




Spectrum of Germline Mutations Within Fanconi Anemia–Associated Genes Across Populations of Varying Ancestry

Sock Hoai Chan , PhD,¹ Ying Ni, PhD,² Shao-Tzu Li, MS, LCGC,¹ Jing Xian Teo , BS,³ Nur Diana Binte Ishak, BS,¹ Weng Khong Lim , PhD,^{3,4,5} Joanne Ngeow, MD, MPH^{1,6,7,*}

¹Division of Medical Oncology, National Cancer Centre Singapore, Cancer Genetics Service, Singapore, ²Genomic Medicine Institute, Cleveland Clinic, Cleveland, OH, USA, ³SingHealth Duke-NUS Institute of Precision Medicine, Singapore, ⁴Cancer and Stem Cell Biology Program, Duke-NUS Medical School Singapore, Singapore, ⁵SingHealth Duke-NUS Genomic Medicine Centre, Singapore, ⁶Oncology Academic Clinical Program, Duke-NUS Medical School, Singapore; and ⁷Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore

*Correspondence to: Joanne Ngeow, MBBS, MRCP, MPH, Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, 11 Mandalay Road, 308232 Singapore (e-mail: joanne.ngeow@ntu.edu.sg).

Abstract

Background: Fanconi anemia (FA) is a rare genetic disorder associated with hematological disorders and solid tumor predisposition. Owing to phenotypic heterogeneity, some patients remain undetected until adulthood, usually following cancer diagnoses. The uneven prevalence of FA cases with different underlying FA gene mutations worldwide suggests variable genetic distribution across populations. Here, we aim to assess the genetic spectrum of FA-associated genes across populations of varying ancestries and explore potential genotype–phenotype associations in cancer. **Methods:** Carrier frequency and variant spectrum of potentially pathogenic germline variants in 17 FA genes (excluding *BRCA1/FANCS*, *BRCA2/FANCD1*, *BRIP1/FANCF*, *PALB2/FANCF*, *RAD51C/FANCO*) were evaluated in 3523 Singaporeans and 7 populations encompassing Asian, European, African, and admixed ancestries from the Genome Aggregation Database. Germline and somatic variants of 17 FA genes in 7 cancer cohorts from The Cancer Genome Atlas were assessed to explore genotype–phenotype associations. **Results:** Germline variants in *FANCA* were consistently more frequent in all populations. Similar trends in carrier frequency and variant spectrum were detected in Singaporeans and East Asians, both distinct from other ancestry groups, particularly in the lack of recurrent variants. Our exploration of The Cancer Genome Atlas dataset suggested higher germline and somatic mutation burden between *FANCA* and *FANCC* with head and neck and lung squamous cell carcinomas as well as *FANCI* and *SLX4/FANCP* with uterine cancer, but the analysis was insufficiently powered to detect any statistical significance. **Conclusion:** Our findings highlight the diverse genetic spectrum of FA-associated genes across populations of varying ancestries, emphasizing the need to include all known FA-related genes for accurate molecular diagnosis of FA.

Fanconi anemia (FA; Online Mendelian Inheritance in Man [OMIM] PS227650) is a rare, predominantly recessive disorder associated with bone marrow failure, hematological and congenital abnormalities, and cancer predisposition. It is clinically heterogeneous, and although many FA patients present overt features at childhood, identifiable phenotype is subtle or absent in 10%–20% of patients (1,2), most of whom remain undetected until adulthood or subsequent to severe chemoradiation therapy-related toxicity following cancer diagnosis. Recognizing these FA patients with subtle features demands a high index of suspicion.

FA is associated with 22 genes (*FANCA*–*FANCW*) of the cellular FA pathway, a critical safeguard for genome integrity (3,4).

Mutations in FA pathway genes are linked to defective DNA repair, specifically the resolution of DNA interstrand cross-links by homologous recombination, resulting in vulnerability to genomic instability and hypersensitivity to DNA cross-linking agents that is characteristic among FA patients (3,5). Germline biallelic inactivation of these genes leads to FA except *FANCB* and *RAD51/FANCR*, which are associated in X-linked recessive and dominant negative inheritance, respectively (1). Notably, monoallelic inactivation of 5 FA genes is associated with hereditary breast and ovarian cancers, namely *BRCA1/FANCS*, *BRCA2/FANCD1*, *BRIP1/FANCF*, *PALB2/FANCF*, and *RAD51C/FANCO*.

In most FA cases worldwide, deleterious germline variants are found in *FANCA*, *FANCC*, and *FANCG* (1,3,6,7). However, there

Received: 19 August 2020; Revised: 24 November 2020; Accepted: 16 December 2020

© The Author(s) 2021. Published by Oxford University Press. All rights reserved. For permissions, please email: journals.permissions@oup.com
This article is published and distributed under the terms of the Oxford University Press, Standard Journals Publication Model (https://academic.oup.com/journals/pages/open_access/funder_policies/chorus/standard_publication_model)

are differences across communities attributable to founder mutations, such as a higher incidence of FANCA c.295C>T among Spanish Gypsy (8), FANCC c.711+4A>T among Ashkenazi Jews (9), and FANCC c.67delG among Mennonites of Dutch ancestry (10,11). Prevalence of other FA genetic subtypes is beginning to be discovered, such as FANCL c.1092G>A founder mutation-linked FANCL FA in South Asia (12) and FANCG FA among Japanese and Koreans associated with 2 FANCG founder mutations (c.307+1G>C, c.1066C>T) (13,14). Such cross-ancestry variability presents barriers to detection and molecular diagnosis of FA in populations that are inadequately represented in literature, potentially challenging interpretations of disease risk (15). Currently, the extent of genetic heterogeneity in FA genes across ancestries worldwide is unclear.

Because our knowledge about the clinical and genetic spectrum of FA is derived predominantly from populations of European ancestry, there exist possibilities of underdetection of FA patients among non-European populations. Here, we aim to profile the genetic variability in FA genes across racially diverse populations, including an ancestrally heterogeneous Asian group represented by Singaporeans. Additionally, we sought to evaluate associations of FA variants with different cancer types using selected cohorts from The Cancer Genome Atlas (TCGA) to explore genotype–phenotype correlations potentially underlying variability in solid malignancies presented by FA patients.

Methods

Participants

Patients suspected of FA were referred to our Cancer Genetics Service at the National Cancer Centre Singapore for physical, clinical, and family history assessment by a clinical geneticist or genetic counselor. FA diagnosis was confirmed by diepoxybutane-mediated chromosomal breakage test and multi-gene panel testing (including 22 FA genes) on peripheral blood performed by Clinical Laboratory Improvement Amendments (CLIA)-certified commercial laboratories accredited by the College of American Pathologists (CAP). Patients were enrolled in this study with written informed consent. This study was approved by SingHealth Centralized Institutional Review Board (CIRB 2011/826/B).

Data Sources

To capture germline variation in the Singaporean population, we analyzed a deidentified institutional control whole-exome dataset derived from peripheral blood of 3523 Singaporeans without known personal history of cancer (SingHealth Exome Consortium [SEC]). Ancestry was self-reported by participants, which was predominantly Chinese (overlapping East Asian group from 1000Genomes Project) (16) followed by smaller groups of non-Chinese ancestries, including Malay and Indian.

For comparison with populations of different ancestry, data representing East Asian (EAS), South Asian (SAS), non-Finnish European (NFE), Finnish European (FIN), Ashkenazi Jewish (ASJ), admixed-Americans or Latino (AMR), and African American (AFR) groups extracted from the non-TCGA subset of whole-exome data from 118 479 individuals publicly available in the Genome Aggregation Database (17) (gnomAD, v2.1.1, accessed October 19, 2019) were used. For analysis of FA gene variants in cancer cohorts, the following TCGA datasets were used: breast invasive carcinoma (TCGA-BRCA), head and neck squamous cell

carcinoma (TCGA-HNSC), lung adenocarcinoma (TCGA-LUAD), lung squamous cell carcinoma (TCGA-LUSC), ovarian serous cystadenocarcinoma (TCGA-OV), uterine corpus endometrial carcinoma (TCGA-UCEC) and uterine carcinosarcoma (TCGA-UCS). To limit confounding effects from different ancestries, only data from individuals with 1) self-reported White race, and 2) 'not Hispanic or Latino' or 'not reported' entry in ethnicity were used. Germline variants were extracted from whole-exome datasets of blood-derived normal or, if unavailable, solid-tissue normal specimens. Somatic variants were derived from Mutect2-annotated whole-exome datasets.

