Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

SNV/indel hypermutator phenotype in biallelic RAD51C variant - Fanconi anemia

Roni Zemet
Baylor College of Medicine
Haowei Du
Baylor College of Medicine
Tomasz Gambin
Warsaw University of Technology: Politechnika Warszawska
Pengfei Liu
Baylor College of Medicine
James R Lupski
Baylor College of Medicine
Paweł Stankiewicz (pawels@bcm.edu)
Baylor College of Medicine https://orcid.org/0000-0002-6456-7490

Research Article

Keywords: de novo variants, genome sequencing, hypermutation, multiple congenital anomalies, prenatal exome sequencing, perizygotic somatic mosaicism

Posted Date: March 2nd, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2628288/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

SNV/indel hypermutator phenotype in biallelic *RAD51C* variant - Fanconi anemia

Roni Zemet[1,*], Haowei Du[1,*], Tomasz Gambin[2,*],
James R Lupski[1], Pengfei Liu[1,3], Paweł Stankiewicz[1]

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX
2) Institute of Computer Science, Warsaw University of Technology, Warsaw, Poland
3) Baylor Genetics, Houston, TX

*Roni Zemet, Haowei Du, and Tomasz Gambin contributed equally to this work.

Correspondence to:

Correspondence: Paweł Stankiewicz; pawels@bcm.edu

Funding

The somatic mosaicism research is funded by Baylor College of Medicine, by grant R01HD087292 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) to PS. RZ is supported by T32 GM07526-44 from the National Institute of General Medical Science (NIH/NIGMS). JRL is supported by the U.S. National Institute of Neurological Disorders and Stroke (NINDS) (R35NS105078). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Acknowledgments

We are thankful to our colleagues who provided their expertise that greatly assisted this research work.

Abstract

We previously reported a fetus with Fanconi anemia (FA), complementation group O due to compound heterozygous variants involving *RAD51C*. Interestingly, the trio exome sequencing analysis also detected eight apparent *de novo* mosaic variants with variant allele fraction (VAF) ranging between 11.5%-37%. Here, using whole genome sequencing and a 'home-brew' variant filtering pipeline and DeepMosaic module, we investigated the number and signature of *de novo* heterozygous and mosaic variants and the rare phenomenon of hypermutation. Eight-hundred-thirty apparent *dn*SNVs and 21 *de novo* indels had VAFs below 37.41% and were considered postzygotic somatic mosaic variants. The VAFs showed a bimodal distribution, with one component with an average VAF of 25% (range: 18.7-37.41%) (n=446), representing potential postzygotic first mitotic events, and the other component with an average VAF of 12.5% (range: 9.55-18.69%) (n=384), describing potential second mitotic events. No increased rate of CNV formation was observed. The mutational pattern analysis for somatic single base substitution showed SBS40, SBS5, and SBS3 as the top recognized signatures. SBS3 is a known signature associated with homologous recombination-based DNA damage repair error. Our data demonstrate that biallelic *RAD51C* variants show evidence for defective genomic DNA damage repair and thereby result in a hypermutator phenotype with the accumulation of postzygotic *de novo* mutations, at least in the prenatal period. This 'genome hypermutator phenomenon' might contribute to the observed hematological manifestations and the predisposition to tumors in patients with FA, and pregnancy loss in general. We propose that other FA groups should be investigated for genome-wide *de novo* variants.

Keywords:

de novo variants; genome sequencing; hypermutation; multiple congenital anomalies; prenatal exome sequencing; perizygotic somatic mosaicism.

Introduction

Fanconi anemia (FA) is a rare chromosomal instability syndrome that affects one in every 136,000 births (Mamrak et al. 2017). FA is a genetically and phenotypically heterogeneous disorder, resulting from perturbations in genes involved in DNA repair and cell cycle regulation. When the cell cannot sufficiently repair its genetic information, and genomic integrity is compromised, various clinical manifestations, such as congenital malformations, early progressive bone marrow failure, and predisposition to hematologic malignancies and solid tumors occur (Auerbach 2009; Kolinjivadi et al. 2020; Mehta and Ebens 2021; Moldovan and D'Andrea 2009; Rageul and Kim 2020). Physical abnormalities, present in approximately 75% of affected individuals with FA, include one or more of the following: endocrine dysfunctions leading to short stature, abnormal skin pigmentation, skeletal deformities of the thumb and forearm, microcephaly, and ophthalmic and genitourinary tract anomalies (Auerbach 2009; Ceccaldi et al. 2016; Kee and D'Andrea 2012; Mehta and Ebens 2021). In addition, FA patients are predisposed to various cancers, including acute myeloid leukemia (AML) and squamous cell carcinoma of the head and neck (Kutler et al. 2016; Kutler et al. 2003).

