

## Original Article

# Fanconi Anemia Patients from an Indigenous Community in Mexico Carry a New Founder Pathogenic Variant in *FANCG*

Pedro Reyes <sup>1,2,3</sup>, Benilde García-de Teresa <sup>1</sup>, Ulises Juárez <sup>1</sup>, Fernando Pérez-Villatoro <sup>1</sup>, Moisés O Fiesco-Roa <sup>1,4</sup>, Alfredo Rodríguez <sup>2, 5</sup>, Bertha Molina <sup>1</sup>, María Teresa Villarreal-Molina <sup>6</sup>, Jorge Meléndez-Zajgla <sup>7</sup>, Alessandra Carnevale <sup>8</sup>, Leda Torres <sup>1,\*</sup> and Sara Frias <sup>1,2,\*</sup>

<sup>1</sup> Laboratorio de Citogenética, Instituto Nacional de Pediatría; Ciudad de México, 04530, Mexico, email: [reyesj87@gmail.com](mailto:reyesj87@gmail.com) (PR); [b.garciadeteresa@gmail.com](mailto:b.garciadeteresa@gmail.com) (BGT); [ehatlujf@gmail.com](mailto:ehatlujf@gmail.com) (UJ); [frpvillatoro@gmail.com](mailto:frpvillatoro@gmail.com) (FPV); [fiescoroa@facmed.unam.mx](mailto:fiescoroa@facmed.unam.mx) (MFR); [bertha\\_molina@yahoo.com.mx](mailto:bertha_molina@yahoo.com.mx) (BM); [ledactorres@gmail.com](mailto:ledactorres@gmail.com) (LT)

<sup>2</sup> Departamento de Medicina Genómica y Toxicología Ambiental, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Ciudad de México, 04510, Mexico. [alfredo.rodriguez@iibio-medicas.unam.mx](mailto:alfredo.rodriguez@iibio-medicas.unam.mx) (AR); [sarafrias@iibio-medicas.unam.mx](mailto:sarafrias@iibio-medicas.unam.mx) (SF)

<sup>3</sup> Doctorado en Ciencias Biomedicas, Universidad Nacional Autónoma de México, 04510, Ciudad de Mexico.

<sup>4</sup> Maestría y Doctorado en Ciencias Médicas, Odontológicas y de la Salud, Universidad Nacional Autónoma de México, Ciudad de Mexico, 04510, México

<sup>5</sup> Instituto Nacional de Pediatría, Ciudad de México, 04530, México.

<sup>6</sup> Laboratorio de Enfermedades Cardiovasculares, Instituto Nacional de Medicina Genómica, Ciudad de México, 14610, Mexico. [mvillareal@inmegen.gob.mx](mailto:mvillareal@inmegen.gob.mx) (MTV)

<sup>7</sup> Laboratorio de Genómica Funcional del Cáncer, Instituto Nacional de Medicina Genómica, Ciudad de México, 14610, Mexico. [jmelendez@inmegen.gob.mx](mailto:jmelendez@inmegen.gob.mx) (JMZ)

<sup>8</sup> Laboratorio de Enfermedades Mendelianas, Instituto Nacional de Medicina Genómica, Ciudad de México, 14610, Mexico. [acarnevale@inmegen.gob.mx](mailto:acarnevale@inmegen.gob.mx) (AC)

\* Correspondence: Sara Frias, [sarafrias@iibio-medicas.unam.mx](mailto:sarafrias@iibio-medicas.unam.mx) and Leda Torres, [ledactorres@gmail.com](mailto:ledactorres@gmail.com). Address: Laboratorio de Citogenética, Instituto Nacional de Pediatría, Av. Insurgentes Sur, 3700-C, Torre de Investigación, 6to piso, Insurgentes Cuicuilco. CP 04530 Coyoacán, Ciudad de México, México.

**Abstract:** Fanconi anemia (FA) is a rare genetic disorder caused by pathogenic variants (PV) in at least 22 genes, which cooperate in the FA/BRCA pathway to maintain genome stability. PV in *FANCA*, *FANCC*, and *FANCG* account for most cases (~90%). This study evaluated the chromosomal, molecular, and phenotypic findings of a novel founder *FANCG* PV, identified in three patients with FA from the *Mixe* community of Oaxaca, Mexico. All patients presented chromosomal instability and a homozygous PV, *FANCG*: c.511-3\_511-2delCA, identified by next-generation sequencing analysis. Bioinformatics predictions suggest that this deletion disrupts a splice acceptor site promoting the exon 5 skipping. Analysis of Cytoscan 750K arrays for haplotyping and global ancestry supported the Mexican origin and founder effect of the variant, reaffirming the high frequency of founder PV in *FANCG*. The degree of bone marrow failure and physical findings (described through the acronyms VACTERL-H and PHENOS) were used to depict the phenotype of the patients. Despite having a similar frequency of chromosomal aberrations and genetic constitution, the phenotype showed a wide spectrum of severity. The identification of a founder PV could help for a systematic and accurate genetic screening of patients with FA suspicion in this population.

**Keywords:** Fanconi anemia; Chromosome instability; *FANCG*; splicing; founder pathogenic variant; *Mixe* indigenous group.

## 1. Introduction

Fanconi anemia (FA) is a rare genetic disorder characterized by chromosomal instability, a high predisposition to physical developmental abnormalities, progressive bone marrow failure, solid tumors, and hematological malignancies

[1,2]. Genetic heterogeneity is a striking feature of FA, with germline pathogenic variants (PV) in 22 *FANC* genes (*FANCA* to *FANCW*) so far associated with the FA phenotype. The protein products of the *FANC* genes participate in the Fanconi Anemia/Breast Cancer (FA/BRCA) pathway, a biochemical network that regulates DNA damage repair in response to DNA interstrand crosslinks, maintains genomic stability during DNA replication, and assists other cellular processes [3-5].

Not only a large number of genes have been associated with the FA phenotype, but also hundreds of unique PV have been reported, making allelic and locus heterogeneity the rule. Worldwide, the more frequent disease-causing PV among patients with FA occur in *FANCA* (~64%), *FANCC* (~12%), and *FANCG* (~8%) [6,7]. Recurrent PV have been identified in specific ethnic backgrounds due to a founder effect [8]. Notable examples, explained by high rates of carrier individuals, include the *FANCA* c.295C>T in Spanish Gypsies [9], *FANCC* c.456+4A>T in Ashkenazi Jews [10], and *FANCC* c.67delG in the Mennonite Community [11,12].

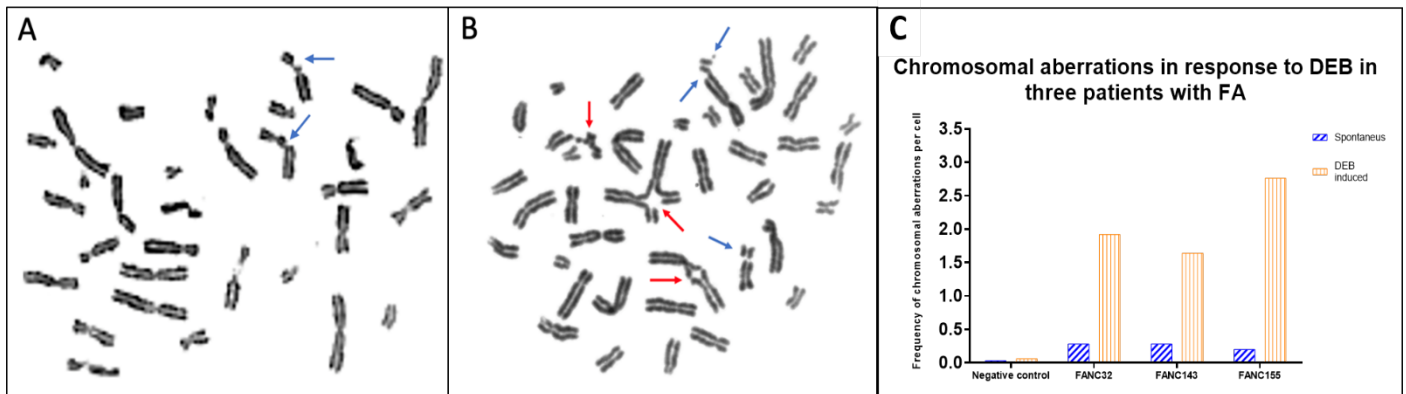
Regarding *FANCG* (OMIM# 602956), founder PV have been reported in Portuguese-Brazilian (c.1077-2A>G), Korean-Japanese (c.307+1G>C), French-Acadian (c.1480+1G>C), and Black South African (c.637\_643del) populations [13,14]. *FANCG* covers around 6 kb that are mapped to chromosome 9p13, it comprises 14 exons and encodes a 2.6 kb mRNA transcript that is translated to a 622 amino acids protein with a molecular weight of 68 kDa [15]. *FANCG* contains seven tetratricopeptide repeat (TPR) motifs, which are required to mediate its protein-protein interactions [15,16]. *FANCG* is an integral component of the FA-core complex, a dynamic E3 ubiquitin ligase protein assembly that catalyzes the mono-ubiquitination of the FANCD2-FANCI heterodimer, a critical event in the recruitment of the repair machinery to the DNA lesion as part of the FA/BRCA pathway activation [17-19].

Here, we report the novel PV in *FANCG*: c.511-3\_511-2delCA, detected in three patients from the *Mixe* indigenous community in Mexico. We used a comprehensive clinical approach to elucidate the consequences of this PV and we provide evidence for a founder effect.

## 2. Results

### 2.1 Chromosomal breakage analysis confirmed a FA diagnosis

Chromosomal breakage analysis with diepoxybutane (DEB) was performed in the three patients to support the suspicion of a clinical FA diagnosis. The number of spontaneous and DEB-induced aberrations, as well as the presence of radial chromosomal figures, were considered as criteria to establish the diagnosis of FA. In all three cases, the frequency of radial exchange figures and chromatid breaks showed an increase in DEB-induced samples, compared to those recorded in the spontaneous aberration cultures. Hypersensitivity to DEB treatment in patients was similar to that observed in the FA-positive control (VU817 cell line), while induced aberrations frequency observed in the patients was higher than the historical average reported in our laboratory for healthy individuals, used as negative controls (Figure 1). These results confirmed the FA diagnosis in the patients, who were identified as FANCG32, FANCG143, and FANCG155 (Figure 1).



**Figure 1.** DEB-induced chromosomal aberrations in lymphocytes from healthy and FA individuals. **(A)** Representative metaphase from a healthy individual (negative control); chromosome aberrations are mainly chromatid breaks. **(B)** Representative metaphase from one of the FA patients; increased chromosome aberrations are observed showing chromatid breaks and radial figures, characteristics of FA patients. Blue arrows show chromatid and isochromatid breaks and red arrows show radial figures. **(C)** Frequency of chromosomal aberrations per cell for each of the three FA patients analyzed in this study.