Variant Filtering and Analysis Criteria

Variant call format files from all datasets were annotated using a customized version of ANNOVAR (18) and filtered for variants 1) from 17 FA genes (excluding hereditary breast and ovarian cancer-associated genes: BRCA1/FANCS, BRCA2/FANCD1, BRIP1/FANCF, PALB2/FANCN, RAD51C/FANCO); 2) that passed quality control; 3) within coding or canonical ± 2 splice site; 4) with less than 3% minor allele frequency in gnomAD exomes database (for gnomAD, TCGA datasets) or the Singaporean cohort (for SEC dataset); and 5) with a minimum alternative allele fraction of 20% (for SEC, TCGA datasets) to exclude somatic mutations. In-frame insertions or deletions and stop-loss variants were excluded. Variants were presumed to be pathogenic and inactivating if they met the following criteria: 1) result in protein truncation (stop-gain, frame-shift); 2) affect splicing (splice acceptor or splice donor site with a dbcsSNV rf score of 0.8 or higher (19), Combined Annotation Dependent Deletion (CADD) phred score of 20 or higher (20), or classified as pathogenic or likely pathogenic in the ClinVar database without conflicting interpretations); or 3) missense variants with a meta-predictor Rare Exome Variant Ensemble Learner (REVEL) score of 0.7 or higher (21). Variants with a computed score exceeding the above-mentioned applied thresholds were interpreted with higher confidence to result in potentially deleterious splice alteration (dbcsSNV rf, CADD) (19,20,22) or predictive of deleterious effect on gene function (REVEL), consistent with our internal calibration and current guidelines applied by other variant interpretation groups (23,24). Finally, a FANCD2 variant (rs750338758) was removed on manual curation because it was too recurrent in the TCGA dataset and is likely an artifact.

Statistical Analysis

Participant characteristics and cohort data were tabulated with descriptive statistics, including proportions for categorical variables. Prevalence of germline variant carriers was expressed as carrier frequency (ie, the fraction of variant carriers over all individuals in the cohort) as a normalized value for comparison across cohorts of different sizes. Two-sided Fisher's exact test for statistical significance was performed for pairwise comparison of proportions, with Bonferroni correction for P values applied for multiple pairwise comparisons. All statistical tests were performed on R 3.6.0 (R Core Team, 2019) using the R packages *fmsb*, *stats*, and heatmaps generated with the *gplots* package using default settings with row-based scaling, unless otherwise specified.

Results

Profile of Germline Variation in FA Genes Among Singaporeans

To understand genetic variation in FA genes within the Singaporean population, we investigated the spectrum and estimated carrier frequency of potentially pathogenic germline variants among 3523 Singaporeans comprising individuals of Chinese ancestry (90.0%), Malay (5.0%), and 5.0% individuals of other racial groups. Collectively, 71 carriers (71 of 3523, 2.0%) across 17 FA genes were identified (Table 1), of whom 1 was a carrier for 2 different variants in *FANCD2* (c.2376G>C, p.Trp792Cys) and *FANCM* (c.2255delC, p.Ser752*). There were no homozygous carriers. Among the carriers, 29 (29 of 71, 40.8%) harbored at least 1 protein truncating variant (PTV), whereas 25 (25 of 71, 35.2%) had a missense variant and 17 (17 of 71, 23.9%) carried a splice-altering change (Table 1). One-half of the 71 carriers harbored a germline variant in *FANCA*, accounting for 1.0% (35 of 3523) of Singaporeans. Among the remaining carriers, germline variants were more frequent in *FANCM* (7 of 3523, 0.2%) followed by *FANCD2*, *FANCG*, and *XRCC2/FANCU* with a carrier frequency of 0.1%, respectively.

Spectrum of Germline Variation in Singaporeans and EAS Is Distinct From Other Ancestry Groups

Next, we compared the genetic spectrum in Singaporeans with other ancestry groups in gnomAD. Although the collective frequency of potentially pathogenic germline variant carriers in other ancestry groups was similar to Singaporeans at 2%-5% (Supplementary Table 1, available online), frequency of carriers aggregated for each FA gene is diverse across ancestries (Figure 1). Notably, all groups harbor a sizeable pool of pathogenic variant carriers in *FANCA* (Figure 1A), especially Singaporeans and gnomAD groups of EAS, NFE, and African American ancestry (Supplementary Table 1, available online).

Overall, the carrier frequency distribution across 17 FA genes in Singaporeans is consistent with gnomAD EAS, whereby carrier frequencies in *FANCD2*, *FANCI*, *FANCL*, and *FANCM* were distinctively lower than groups of European-descent (NFE, FIN, ASJ) and AMR (Figure 1, B; Supplementary Table 2, available online). This is partially due to higher carrier rates in a few recurrent variants among the European-descent groups, as reflected by the higher allele frequencies (Supplementary Table 3, available online) and a greater overall carrier-to-variant ratio in specific genes (Figure 2). For instance, the recurrent *FANCL* c.1096_1099dup (p.Thr367Asnfs*13) variant is consistently over 10 times more frequent than other *FANCL* variants among NFE, FIN, ASJ, and AMR (Supplementary Table 3, available online), contributing to the statistically significantly higher *FANCL* carrier frequency than EAS and Singaporeans. Such a distinct effect of recurrent variants was absent in Singaporean and EAS datasets, except for *FANCA* c.3068A>G (p.Glu1023Gly) with a slightly elevated allele frequency of 2-3 times more than other *FANCA* variants within both groups. Also noteworthy is that for certain genes, namely *FANCA*, *FANCG*, *FANCM*, and *XRCC2/FANCU*, proportionally more unique variants were detected in Singaporeans (Supplementary Table 4, available online).

Among the 71 Singaporean carriers, 51 unique variants were identified, of which 36 (36 of 51, 71%) were not seen in other ancestries; 33 of these occurred in single individuals (Table 1). The predominance of private variants in most FA genes is

consistent across all ancestries (Supplementary Table 4, available online). Of the unique variants, 15 (15 of 51, 29.4%) were reported in ClinVar or Fanconi Anemia Mutation Database (4), of which 11 were classified as pathogenic or likely pathogenic or were identified in FA patients and comprise mostly *FANCA* variants (Table 1). Furthermore, approximately one-half of the variants unique to Singaporeans were rare PTVs.

Of the remaining 15 Singaporean variants, 9 overlapped exclusively with gnomAD EAS, whereas 6 were also detected in other groups (Table 1). Using genes with the most germline variants as an example, namely *FANCA*, *FANCD2*, *FANCI*, and *ERCC4*, we demonstrated that the spectrum of variants harbored by Singaporeans and EAS was mostly distinct from other groups, especially of European ancestry (Figure 3). For instance, *FANCA* c.3068A>G (p.Glu1023Gly) and *ERCC4* c.2169C>A (p.Cys723*) occurred exclusively among Singaporean and EAS at allele frequencies less than 0.1%, suggesting that both are rare EAS-specific polymorphisms. Additionally, allele or codon changes distinct to EAS were also observed, notably *FANCA* c.1874G>A (p.Cys625Tyr) vs *FANCA* c.1874G>C (p.Cys625Ser) in other groups, and *FANCI* c.84 + 1G>A vs *FANCI* c.84 + 1G>C in NFE (Figure 3).

Profile of FA Gene Variations in FA-Associated Cancer Cohorts

Approximately 26% of 417 young head and neck SCC (HNSCC) patients were previously reported to harbor predicted deleterious germline FA gene variants, with an increased mutation burden observed in *FANCD2*, *FANCE*, and *FANCL* (25). Using the White racial subpopulation of the TCGA dataset, we sought to explore whether such an association with specific FA genes may occur for cancers commonly seen in FA.