To date, variants in more than 20 genes (*FANCA* to *FANCW*) have been found to cause FA as an autosomal recessive (AR) trait, in *FANCB* exhibits X-linked (XL) inheritance, and in FANCR/RAD51, which have been associated with an autosomal dominant (AD) rare disease trait (Ameziane et al. 2015; Badra Fajardo et al. 2022; Moreno et al. 2021; Wang et al. 2015). All proteins form an integrated protein network, known as the FA/BRCA repair pathway (Ceccaldi et al. 2016). The network is active during DNA replication when converging replisomes encounter damage that covalently binds the two

DNA strands (interstrand crosslinking; ICL). Sequential recruitment of FA proteins and associated partners onto chromatin unhooks the ICL and coordinates the repair through homologous monoubiquitination and recombination (Badra Fajardo et al. 2022; Moreno et al. 2021; Nalepa and Clapp 2018). Emerging evidence suggests that independent of ICL and homologous recombination (HR) repair, FA proteins also regulate cell-cycle checkpoints and/or promote replication fork remodeling in response to replication stress, redefining the FA pathway as a cardinal mechanism to preserve genome integrity throughout the entire replication process (Badra Fajardo et al. 2022).

Alterations in FA genes have been found to incite genomic instability and contribute to tumorigenesis (Badra Fajardo et al. 2022). Accordingly, cells from FA patients are hypersensitive to ICL-inducing agents such as diepoxybutane (DEB) and mitomycin C (MMC), which cause high levels of chromosomal aberrations, including chromosomal breaks and quadriradial formation. In addition to identifying pathogenic variants in the FA genes, this unique and characteristic cellular phenotype is still employed for the objective clinical laboratory diagnostics of FA patients using a DEB and MMC-induced chromosome breakage test of lymphocytes (Rageul and Kim 2020).

FA proteins are classified into three general groups based on the biochemical activities and functions in ICL repair: FA core complex, which recognizes the damage and ubiquitylates the heterodimer (FANCI-FANCD2), ID2 heterodimeric complex formation, which recruits repair factors, such as nucleases (FANCP and FANCQ), trans-lesion synthesis polymerases (FANCV and polymerase ζ), and HR factors (FANCD1, FANCJ, FANCN, FANCO, FANCR, FANCS, and FANCU) (Duxin and Walter 2015; Kolinjivadi et al. 2020; Milletti et al. 2020). Several FA proteins from the third group participate in

HR, a double-strand break (DSB) repair pathway, which is active during the S and G2 cell-cycle phases; HR utilizes extensive sequence homology from a donor template for error-free repair of broken DNA ends. Furthermore, monoallelic mutations in the FA genes within the HR pathway, including *RAD51C* (FANCO), *BRCA2* (FANCD1), *BRCA1* (FANCS), *BRIP1* (FANCJ), or *PALB2* (FANCN), have been associated with familial predisposition to breast, ovarian, and pancreatic cancers (Ceccaldi et al. 2016; D'Andrea 2010; Kottemann and Smogorzewska 2013; Rageul and Kim 2020). RAD51C has also been shown to be involved in intra-S-phase checkpoint regulation through CHK2 activation in response to DNA damage (Somyajit et al. 2012; Vaz et al. 2010). *Rad51c* deficient mice show early embryonic lethality (Kuznetsov et al. 2009), whereas mice carrying a null and a hypomorphic allele show infertility with partial penetrance (Kuznetsov et al. 2007).

Previous whole genome sequencing (WGS) and exome sequencing (ES) studies have provided insights into the scale of *de novo* variants in the normotypical population and as a cause of genetic diseases (Acuna-Hidalgo et al. 2016; Veltman and Brunner 2012). The mutation rate of single nucleotide variants (SNVs) has been estimated at $1.0\text{-}1.8\text{x}10^{-8}$ variants per base per generation, giving rise to 60-70 *de novo* variants per genome, with one to two affecting the coding sequence (Campbell and Eichler 2013; Goldmann et al. 2016; Kong et al. 2012; Roach et al. 2010). It is estimated that locus-specific spontaneous mutation rates for copy-number variants (CNVs) are approximately hundreds or thousands fold higher than that of *de novo* SNVs (*dn*SNVs), *i.e.* $\sim 10^{-6}$ to 10^{-4} per generation, resulting in 0-1 *de novo* CNV per genome (Lupski 2007; Turner et al. 2008).