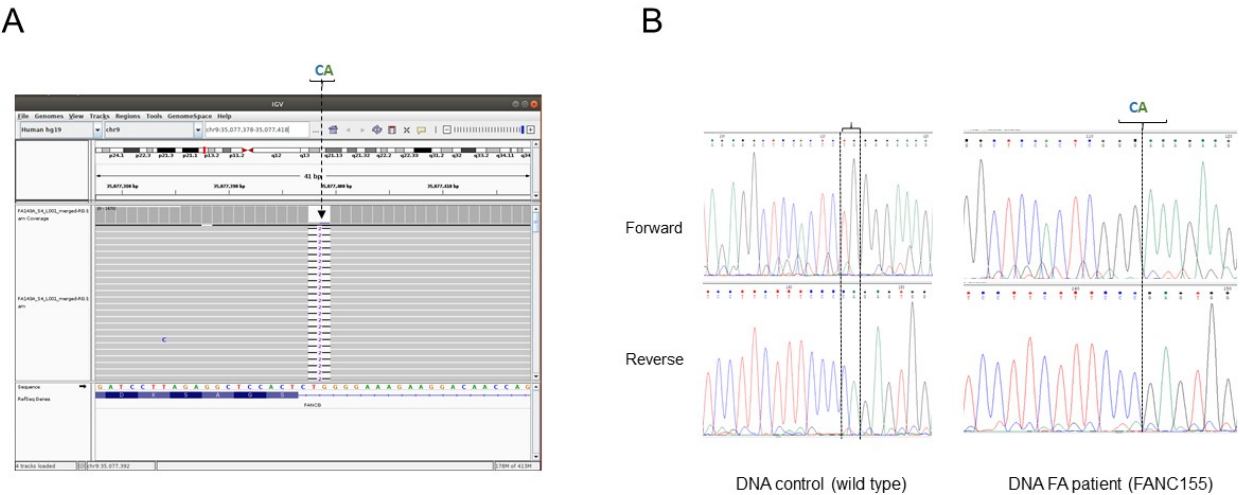
## 2.2 Pathogenic variant found in FANCG: c.511-3\_511-2delCA is predicted to induce exon 5 skipping

Targeted massively parallel DNA sequencing from the three patients with FA, revealed a homozygous two-base deletion on FANCG (Figure 2A and Table 1). This alteration is located 2 bp downstream from the nearest acceptor splice site between intron 4 and exon 5. The genotype was subsequently validated via Sanger sequencing (Figure 2B). According to gnomAD (<https://gnomad.broadinstitute.org/>) this variant (rs1491369358) has only been observed in 1/251,480 alleles, corresponding to a worldwide frequency of 0.00000398. Further analysis of the effect of this novel variant, following the recommended American College of Medical Genetics (ACMG) criteria, supports its pathogenicity.

**Table 1.** Features of the pathogenic variant found in FANCG.

Reference sequence	DNA change (genomic, hg19)	ACMG criteria	HGVS nomenclature
NG_007312.1	g.35077398_35077399del	PVS1, PM2, PP5	FANCG(NM_004629.2): c.511-3_511-2delCA

Criteria for classifying pathogenic variants according to the ACMG: American College of Medical Genetics; HGVS: Human Genome Variation Society; PVS1: Pathogenic very strong 1; PM2: Pathogenic Moderate 2; PP5: Pathogenic Supporting 5. For a complete description of these criteria refer to [20].



**Figure 2. Detection of a novel homozygous PV in FANCG.** (A) The homozygous deletion *FANCG*: c.511-3\_511-2delCA was detected by Haloplex-NGS. Visual exploration of this change in FANC143.bam file with the Integrative Genomics Viewer software (<http://software.broadinstitute.org/software/igv/>) is displayed. (B) PV *FANCG*: c.511-3\_511-2delCA was corroborated by bidirectional Sanger sequencing. Representative electropherograms of a healthy control and FANC155 are shown.

Inspection of the genomic region where the variant was found suggests that this change could disrupt messenger RNA processing. *In silico* analysis using The Human Splice Finder tool revealed that *FANCG*: c.511-3\_511-2delCA impairs the wild-type acceptor site, most probably affecting splicing. MaxEntScan rated this PV with a score of 1.1, with respect to a reference score of 9.63, predicting a deleterious modification of this splice site. Finally, CRYP–SKYP tool indicated that the deletion may promote exon 5 skipping (probability of cryptic site activation = 0.40). These results consensually predict that this PV in the acceptor splice site leads to loss of *FANCG* exon 5.

2.3 Reduced genetic variation in locus g.29754068-38771831 supporting a founder effect for *FANCG*: c.511-3\_511-2delCA in the Mixe population

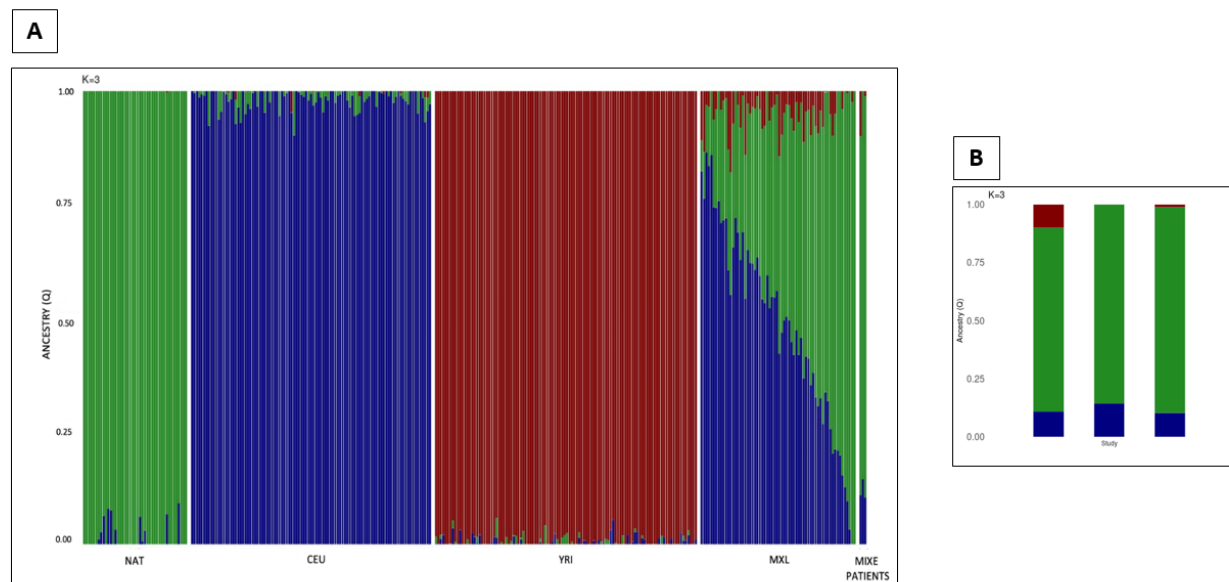
The three patients with FA described in this study come from an isolated mountainous region in the state of Oaxaca (southern Mexico), they self-identified as belonging to the *Mixe* indigenous group and shared a surname. Although a kin relationship was unknown, we investigated the possibility of a common haplotype using high-density microarrays (CytoScan 750K array). Several long-contiguous stretches of homozygosity (LCSH, >3 Mb) were shared in autosomal regions by these three patients with FA (Table 2). Some of these LCSH have previously been reported in the HapMap and the 1000 Genomes Project (1KGP) in populations of Asian (CHB), Northern European (CEU) and American (AMR) origin (Supplementary Table 1) [21–24]. We found two shared zones among the LCSH regions not previously reported. The first one in chromosome 5q spanning 2935 Kbp (g. 101035497-98100800) and the other one in chromosome 9p spanning 9018 Kbp (g.29754068-38771831), the latter, importantly, including the *FANCG* locus. Inbreeding coefficient (F), calculated with the total LCSH extension and used to estimate the degree of kinship, suggested a possible parental relationship between the three patients (Table 2).

**Table 2.** Native component, homozygosity, and coefficient of inbreeding (F) in the three *Mixe* patients with FA.

Patient ID	Total autosomal LCSH >3Mb (Kb)	Percentage of Native Component	Autosomal homozygosity (%)	F	Probable parental relationship
FANC032	140214361	79.17%	4.9%	0.0487	Fourth-degree
FANC143	203913825	85.64%	7.1%	0.0708	Third-degree
FANC155	249262167	88.53%	8.7%	0.0865	Third-degree

When we analyzed the g.29754068-38771831 region, we discovered that 439 out of 544 single nucleotide variants (SNVs) (80.7%) included in the CytoScan 750K array were common among the three patients. We compared the allelic frequency of these SNVs with the frequency reported in the 1KGP for the global and the Mexican population (MXL). We found that 99.32% of the alleles of our *patients* are coincident frequencies, compared to the global frequencies and the MXL group, whereas the remaining 0.68% has a different frequency. These results confirm that patients share a common haplotype of 30 SNVs around *FANCG*, supporting the discovery of a founder effect (Supplementary Table 2).

Finally, to ascertain that our patients share the Mexican population (MXL) genetic background, as reported in the 1KGP, a global ancestry analysis was performed. In this context and considering that Mexicans present a remarkable genetic diversity due to population admixture, we decided to contrast all SNVs data of these 3 *Mixe* patients against other populations beside the MXL, including Native-Americans (NAT), Utah residents with Northern and Western European ancestry (CEU), and Yoruba in Ibadan, Nigeria (YRI); similar to how the Mexican population structure has been analyzed in previous studies [25,26]. Using this approach, it was observed that the MXL population presents an ancestry pattern with native and european components predominance (NAT:46.59% and CEU:48.44%, respectively). On the other hand, our results indicate that the patients have a proportion of the native component (NAT) as high as 79 to 88% (Table 2), with lower CEU and YRI contributions, when compared to the MXL population although the diversity described in the latter is conserved in both (Figure 3).