Overall, TCGA-HNSC and TCGA-LUSC cohorts trend similarly in having proportionally more pathogenic germline variant carriers in *FANCA*, *FANCC*, and *FANCG* compared with other cohorts but less in genes, including *FANCF*, *SLX4/FANCP* and *XRCC2/FANCU* (Supplementary Table 5; Supplementary Figure 1, available online). The converse is true for TCGA-UCEC, where germline variants were more frequent in *SLX4/FANCP* and *FANCF* than *FANCA*, *FANCC*, or *FANCG*. Interestingly, germline carriers of rare *FANCI* variants (allele frequency <0.002% in gnomAD NFE) were more frequent in both TCGA-UCEC and TCGA-HNSC cohorts, although these associations were not statistically significant (Table 2; Supplementary Table 6, available online). Consistent with emerging data suggesting predisposition to breast cancer among *FANCM* (26,27) and *XRCC2/FANCU* (28) monoallelic carriers, we observed proportionally more pathogenic germline *FANCM* and *XRCC2/FANCU* variant carriers in the TCGA-BRCA cohort (Table 2). Association with an increased breast cancer risk has been reported for several variants, including *FANCM* c.5101C>T (p.Gln1701*) (26), c.5791C>A (p.Arg1931*) (29), c.1972C>T (p.Arg658*) (30), and *XRCC2* c.96del (p.Phe32Leufs*30) (31). Although these broad trends were less consistent on survey of somatic mutation burden across the TCGA cohorts, we observed more mutations in *FANCA* with TCGA-LUSC, *FANCC* with TCGA-HNSC and TCGA-LUSC as well as *FANCI* and *SLX4/FANCP* with TCGA-UCEC (Supplementary Table 7, available online).

The spectrum of germline variants is variable for FA genes across TCGA cohorts. Notably, proportionally more PTVs were observed with TCGA-HNSC and TCGA-LUSC in *FANCA* and with TCGA-BRCA in *FANCM* (Table 2). Many of these PTVs occur

Table 1. Potentially pathogenic germline variants in the 17 FA genes found in 71 Singaporeans

Gene	Reference transcript	Variant type	Genomic change (hg19)	Coding DNA change	Protein change	Allele count	Allele frequency ^a	Observed in other ancestry groups?	Observed in ClinVar/FADB? (classification)	Occurred in FA patient?
FANCA	NM_000135	PT	chr16: g.89862330_89862333del	c.987_990del	His330Alafs*4	1	1.4 × 10 ⁻⁴	Yes (EAS, SAS, NFE, AMR)	Yes (P)	Yes
		PT	chr16: g.89857830G>C	c.1340C>G	Ser447*	1	1.4 × 10 ⁻⁴	No	Yes (P)	—
		PT	chr16: g.89833622_89833623dup	c.2527_2528dup	Ser844Thrfs*46	1	1.4 × 10 ⁻⁴	No	No	—
		PT	chr16: g.89831367del	c.2709delG	Trp903Cysfs*18	1	1.4 × 10 ⁻⁴	Yes (EAS)	No	—
		PT	chr16: g.89809334_89809335del	c.3638_3639del	Pro1213Argfs*64	1	1.4 × 10 ⁻⁴	No	Yes (P)	Yes
		PT	chr16: g.89806418dup	c.3918dup	Gln1307Serfs*6	1	1.4 × 10 ⁻⁴	Yes (EAS)	Yes (P)	Yes
		Splice	chr16: g.89845259C>G	c.1777-1G>C	n.a.	5	7.1 × 10 ⁻⁴	Yes (EAS)	Yes (LP)	—
		Splice	chr16: g.89838223C>G	c.2015-1G>C	n.a.	1	1.4 × 10 ⁻⁴	No	No	—
		Splice	chr16: g.89838084A>G	c.2151 + 2T>C	n.a.	1	1.4 × 10 ⁻⁴	No	Yes (LP)	—
		Splice	chr16: g.89815176C>T	c.3240-1G>A	n.a.	1	1.4 × 10 ⁻⁴	No	Yes (C)	—
FANCC	NM_000136	Missense	chr16: g.89837005A>G	c.2189T>C	Leu730Pro	1	1.4 × 10 ⁻⁴	No	Yes (P)	Yes
		Missense	chr16: g.89831437C>T	c.2639G>A	Arg880Gln	1	1.4 × 10 ⁻⁴	Yes (SAS, NFE, AMR)	Yes (C)	Yes
		Missense	chr16: g.89828413C>G	c.2796G>C	Trp932Cys	1	1.4 × 10 ⁻⁴	No	No	—
		Missense	chr16: g.89816309T>C	c.3068A>G	Glu1023Gly	12	1.7 × 10 ⁻³	Yes (EAS)	No	—
		Splice	chr9: g.97887469T>G	c.897-2A>C	n.a.	1	1.4 × 10 ⁻⁴	No	No	—
		Splice	chr9: g.97876993C>G	c.1073-1G>C	n.a.	1	1.4 × 10 ⁻⁴	No	Yes (LP)	—
		PT	chr3: g.10084725_10084737del	c.889-9_892del	n.a.	1	1.4 × 10 ⁻⁴	No	No	—
		PT	chr3: g.10138111T>A	c.4140T>A	Cys1380*	1	1.4 × 10 ⁻⁴	No	No	—
		Splice	chr3: g.10107179G>T	c.2269 + 1G>T	n.a.	1	1.4 × 10 ⁻⁴	No	No	—
		Missense	chr3: g.10107654G>C	c.2376G>C	Trp792Cys	2	2.8 × 10 ⁻⁴	Yes (EAS)	No	—
FANCG	NM_004629	PT	chr9: g.35075661del	c.1222_1234del	Gly408Metfs*3	1	1.4 × 10 ⁻⁴	No	No	—
		PT	chr9: g.35074472dup	c.1656dup	His553Serfs*14	2	2.8 × 10 ⁻⁴	No	No	—
		Splice	chr9: g.35077397C>G	c.511-1G>C	n.a.	1	1.4 × 10 ⁻⁴	Yes (NFE)	No	—

(continued)

Table 1. (continued)

Gene	Reference transcript	Variant type	Genomic change (hg19)	Coding DNA change	Protein change	Allele count	Allele frequency ^a	Observed in other ancestry groups?	Observed in ClinVar/FADDB? (classification)	Occurred in FA patient?
FANCI	NM_001113378	Splice	chr9: g.35075274A>C	c.1480 + 2T>G	n.a.	1	1.4×10^{-4}	No	No	—
		PT	chr15: g.89850869dup	c.3616dup	Thr1206Asnfs*14	1	1.4×10^{-4}	No	No	—
FANCL	NM_018062	Missense	chr15: g.89821942A>T	c.1318A>T	Ile440Phe	1	1.4×10^{-4}	No	No	—
		Splice	chr2: g.58393011T>G	c.541-2A>C	n.a.	1	1.4×10^{-4}	No	No	—
FANCM	NM_020937	Splice	chr2: g.58388655A>C	c.1020 + 2T>G	n.a.	1	1.4×10^{-4}	No	No	—
		PT	chr14: g.45620696del	c.1015del	Asp339Ilefs*18	2	2.8×10^{-4}	No	No	—
		PT	chr14: g.45642352del	c.2255del	Ser752*	1	1.4×10^{-4}	No	No	—
		PT	chr14: g.45645552del	c.3595del	Asn1199Ilefs*16	1	1.4×10^{-4}	No	No	—
		PT	chr14: g.45646168_45646169del	c.4211_4212del	His1404Leufs*3	1	1.4×10^{-4}	No	No	—
SLX4	NM_032444	Splice	chr14: g.45650839G>A	c.4318-1G>A	n.a.	1	1.4×10^{-4}	Yes (SAS)	No	—
		PT	chr14: g.45653106G>C	c.4515 + 1G>C	n.a.	1	1.4×10^{-4}	No	No	—
		PT	chr16: g.3640867del	c.2772del	Gln925Serfs*76	1	1.4×10^{-4}	No	No	—
ERCC4	NM_005236	PT	chr16: g.3634816dup	c.4693dup	Gln1565Profs*16	2	2.8×10^{-4}	Yes (EAS)	No	—
		PT	chr16: g.14014035C>T	c.13C>T	Gln5*	1	1.4×10^{-4}	No	No	—
RAD51	NM_001164269	PT	chr16: 14041622C>A	c.2169C>A	Cys723*	1	1.4×10^{-4}	Yes (EAS)	Yes (VUS)	—
		PT	chr15: g.40994060C>A	c.282C>A	Cys94*	1	1.4×10^{-4}	No	No	—
UBE2T	NM_014176	PT	chr1: g.202302381dup	c.368dup	Leu124Alafs*4	2	2.8×10^{-4}	Yes (EAS)	No	—
		PT	chr1: g.202302185T>A	c.421A>T	Lys141*	1	1.4×10^{-4}	No	No	—
XRCC2	NM_005431	PT	chr7: g.152346290dup	c.280dup	Thr94Asnfs*4	1	1.4×10^{-4}	No	No	—
		PT	chr7: g.152346220dup	c.350dup	Leu117Phefs*6	1	1.4×10^{-4}	Yes (EAS, SAS, NFE, ASJ)	Yes (LP)	—
RFD3	NM_018124	Missense	chr7: g.152346299G>A	c.271C>T	Arg91Trp	1	1.4×10^{-4}	No	Yes (VUS)	—
		Missense	chr7: g.152345743C>G	c.827G>C	Gly276Ala	1	1.4×10^{-4}	No	No	—
		PT	chr16: g.74695115_74695118del	c.230_233del	Ser77Leufs*3	1	1.4×10^{-4}	No	No	—