Interestingly, Liu et al. (Liu et al. 2017) and Du et al. (Du et al. 2022) described a novel type of constitutional genome instability with an unusually large number of *de novo*

mutations (DNMs) for multiple *de novo* CNVs (M*dn*CNV) occurring during perizygotic mutagenesis; this M*dn*CNV phenomenon showed evidence for regional SNV hypermutagenesis in a 4 Mb 'window' surrounding the CNV breakpoint junctions consistent with replicative recombination repair involving an error prone polymerase (Kaplanis et al. 2022; Liu et al. 2017). Most M*dn*CNVs were arranged as large tandem duplications (~1 Mb in size) with microhomology and microhomeology at the breakpoints and *dn*SNVs in their vicinity. Genetic marker studies revealed the M*dn*CNV arose in a perizygotic time interval of organismal development, thus affecting all cells of the human body.

In a more recent study, germline hypermutation of *dn*SNVs was identified in genome-wide studies from 12 individuals out of 21,879 families with rare genetic diseases. The number of *dn*SNVs for each individual with hypermutation ranged from 110 to 425, correlating to a 1.7-6.5 fold increase compared with the median number of *dn*SNVs in the general population. Two of these individuals also had a significantly increased number of *de novo* insertion/deletions (indels) (Kaplanis et al. 2022). Constitutional new variants have been considered to primarily arise from germline or zygotic events; however, more recent data suggest postzygotic new variants are an under-recognized source of *de novo* genomic variations (Acuna-Hidalgo et al. 2015; Rahbari et al. 2016). Postzygotic events contribute to the formation of mosaicism, and recent advances in genomic technologies have enhanced our ability to detect and characterize low-level mosaicism (Contini et al. 2015; Doan et al. 2021; Lannoy and Hermans 2020; Uchiyama et al. 2016).

In 2018, we reported a newborn female with an expanded phenotype of Fanconi anemia, complementation group O (FANCO) (Jacquinet et al. 2018). She was diagnosed

prenatally with several congenital anomalies: bilateral ventriculomegaly, absence or fenestration of the septum pellucidum and fusion of the fornices anteriorly, thick and echogenic corpus callosum, cleft lip and palate, overlapping fingers, heart anomalies, symmetric fetal growth restriction, and suspected ambiguous genitalia. Of note, the family history was significant for breast cancer in the paternal grandmother and great-grandmother. Chromosomal microarray analysis using Affymetrix CytoScan HD SNP array performed on DNA from amniotic fluid was normal. Prenatal trio ES on DNA isolated from cultured amniocytes revealed inherited compound heterozygous variant alleles in *RAD51C*: (NC_00017.10(NM_058216.2)): c.935G>A (p.Arg312Gln) and c.571+5G>A that were interpreted as likely pathogenic. In addition, trio ES analysis detected eight apparent *de novo* mosaic variants in the fetus with variant allele fractions (VAF) ranging between 11.5% and 37% (Jacquinet et al. 2018). Chromosome breakage studies confirmed the diagnosis of FANCO, while the cleft lip and palate and the lobar holoprosencephaly were considered an expansion of the phenotypic spectrum of FANCO. The child died soon after birth.

Here, we describe the results of subsequent trio WGS studies in the family, which revealed SNV hypermutagenesis as evidenced by a large number of apparent *dn*SNVs and indels.

Materials and Methods

Genomic sequencing

Fetal DNA was extracted from amniotic fluid and parental DNA was extracted from peripheral blood. The prenatal trio ES and WGS were performed on the Illumina HiSeq platform following standard protocols as previously described (Liu et al. 2017; Normand et al. 2018; Yang et al. 2013). The total mean autosomal sequencing read depth-of-coverage in WGS ranged between 42-58x per sample (**Supplementary Table 1**).

Selection criteria for candidate* de novo *variants

A custom bioinformatics script was utilized to detect and filter apparent *de novo* SNVs or small indels in the trio WGS data (Gambin et al. 2020). We analyzed the VCF file to select variants for which the proband was found to be heterozygous by calculating the VAFs. We have previously shown that more than 95% of apparent *de novo* autosomal SNVs and X-linked SNVs in females in ES analyses have VAF ranging between 37.41-62.6% (Cao et al. 2019). We have now used more stringent criteria to eliminate genotype calls erroneously classified as heterozygous and removed variants with VAF above 70%, variants with a total depth of coverage below 20x in any sample from the trio, and DNMs overlapping known segmental duplications, centromeres, or *Alu* repetitive elements. We have included variants (SNVs or indels) with ≥2 alternative reads in the proband and absent in both parents. To further reduce the number of false positives and technical artifacts, we have removed all variants present in gnomAD v3.1 database [https://gnomad.broadinstitute.org]. For each selected variant, we have retrieved pileup information from the proband and parental BAM files that enabled obtaining more precise data on read depth and VAF in these samples. The analyses of the pileup data were performed using Samtools version 1.13 (Danecek et al. 2021). We used hard filtering based on https://gatk.broadinstitute.org/hc/en-us/articles/360035890471-Hard-filtering-germline-short-variants [RMSMappingQuality >