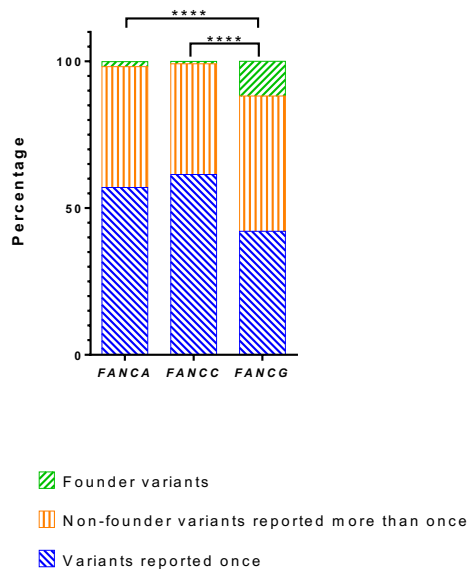


**Figure 3. Mixe patients with FA present an ancestry pattern that have the hallmarks of the genetic diversity of the Mexican population.** (A) Barplot showing the inference of global ancestry for the Mixe and Mexican individuals using ADMIXTURE, population reference panels include NAT, CEU, and YRI (K=3 model). Each individual is depicted as a vertical bar. Colors represent the percentage of ancestry assigned to each cluster for each individual. Green, blue, and red colors indicate Native-Americans (NAT), Northern Europeans (CEU), and Yorubas (YRI) populations, respectively. The two panels located at the extreme right of the figure show the three-way ancestry components for the admixed Mexican (MXL) population and the Mixe patients (last three bars). (B) Detailed barplot of the three Mixe patients with FA

#### 2. 4 The fraction of pathogenic variants linked to a founder effect in FANCG is the highest among the most common FANC genes

Most founder variants described so far in FA are clustered together in the three most frequently mutated FANC genes (*i.e.* FANCA, FANCC, and FANCG). The PV described here adds a new founder PV (FPV) in FANCG (Table 3). The literature analysis indicates that the number of FPV, validated through haplotype analysis, was 11, three, and eight for FANCA, FANCC, FANCG, respectively. FANCG has the lowest number of unique variants among the most frequently affected genes in FA. There were differences between the proportion of founder variants regarding non-founder among these genes (Figure 4). The Fisher's exact test to compare the proportion of FPV between FANCG and FANCA was  $p < 0.0001$ , whereas between FANCG and FANCC was  $p < 0.0001$  (Figure 4).





**Figure 4.** The proportion of variants with founder effect in *FANCG* is the highest among the most frequently reported genotypes. Proportions of founder pathogenic variants, non-founder pathogenic variants reported more than once, and pathogenic variants reported just once in *FANCA*, *FANCC*, and *FANCG*. The significance level of the Fisher test is represented by four asterisks ( $p<0.0001$ ).

**Table 3.** Founder pathogenic variants reported in *FANCG*

Pathogenic Variant	Location	Effect	Geographic/ Ethnic background	Reference
c.307+1G>C	Intron 3	Aberrant splicing	Japanese-Korean	27
c.313G>T (p.Glu105Ter)	Exon 4	Truncated protein, null variant	German	28
c.637_643delTACCGCC (p.Tyr213LysfsTer6)	Exon 5	Truncated protein, null variant	South African	14
c.1066C>T (p.Gln356Ter)	Exon 8	Truncated protein, null variant	Japanese	29
c.1077-2A>C	Intron 8	Aberrant splicing	Portuguese-Brazilian	13
c.1480+1G>C	Intron 11	Aberrant splicing?	French-Acadian	13

c.1589_1591delATA p.(Asp530_Thr531del insAla)	Exon 12	Reduce protein activity, Hypomorphic variant	Korean	29
c.1649delC (p.Thr550IlefsTer9)	Exon 13	Truncated protein, null variant	Turkish	28
c.511-3_511-2delCA	Intron 4	Aberrant splicing?	Mixe (Mexican)	Present study

### 2.5 Evaluation of the phenotype of Mixe patients with FA in the framework of the reported FANCG pathogenic variant

Clinical records of each patient were analyzed to identify physical findings, mainly those described in VACTERL-H (Vertebral, Anal, Cardiac, Tracheo-esophageal fistula, Esophageal atresia, Renal, upper Limb, and Hydrocephalus) association and PHENOS (skin Pigmentation, small Head, small Eyes, Nervous system, Otology, Short stature) acronym [30]. Likewise, hematologic data analysis was used to classify the severity of the bone marrow failure.

**FANC32** was the third of four children, born to non-consanguineous parents. His older sister had FA diagnosis and died due to bone marrow failure (BMF). The patient's pregnancy was uneventful, he was delivered vaginally at 40 weeks of gestation (WG). His weight at birth was 2,500 g (<3<sup>rd</sup> percentile, HP:0001518) and he had hypotonia (HP:0001252). Developmental milestones were delayed (HP:0012758), beginning to walk at 36 months. He was sent to our clinic at seven years old (y.o.) due to BMF with a history of recurrent upper respiratory tract infections. Bone marrow aspiration and biopsy performed at age nine y.o. reported 50% cellularity with erythroid series and megakaryocyte predominance and no signs of fibrosis. He was initially treated with steroids and transfusions when required. He later started evaluation for hematopoietic stem-cell transplantation (HSCT); unfortunately, the patient stopped attending follow-up visits. At 14 y.o. anthropometric evaluation showed low weight (-2.31 standard deviation score (SDS), HP:0004325), height (-2.73 SDS, HP:0004322), and head circumference (-2.07 SDS, HP:0000252) for his age and sex. Physical examination revealed thenar hypoplasia of both hands (HP:0001245), generalized hyperpigmentation (HP:0000953), *cafe au lait* spots (HP:0000957), and melanonychia (HP:0100644). Cardiac and renal structural abnormalities were not found through sonographic evaluation.

**FANC143** was the first child of non-consanguineous parents. Her pregnancy was complicated by fetal distress and oligohydramnios (HP: 0001562) which warranted pregnancy termination by C-section at 36 WG. The birth weight was 1,590 g (<3<sup>rd</sup> percentile, HP:0001518), she remained hospitalized her first month of life. Bilateral radial ray alteration (HP:0410049) and left hip dysplasia (HP: 0001385) were identified. She had developmental delay (HP:0012758), particularly affecting language.

The patient was referred to our clinic when she was nearly seven y.o. due to BMF. She was initially treated with transfusions when needed and steroids. Physical examination revealed low weight (-2.15 SDS, HP:0004325), short stature (-2.90 SDS, HP:0004322), and microcephaly (-3.23 SDS, HP:0000252). She had up slanted palpebral fissures (HP:0000582), bifid uvula (HP: 0000193), bilateral microtia (HP: 0008551), with folded pinnae (HP:0000396) of the right ear, agenesis of the right thumb (HP:0009777), floating hypoplastic left thumb (HP:0009601), generalized skin



hyperpigmentation (HP:0000953), and *cafe au lait* spots (HP:0000957). Tympanometry and brainstem auditory evoked potential showed right unilateral hearing loss (HP:0000365). Spinal X-rays showed spina bifida in S1 (HP:0004614) and a lumbosacral hemivertebra (HP:0008439). Renal and cardiac malformations were ruled out with imaging methods. Central nervous system (CNS) magnetic resonance imaging (MRI) revealed cortico-subcortical atrophy (HP:0012444), supratentorial ventriculomegaly (HP:0002119), and encephalomalacia (HP:0040197). The patient underwent HSCT at 8.6 y.o. and died a month later due to graft versus host disease complications and septic shock.

**FANC155** was the younger of two children, born to non-consanguineous parents. Her family health history revealed that two great-aunts died of breast cancer in their early thirties. She had a maternal uncle with a history of infertility and vertebral defects. The mother reported an uneventful pregnancy. She was vaginally delivered after 40 weeks, with a birth weight of 2,400 g (<3<sup>rd</sup> percentile, HP:0001518). She had developmental delay (HP:0012758) with language disability (HP:0002463). At six y.o. she was evaluated at a local hospital due to a history of recurring epistaxis and pallor. After evaluation, a diagnosis of aplastic anemia was reached. She received steroid therapy and supportive care including blood transfusions when required. At 12 y.o., she was referred to our clinic for HSCT. Physical examination revealed weight (-1.26 SDS) and head circumference (-0.86 SDS) within normal range but low height (-2.18 SDS, HP:0004322) for her age and sex. She had upslanted palpebral fissures (HP:0000582), ptosis (HP:0000508), simple myopic astigmatism (HP:0500041), hypoplastic helix (HP:0011039) in both ears, left hypoplastic thenar region (HP:0001245) with ipsilateral thumb limited flexion, generalized hyperpigmentation (HP:0000953), and *café au lait* spots (HP:0007565). Audiometric findings were unremarkable. Brain MRI, renal ultrasonography, and echocardiography did not show alterations. She underwent HSCT from a non-related donor but died a couple of days after the procedure due to septic shock.

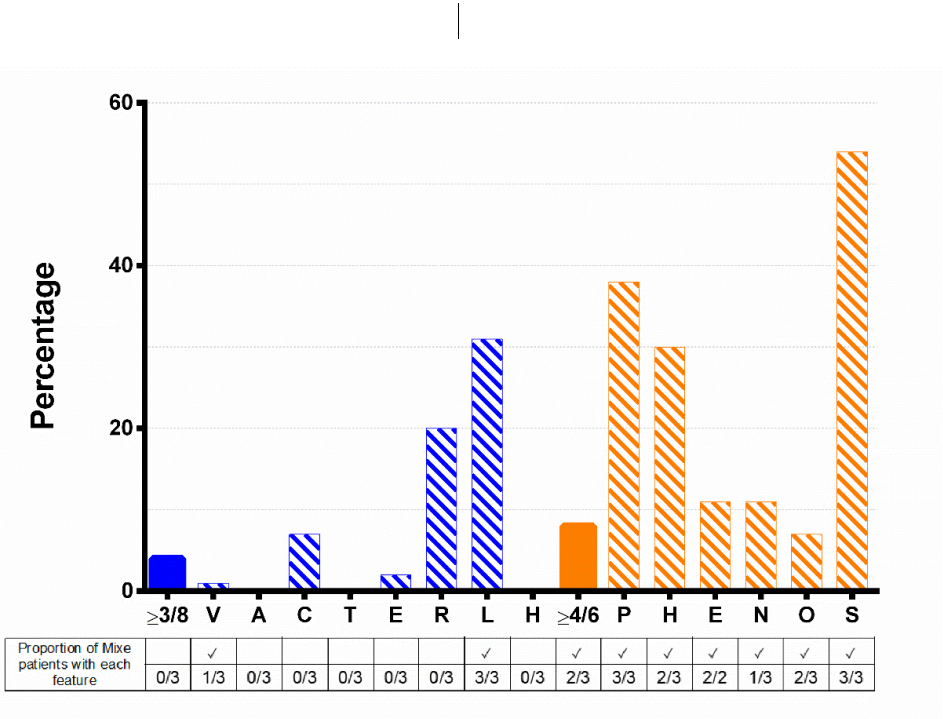
Table 4 summarizes the three patients’ clinical phenotypes. The mean age at FA diagnosis was 6.45 years. All patients had short stature, microcephaly, skin pigmentation changes, and radial ray anomalies. None of the patients met criteria for VACTERL-H association (≥3/8 features). Nevertheless, the FANC143 patient had four PHENOS features. All patients had data associated with progressive BMF.

