<td></td>
</tr>
<tr>
<td>TKFA33</td>
<td>c.1092G&gt;A; c.1092G&gt;A</td>
<td>Pakistan</td>
<td>Not known</td>
<td>5</td>
<td>Severe AA (CBT @ 6 years)</td>
<td>Hyperpigmentation; horseshoe kidney</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>FA17(^b)</td>
<td>c.1092G&gt;A; c.1092G&gt;A</td>
<td>Bihar, India</td>
<td>None</td>
<td>6.5</td>
<td>AA (BMT @ 9 years)</td>
<td>Café-au lait spots; microcephaly; patent ductus arteriosus (PDA)</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>FA91(^a)</td>
<td>c.1092G&gt;A; c.1092G&gt;A</td>
<td>Gujarat, India</td>
<td>2nd degree</td>
<td>5</td>
<td>Severe AA (BMT @ 7 years)</td>
<td>Café-au lait spots</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: AA, aplastic anemia; BMT, bone marrow transplantation; CBT, cord blood transplantation; VACTERL-H, vertebral, anal, cardiac, tracheal-esophageal, renal, limbs, hydrocephaly.

\(^a\)The cytopenia severity was scored as described (Svahn et al., 2016); bone marrow failure status was evaluated by hematologists while referring the samples for the study.

\(^b\)SNP array genotyping was performed on DNA from these patients.
indicating a defect in an upstream protein as the source of the disorder (Figure 1a,b).

High-throughput sequencing of the patients’ DNA revealed the FANCL variant, NG_007418.1(NM_018062.3):g.58387243C>T:c.1092G>A, in FA17, FA91, TKFA33, IN05, IN06, IN07, and IN10. All were homozygous (Figure 1c-II) except for IN10 which was compound heterozygous, c.1092G>A; c.592delA (Figure 1c-III). The unusual prevalence of the variant led to the screening of additional FA patients, via Sanger sequencing, who had not been molecularly analyzed and were known to not have FA patients, via Sanger sequencing, who had not been molecularly analyzed. The unusual prevalence of the variant led to the screening of additional FA patients, via Sanger sequencing, who had not been molecularly analyzed. The unusual prevalence of the variant led to the screening of additional FA patients, via Sanger sequencing, who had not been molecularly analyzed.

The c.1092G>A variant in FANCL (amino acid residues 104–294; Cole, Lewis, & Walden, 2010) was located at the last nucleotide position of exon 13. Three in silico methods produced a consensus prediction that this variant affects splicing with high probability (Table S2). According to gnomAD (https://gnomad.broadinstitute.org/variant/2–58387243-C-T), this variant (rs57706314) is only observed in 5/250,742 alleles, corresponding to a worldwide frequency of 0.00001994. Interestingly, all five occurrences are heterozygous carriers from South Asia, where the allele frequency is 0.0001634 (5/30,596 alleles). Analysis of complementary DNA (cDNA) showed that the variant does, indeed, induce aberrant mRNA splicing, c.1021_1092del, skipping exon 13 (Figure 1c-IV) and removing 24 amino acids from the protein product, p.W341_K364del.

To observe the functional defect caused by the c.1021_1092del variant at the protein level, we evaluated the cell line from TKFA33. The cells were transduced with FANCL-FLAG wild-type (WT) lentivirus. In the TKFA33-hTERT cells, the mono-ubiquitinated long-form of FANC2D protein was undetectable even after MMC treatment (Figure 1b-l, lanes 3 and 4). However, when transduced with lentivirus encoding FANCL WT, the MMC-induced long form of FANC2D was restored (Figure 1b-I, lanes 5 and 6). Furthermore, TKFA33 cells showed increased levels of G2/M phase accumulation compared with the complemented cells (Figure 1b-II). Loss of FANC2D foci formation was also reversed by the introduction of WT-FANCL (Figure 1b-III). Collectively, we confirmed that the FANCL c.1021_1092del variant results in an FA cellular phenotype.

We also wanted to evaluate the nuances of the conformational change caused by the p.W341_K364del variation through modeling predictions. FANCL is made up of three domains: the N-terminal or ELF (amino acid residues 1–110), the central-DRD2WW (amino acid residues 104–294), and the RING domain (amino acid residues 307–365; Cole, Lewis, & Walden, 2010). The mutant (p.W341_K364del) protein lacks part of the RING domain. Low sequence identity (22%) of FANCL with Drosophila melanogaster protein prompted us to, instead, use GenTHREADER, a fold recognition approach, to generate 3D models of the WT and mutant FANCL proteins. The 3D models were validated using PROSA, which revealed that most of the regions were stable wherein the graph value is below zero (Figure S1a,b). Superimposition of WT with the mutant shows better structural superimposition in the N-terminal domain, as expected (Figure S1c). The deleted 24 residues in the RING domain represent a short helix in antiparallel beta sheet associated with the UBE2T Complex (Hudson, Purkiss, Miles, & Walden, 2014). This loss of secondary structural element from the RING domain of the mutant results in random coil formation, expected to affect the ubiquitination function of the protein.