^aAllele frequency reported is a fraction of 3523 Singaporeans (number of alleles = 7046). — = not found in FADDB or published reports of FA cases in the literature; AFR = African American; AMR = admixed-American or Latino; ASJ = Ashkenazi Jewish; C = conflicting interpretations of pathogenicity; EAS = East Asian; FADDB = Fanconi Anemia Mutation Database; FIN = Finnish European; LP = likely pathogenic; n.a. = not applicable; NFE = non-Finnish European; P = pathogenic; PT = protein truncating; SAS = South Asian; VUS = variant of uncertain significance.

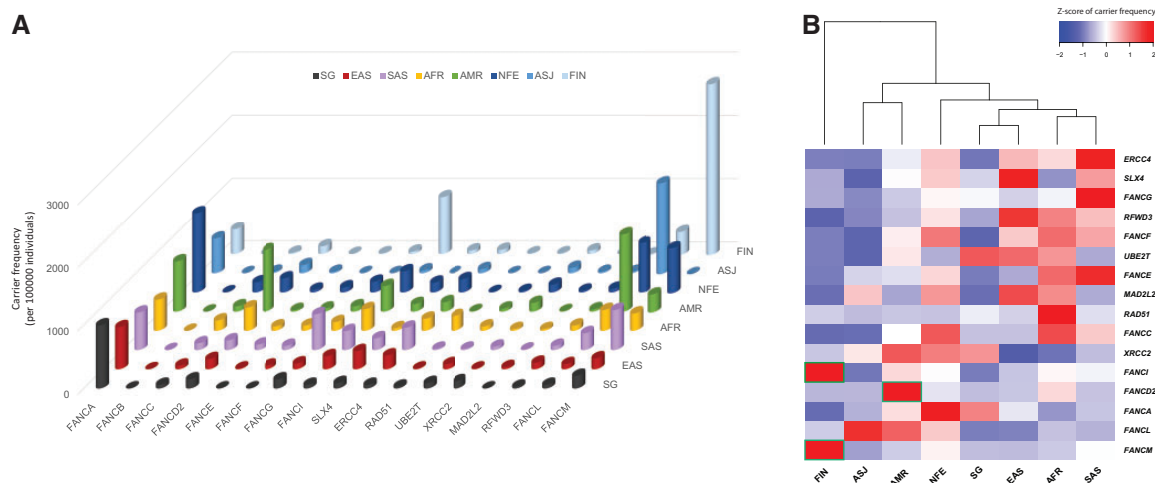


Figure 1. Frequency of potentially pathogenic germline variant carriers in 17 Fanconi anemia (FA) genes across ancestry groups. (A) Carrier frequency (per 100 000 individuals) aggregated for each of the 17 FA genes visualized for Singaporeans (SG) and 7 selected ancestry groups from gnomAD. (B) Heatmap of normalized carrier frequencies in the 8 populations for each FA gene. The dendrogram represents hierarchical clustering of the columns based on Euclidean distance. Pairwise differences in proportion were evaluated by Fisher's exact test, with Bonferroni correction for multiple testing (448 tests). Adjusted $P < .05$ was considered statistically significant, with **green boxes** indicating ancestry groups with carrier frequencies statistically significantly higher than all other compared groups. **Color scale** represents row-wise Z-scores of carrier frequencies. AFR = African American; AMR = admixed-American or Latino; ASJ = Ashkenazi Jewish; EAS = East Asian; FIN = Finnish European; NFE = non-Finnish European; SAS = South Asian.

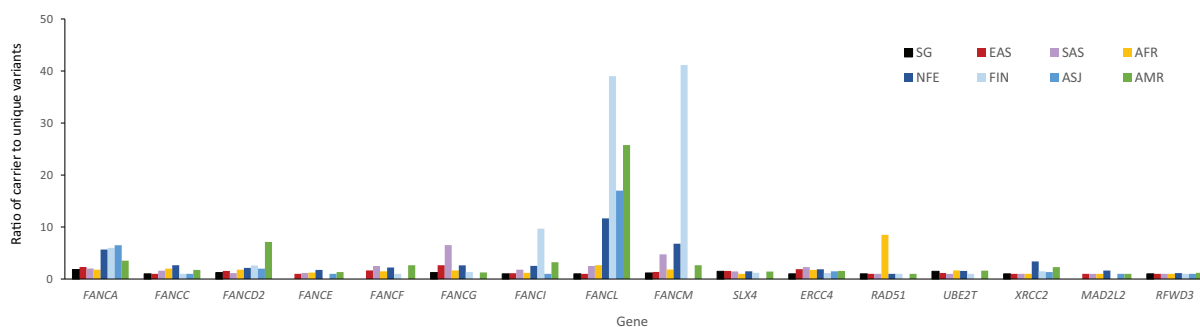


Figure 2. Recurrent germline variants in Fanconi anemia (FA) genes are less common among East Asian (EAS) and Singaporean (SG) populations. Enrichment of recurrent variants by gene is depicted by a relatively higher ratio of carriers to unique variants for each gene and is more commonly observed among groups of European descent compared with Asians. AFR = African American; AMR = admixed-American or Latino; ASJ = Ashkenazi Jewish; FIN = Finnish European; NFE = non-Finnish European; SAS = South Asian.

closer to the protein N-terminal (Figure 4) and coincide with domains of critical function, such as FANCA-FANCG interaction (32) and FANCM helicase activity (33). Separately, some of the recurrent variants identified within the gnomAD NFE, such as FANCL c.1096_1099dup (p.Thr367Asnfs*13) and FANCA c.1874G>C (p.Cys625Ser), were seen in almost all TCGA cohorts analyzed. However, we were unable to assess enrichment of the variants in respective cohorts because the number of observations were insufficiently powered for statistical significance.

Genotype and Phenotype Spectrum of Singaporean FA Patients

At our clinic, 4 individuals were referred for suspected FA, of whom 2 presented HNSCC, 1 with lung SCC, and 1 uterine cancer (Table 3). Multi-gene panel testing revealed germline biallelic FANCA variants in 1 HNSCC and lung SCC patients, and biallelic FANCI variants in the other HNSCC and uterine cancer patients. Notably, the genotype associated with cancer type

presented in all 4 FA patients was consistent with the associations observed within TCGA cohorts. Furthermore, of the 6 unique variants detected in our 4 FA patients, 3 were rare variants found only in Asians, 2 were novel, and 1 (FANCA exon21 deletion) was previously reported in FA patients (34,35). The Asian specificity of these identified variants is congruent to our findings, supporting a distinction in variant spectrum between Singaporeans and EAS with other ancestries.