**Table 4.** Summary of the clinical phenotype of the patients described in this study.

Patient ID	Sex	Family history of FA	Age at diagnosis	Hematologic phenotype at initial evaluation in the INP					VACTERL-H features	PHENOS features
				Hb (g/dl)	MCV (fl)	ANC (cells/ $\mu$ l)	PLT (cells/ $\mu$ l)	BMF status		
FANC32	M	Sister	6.45	11.4	100.0	700	26,000	Moderate BMF	L	PS
FANC143 <sup>†</sup>	F	No	3.6	8.9	94.1	900	15,000	Moderate BMF	VL	PHNOS
FANC155 <sup>†</sup>	F	No	9.3	11.1	99.5	220	19,000	Severe BMF	L	POS

**Abbreviations:** ANC: Absolut Neutrophils Counts; BMF: Bone marrow failure. Hb: Hemoglobin; MCV: Mean Corpuscular Volume; PLT: Platelets; PHENOS: skin Pigmentation, small Head, small Eyes, Nervous system, Otology, Short stature; VACTERL-H: Vertebral, Anal, Cardiac, Tracheo-esophageal fistula, Esophageal atresia, Renal, upper Limb, and Hydrocephalus; URTI: Upper respiratory tract infection. † Patient deceased.

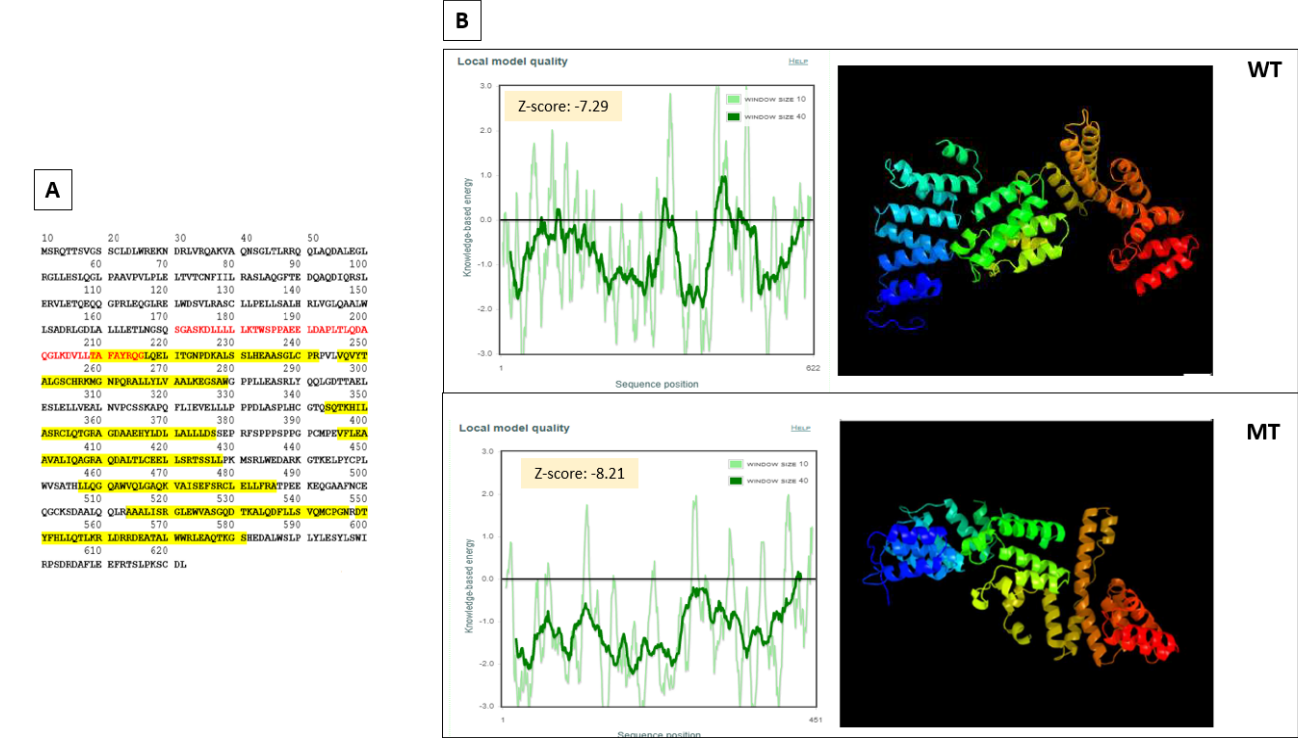
We compared the clinical manifestations observed in the patients reported here to those described in the literature for cases with PV in *FANCG*, we excluded reports in which the phenotypes were not mentioned or individually described. Remarkably, none of these had any of the 8 founder-effect PV listed in Table 3. We identified 84 cases with PV in *FANCG*. The most frequent physical characteristics were short stature (54%), skin pigmentation changes (38%), upper limb (radial ray) abnormalities (31%), microcephaly (29%), and renal malformations (20%). The rest of abnormalities were present in less than 20%. The type and frequency of phenotypic anomalies are shown in Figure 5. Three out of 84 patients met criteria for VACTERL-H ( $\geq 3/8$  features), 7/84 had  $\geq 4/6$  PHENOS features, and only one patient had VACTERL-H plus PHENOS. Regarding the time of diagnosis, the median age was 6 y.o. and the median age at report was 9 y.o. The male to female ratio was 1:0.64 ( $p = 0.5$ ).



**Figure 5. VACTERL-H and PHENOS features in 84 patients from the literature with a *FANCG* genotype.** Upper limb includes abnormal thumb +/- abnormal radius. Nervous system includes structural brain malformations other than hydrocephalus. Otology comprises ear malformations and/or hearing loss. Solid blue or orange: VACTERL-H or PHENOS; oblique stripes: individual findings. Horizontal axis: abnormalities analyzed; vertical axis: percent of total cases with that abnormality. VACTERL-H association ( $\geq 3/8$  features) was present in 4% and PHENOS ( $\geq 4/6$  features) in 8% of patients. The proportion of *Mixe* patients with each feature. ✓: This finding was found among *Mixe* patients.

2.6 Protein modeling of mutant FANCG

We aimed to evaluate the impact of *FANCG: c.511-3\_511-2delCA* at the protein level, however the crystal structure of wild type FANCG protein (NP\_004620.1) has not been reported, therefore we recurred to 3D modeling predictions using a fold based-template search using the Phyre2 server. The protein model generated for FANCG wild type (WT) showed a superhelicoidal structure consisting of a single domain, composed of repeated alpha-helices assembled in an antiparallel fashion and containing a ligand-binding groove (Figure 6B), a prototypical assembly observed in the TPR-like domain superfamily. In contrast, the prediction of *FANCG: c.511-3\_511-2delCA*, leading to exon 5 skipping, includes the lost of 45 aminoacids in the FANCG protein, including the first 8 residues of the TPR1 motif (Figure 6A). The mutant protein (p.S171\_216Gdel) presents two domains of different sizes, in which the binding groove is shorter and oriented towards the N-terminus regarding FANCG WT protein. The loss of secondary and tertiary structure in the mutant protein is expected to affect its original conformation and disturb its interaction with other proteins. Both models were supported by the ProSA tool, which revealed that most of the regions were stable with a graph value (Z) below zero.



**Figure 6: In silico 3D structure analysis of the effect of exon 5 skipping on FANCG protein, predicted by Phyre-2 server. (A)** Amino acids (AAs) encoded by *FANCG*. AAs encoded by exon 5 are shown in red. AAs that are part of the TPR motifs (1-7) appear highlighted in yellow, each TPR motif is composed by 34 amino acids. **(B)** 3D prediction of the wild type (WT) and mutant (MT) FANCG protein as predicted by Phyre-2 and ratified by ProSA.

3. Discussion

FA is a disorder characterized by high genetic, allelic, and locus heterogeneity. The simultaneous study of several genes using NGS-based targeted panels has advanced the recognition of variants not previously reported within known *FANC*

genes. Applying this NGS strategy, we diagnosed three patients from the same geographical region with a classic FA phenotype. We found a non-previously reported PV: *FANCG*:c.511-3\_511-2delCA, consisting of a two base-pair deletion affecting the splicing acceptor site of *FANCG* exon 5. This deletion is predicted to lead to an aberrant splicing site, possibly affecting the function of *FANCG* within the FA core complex. Confirmation of a founder effect was obtained by haplotype and ancestry analysis. Overall, our results represent an integrative approach to evaluate the effect of this change in *FANCG* and provide useful information to guide a potential targeted FA screening on the *Mixe* population, a well-defined ethnic group in Mexico.

Besides having the same variant, these patients also share the same geographic and ethnic origin, as well as one of their surnames. These data, in addition to the very low reported frequency of this change (rs1491369358), which has been found in a heterozygous state only in one non-FA individual, led us to explore the possibility of a founder effect for PV *FANCG*:c. 511-3\_511-2delCA. The analysis of the degree of inbreeding (F value), the percentage of homozygosity and the regions with LCSH allowed us to corroborate this inference, since all the patients have large stretches of homozygosity in their genomes due to inheritance of identical ancestral genomic segments from both parents. A consequence of this genetic structure is an increased incidence of recessive diseases, like FA. Additionally, the inferred haplotype in the boundaries of the *FANCG* locus supports the hypothesis of a founder effect in this novel PV. Moreover, the high contribution of the native component of around 80% in these three patients contrasts with the average 46% found in the MXL group and supports that these subjects who belong to the *Mixe* group have reduced genetic diversity (Figure 3). While only two of these patients have a native contribution of over 85%, sufficient to name them Mexican natives (according to [31]), the third one who does not reach this threshold also has the lower autosomic homozygosity and the more distant familial relationship, reflecting more mixture and illustrating the origin of mestizo-Mexicans [23].

*FANCG* c.511-3\_511-2delCA is the first founder variant reported in patients with FA of Mexican origin and is consistent with the demographic history of the *Mixes*, an original population that has resisted several attempts to occupy their lands since the pre-Hispanic era and the Spanish conquest. Even though the changes that have taken place in the political-administrative organization of the region in which they reside since then, this isolated ethnic group has preserved its language, social structure, and territory [32]. These facts also highlight their sedentary nature, which has led to limited genetic admixture [33]. Considering the appraised size of the community where they come from [34], the calculated prevalence of FA is 7.5/1,000,000 which falls in the upper rank of the estimated worldwide FA prevalence in the general population (1-9/1,000,000) [35]. Of note, founder PV constitute more than 10% of the total PV discovered so far in *FANCG*, the highest proportion among the most prevalent *FANC* genes (Figure 4). Although most of the PV in FA are concentrated in *FANCA*, *FANCC*, and *FANCG*, a study suggest that they do not appear to be intrinsically more mutable than the other *FANC* genes [36], therefore another explanation should be sought to understand the higher prevalence of PV in these genes. On the other hand, it is unlikely that the rate of FPV in *FANCG* could confer any advantage to carriers, since half of them generate a truncated protein and others a dysfunctional one, so it is likely that they may occur due to chance. Certainly, this observation deserves further investigation.