DNA from eight patients of Indian origin, IN01, IN02, IN03, IN04, IN05, IN06, FA17, and FA91, along with the parents of IN02, IN03, and FA17, were genotyped using Illumina SNP arrays. We used the genotype data to accomplish three goals: (a) confirm the South Asian ancestral origin of the patients, (b) estimate potential relatedness between individuals, and (c) identify a possible haplotype shared between the patients that could be attributed to an ancestral allele.

To confirm the South Asian ancestral origin of the patients, we performed a multidimensional scaling (MDS) analysis using KING (Manichaikul et al., 2010) and included data from 2,501 individuals from the 1000 Genomes Project Phase 3 (Genomes Project Consortium et al., 2015) cohort, including 489 individuals from South Asia (SAS). All 14 genotyped individuals clustered with the SAS population (Figure S2). We also used KING to infer relatedness between individuals, which confirmed the consanguinity reported for trios IN02 and IN03 and not for FA17 (none reported) in Table 1 and determined there was no evidence of any close relatedness between patients within three degrees, the extent to which the program could reliably predict. Patient IN06, the only genotyped individual whose parental consanguinity was unknown, was estimated to be born to parents with 2nd-degree relatedness by evaluating the genome-wide distribution of autozygosy and obtaining the “Finbred” value using KING.

The genome-wide SNP data corresponded with the reported consanguinity from Table 1, showing extensive autozygosy for the seven patients born to closely related parents. When we evaluated the region on chromosome 2 where FANCL resides, g.58,386,378–58,468,515, we observed a large region of homozygosy in all eight patients, and a shared region of homozygosy extending ~4.8 MB (Figure 1d-I), g.55,357,427–60,167,206. Further analysis of this common region of homozygosy revealed a region of shared identity, extending ~666 kb and including 160 SNPS, among all eight patients (Figure 1d-II and Table S3), g.57,941,416–58,607,850.

We estimated the variant’s age, or the time since the most recent common ancestor, using the Gamma method (Gandolfo, Bahlo, & Speed, 2014), which analyzes the contribution from the ancestral chromosome in each individual. We identified the ancestral segment lengths by continuous haplotype sharing between at least two of the patients, with the longest haplotype extending up to ~3.83 Mb and the shortest haplotype spanning ~701 kb (Figure 1d-I and Table S3). If we assume an independent genealogy, then the estimated age is equal to 91.6 generations with a 95% confidence interval (CI) [55.8–151], and if we assume that one generation is equal to 28 years (Moorjani et al., 2016), then the age of the most recent common ancestor is ~2,691 years with a 95% CI [1,562–4,228 years].

In summary, we present 13 South Asian FA cases caused by a founder variant in FANCL, a gene that is generally a rare cause of FA worldwide. Although the variant was initially predicted to be
synonymous, c.1092G>A;p.K364=. in silico splicing predictors, conveyed a high probability to affect splicing, and analysis of cDNA confirmed aberrant splicing that skips exon 13 and removes 24 amino acids, c.1021_c.1092del;p.W341_K364del. The variant is isolated to South Asia, is influenced by a founder effect, and carriers share a common haplotype from an ancestral allele that dates back ~2,700 years. As more individuals carrying this variant are discovered, we will be able to better identify the specific subpopulations most at risk. The discovery of population-specific disease-causing variants with increased prevalence due to founder effects are important for accurate and efficient genetic screening of FA patients and ensures that we appropriately raise risk awareness and provide proper genetic counseling to FA families.

ACKNOWLEDGMENTS

We (F. X. D., R. R.-B., and S. C. C.) acknowledge the support from the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health, Studies from the IFAR are supported in part by the National Center for Advancing Translational Sciences, National Institutes of Health, through Rockefeller University, Grant No. U11 TR001866, and by RO1HL120922. A. S. is an HHMI faculty scholar. We also thank Adebowale Adegemo for careful reading of the manuscript and advice on the presentation of genotype data. The study was carried out under the Indo-JSPS program (DST/INT/JSPS/P-206/2015; B. R. V. and M. T.), DST/SERB (Grant Number EEE/2016/000510; B. R. V.), JSPS KAKENHI (Grant Number JP15H01738; M. T.), and grants from the Ministry of Health, Labor, and Welfare (S. K. and E. I.).

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest.

ORCID

Frank X. Donovan https://orcid.org/0000-0002-7712-2335
Minako Mori https://orcid.org/0000-0003-4274-8714
Ramanagouda Ramanagoudr-Bhojappa https://orcid.org/0000-0002-3122-3827
Agata Smogorzewska https://orcid.org/0000-0001-6285-1562
Arleen D. Auerbach https://orcid.org/0000-0002-6911-8379
Minoru Takata https://orcid.org/0000-0002-4926-3675
Settara C. Chandrasekharappa https://orcid.org/0000-0002-8084-0530

REFERENCES