Discussion

The extensive clinical and genotypic heterogeneity in FA has made genotype–phenotype correlation challenging. A few studies have suggested associations between certain clinical phenotypes with FA genes or variants (2,6,36), such as the occurrence of classical congenital abnormalities, VACTERL-H (vertebral abnormalities, anal atresia, cardiac defects, tracheoesophageal fistula, esophageal atresia, renal and radial abnormalities, limb abnormalities, with hydrocephalus; OMIM 276950), among

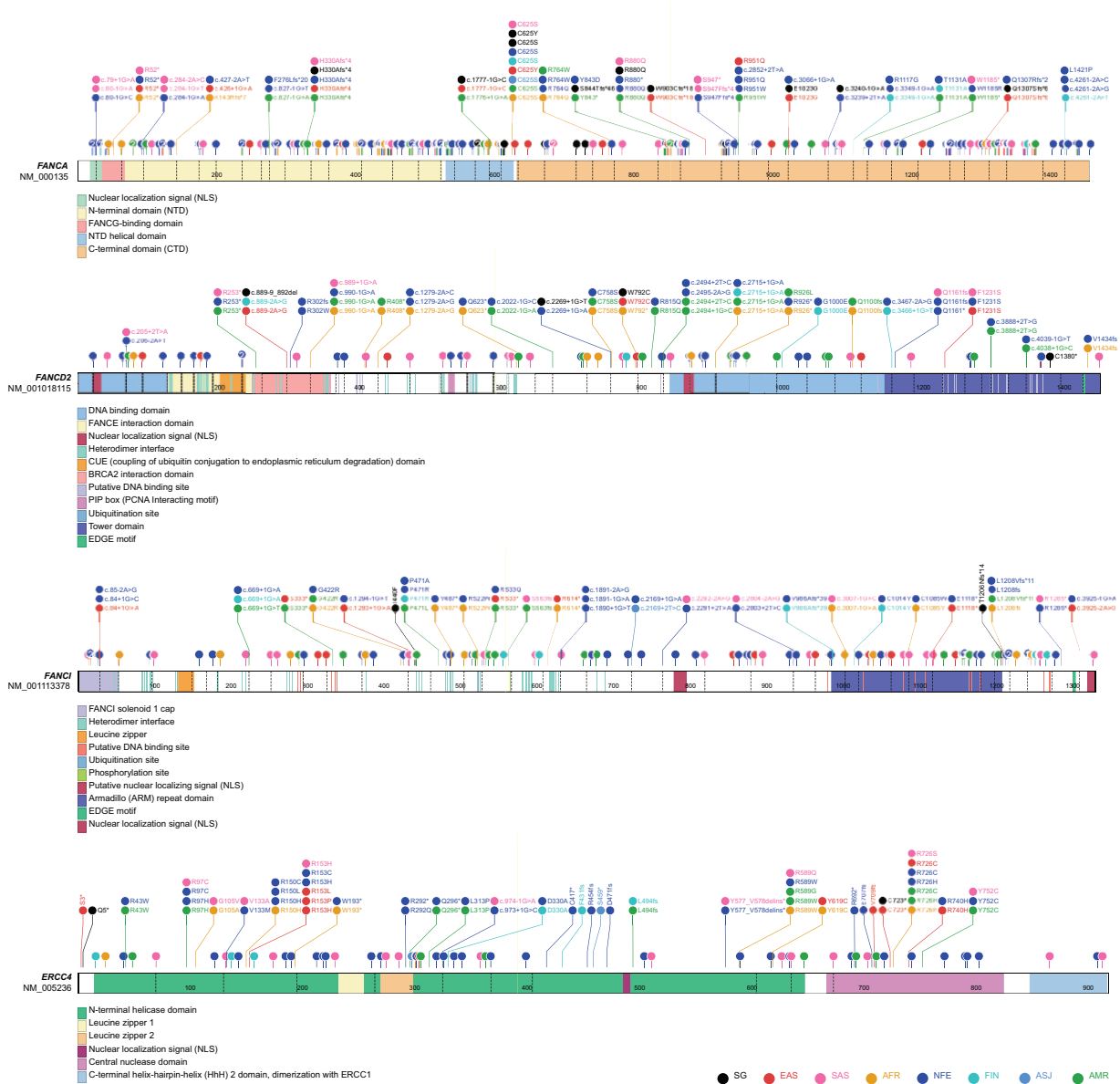


Figure 3. Spectrum of potentially pathogenic germline variants in *FANCA*, *FANCD2*, *FANCI*, and *ERCC4* across the 8 ancestry groups. Variant frequencies in each ancestry group are not represented in the schematic diagrams. AFR = African American; AMR = admixed-American or Latino; ASJ = Ashkenazi Jewish; EAS = East Asian; FIN = Finnish European; NFE = non-Finnish European; SAS = South Asian; SG = Singaporean.

carriers of null *FANCB* and *FANCI* variants (2,37). A plausible conjecture is that variability in carrier frequencies across FA genes in different populations may underlie the heterogeneity in clinical phenotype reported globally.

In this study, we explored the frequency and spectrum of potentially pathogenic germline variant carriers in 17 FA genes in Singapore and several selected populations of varying ancestry. Consistent with the higher prevalence of *FANCA* FA cases worldwide (4,38-41), germline *FANCA* variant carriers were more frequently identified in all populations, with aggregated carrier frequencies ranging from 0.4% to 1.3%. Importantly, we showed that there are differences in carrier frequencies and the variant spectrum across populations, particularly distinguishing populations of EAS ancestry from European-descent groups. This has implications for clinical practices aimed at detection and diagnosis of FA individuals. First, given the variable distribution of

germline pathogenic variant carriers across different ancestries, genetic testing for molecular diagnosis should not be restricted to *FANCA*, *FANCC*, and *FANCG*, which are reported more frequently in FA patients of European descent, but should include all known FA genes. Second, although existing databases of FA-related genetic variants are useful references, clinicians and genetic professionals should be cognizant of ancestry-specific alterations when interpreting genetic results of FA suspects. As demonstrated in our series of Singaporean FA cases, one-half of the identified variants were rare Asian variants, all of which were classified as variants of uncertain significance. This highlights the importance for interpreting genetic findings within the context of the patient's clinical phenotype and genetic ancestry. Furthermore, it reflects the limitations of benchmarking non-European variants to reference databases derived

Table 2. Spectrum of potentially pathogenic germline variants in the 17 FA genes identified among patients of White (non-Hispanic) racial subgroup in selected TCGA cohorts

Gene	TCGA cohort	Variant type	cDNA change	Protein change	Carriers, No.	Allele frequency within cohort	Allele frequency within gnomAD NFE	Mean allele fraction (range) ^a
FANCA	BRCA	Missense	c.1874G>C	Cys625Ser	8	1.2×10^{-2}	3.9×10^{-3}	0.50 (0.40-0.54)
		Missense	c.2852G>A	Arg951Gln	1	1.4×10^{-3}	1.0×10^{-4}	0.49
		Missense	c.3206T>G	Leu1069Arg	1	1.4×10^{-3}	8.8×10^{-6}	0.52
		Splice site	c.3349-1G>A	-	1	1.4×10^{-3}	4.7×10^{-5}	0.44
		Missense	c.3519G>C	Trp1173Cys	1	1.4×10^{-3}	9.5×10^{-6}	0.55
		PT	c.652del	Cys218Valfs*27	1	2.3×10^{-3}	8.9×10^{-6}	0.62
		PT	c.984del	Thr329Leufs*6	1	2.3×10^{-3}	8.8×10^{-6}	0.53
		PT	c.1115_1118del	Val372Alafs*42	1	2.3×10^{-3}	1.0×10^{-4}	0.43
		Missense	c.1874G>C	Cys625Ser	5	1.2×10^{-2}	3.9×10^{-3}	0.42 (0.29-0.53)
		Missense	c.1490C>G	Pro497Arg	1	2.6×10^{-3}	8.8×10^{-6}	0.50
LUAD	Missense	c.1874G>C	Cys625Ser	4	1.0×10^{-2}	3.9×10^{-3}	0.50	
		Missense	c.2345T>C	Leu782Pro	1	2.6×10^{-3}	8.9×10^{-6}	0.49
LUSC	Missense	c.1874G>C	Cys625Ser	3	8.7×10^{-3}	3.9×10^{-3}	0.42 (0.39-0.63)	
		Splice site	c.2222 + 1G>T	-	1	2.9×10^{-3}	0	0.4
FANCC	BRCA	PT	c.2638C>T	Arg880*	1	2.9×10^{-3}	8.8×10^{-6}	0.32
		PT	c.3295C>T	Gln1099*	1	2.9×10^{-3}	8.8×10^{-6}	0.41
		Splice site	c.190-1G>T	-	1	2.6×10^{-3}	8.8×10^{-6}	0.49
		Missense	c.1874G>C	Cys625Ser	2	5.2×10^{-3}	3.9×10^{-3}	0.48 (0.44-0.52)
		Splice site	c.3934 + 2T>C	-	1	2.6×10^{-3}	0	0.72
		Missense	c.1874G>C	Cys625Ser	2	5.5×10^{-3}	3.9×10^{-3}	0.42 (0.40-0.44)
		PT	c.553C>T	Arg185*	1	1.4×10^{-3}	1.0×10^{-4}	0.45
		PT	c.67del	Asp231lefs*23	1	2.3×10^{-3}	3.0×10^{-4}	0.49
		PT	c.1642C>T	Arg548*	1	2.3×10^{-3}	7.0×10^{-5}	0.49
		PT	c.553C>T	Arg185*	1	2.9×10^{-3}	1.0×10^{-4}	0.42
FANCD2	BRCA	Missense	c.2444G>A	Arg815Gln	1	1.4×10^{-3}	8.8×10^{-5}	0.48
		Splice site	c.2715 + 1G>A	-	2	2.9×10^{-3}	3.0×10^{-4}	0.49 (0.47-0.51)
FANCF	HNSC	Missense	c.546G>T	Trp182Cys	1	2.3×10^{-3}	1.8×10^{-5}	0.53
		Splice site	c.2715 + 1G>A	-	1	2.6×10^{-3}	3.0×10^{-4}	0.48
		Missense	c.1250T>C	Leu417Pro	1	2.9×10^{-3}	0	0.25
		Splice site	c.2715 + 1G>A	-	2	5.2×10^{-3}	3.0×10^{-4}	0.46 (0.44-0.49)
FANCG	HNSC	PT	c.484_485del	Leu162Aspfs*103	1	2.6×10^{-3}	8.8×10^{-5}	0.42
		PT	c.484_485del	Leu162Aspfs*103	1	2.8×10^{-3}	8.8×10^{-5}	0.73
FANCD1	LUAD	Splice site	c.1144-1G>T	-	1	2.3×10^{-3}	5.4×10^{-5}	0.44
		PT	c.1483del	Ala495Glnfs*23	1	2.3×10^{-3}	8.8×10^{-6}	0.33
		Splice site	c.1480 + 1G>T	-	1	2.6×10^{-3}	4.4×10^{-5}	0.52
		Missense	c.46T>C	Trp16Arg	1	2.9×10^{-3}	0	0.46
		PT	c.1795_1804del	Trp599Profs*49	1	2.9×10^{-3}	5.3×10^{-5}	0.44
		PT	c.1354dupG	Val452Glyfs*23	1	2.6×10^{-3}	0	0.32
UCSC	UCSC	PT	c.1182_1192delinsG	Glu395Trpfs*5	1	2.8×10^{-3}	0	0.32
		PT	c.313G>T	Glu105*	1	1.8×10^{-2}	2.7×10^{-5}	0.52 (continued)