The use of *in silico* tools anticipates that the PV *FANCG*:c. 511-3\_511-2delCA affects a splicing acceptor site, which would lead to exon 5 skipping in the processed mRNA. The analysis of this deletion showed that this in-frame change does not generate a stop codon, therefore the synthesis of a mis-spliced transcript is expected [37,38]. The splicing impairment potential of this PV is further supported by the fact that the affected site is *cis*-acting canonical, which usually promotes the loss of single exons [38].

Although it is known that FANCG consists of at least 7 TPR motifs that cover most of its surface and facilitate its protein interplays, it has not yet been possible to isolate it individually. For this reason, we chose a template-folding modeling program (Phyre2), capable of integrating comparative analysis, homology, and fold recognition strategies, to recreate the effect of exon 5 skipping on FANCG and put into perspective the differences between the wild-type and mutant proteins. Our 3D model predicts a conformational change in the mutant protein in which the size of the binding groove is reduced due to the shortening of the TPR1 motif (Figure 6). Several reports indicate that changes in this motif, such as phosphorylation of serine 7 and conservation of the eighth residue, are critical in regulating its interactions with other proteins of the FA/BRCA pathway, such as FANCA, FANCD1 (BRCA2), and FANCF [15,39]. The analysis of the stability of the mutant model is consistent with the feasibility of producing a shortened protein from a transcript that is not degraded through non-sense mediated decay (NMD). This evidence suggests that alterations in TPR1 jeopardize their interaction, so the mutant protein must have compromised functionality, which could explain the FA phenotype.

The three patients presented here have a classic FA phenotype, consisting of BMF and physical abnormalities [40]. The acronyms VACTERL-H and PHENOS include the most frequently reported physical abnormalities in patients with FA [30,41]. Among these 14 features, only five are present in over 25% of patients: short stature (43%), radial ray defects (40%), changes in skin pigmentation (37%), microcephaly (27%), and renal malformations (27%) [41]. Figure 5 shows that 4% of patients reported in the literature with PV in *FANCG* met VACTERL-H criteria; while if we consider all genotypes, this percentage rises to 12% [41]. The percentage of patients reported in the literature with PV in *FANCG* who had  $\geq 4/6$  PHENOS features were similar considering all genotypes (8.3 v.s. 9%). None of our patients had VACTERL-H and only one had  $\geq 4/6$  PHENOS features. All of them share height, radial and skin abnormalities, two also had microcephaly and ultrasonographic evaluations rule out renal malformations. As expected, although these three patients share the same pathogenic *FANCG* genotype and a similar genetic background, their phenotypic presentation was not homogeneous, as has been evidenced in other cases [42]. Yet, they illustrate that clinical presentation in FA has a broad spectrum of severity that in this case goes from a florid presentation in FANCG143 which includes five PHENOS features and uncommon findings like bifid uvula, to a more discreet presentation in patient FANCG155 in whom her unilateral radial abnormality was only recorded after intentional evaluation by a dysmorphologist (Table 4). Concerning the hematologic phenotype, patients with *FANCG* PV have been found to have more severe cytopenias as well as a higher frequency and an earlier diagnosis of AML or MDS than patients with *FANCA* or *FANCC* genotypes [43]. All three patients have a hematologic phenotype that merited HSCT, yet platelets and neutrophils appeared to be more severely affected than red cells. Oncologic manifestations were not found in these patients, but this observation is certainly limited not only because of the small number of individuals but also due to their young age.

To our knowledge, the only available phenotypic descriptions of cohorts of patients with FA that have a FPV *FANCG* genotype refer to Black South African individuals, with the homozygous variant c.637\_643del [14]. The analysis of clinical data from 35 of these patients could not definitely establish a particular phenotype for this genotype, yet it was recognized that they had a high frequency of skin pigmentation alterations (97%) and that anomalies of the upper limbs although frequent (>70%) were subtle (no radial hypoplasia). Also, the frequency of renal abnormalities (37%) was higher than reported elsewhere [44]. A second study that included 24 patients and focused in endocrinologic data, showed that 33% had microcephaly, 45.8% had short stature for age and 41.7% had low weight for age [45]. When comparing the phenotype of the three *Mixe* patients to the Black South African patients, in whom the genotype leads to a truncated protein (p.Tyr213Lysfs\*6), the anthropometric alterations appear to be more severe [45]. We take these



observations with caution, both because of the small size of the *Mixe* cohort and the influences of environmental factors and additional *loci* in these multifactorial traits cannot be overlooked. This is why a study in this geographical region should be warranted.

In conclusion, in this study we present the ninth FPV in the *FANCG* gene, thereby contributing to reaffirm that this gene has the highest proportion of FPV among the most prevalent *FANC* genes. Despite having similar genetic and environmental backgrounds, the phenotype of these three *Mixe* patients agrees with the highly variable dysmorphological phenotype of patients with PV in *FANCG* described in the literature. The identification of the founder *FANCG*: c.511-3\_511-2delCA PV in these patients with FA, constitutes a fundamental step for the systematic and accurate genetic screening of patients with FA in the *Mixe* indigenous group, this could lead to carrier identification, appropriate genetic counseling for families with FA, and optimization of the search for potential bone marrow transplant donors. Finally, the detection of a high-prevalence population-specific disease-causing PV, represents an opportunity to study the phenotypic effect of the variant in a population with a homogeneous genetic background.

## 4. Materials and Methods

### 4.1 Editorial Policies and Ethical considerations.

This study was approved by the Ethics and Research Committees of the Instituto Nacional de Pediatría (INP) in Mexico City. All patients provided written informed consent to participate in this study (INP 041-2014) and blood samples of each participant were collected by peripheral venipuncture.

### 4.2 Patients.

Three patients with cytopenias were independently referred for a medical evaluation to the INP. These patients came from the same geographic region in the southern part of Mexico (Oaxaca State). All patients underwent a comprehensive physical examination by trained medical geneticists. Clinical data including age, sex, parental consanguinity, family history, and anthropometric measurements (weight, height, and head circumference) were recorded. Hematological information, as full blood counts at presentation and subsequent monitoring, bone marrow biopsy results, age of development of bone marrow failure, and transfusions were documented through retrospective review of each patient's medical file. Available imaging studies were evaluated.

### 4.3 Chromosomal breakage test.

A diepoxybutane (DEB) test was performed in all patients. Two-paired lymphocyte cultures per blood sample were cultured in RPMI-1640 medium (Gibco, BRL, Grand Island, NY) and stimulated with phytohemagglutinin (Gibco, BRL, Grand Island, NY). To induce DNA damage, 0.1 µg/ml of DEB (Sigma, St Louis, MO) was added from the start of the incubation to half of the cultures; the untreated cultures were used for the analysis of spontaneous chromosomal breakage. In parallel, cultures of blood samples from healthy individuals as negative controls and the FA VU817 cell line as a positive control were performed. Harvesting of all cultures was carried out after of 72 h at 37°C in a 5% CO<sub>2</sub> incubator. Metaphase spreads and staining were performed according to standard protocols [46]. We evaluated the frequency of chromosome aberrations including breaks, fragments, dicentrics, rings, and radial figures. The frequency of aberrations per cell, as well as the percentage of aberrant cells, were calculated.

### 4.4 Genotyping.



#### 4.4.1 Genomic DNA extraction

The Gentra Puregene Kit (Qiagen, Venlo, Limburg, NL) was used for the extraction of total genomic DNA from peripheral blood samples. The concentrations of the DNA were determined on a Nanodrop spectrophotometer (Nanodrop Technologies, Wilmington, DE, USA) and DNA integrity was verified by agarose gel electrophoresis.

#### 4.4.2 Targeted Next-Generation Sequencing

A customized HaloPlex panel (Agilent Technologies, Santa Clara, CA, USA) was designed to target 16 FA genes (*FANCA*, *B*, *C*, *D1*, *D2*, *E*, *F*, *G*, *I*, *J*, *L*, *M*, *N*, *O*, *P* and *Q*). The final design included all coding exons and 50 bp of their flanking 5' and 3' intronic regions, with a coverage of 99.5%. Agilent's HaloPlex Enrichment System protocol was followed for library preparation and paired-end sequencing was performed in the MiSeq System platform (Illumina, San Diego, CA, USA). Raw reads were filtered and mapped to the GRCh37/hg19 reference genome with the Burrows-Wheeler Aligner (BWA). The overall mean sequencing depth of the samples was 1000X. Alignment and variant calling were performed using the Genome Analysis Toolkit (GATK), version 4.0.3. A pipeline was designed to maximize the accuracy of variant calls, according to GATK Best Practices recommendations [47, 48]. Variants were described following the guidelines proposed by the HGVS nomenclature [49]. The classification of variants, according to the five-tier criteria of the ACMG guidelines [20] as well as the estimated global frequency, were assessed using the online tool VarSome [50].

#### 4.4.3 Sanger sequencing

*FANCG* variant-specific primers were designed to encompass the splicing acceptor site between intron 4 and exon 5, using the Primer3 software (<http://bioinfo.ut.ee/primer3-0.4.0/>). The selected primers were as follows: forward 5'-GACCTTGGCGGTAGGCAAA; and reverse 5'-ATTGGGGGAACTACAGGCA. PCR amplification was carried out from 100 ng of DNA template. PCR products were purified with QIAquick kit (QIAGEN, Venlo, Limburg, NL) according to manufacturer instructions. Purified amplicons were bidirectionally sequenced using the Big Dye Terminator sequencing kit and resolved on an Applied Biosystems 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). The electropherograms were analyzed with Chromas V2.6.6 software ([www.technelysium.com.au](http://www.technelysium.com.au)) and target sequences were compared to the corresponding reference sequences *FANCG* (NG\_007312.1) from GenBank (<https://www.ncbi.nlm.nih.gov>).

#### 4.5 *In silico* analysis.

##### 4.5.1 Splice site analysis

Human Splicing Finder (HSF) (<http://www.umd.be/HSF/>), MaxEntScan ([http://hollywood.mit.edu/burge-lab/maxent/Xmaxentscan\\_scoreseq\\_acc.html](http://hollywood.mit.edu/burge-lab/maxent/Xmaxentscan_scoreseq_acc.html)) and CRYP-SKIP (<https://cryp-skip.img.cas.cz/>) were the three bioinformatics tools used to predict the effect of the variant on splicing signals.