30 (for SNPs), ReadPosRankSum >-8 (SNPs) or ReadPosRankSum >-20 (InDels), and FisherStrand < 60 (SNPs) or FisherStrand < 200 (InDels), QualByDepth > 2]. All *de novo* SNVs and indels were manually curated via Integrative Genomics Viewer (IGV, v2.3) software (Robinson et al. 2011; Thorvaldsdottir et al. 2013) with the previously described criteria (Du et al. 2022).

To validate the customized variant filtering, we have used the recently published prediction module DeepMosaic, which combines an image-based visualization for mosaic SNVs with a convolutional neural network-based classification for mosaic variants detection. This pipeline has an increased sensitivity (using HaplotypeCaller with ploidy=50) and fully automated filtration mechanism (Yang et al. 2023). The *de novo* variants observed in the trio WGS of the FA patient was compared to the *de novo* variants in the control trio WGS using both pipelines.

De novo substitution mutational signature pattern analysis

The R package MutationalPatterns (Manders et al. 2022) was used for *de novo* substitution mutational signature analysis. The tri-nucleotide and pan-nucleotide mutational contexts were extracted and visualized with 'mut_matrix' and 'plot_96_profile' functions from the MutationalPattern R package. Known signatures from COSMIC (v3.2) were refitted using the 'backwards' method. The method starts by achieving an optimal reconstruction via 'fit_to_signatures.' The signature with the lowest contribution is then removed and refitting is repeated iteratively. Each time the cosine similarity between the original and reconstructed profile is calculated using refitting with 'backwards' method.

Refitting bimodal VAF distribution

To infer the timing within the life cycle of the *de novo* variants event, we have evaluated their VAF distribution (density plot). We have used a custom R code to identify the boundaries corresponding to the first, second, and third cell division, looking for the points of intersection between different distributions. To this end, we have sampled 10,000 values from theoretical densities of binomial distributions of VAFs corresponding to each cell division using the cbinom R package (https://cran.r-project.org/web/packages/cbinom/cbinom.pdf). Next, we have estimated the boundaries between the consecutive pairs of distributions by identifying the positions closest to the intersection points.

CNV calling and visualization

CNVs were called using Illumina Dragen Bio-IT Platform (v3.4.15). The read depth was calculated with mosdepth (v 0.3.4) (Pedersen and Quinlan 2018) and visualized with the in-house visualization tool VizCNV (https://github.com/BCM-Lupskilab/VizCNV) that allows for normalized read depth plotting of the proband and both parents to help with manual inspection of potential *de novo* CNVs larger than 3 kb.

Results

De novo variants identification in ES and WGS data

Computational CNV analyses followed by manual read-depth visualization did not reveal any increased rate of CNV formation (**Supplementary Figure 1**). Using the customized variant filtering pipeline, genome-wide analyses of the trio WGS data for the FA patient revealed 45 *de novo* variants with VAFs ranging between 37.41% and 62.6%, considered

as germline variant alleles. In the control trio WGS, 34 *de novo* variants had the same germline VAF range.

For VAFs below 37.41%, considered as likely postzygotic somatic mosaic events, after removal of the false-positive variants during IGV inspection, an unexpectedly high number of apparent dnSNVs (n=830) and indels (n=21) were detected, compared to only 11 dnSNVs and 3 indels in the control case (**Table 1a**). Using the DeepMosaic module, analyses of trio WGS data from the FA patient identified 44 non-mosaic heterozygous variants, and 850 dnSNVs and 25 indels with VAF below 37.41%, compared to 34 heterozygous variants and 38 dnSNVs and 8 indels with VAF below 37.4% in the control sample (**Table 1b**) (**Supplementary Table 2**).

Re-analysis of prenatal trio ES data confirmed eight apparent *de novo* mosaic variants with VAF ranging between 11.5% and 37% and revealed an additional 15 apparent *de novo* mosaic variants with VAFs ranging between 9% and 37% (**Supplementary Table 3**). The novel variants are mainly located in non-coding regions, close to the exon boundaries. Although the number of *de novo* mosaic variants detected through the exome analysis is not high, the depth-of-coverage is sufficient to consider the VAFs of the mosaic variants as solid.