Table 2. (continued)

Gene	TCGA cohort	Variant type	cDNA change	Protein change	Carriers, No.	Allele frequency within cohort	Allele frequency within gnomAD NFE	Mean allele fraction (range) ^a
FANCI	BRCA	Missense	c.1264G>A	Gly422Arg	2	2.9×10^{-3}	2.0×10^{-4}	0.49 (0.48-0.50)
	HNSC	PT	c.849T>A	Tyr283*	1	2.3×10^{-3}	8.8×10^{-6}	0.48
	LUAD	Splice site	c.3538-2A>G		1	2.3×10^{-3}	0	0.68
	UCEC	Missense	c.1264G>A	Gly422Arg	1	2.6×10^{-3}	2.0×10^{-4}	0.41
	BRCA	PT	c.1461T>A	Tyr487*	1	2.6×10^{-3}	5.3×10^{-5}	0.53
	HNSC	PT	c.2345_2346del	Ser782*	1	2.8×10^{-3}	1.8×10^{-5}	0.56
	LUAD	PT	c.2509G>T	Glu837*	1	2.8×10^{-3}	1.6×10^{-5}	0.48
	UCEC	PT	c.1096_1099dup	Thr367Asnfs*13	5	7.2×10^{-3}	3.5×10^{-3}	0.39 (0.34-0.44)
	HNSC	PT	c.1096_1099dup	Thr367Asnfs*13	2	4.7×10^{-3}	3.5×10^{-3}	0.30 (0.29-0.31)
	LUAD	PT	c.1096_1099dup	Thr367Asnfs*13	4	1.0×10^{-2}	3.5×10^{-3}	0.38 (0.32-0.50)
	LUAD	PT	c.1096_1099dup	Thr367Asnfs*13	2	5.8×10^{-3}	3.5×10^{-3}	0.39
	OV	PT	c.1096_1099dup	Thr367Asnfs*13	6	1.6×10^{-2}	3.5×10^{-3}	0.38 (0.34-0.44)
	UCEC	PT	c.1096_1099dup	Thr367Asnfs*13	2	5.5×10^{-3}	3.5×10^{-3}	0.40 (0.33-0.41)
	BRCA	PT	c.333_334insTC TG	Pro112Serfs*74	1	1.4×10^{-3}	8.8×10^{-6}	0.57 (0.35-0.44)
	LUAD	Splice site	c.759 + 1G>A	-	1	1.4×10^{-3}	1.8×10^{-5}	0.42
	UCEC	PT	c.1491dup	Gln498Thrfs*7	1	1.4×10^{-3}	5.3×10^{-5}	0.46
	BRCA	Missense	c.1597C>T	Arg533Cys	1	1.4×10^{-3}	6.0×10^{-4}	0.56
	LUAD	PT	c.1972C>T	Arg58*	2	2.9×10^{-3}	1.0×10^{-4}	0.53 (0.51-0.55)
	LUAD	PT	c.4537del	Asp1513Metfs*34	1	1.4×10^{-3}	8.9×10^{-6}	0.38
	LUAD	PT	c.5101C>T	Gln1701*	1	1.4×10^{-3}	1.0×10^{-3}	0.44
	LUAD	PT	c.5791C>T	Arg1931*	1	1.4×10^{-3}	1.1×10^{-3}	0.33
	LUAD	PT	c.1972C>T	Arg658*	1	2.3×10^{-3}	1.0×10^{-4}	0.51
	LUAD	PT	c.1985_1986insAT	Phe663Serfs*8	1	2.3×10^{-3}	8.8×10^{-6}	0.47
	LUAD	PT	c.5101C>T	Gln1701*	1	2.3×10^{-3}	1.0×10^{-3}	0.41
	LUAD	PT	c.5791C>T	Arg1931*	2	5.2×10^{-3}	1.1×10^{-3}	0.41 (0.34-0.47)
	LUAD	PT	c.5791C>T	Arg1931*	1	2.9×10^{-3}	1.1×10^{-3}	0.43
	LUAD	PT	c.5791C>T	Arg1931*	1	2.8×10^{-3}	1.1×10^{-3}	0.48
	LUAD	Missense	c.4523C>A	Ser1508*	1	1.4×10^{-3}	1.8×10^{-5}	0.43
	LUAD	PT	c.4668del	Val1558Cysfs*3	1	2.8×10^{-3}	8.8×10^{-6}	0.42
	LUAD	Splice site	c.4739 + 2T>C	-	1	2.8×10^{-3}	8.8×10^{-6}	0.43
	LUAD	Missense	c.2233T>C	Cys745Arg	1	2.3×10^{-3}	8.8×10^{-6}	0.49
	LUAD	PT	c.175G>T	Glu59*	1	2.6×10^{-3}	8.9×10^{-6}	0.33
	LUAD	PT	c.96del	Phe32Leufs*30	4	5.8×10^{-3}	1.0×10^{-4}	0.48 (0.42-0.59)
	LUAD	PT	c.829del	Val277Leufs*20	1	2.6×10^{-3}	9.1×10^{-6}	0.45
	LUAD	PT	c.1700_1701del	Tyr567*	1	2.9×10^{-3}	8.8×10^{-6}	0.49

^a Allele fraction: ratio of alternative allele counts to the sum of reference and alternative allele counts for the variant. For variants with more than 1 carrier, the allele fraction is indicated by the mean allele fraction and range of allele fraction values. BRCA = breast invasive carcinoma; HNSC = head and neck squamous cell carcinoma; LUAD = lung adenocarcinoma; LUSC = lung squamous cell carcinoma; NFE = non-Finnish European; OV = ovarian serous cystadenocarcinoma; PT = protein truncating; TCGA = The Cancer Genome Atlas; UCEC = uterine corpus endometrial carcinoma; UCS = uterine carcinosarcoma.