#### 4.5.2 Modeling of FANCG wildtype and mutant proteins

Due to the lack of X-ray crystal structure of FANCG in the Protein Data Bank (PDB), we used the Phyre2 server [51; <http://www.sbg.bio.ic.ac.uk/~phyre2/>] to generate 3D models of wildtype (WT) and mutant (MT) FANCG proteins. This server applies Hidden Markov Method to produce alignments of submitted protein sequences against proteins with published structure. The resulting alignments are later used to generate 3D homology-based structures of the inquiry sequences. Most suitable 3D models were selected for final validation, using Z-scores derived from the ProSA-web-server [52; <https://prosa.services.came.sbg.ac.at/prosa.php>]

#### 4.6 High-resolution microarray analysis.

We performed the analysis of the gDNA with Affymetrix CytoScan 750K Arrays (Affymetrix, Santa Clara, USA) on the three patients. This array provides a genome-wide coverage, including 550,000 markers for detecting copy number variation and 200,436 SNP probes. We performed the procedures for DNA digestion, ligation, PCR amplification, fragmentation, labelling, denaturing and hybridization into the array according to the protocols and QC guidelines provided by the supplier. Arrays were then stained and washed in the affymetrix GeneChip Fluidics Station 450 and scanned using an Affymetrix GeneChip Scanner 3000 7G (Affymetrix, Santa Clara, USA); we analyzed the files obtained with the appropriate bioinformatics tools.

##### 4.6.1 Analysis of long-contiguous stretch of homozygosity (LCSH).

We visualized the LCSH in the software Chromosome Analysis Suite (ChAS) software version 4.1, provided by Affymetrix (Affymetrix, Santa Clara, CA, USA). For the analysis, we used the NetAffx 33 hg19 annotation files (<https://www.affymetrix.com/analysis/index.affx>). For LCSH>3Mb, the analysis configuration was set at LOH with marker count = 50 and size = 3000 Kbp, and for LCSH>5Mb was set at marker count = 50 and size = 5000 Kbp.

##### 4.6.2 Estimation of coefficient of inbreeding (F)

Individual inbreeding coefficients (F) were estimated using LCSH>3Mb data; F was the total length of autosomal LCSH in Kbp divided by the total autosomal size covered by the Cytoscan 750 array (2,881,033,286 Kbp for hg19). The F value of 0.25 could reflect a first-degree parental relationship, 0.125 a second-degree, 0.0625 a third-degree and 0.03125 a fourth-degree [53].

##### 4.6.3 Haplotype inference

We used the LCSH>5Mb data to perform a haplotype inference approach, specifically the LCSH region shared by the three patients that include *FANCG*. The number of single nucleotide variants (SNVs) identical in the three patients was calculated, and the allele frequency was reviewed to infer the haplotype. The SNVs was compared with those reported in global population and the unrelated Mexican individuals (86 samples, MXL population) found in the 1000 Genomes Project (phase 3) database (1KGP) (<ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase3/data>), using the LDhap tool (<https://ldlink.nci.nih.gov/?tab=ldhap>).

#### 4.7 Ancestry analysis.

For the ancestry study, .CEL files derived from the arrays were analyzed with the ChAS software to obtain the SNV genotyping calls, generated from the use of the BRLMM-P-plus algorithm. Subsequently, these results were contrasted with genotype information derived from Affymetrix high-density arrays present in the 1KGP. PLINK 1.9 software (<https://zzz.bwh.harvard.edu/plink/download.shtml/>) was used to filter and mapping the SNVs. SNVs with discordant strings or genotypes were corrected or removed. Finally, the global ancestry of the MXL population and the three patients was deduced through a supervised maximum likelihood ADMIXTURE approach, from K=2 to K=3 ancestral components. For this analysis, genotypic frequencies of Northern European (CEU), Yoruba (YRI), and Native American (NAT), reported in the 1KGP, were considered as reference parental populations.

#### 4.8 Fanconi Anemia Variant Database Analysis.

We searched the Fanconi Anemia Mutation Database (displayed using Leiden Open Variation Database [LOVD, v.3.0], <https://www2.rockefeller.edu/fanconi/genes/>) for *FANCA*, *FANCC*, and *FANCG* as of November 30<sup>th</sup>, 2021. We first categorized the variant according to the number of times it has been reported. In those where five or more patients were acknowledged, we reviewed the cited publications in order to establish if a haplotype analysis demonstrated a founder effect. We complemented this with a PubMed/MEDLINE search using the terms: “*FANCA* OR *FANCC* OR *FANCG*” AND “mutation OR variant” AND “founder OR haplotype”. We classified the variants as founder PV (FPV), when demonstrated by haplotype analysis, repeatedly reported non-founder PV (recurrent, but not demonstrated as FPV by haplotype analysis), and those reported just once. We compared the proportion of these variant categories between *FANCG* and the other two genes *FANCA* and *FANCC*. Fisher's exact test ( $p$ -values < .05) was used to statistically test these differences.

#### 4.9 Phenotype Analysis

We searched PubMed for publications limited to human subjects through December 1, 2020 using the terms “Fanconi” AND “anemia”. We collected and analyzed all cases with a *FANCG* genotype in whom individual patient's phenotypes were detailed. We described frequencies of VACTERL-H (Vertebral, Anal, Cardiac, Tracheo-esophageal fistula, Esophageal atresia, Renal, upper Limb, and Hydrocephalus) and PHENOS (skin Pigmentation, small Head, small Eyes, Nervous system, Otology, Short stature) features. VACTERL-H association was considered if the patient had  $\geq 3/8$  features, and for PHENOS  $\geq 4/6$  features were needed. We extracted clinical information from publications reporting in *FANCG* FPV (as described in 4.8 section) to compare it to the phenotype of the patients reported here. The statistical analyses were performed using Microsoft Excel Office 365 (Microsoft, Redmond, WA, USA) and R (R Core Team [2013]. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>). We used Fisher's Exact and  $p$ -values < 0.05 were significant.

**Supplementary Materials:** The following are available online at [www.mdpi.com/xxx/s1](http://www.mdpi.com/xxx/s1).

Supplementary Table S1. Long-contiguous stretch of homozygosity (LCSH), previously reported as commons in different population.

Supplementary Table S2. Estimation of the haplotype of FA patients from SNVs analysis.

**Author Contributions:**

Conceptualization, Benilde García-deTeresa, Leda Torres and Sara Frias; Data curation, Benilde García-deTeresa, Ulises Juárez, Moisés Fiesco-Roa, Leda Torres and Sara Frias; Formal analysis, Pedro Reyes, Benilde García-deTeresa, Ulises Juárez, Fernando Pérez-Villatoro, Moisés Fiesco-Roa, Alfredo Rodríguez, Bertha Molina, Maria Teresa Villarreal-Molina, Jorge Melendez-Zajgla, Leda Torres and Sara Frias; Funding acquisition, Alessandra Carnevale, Leda Torres and Sara Frias; Investigation, Sara Frias; Methodology, Pedro Reyes, Moisés Fiesco-Roa, Bertha Molina, Maria Teresa Villarreal-Molina and Leda Torres; Project administration, Leda Torres and Sara Frias; Resources, Leda Torres and Sara Frias; Software, Ulises Juárez, Fernando Pérez-Villatoro and Leda Torres; Supervision, Sara Frias; Validation, Pedro Reyes; Writing – original draft, Pedro Reyes and Leda Torres; Writing – review & editing, Benilde García-deTeresa, Ulises Juárez, Fernando Pérez-Villatoro, Moisés Fiesco-Roa, Alfredo Rodríguez, Bertha Molina, Maria Teresa Villarreal-Molina, Jorge Melendez-Zajgla, Alessandra Carnevale, Leda Torres and Sara Frias.

**Funding:** This research was partially funded by Consejo Nacional de Ciencia y Tecnología (CONACyT), grant number SALUD-2014-1-233721, and by Instituto Nacional de Pediatría, Recursos Fiscales E022 project INP 2014/41. Pedro Reyes acknowledges for the support of CONACYT, through the scholarship 899784, to develop his PhD in Biomedical Sciences at UNAM.

**Institutional Review Board Statement:** This study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of Instituto Nacional de Pediatría (INP), Ciudad de Mexico, Mexico (protocol code: INP 2014/041; date of approval: July 17<sup>th</sup>, 2014).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study to publish this paper.

**Acknowledgments:** We would like to thank the patients for their consent to participate in this study. We would like to acknowledge Yañez J., Paul Gaytan P., Lopez E., and Becerra S. from the Sequencing and Synthesis Unit of the Instituto de Biotecnología, UNAM, for primer synthesis and sequencing services.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Abbreviations

1KGP	1000 Genomes Project
ACMG	American College of Medical Genetics
AMR	Ad Mixed American (1KGP super population)
ANC	Absolut Neutrophils Counts
BMF	Bone marrow failure
CNS	Central nervous system
CEU	Utah residents (CEPH) with Northern and Western European ancestry (1KGP population)
ChAS	Chromosome Analysis Suite
CHB	Han Chinese in Beijing, China (1KGP population)
DEB	Diepoxibutane

---

FA	Fanconi anemia
FA/BRCA	Fanconi Anemia/Breast Cancer pathway
FPV	Founder pathogenic variant
Hb	Hemoglobin
HGVS	Human Genome Variation Society
HSCT	Hematopoietic stem-cell transplantation
HSF	Human Splicing Finder
INP	Instituto Nacional de Pediatría
LCSH	Long-contiguous stretches of homozygosity
MCV	Mean Corpuscular Volume
MXL	Mexican Ancestry in Los Angeles CA, USA (1 KGP population)
MRI	Magnetic resonance imaging
MT	Mutant
NAT	Native Americans (1KGP population)
NMD	Non-sense mediated decay
OMIM	Online Mendelian Inheritance in Man
PCR	Polymerase chain reaction
PHENOS	Skin Pigmentation, small Head, small Eyes, Nervous system, Otology, Short stature
PLT	Platelets
PVS1	Pathogenic very strong 1
PM2	Pathogenic Moderate 2
PP5	Pathogenic Supporting 5
PDB	Protein Data Bank
PV	Pathogenic variants
SNV	Single nucleotide variant
SDS	Standard deviation score
TPR	Tetratricopeptide repeat
URTI	Upper respiratory tract infection
VACTERL-H Hydrocephalus	Vertebral, Anal, Cardiac, Tracheo-esophageal fistula, Esophageal atresia, Renal, upper Limb, and Hydrocephalus
WG	Weeks of gestation
WT	Wild type
YRI	Yoruba in Ibadan, Nigeria (1KGP population)

## References

1. Auerbach, A. D. Fanconi Anemia and Its Diagnosis. *Mutat. Res.* 2009, 668 (1–2), 4–10. <https://doi.org/10.1016/j.mrfmmm.2009.01.013>.
2. Alter, B. P.; Giri, N.; Savage, S. A.; Rosenberg, P. S. Cancer in the National Cancer Institute Inherited Bone Marrow Failure Syndrome Cohort after Fifteen Years of Follow-Up. *Haematologica* 2018, 103 (1), 30–39. <https://doi.org/10.3324/haematol.2017.178111>.
3. Savage, S. A.; Walsh, M. F. Myelodysplastic Syndrome, Acute Myeloid Leukemia, and Cancer Surveillance in Fanconi Anemia. *Hematol. Oncol. Clin. North Am.* 2018, 32 (4), 657–668. <https://doi.org/10.1016/j.hoc.2018.04.002>.