De novo variant allele frequency distribution

The VAF pattern of apparent *dn*SNVs **(Figure 1)** suggests their bimodal distribution, with one component with an average VAF of 25% (ranging between 18.7-37.41%) (n=446), representing potential postzygotic first mitotic events, and the other with an average VAF of 12.5% (ranging between 9.55-18.69%) (n=384), representing potential second mitotic events.

Mutational signature of the de novo *substitutions*

Genome-wide distribution of the somatic *de novo* substitutions shows clusters (genomic distance <50 bp) spreading across multiple chromosomes (**Figure 2a**). Clustered variants are likely generated by the same mutagenic events and may provide insights into the somatic mutational processes (Alexandrov et al. 2020; Supek and Lehner 2017). Distinct patterns were observed after reviewing the pileup of *de novo* variants in IGV. We have classified the *de novo* substitutions into four categories: single-base substitution (SBS, n=799), tandem dinucleotide or duplet-base substitution (DBS, n=9), multiple (n>1) single-base substitution within 50 bp in distance or clustered-base substitution (CBS, n=13), and biallelic base substitution (BBS, n=3), in which two substitutions are present at the same location **(Figure 2 b-e**). Of note, a higher number of SBS, DBS, and CBS were observed in the FA patient compared to the control case (**Table 2**). The CBS does not seem to represent the kataegis phenomenon, where a much large number of base pair mutations occur in a small region with enrichment of transition substitutional pattern (C>T or G>A) (54). Most of the CBS observed in the FA patient were in two base pair variants in a distance of 1 bp and enriched for transversion over the transition (Ti:Tv = 0.04). We compared SBS to the COSMIC context to infer potential somatic mutational processes. The mutational pattern analyses of the SBS revealed enrichment of C>G, C>A, and C>T variants (**Figure 3**). The mutational pattern analysis for somatic SBS shows SBS40, SBS5, and SBS3 as the top signatures (**Figure 4**). The SBS40 is a flat signature similar to SBS5, of which the underlying etiology is uncertain. SBS3 is a known signature associated with HR-based DNA damage repair error, often due to *BRCA1* or *BRCA2* inactivation (Nik-Zainal et al. 2012).

Discussion

RAD51 is a RecA-like DNA recombinase that initiates HR upon DNA damage by replacing Replication protein A (RPA) and catalyzing strand transfer between the broken sequence and its undamaged homolog (Boni et al. 2022). RAD51C is a member of the RAD51 family, RAD51B, RAD51C, RAD51D, XRCC2, and XRCC3, required for efficient DNA double-strand break repair by HR (Chun et al. 2013). Furthermore, RAD51 paralogs have been shown also to play cardinal roles in protecting the replicative fork during DNA synthesis (Somyajit K 2015), avoiding unrestrained fork progression, and promoting efficient restart (Berti et al. 2020). Depletion of RAD51 paralog genes in human cell lines has been associated with impaired HR, reduced genome stability, increased DSBs, and growth defects (Boni et al. 2022). The collection of replication stress-associated DNA lesions leads to genomic instability and may predispose to cancerogenesis, also in the heterozygous state, *e.g.,* cancers arising mainly in the breast and ovary (Badra Fajardo et al. 2022; Boni et al. 2022; Kottemann and Smogorzewska 2013).

The first pathogenic variant in *RAD51C* associated with a human disorder was described in three siblings from a consanguineous Pakistani family with clinical features suggestive of FA (Vaz et al. 2010). Functional analyses of the homozygous missense mutation (NM_058216.2:c.773G>A (p.Arg258His)) demonstrated an increase in G2 arrest in response to MMC in patient cultured lymphocytes compared to controls, a decrease in RAD51C focus formation in response to MMC in patient fibroblasts, a modest radiosensitivity, and an increased sensitivity to the topoisomerase I inhibitor, camptothecin (Somyajit et al. 2012; Vaz et al. 2010). Since the report of this family, only the presented

individual with FANCO has been reported in the literature, thus the phenotypic spectrum associated with *RAD51C* mutations remains obscure.