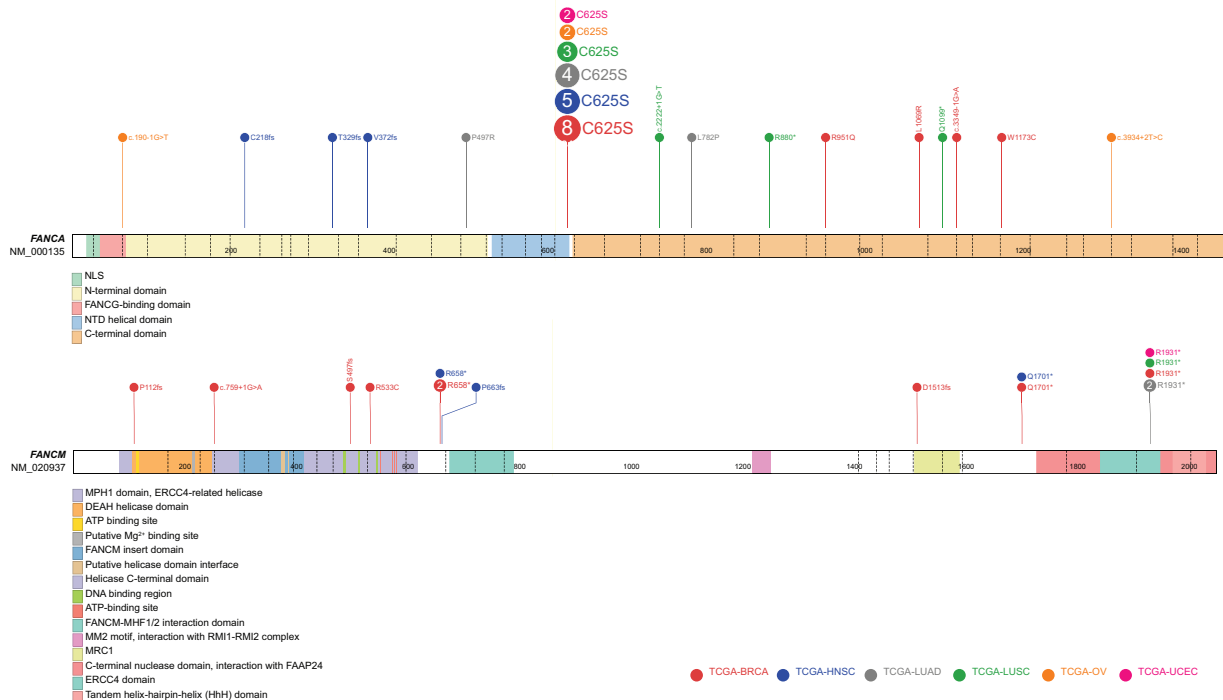


Figure 4. Spectrum of FANCA and FANCM germline variants identified in White (non-Hispanic) racial subgroup of patients in selected The Cancer Genome Atlas (TCGA) cohorts. TCGA-BRCA = breast invasive carcinoma; TCGA-HNSC = head and neck squamous cell carcinoma; TCGA-LUAD = lung adenocarcinoma; TCGA-LUSC = lung squamous cell carcinoma; TCGA-OV = ovarian serous cystadenocarcinoma; TCGA-UCEC = uterine corpus endometrial carcinoma.

predominantly from Europeans and underscores the value of ancestry-specific references toward more optimal patient care.

The absence of physical anomalies occurs in 20% of FA patients (1,2), and many remain undetected until triggered by cancer diagnosis or treatment-related toxicity at adulthood. Timely identification and accurate diagnosis of these individuals are critical because it immediately informs the appropriate treatment modalities important for mitigating potentially fatal therapy-related toxicities. FA patients have a higher-than-population risk of HNSCC, leukemia, and esophageal and vulvar cancers (42), but correlations with the FA genotype have been elusive (6). Our exploration using the TCGA dataset suggests a higher likelihood of FANCA and FANCC PTVs in HNSCC and lung SCC as well as FANCI PTVs in HNSCC and uterine cancer. This finding was interestingly consistent with the cancer types presented by our Singaporean adult FA patients; however, we caution that the number of observations was small and not statistically significant. Also noteworthy is that a different variant effect for the same gene may account for phenotypic heterogeneity among FA patients (2,6). Contrasting the adult-onset cancers of our FANCI FA patients are 2 Japanese FANCI FA cases reported with childhood-onset malignancies, 1 of whom with severe features of VACTERL-H (38). Indeed, the VACTERL-H Japanese patient harbored biallelic null variants resulting in early truncation of FANCI (c.158-2A>G [p.Ser54Phefs*5], c.288G>A [p.Cys56Phefs*8]) compared with variants in the remaining Japanese and Singaporean patients with a predicted protein truncating effect closer to the protein C-terminal (38), indicating that phenotypic heterogeneity is potentially driven by variant-level specificity. Furthermore, the Japanese study also showed that the ancestry-specific polymorphism *ALDH2**504Lys may modify clinical severity in patients with

identical FA variants (38), accentuating the importance of interpreting molecular diagnosis in an ancestry-dependent context.

Because certain sample sizes were small and some observations rare, our study was limited in power to detect statistically significant associations and may be subjected to ascertainment bias. Comparisons of datasets derived from varying sources are limited by technical differences such as sequencing methods and data quality, which may confound our variant analysis and carrier estimates. Additionally, predictions for variant pathogenicity are constrained by our filtering criteria and assumption of a loss-of-function mechanism for all FA genes, which could under- or overestimate the true functional impact of variants limited to small alterations, such as single nucleotide changes and small insertions or deletions. However, continuous reporting of FA cases from various ancestry groups (38,41) and the progressive evolution of prediction tools and variant characterization studies will refine our genotype-phenotype profiling for FA in time.

In summary, we have highlighted the diverse mutational profile in FA-associated genes across ancestries, emphasizing that molecular diagnosis of FA should include interrogation of all known FA genes. Furthermore, clinicians suspecting FA should be cognizant of the interplay between ancestry-specific genetic variation and phenotypic variability to avoid missed diagnoses. This is especially critical in FA individuals first presenting with adult-onset malignancies, because there is an opportunity to reduce incidence of life-threatening treatment-related toxicities.

Funding

This study was supported by the National Medical Research Council Singapore Clinician-Scientist Award (NMRC/GSA-

Table 3. Characteristics of Singaporean FA patients^a

Patient ID	Race	Physical/clinical features	Malignancy (age at diagnosis, y)	Chromosomal breakage test	Gene	DNA change	Protein change	Zygosity [inheritance]	Variant classification	MAF in Singaporeans	MAF in gnomAD
CGS0026-1	CH	Short stature Microphthalmia Elfn facies Subfertility Hypothyroidism	Uterine serous papillary carcinoma (32) Chemotherapy-induced pancytopenia and BMF (32)	Positive (>70%)	FANCI FANCI	c.445G>A c.3924+1G>A	Gly149Ser -	Het [Pat] Het [Mat]	VUS LP	2.8 × 10 ⁻⁴ 0	(EAS) 1.6 × 10 ⁻⁴ nil
CGS0026-2	CH	Short stature Microphthalmia Duplicated left thumb Café au lait spots Low platelet level	Well-differentiated SCC of tongue (32)	Positive (>70%)	FANCI FANCI	c.445G>A c.3924+1G>A	Gly149Ser -	Het [Pat] Het [Mat]	VUS LP	2.8 × 10 ⁻⁴ 0	(EAS) 1.6 × 10 ⁻⁴ nil
CGS1249-1	MY	Small head circumference Epicanthal folds Café au lait spots	MDS refractory cytopenia with tri-lineage dysplasia without excess blast (32)	Positive (>70%)	FANCA FANCA	c.3085G>T Deletion exon 21	Glu1029* -	Het [Unk] Het [Unk]	P P	0 0	nil nil
CGS1253-1	CH	Short stature Small head circumference Duplicated left thumb	SCC of lung (32) SCC of hypopharynx (39)	Positive (>70%)	FANCA FANCA	c.2959G>C c.2959G>T	Ala987Pro Ala987Ser	Het [Unk] Het [Unk]	VUS VUS	0 0	(EAS) 5.4 × 10 ⁻⁵ (SAS) 3.3 × 10 ⁻⁵

^aBMF = bone marrow failure; CH = Chinese; EAS = East Asian; F = female; Het = heterozygous; LP = likely pathogenic; M = male; MAF = minor allele frequency; Mat = maternal; MDS = myelodysplastic syndrome; MY = Malay; nil = variant is not observed in gnomAD; P = pathogenic; Pat = paternal; SAS = South Asian; SCC = squamous cell carcinoma; Unk = unknown; VUS = variant of uncertain significance.