4. Kottemann, M. C.; Smogorzewska, A. Fanconi Anaemia and the Repair of Watson and Crick DNA Crosslinks. *Nature* 2013, 493 (7432), 356–363. <https://doi.org/10.1038/nature11863>.
5. Rodríguez, A.; D'Andrea, A. Fanconi Anemia Pathway. *Curr. Biol.* 2017, 27 (18), R986–R988. <https://doi.org/10.1016/j.cub.2017.07.043>.
6. Fanconi Anemia Clinical Care Guidelines. Fifth Edition; Sroka, I; Frohnmayer, L; Van Ravenhorst, S; Wirkkula, L. Fanconi Anemia Research Found, 2020. URL: <https://www.fanconi.org/explore/clinical-care-guidelines>
7. Niraj, J.; Färkkilä, A.; D'Andrea, A. D. The Fanconi Anemia Pathway in Cancer. *Annu. Rev. Cancer Biol.* 2019, 3 (1), 457–478. <https://doi.org/10.1146/annurev-cancerbio-030617-050422>.
8. Gille, J. J. P.; Floor, K.; Kerkhoven, L.; Ameziane, N.; Joenje, H.; de Winter, J. P. Diagnosis of Fanconi Anemia: Mutation Analysis by Multiplex Ligation-Dependent Probe Amplification and PCR-Based Sanger Sequencing. *Anemia* 2012, 2012, 603253. <https://doi.org/10.1155/2012/603253>.
9. Callén, E.; Casado, J. A.; Tischkowitz, M. D.; Bueren, J. A.; Creus, A.; Marcos, R.; Dasí, A.; Estella, J. M.; Muñoz, A.; Ortega, J. J.; de Winter, J.; Joenje, H.; Schindler, D.; Hanenberg, H.; Hodgson, S. V.; Mathew, C. G.; Surrallés, J. A Common Founder Mutation in FANCA Underlies the World's Highest Prevalence of Fanconi Anemia in Gypsy Families from Spain. *Blood* 2005, 105 (5), 1946–1949. <https://doi.org/10.1182/blood-2004-07-2588>.
10. Whitney, M. A.; Saito, H.; Jakobs, P. M.; Gibson, R. A.; Moses, R. E.; Grompe, M. A Common Mutation in the FACC Gene Causes Fanconi Anaemia in Ashkenazi Jews. *Nat. Genet.* 1993, 4 (2), 202–205. <https://doi.org/10.1038/ng0693-202>.
11. de Vries, Y., Lwiwski, N., Levitus, M., Kuyt, B., Israels, S., Arwert, F., Zwaan, M., Greenberg, C., Alter, B., Joenje, H. and Meijers-Heijboer, H. A Dutch Fanconi Anemia FANCC Founder Mutation in Canadian Manitoba Mennonites. *Anemia*, 2012, pp.1-6. <https://doi.org/10.1155/2012/865170>
12. García-de Teresa, B.; Frias, S.; Molina, B.; Villarreal, M. T.; Rodríguez, A.; Carnevale, A.; López-Hernández, G.; Vollbrechtshausen, L.; Olaya-Vargas, A.; Torres, L. FANCC Dutch Founder Mutation in a Mennonite Family from Tamaulipas, México. *Mol. Genet. Genomic Med.* 2019, 7 (6), e710. <https://doi.org/10.1002/mgg3.710>.
13. Auerbach, A. D.; Greenbaum, J.; Pujara, K.; Batish, S. D.; Bitencourt, M. A.; Kokemohr, I.; Schneider, H.; Lobitz, S.; Pasquini, R.; Giampietro, P. F.; Hanenberg, H.; Levrán, O.; International Fanconi Anemia Registry. Spectrum of Sequence Variation in the FANCG Gene: An International Fanconi Anemia Registry (IFAR) Study. *Hum. Mutat.* 2003, 21 (2), 158–168. <https://doi.org/10.1002/humu.10166>.
14. Morgan, N. V.; Essop, F.; Demuth, I.; de Ravel, T.; Jansen, S.; Tischkowitz, M.; Lewis, C. M.; Wainwright, L.; Poole, J.; Joenje, H.; Digweed, M.; Krause, A.; Mathew, C. G. A Common Fanconi Anemia Mutation in Black Populations of Sub-Saharan Africa. *Blood* 2005, 105 (9), 3542–3544. <https://doi.org/10.1182/blood-2004-10-3968>.
15. Wilson, J. B.; Blom, E.; Cunningham, R.; Xiao, Y.; Kupfer, G. M.; Jones, N. J. Several Tetratricopeptide Repeat (TPR) Motifs of FANCG Are Required for Assembly of the BRCA2/D1-D2-G-X3 Complex, FANCD2 Monoubiquitylation and Phleomycin Resistance. *Mutat. Res.* 2010, 689 (1–2), 12–20. <https://doi.org/10.1016/j.mrfmmm.2010.04.003>.
16. Blom, E.; van de Vrugt, H. J.; de Vries, Y.; de Winter, J. P.; Arwert, F.; Joenje, H. Multiple TPR Motifs Characterize the Fanconi Anemia FANCG Protein. *DNA Repair (Amst.)* 2004, 3 (1), 77–84. <https://doi.org/10.1016/j.dnarep.2003.09.007>.
17. Shakeel, S.; Rajendra, E.; Alcón, P.; O'Reilly, F.; Chorev, D. S.; Maslen, S.; Degliesposti, G.; Russo, C. J.; He, S.; Hill, C. H.; Skehel, J. M.; Scheres, S. H. W.; Patel, K. J.; Rappsilber, J.; Robinson, C. V.; Passmore, L. A. Structure of the Fanconi Anaemia Monoubiquitin Ligase Complex. *Nature* 2019, 575 (7781), 234–237. <https://doi.org/10.1038/s41586-019-1703-4>.
18. Ceccaldi, R.; Sarangi, P.; D'Andrea, A. D. The Fanconi Anaemia Pathway: New Players and New Functions. *Nat. Rev. Mol. Cell Biol.* 2016, 17 (6), 337–349. <https://doi.org/10.1038/nrm.2016.48>
19. García-de-Teresa, B.; Rodríguez, A.; Frias, S. Chromosome Instability in Fanconi Anemia: From Breaks to Phenotypic Consequences. *Genes (Basel)* 2020, 11 (12), 1528. <https://doi.org/10.3390/genes11121528>



20. Richards, S.; Aziz, N.; Bale, S.; Bick, D.; Das, S.; Gastier-Foster, J.; Grody, W. W.; Hegde, M.; Lyon, E.; Spector, E.; Voelkerding, K.; Rehm, H. L.; ACMG Laboratory Quality Assurance Committee. Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* 2015, 17 (5), 405–424. <https://doi.org/10.1038/gim.2015.30>.
21. Li, L.-H.; Ho, S.-F.; Chen, C.-H.; Wei, C.-Y.; Wong, W.-C.; Li, L.-Y.; Hung, S.-I.; Chung, W.-H.; Pan, W.-H.; Lee, M.-T. M.; Tsai, F.-J.; Chang, C.-F.; Wu, J.-Y.; Chen, Y.-T. Long Contiguous Stretches of Homozygosity in the Human Genome. *Hum. Mutat.* 2006, 27 (11), 1115–1121. <https://doi.org/10.1002/humu.20399>.
22. Pajusalu, S.; Žilina, O.; Yakoreva, M.; Tammur, P.; Kuuse, K.; Mölter-Väär, T.; Nõukas, M.; Reimand, T.; Õunap, K. The Diagnostic Utility of Single Long Contiguous Stretches of Homozygosity in Patients without Parental Consanguinity. *Mol. Syndromol.* 2015, 6 (3), 135–140. <https://doi.org/10.1159/000438776>.
23. Wang, J.-C.; Ross, L.; Mahon, L. W.; Owen, R.; Hemmat, M.; Wang, B. T.; El Naggar, M.; Kopita, K. A.; Randolph, L. M.; Chase, J. M.; Matas Aguilera, M. J.; Siles, J. L.; Church, J. A.; Hauser, N.; Shen, J. J.; Jones, M. C.; Wierenga, K. J.; Jiang, Z.; Haddadin, M.; Boyar, F. Z.; Anguiano, A.; Strom, C. M.; Sahoo, T. Regions of Homozygosity Identified by Oligonucleotide SNP Arrays: Evaluating the Incidence and Clinical Utility. *Eur. J. Hum. Genet.* 2015, 23 (5), 663–671. <https://doi.org/10.1038/ejhg.2014.153>.
24. Chaves, T. F.; Oliveira, L. F.; Ocampos, M.; Barbato, I. T.; de Luca, G. R.; Barbato Filho, J. H.; de Camargo Pinto, L. L.; Bernardi, P.; Maris, A. F. Long Contiguous Stretches of Homozygosity Detected by Chromosomal Microarrays (CMA) in Patients with Neurodevelopmental Disorders in the South of Brazil. *BMC Med. Genomics* 2019, 12 (1), 50. <https://doi.org/10.1186/s12920-019-0496-5>.
25. Silva-Zolezzi, I.; Hidalgo-Miranda, A.; Estrada-Gil, J.; Fernandez-Lopez, J. C.; Uribe-Figueroa, L.; Contreras, A.; Balam-Ortiz, E.; del Bosque-Plata, L.; Velazquez-Fernandez, D.; Lara, C.; Goya, R.; Hernandez-Lemus, E.; Davila, C.; Barrientos, E.; March, S.; Jimenez-Sanchez, G. Analysis of Genomic Diversity in Mexican Mestizo Populations to Develop Genomic Medicine in Mexico. *Proc. Natl. Acad. Sci. U. S. A.* 2009, 106 (21), 8611–8616. <https://doi.org/10.1073/pnas.0903045106>.
26. Spear, M. L.; Diaz-Papkovich, A.; Ziv, E.; Yracheta, J. M.; Gravel, S.; Torgerson, D. G.; Hernandez, R. D. Recent Shifts in the Genomic Ancestry of Mexican Americans May Alter the Genetic Architecture of Biomedical Traits. *Elife* 2020, 9. <https://doi.org/10.7554/eLife.56029>.
27. Yagasaki, H.; Hamanoue, S.; Oda, T.; Nakahata, T.; Asano, S.; Yamashita, T. Identification and Characterization of Novel Mutations of the Major Fanconi Anemia Gene FANCA in the Japanese Population. *Hum. Mutat.* 2004, 24 (6), 481–490. <https://doi.org/10.1002/humu.20099>.
28. Demuth, I.; Wlodarski, M.; Tipping, A. J.; Morgan, N. V.; de Winter, J. P.; Thiel, M.; Gräsl, S.; Schindler, D.; D'Andrea, A. D.; Altay, C.; Kayserili, H.; Zatterale, A.; Kunze, J.; Ebell, W.; Mathew, C. G.; Joenje, H.; Sperling, K.; Digweed, M. Spectrum of Mutations in the Fanconi Anaemia Group G Gene, FANCG/XRCC9. *Eur. J. Hum. Genet.* 2000, 8 (11), 861–868. <https://doi.org/10.1038/sj.ejhg.5200552>.
29. Park, J.; Kim, M.; Jang, W.; Chae, H.; Kim, Y.; Chung, N.-G.; Lee, J.-W.; Cho, B.; Jeong, D.-C.; Park, I. Y.; Park, M. S. Founder Haplotype Analysis of Fanconi Anemia in the Korean Population Finds Common Ancestral Haplotypes for a FANCG Variant: Ancestral Haplotypes for AFANCGVariant. *Ann. Hum. Genet.* 2015, 79 (3), 153–161. <https://doi.org/10.1111/ahg.12097>.
30. Alter, B. P.; Giri, N. Thinking of VACTERL-H? Rule out Fanconi Anemia According to PHENOS: VACTERL-H, PHENOS, and Fanconi Anemia. *Am. J. Med. Genet. A* 2016, 170 (6), 1520–1524. <https://doi.org/10.1002/ajmg.a.37637>.
31. Nuevo 31. Aguilar-Ordoñez, I.; Pérez-Villatoro, F.; García-Ortiz, H.; Barajas-Olmos, F.; Ballesteros-Villascán, J.; González-Buenfil, R.; Fresno, C.; Garcíaarrubio, A.; Fernández-López, J. C.; Tovar, H.; Hernández-Lemus, E.; Orozco, L.; Soberón, X.; Morett, E. Whole Genome Variation in 27 Mexican Indigenous Populations, Demographic and Biomedical Insights. *PLoS One* 2021, 16 (4), e0249773. <https://doi.org/10.1371/journal.pone.0249773>.
32. Torres-Cisneros, G. Mixes. In *Pueblos Indigenas del Mexico Contemporaneo*. Comisión Nacional para el Desarrollo de los Pueblos Indígenas:PNUD, Mexico, 2004. pp 1-43.