The 'mutator phenotype' was first described in *Drosophila* and in bacteria (Liu et al. 2017; Miyake 1960; Plough 1941). In the last two decades, it has been appreciated also in cancers (Nicolaides et al. 1998). The variants have been found to be usually driven by an intrinsic source, such as a defective mismatch repair gene (Loeb 2001). Therefore, the mutations are expected to develop over time, and accumulate in a somatic mosaic state (Kilpivaara and Aaltonen 2013). In contrast, the non-cancer constitutional CNV mutator phenotype described by Liu et al. (Liu et al. 2017) and most recently further expanded by Du et al. (Du et al. 2022) seems to be restricted to a specific time-interval in early embryonic development and no longer operating after early postzygotic division, as evidenced by the lack of mosaicism for any *dn*CNVs over time. Nevertheless, several features of the constitutional M*dn*CNV phenomenon suggested that these mutations arise due to a faulty replicative repair process, resembling chromosomal instability in various cancers. The presence of complex genomic rearrangements, microhomology and microhomeology observed at breakpoint junctions, templated insertion at the breakpoints, and *dn*SNVs in the proximity (~1-4 Mb) to breakpoint junctions are characteristics of the fork stalling and template switching (FoSTeS)/microhomology-mediated break-induced replication (MMBIR) mechanisms (Bahrambeigi et al. 2019; Beck et al. 2019; Carvalho and Lupski 2016; Carvalho et al. 2013; Lee et al. 2007).

Recently, Kaplanis et al. proposed three potential sources of germline hypermutation identified in 12 individuals: (i) defects in paternal DNA-repair genes (*XPC*, involved in the early stages of the NER pathway, and *MPG*, involved in recognition of base

lesions, including alkylated and deaminated purines, and initiation of the base-excision repair pathway), (ii) paternal exposure to chemotherapeutics, and (iii) postzygotic mutational factors (Kaplanis et al. 2022). One patient with postzygotic mutational factors had several blood-related clinical phenotypes, including myelodysplasia, and the observations were assumed to be related to clonal hematopoiesis, leading to a large number of somatic mutations in the child's blood. The hematopoietic presentation is shared with FA patients, although the DNA repair process differs. There were no similar blood-related phenotypes in the second patient with postzygotic mutational factors, although the child was only one year old at recruitment.

The distribution pattern of the VAF for the mosaic *dn*SNVs showed a bimodal distribution, with most of the VAFs oscillating around 25% and 12.5%, indicating these variants likely arose during the first and second mitotic division (Ju et al. 2017; Oron and Ivanova 2012). Relatively low read coverage in WGS data and lack of postnatal proband DNA precluded us from analyzing lower-level mosaic variants potentially corresponding to subsequent mitotic divisions. More systematic studies of ultra-high coverage may better inform on a complex interplay of mechanisms surveilling early postzygotic genome stability, *e.g.* protective maternal cytoplasmatic proteins (Schulz and Harrison 2019) and more efficient DNA-repair leading to a reduced mutation rate proposed to be established from the 4-cell stage onwards (Coorens et al. 2021; Park et al. 2021).

Distinctive mutational patterns, termed 'mutational signatures,' are associated with different mutational mechanisms (Alexandrov et al. 2013). There are more than 100 somatic mutational signatures defined across various cancers, of which some have been attributed to endogenous mutagenic processes (Phillips 2018). Most germline mutations

can be accounted for by two signatures, SBS1 and SBS5. Both signatures are ubiquitous among normal and cancer cells (Moore et al. 2021). The mutational signatures identified in our patient exhibited known single base substitution signatures (SBS1, SBS5), including one associated with HR-based DNA damage repair error (SBS3), consistent with *RAD51C* role in HR-mediated repair of ICL and double-strand DNA breaks (Somyajit et al. 2012; Vaz et al. 2010). In Kaplanis et al., the two individuals with postzygotic mutational factors shared a large contribution from SBS1 (Kaplanis et al. 2022).

We propose that biallelic *RAD51C* mutations and defects in DNA damage response result in a 'hypermutator phenotype' with the accumulation of postzygotic *de novo* mutations. Recently reported mutational spectrum of squamous cell carcinoma from individuals with FA showed a collection of somatic SV, with exposure to tobacco and alcohol mutagens overwhelming the ability of a defective DNA repair pathway (Webster et al. 2022). The absence of major SV in our patient could be due to the more stringent selection pressure that occurs during embryonic development compared to selection for cancer cells. We cannot exclude the possibility that somatic mosaic SVs could contribute to the clinical phenotype of this patient later in life, e.g. AML or squamous cell carcinoma of the head and neck.