INV/0017/2017) and National Cancer Centre Research Fund Terry Fox Grant (NCCRF-YR2018-NOV-1) to JN.

Notes

Role of the funder: The funding source had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, and approval of the manuscript; and the decision to submit the manuscript for publication.

Disclosures: JN receives funding from AstraZeneca for ovarian cancer research. All other authors have no disclosures.

Author contributions: Conceptualization and Methodology, SHC and JN; Formal analysis, SHC and WKL; Investigation, SHC, JXT, STL, NDBI and YN; Resources, WKL and JN; Data curation: SHC, JXT and WKL; Writing—Original draft, SHC; Writing—Review and Editing, SHC, STL, YN, WKL and JN; Funding acquisition and Supervision, JN.

Acknowledgments: We would like to thank all our FA patients and their families for their participation and support to our study. We also acknowledge contributors to the SEC for the Singaporean dataset.

Data Availability

The data that support findings of this study are available from the corresponding author upon reasonable request.

References

- Frohnmayr D, Frohnmayr L, Guinan E, Kennedy T, Larsen K, eds. *Fanconi Anemia: Guidelines for Diagnosis and Management*. 4th ed. Eugene, OR: Fanconi Anemia Research Fund, Inc; 2014.
- Fiesco-Roa MO, Giri N, McReynolds LJ, Best AF, Alter BP. Genotype-phenotype associations in Fanconi anemia: a literature review. *Blood Rev*. 2019;37:100589.
- Nalepa G, Clapp DW. Fanconi anaemia and cancer: an intricate relationship. *Nat Rev Cancer*. 2018;18(3):168–185.
- Auerbach AD, Smogorzewska A. International Fanconi Anemia Registry (IFAR). Fanconi Anemia Mutation Database. www2.rockefeller.edu/fanconi/. Accessed September 30, 2019.
- Niraj J, Färkkilä A, D'Andrea AD. The Fanconi anemia pathway in cancer. *Annu Rev Cancer Biol*. 2019;3(1):457–478.
- Neveling K, Endt D, Hoehn H, Schindler D. Genotype-phenotype correlations in Fanconi anemia. *Mutat Res*. 2009;668(1-2):73–91.
- 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.
- Callén E, Casado JA, Tischkowitz MD, et al. 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.
- Kutler DI, Auerbach AD. Fanconi anemia in Ashkenazi Jews. *Fam Cancer*. 2004;3(3-4):241–248.
- de Vries Y, Lwiwski N, Levitus M, et al. A Dutch Fanconi anemia FANCC founder mutation in Canadian Manitoba Mennonites. *Anemia*. 2012;2012:1–6.
- García-de Teresa B, Frias S, Molina B, et al. FANCC Dutch founder mutation in a Mennonite family from Tamaulipas, México. *Mol Genet Genomic Med*. 2019;7(6):e710.
- Donovan FX, Solanki A, Mori M, et al. A founder variant in the South Asian population leads to a high prevalence of FANCL Fanconi anemia cases in India. *Hum Mutat*. 2020;41(1):122–128.
- Park J, Kim M, Jang W, et al. Founder haplotype analysis of Fanconi anemia in the Korean population finds common ancestral haplotypes for a FANCG variant. *Ann Hum Genet*. 2015;79(3):153–161.
- 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.
- Sirugo G, Williams SM, Tishkoff SA. The missing diversity in human genetic studies. *Cell*. 2019;177(1):26–31.
- Wu D, Dou J, Chai X, et al. Large-scale whole-genome sequencing of three diverse Asian populations in Singapore. *Cell*. 2019;179(3):736–749.e15.
- Karczewski KJ, Francioli LC, Tiao G, et al.; Genome Aggregation Database Consortium. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434–443.
- Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
- Jian X, Boerwinkle E, Liu X. In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Res*. 2014;42(22):13534–13544.
- Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*. 2019;47(D1):D886–D894.
- Ioannidis NM, Rothstein JH, Pejaver V, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am J Hum Genet*. 2016;99(4):877–885.
- Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46(3):310–315.
- Oza AM, DiStefano MT, Hemphill SE, et al.; ClinGen Hearing Loss Clinical Domain Working Group. Expert specification of the ACMG/AMP variant interpretation guidelines for genetic hearing loss. *Hum Mutat*. 2018;39(11):1593–1613.
- Morales A, Kinnamon DD, Jordan E, et al. Variant Interpretation for Dilated Cardiomyopathy: refinement of the American College of Medical Genetics and Genomics/ClinGen Guidelines for the DCM Precision Medicine Study. *Circ Genom Precis Med*. 2020;13(2):e002480.
- Chandrasekharappa SC, Chinn SB, Donovan FX, et al. Assessing the spectrum of germline variation in Fanconi anemia genes among patients with head and neck carcinoma before age 50. *Cancer*. 2017;123(20):3943–3954.
- Kiiski JI, Peltari LM, Khan S, et al. Exome sequencing identifies FANCM as a susceptibility gene for triple-negative breast cancer. *Proc Natl Acad Sci USA*. 2014;111(42):15172–15177.
- Nguyen-Dumont T, Myszkowski A, Karpinski P, et al. FANCM and RECQL genetic variants and breast cancer susceptibility: relevance to South Poland and West Ukraine. *BMC Med Genet*. 2018;19(1):12.
- Park DJ, Lesueur F, Nguyen-Dumont T, et al. Rare mutations in XRCC2 increase the risk of breast cancer. *Am J Hum Genet*. 2012;90(4):734–739.
- Kiiski JI, Tervasmäki A, Peltari LM, et al. FANCM mutation c.5791C>T is a risk factor for triple-negative breast cancer in the Finnish population. *Breast Cancer Res Treat*. 2017;166(1):217–226.
- *30. Figlioli G, Bogliolo M, Catucci I, et al.; ABCTB Investigators. The FANCM:p.Arg658 truncating variant is associated with risk of triple-negative breast cancer. *NPJ Breast Cancer*. 2019;5(1):38.
- Kluźniak W, Wokolorczyk D, Rusak B, et al.; the Polish Hereditary Breast Cancer Consortium. Inherited variants in XRCC2 and the risk of breast cancer. *Breast Cancer Res Treat*. 2019;178(3):657–663.
- Jeong E, Lee S-G, Kim H-S, et al. Structural basis of the Fanconi anemia-associated mutations within the FANCA and FANCG complex. *Nucleic Acids Res*. 2020;48(6):3328–3342.
- Meetei AR, Medhurst AL, Ling C, et al. A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. *Nat Genet*. 2005;37(9):958–963.
- Gille JJP, Floor K, Kerkhoven L, Ameziene N, Joenje H, de Winter JP. Diagnosis of Fanconi anemia: mutation analysis by multiplex ligation-dependent probe amplification and PCR-based Sanger sequencing. *Anemia*. 2012;2012:1–13.
- Tsangaris E, Klaassen R, Fernandez CV, et al. Genetic analysis of inherited bone marrow failure syndromes from one prospective, comprehensive and population-based cohort and identification of novel mutations. *J Med Genet*. 2011;48(9):618–628.
- Faiyre L, Guardiola P, Lewis C, et al. Association of complementation group and mutation type with clinical outcome in Fanconi anemia. European Fanconi Anemia Research Group. *Blood*. 2000;96(13):4064–4070.
- Alter BP, Rosenberg PS. VACTERL-H association and Fanconi anemia. *Mol Syndromol*. 2012;4(1-2):87–93.
- Mori M, Hira A, Yoshida K, et al. Pathogenic mutations identified by a multi-modality approach in 117 Japanese Fanconi anemia patients. *Haematologica*. 2019;104(10):1962–1973.
- Li N, Ding L, Li B, Wang J, D'Andrea AD, Chen J. Functional analysis of Fanconi anemia mutations in China. *Exp Hematol*. 2018;66:32–41.e8.
- Kutler DI, Singh B, Satagopan J, et al. A 20-year perspective on the International Fanconi Anemia Registry (IFAR). *Blood*. 2003;101(4):1249–1256.
- Steinberg-Shemer O, Goldberg TA, Yacobovich J, et al. Characterization and genotype-phenotype correlation of patients with Fanconi anemia in a multi-ethnic population. *Haematologica*. 2020;105(7):1825–1834.
- Alter BP, Giri N, Savage SA, Rosenberg PS. Cancer in the National Cancer Institute inherited bone marrow failure syndrome cohort after fifteen years of follow-up. *Haematologica*. 2018;103(1):30–39.