33. Quinto-Cortés, C. D.; Arriola, L. A.; García-Hughes, G.; García-López, R.; Molina, D. P.; Flores, M.; Palacios, R.; Piñero, D. Genetic Characterization of Indigenous Peoples from Oaxaca, Mexico, and Its Relation to Linguistic and Geographic Isolation. *Hum. Biol.* 2010, 82 (4), 409–432. <https://doi.org/10.3378/027.082.0405>.
34. Cuéntame INEGI.  
[http://www.cuentame.inegi.org.mx/monografias/informacion/oax/territorio/div\\_municipal.aspx?tema=me&e=20](http://www.cuentame.inegi.org.mx/monografias/informacion/oax/territorio/div_municipal.aspx?tema=me&e=20)  
(Accessed Dec, 8, 2021).
35. Orphanet: Anemia de Fanconi [https://www.orpha.net/consor/cgi-bin/OC\\_Exp.php?Lng=ES&Expert=84](https://www.orpha.net/consor/cgi-bin/OC_Exp.php?Lng=ES&Expert=84) (accessed Dec 19, 2021).
36. Rogers, K. J.; Fu, W.; Akey, J. M.; Monnat, R. J., Jr. Global and Disease-Associated Genetic Variation in the Human Fanconi Anemia Gene Family. *Hum. Mol. Genet.* 2014, 23 (25), 6815–6825. <https://doi.org/10.1093/hmg/ddu400>.
37. Singh, R. K.; Cooper, T. A. Pre-mRNA Splicing in Disease and Therapeutics. *Trends Mol. Med.* 2012, 18 (8), 472–482. <https://doi.org/10.1016/j.molmed.2012.06.006>.
38. Abramowicz, A.; Gos, M. Splicing Mutations in Human Genetic Disorders: Examples, Detection, and Confirmation. *J. Appl. Genet.* 2018, 59 (3), 253–268. <https://doi.org/10.1007/s13353-018-0444-7>.
39. Wilson, J. B.; Yamamoto, K.; Marriott, A. S.; Hussain, S.; Sung, P.; Hoatlin, M. E.; Mathew, C. G.; Takata, M.; Thompson, L. H.; Kupfer, G. M.; Jones, N. J. FANCG Promotes Formation of a Newly Identified Protein Complex Containing BRCA2, FANCD2 and XRCC3. *Oncogene* 2008, 27 (26), 3641–3652. <https://doi.org/10.1038/sj.onc.1211034>.
40. Shimamura, A.; Alter, B. P. Pathophysiology and Management of Inherited Bone Marrow Failure Syndromes. *Blood Rev.* 2010, 24 (3), 101–122. <https://doi.org/10.1016/j.blre.2010.03.002>.
41. Fiesco-Roa, M. O.; Giri, N.; McReynolds, L. J.; Best, A. F.; Alter, B. P. Genotype-Phenotype Associations in Fanconi Anemia: A Literature Review. *Blood Rev.* 2019, 37 (100589), 100589. <https://doi.org/10.1016/j.blre.2019.100589>.
42. Jung, M.; Mehta, P. A.; Jiang, C. S.; Rosti, R. O.; Usleaman, G.; Correa da Rosa, J. M.; Lach, F. P.; Goodridge, E.; Auerbach, A. D.; Davies, S. M.; Smogorzewska, A.; Boulad, F. Comparison of the Clinical Phenotype and Haematological Course of Siblings with Fanconi Anaemia. *Br. J. Haematol.* 2021, 193 (5), 971–975. <https://doi.org/10.1111/bjh.17061>.
43. Faivre, L.; Guardiola, P.; Lewis, C.; Dokal, I.; Ebell, W.; Zatterale, A.; Altay, C.; Poole, J.; Stones, D.; Kwee, M. L.; van Weel-Sipman, M.; Havenga, C.; Morgan, N.; de Winter, J.; Digweed, M.; Savoia, A.; Pronk, J.; de Ravel, T.; Jansen, S.; Joenje, H.; Gluckman, E.; Mathew, C. G. Association of Complementation Group and Mutation Type with Clinical Outcome in Fanconi Anemia. European Fanconi Anemia Research Group. *Blood* 2000, 96 (13), 4064–4070.
44. Feben, C.; Kromberg, J.; Wainwright, R.; Stones, D.; Sutton, C.; Poole, J.; Haw, T.; Krause, A. Phenotypic Consequences in Black South African Fanconi Anemia Patients Homozygous for a Founder Mutation. *Genet. Med.* 2014, 16 (5), 400–406. <https://doi.org/10.1038/gim.2013.159>.
45. Dillon, B.; Feben, C.; Segal, D.; du Plessis, J.; Reynders, D.; Wainwright, R.; Poole, J.; Krause, A. Endocrine Profiling in Patients with Fanconi Anemia, Homozygous for a FANCG Founder Mutation. *Mol. Genet. Genomic Med.* 2020, 8 (8), e1351. <https://doi.org/10.1002/mgg3.1351>.
46. Esmer, C.; Sánchez, S.; Ramos, S.; Molina, B.; Frias, S.; Carnevale, A. DEB Test for Fanconi Anemia Detection in Patients with Atypical Phenotypes: AF in Patients With Atypical Phenotypes. *Am. J. Med. Genet. A* 2004, 124A (1), 35–39. <https://doi.org/10.1002/ajmg.a.20327>.
47. DePristo, M. A.; Banks, E.; Poplin, R.; Garimella, K. V.; Maguire, J. R.; Hartl, C.; Philippakis, A. A.; del Angel, G.; Rivas, M. A.; Hanna, M.; McKenna, A.; Fennell, T. J.; Kernysky, A. M.; Sivachenko, A. Y.; Cibulskis, K.; Gabriel, S. B.; Altshuler, D.; Daly, M. J. A Framework for Variation Discovery and Genotyping Using Next-Generation DNA Sequencing Data. *Nat. Genet.* 2011, 43 (5), 491–498. <https://doi.org/10.1038/ng.806>.
48. Van der Auwera, G. A.; Carneiro, M. O.; Hartl, C.; Poplin, R.; Del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; Banks, E.; Garimella, K. V.; Altshuler, D.; Gabriel, S.; DePristo, M. A. From FastQ Data to High Confidence Variant

- 
- Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Curr. Protoc. Bioinformatics* 2013, 43 (1), 11.10.1-11.10.33. <https://doi.org/10.1002/0471250953.bi1110s43>.
49. den Dunnen, J. T.; Dalgleish, R.; Maglott, D. R.; Hart, R. K.; Greenblatt, M. S.; McGowan-Jordan, J.; Roux, A.-F.; Smith, T.; Antonarakis, S. E.; Taschner, P. E. M. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum. Mutat.* 2016, 37 (6), 564–569. <https://doi.org/10.1002/humu.22981>.
50. Kopanos, C.; Tsiolkas, V.; Kouris, A.; Chapple, C. E.; Albarca Aguilera, M.; Meyer, R.; Massouras, A. VarSome: The Human Genomic Variant Search Engine. *Bioinformatics* 2019, 35 (11), 1978–1980. <https://doi.org/10.1093/bioinformatics/bty897>.
51. Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J. E. The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. *Nat. Protoc.* 2015, 10 (6), 845–858. <https://doi.org/10.1038/nprot.2015.053>.
52. Wiederstein, M.; Sippl, M. J. ProSA-Web: Interactive Web Service for the Recognition of Errors in Three-Dimensional Structures of Proteins. *Nucleic Acids Res.* 2007, 35 (Web Server issue), W407–10. <https://doi.org/10.1093/nar/gkm290>.
53. Kearney, H. M.; Kearney, J. B.; Conlin, L. K. Diagnostic Implications of Excessive Homozygosity Detected by SNP-Based Microarrays: Consanguinity, Uniparental Disomy, and Recessive Single-Gene Mutations. *Clin. Lab. Med.* 2011, 31 (4), 595–613, ix. <https://doi.org/10.1016/j.cll.2011.08.003>.