Heterozygous carriers of variants involving the FA genes from the HR pathway, including *RAD51C*, have been associated with familial predisposition to breast-ovarian cancer (D'Andrea 2010; Rageul and Kim 2020). Germline mutations in *RAD51C* have been identified in about 1% of hereditary breast and ovarian cancer families (the lifetime risk of ovarian cancer is approximately 9%) and in families with a history of breast cancer alone (Sopik et al. 2015). Importantly, in the present case, the family history was significant for

breast cancer in female family members of the proband's father with the heterozygous splicing variant c.571+5G>A in *RAD51C*. Interestingly, the same variant was identified in a Chinese individual among 273 BRCA1/2-negative familial breast cancer cases (Pang et al. 2011). Segregation analysis of the paternally inherited splicing variant (c.571+5G>A) was recommended to better counsel the family about the increased risk for breast and ovarian cancer in female carriers. Similar segregation should also be considered for the maternally inherited missense variant, although it was not previously described in familial breast cancer cases. Of note, reproductive options, such as prenatal diagnosis and preimplantation genetic testing, were discussed with the parents due to 25% risk of recurrence of FA.

Finally, we propose that in addition to chromosome aneuploidies, common during early human embryogenesis (Cavazza et al. 2021; Vanneste et al. 2009), the hypermutation phenomenon, such as observed in our patient with FA, might represent the underrecognized mechanism responsible for early pregnancy losses.

Conclusions

Our data demonstrate that biallelic *RAD51C* variants affecting DNA damage repair process result in an SNV hypermutator phenotype, leading to the accumulation of postzygotic *de novo* variants, at least in the prenatal period. This phenomenon might contribute to the hematological manifestations and the predisposition to cancer in patients with FA. Investigating this phenomenon may enhance our knowledge about the phenotype of FA and cancer biology. We propose that additional FA groups be analyzed genome-wide for *de novo* variants and better genotype-phenotype correlations. As most DNA-repair

disorders, including FA, are recessive disorders, we suggest that genetic causes of hypermutation are more likely to be found at higher frequencies in populations with increased consanguinity rates. Importantly, hypermutation phenotype may be an under-ascertained factor responsible for early pregnancy losses.

References

Acuna-Hidalgo R, Bo T, Kwint MP, van de Vorst M, Pinelli M, Veltman JA, Hoischen A, Vissers LE, Gilissen C (2015) Post-zygotic Point Mutations Are an Underrecognized Source of De Novo Genomic Variation. Am J Hum Genet 97: 67-74. doi: 10.1016/j.ajhg.2015.05.008

Acuna-Hidalgo R, Veltman JA, Hoischen A (2016) New insights into the generation and role of de novo mutations in health and disease. Genome Biol 17: 241. doi: 10.1186/s13059-016-1110-1

Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, Boot A, Covington KR, Gordenin DA, Bergstrom EN, Islam SMA, Lopez-Bigas N, Klimczak LJ, McPherson JR, Morganella S, Sabarinathan R, Wheeler DA, Mustonen V, Group PMSW, Getz G, Rozen SG, Stratton MR, Consortium P (2020) The repertoire of mutational signatures in human cancer. Nature 578: 94-101. doi: 10.1038/s41586-020-1943-3

Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR (2013) Deciphering signatures of mutational processes operative in human cancer. Cell Rep 3: 246-59. doi: 10.1016/j.celrep.2012.12.008

Ameziane N, May P, Haitjema A, van de Vrugt HJ, van Rossum-Fikkert SE, Ristic D, Williams GJ, Balk J, Rockx D, Li H, Rooimans MA, Oostra AB, Velleuer E, Dietrich R, Bleijerveld OB, Maarten Altelaar AF, Meijers-Heijboer H, Joenje H, Glusman G, Roach J, Hood L, Galas D, Wyman C, Balling R, den Dunnen J, de Winter JP, Kanaar R, Gelinas R, Dorsman JC (2015) A novel Fanconi anaemia subtype associated with a dominant-negative mutation in RAD51. Nat Commun 6: 8829. doi: 10.1038/ncomms9829

Auerbach AD (2009) Fanconi anemia and its diagnosis. Mutat Res 668: 4-10. doi: 10.1016/j.mrfmmm.2009.01.013

Badra Fajardo N, Taraviras S, Lygerou Z (2022) Fanconi anemia proteins and genome fragility: unraveling replication defects for cancer therapy. Trends Cancer 8: 467-481. doi: 10.1016/j.trecan.2022.01.015

Bahrambeigi V, Song X, Sperle K, Beck CR, Hijazi H, Grochowski CM, Gu S, Seeman P, Woodward KJ, Carvalho CMB, Hobson GM, Lupski JR (2019) Distinct patterns of complex rearrangements and a mutational signature of microhomeology are frequently observed in PLP1 copy number gain structural variants. Genome Med 11: 80. doi: 10.1186/s13073-019-0676-0

Beck CR, Carvalho CMB, Akdemir ZC, Sedlazeck FJ, Song X, Meng Q, Hu J, Doddapaneni H, Chong Z, Chen ES, Thornton PC, Liu P, Yuan B, Withers M, Jhangiani SN, Kalra D, Walker K, English AC, Han Y, Chen K, Muzny DM, Ira G, Shaw CA, Gibbs RA, Hastings PJ, Lupski JR (2019) Megabase Length Hypermutation Accompanies Human Structural Variation at 17p11.2. Cell 176: 1310-1324 e10. doi: 10.1016/j.cell.2019.01.045

Berti M, Teloni F, Mijic S, Ursich S, Fuchs J, Palumbieri MD, Krietsch J, Schmid JA, Garcin EB, Gon S, Modesti M, Altmeyer M, Lopes M (2020) Sequential role of RAD51 paralog complexes in replication fork remodeling and restart. Nat Commun 11: 3531. doi: 10.1038/s41467-020-17324-z

Boni J, Idani A, Roca C, Feliubadalo L, Tomiak E, Weber E, Foulkes WD, Orthwein A, El Haffaf Z, Lazaro C, Rivera B (2022) A decade of RAD51C and RAD51D germline variants in cancer. Hum Mutat 43: 285-298. doi: 10.1002/humu.24319

Campbell CD, Eichler EE (2013) Properties and rates of germline mutations in humans. Trends Genet 29: 575-84. doi: 10.1016/j.tig.2013.04.005

Cao Y, Tokita MJ, Chen ES, Ghosh R, Chen T, Feng Y, Gorman E, Gibellini F, Ward PA, Braxton A, Wang X, Meng L, Xiao R, Bi W, Xia F, Eng CM, Yang Y, Gambin T, Shaw C, Liu P,

Stankiewicz P (2019) A clinical survey of mosaic single nucleotide variants in disease-causing genes detected by exome sequencing. Genome Med 11: 48. doi: 10.1186/s13073-019-0658-2

Carvalho CM, Lupski JR (2016) Mechanisms underlying structural variant formation in genomic disorders. Nat Rev Genet 17: 224-38. doi: 10.1038/nrg.2015.25

Carvalho CM, Pehlivan D, Ramocki MB, Fang P, Alleva B, Franco LM, Belmont JW, Hastings PJ, Lupski JR (2013) Replicative mechanisms for CNV formation are error prone. Nat Genet 45: 1319-26. doi: 10.1038/ng.2768

Cavazza T, Takeda Y, Politi AZ, Aushev M, Aldag P, Baker C, Choudhary M, Bucevicius J, Lukinavicius G, Elder K, Blayney M, Lucas-Hahn A, Niemann H, Herbert M, Schuh M (2021) Parental genome unification is highly error-prone in mammalian embryos. Cell 184: 2860-2877 e22. doi: 10.1016/j.cell.2021.04.013

Ceccaldi R, Sarangi P, D'Andrea AD (2016) The Fanconi anaemia pathway: new players and new functions. Nat Rev Mol Cell Biol 17: 337-49. doi: 10.1038/nrm.2016.48

Chun J, Buechelmaier ES, Powell SN (2013) Rad51 paralog complexes BCDX2 and CX3 act at different stages in the BRCA1-BRCA2-dependent homologous recombination pathway. Mol Cell Biol 33: 387-95. doi: 10.1128/MCB.00465-12

Contini E, Paganini I, Sestini R, Candita L, Capone GL, Barbetti L, Falconi S, Frusconi S, Giotti I, Giuliani C, Torricelli F, Benelli M, Papi L (2015) A Systematic Assessment of Accuracy in Detecting Somatic Mosaic Variants by Deep Amplicon Sequencing: Application to NF2 Gene. PLoS One 10: e0129099. doi: 10.1371/journal.pone.0129099

Coorens THH, Moore L, Robinson PS, Sanghvi R, Christopher J, Hewinson J, Przybilla MJ, Lawson ARJ, Spencer Chapman M, Cagan A, Oliver TRW, Neville MDC, Hooks Y, Noorani A, Mitchell TJ, Fitzgerald RC, Campbell PJ, Martincorena I, Rahbari R, Stratton MR (2021) Extensive phylogenies of human development inferred from somatic mutations. Nature 597: 387-392. doi: 10.1038/s41586-021-03790-y

D'Andrea AD (2010) Susceptibility pathways in Fanconi's anemia and breast cancer. N Engl J Med 362: 1909-19. doi: 10.1056/NEJMra0809889

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H (2021) Twelve years of SAMtools and BCFtools. Gigascience 10: giab008. doi: 10.1093/gigascience/giab008

Doan RN, Miller MB, Kim SN, Rodin RE, Ganz J, Bizzotto S, Morillo KS, Huang AY, Digumarthy R, Zemmel Z, Walsh CA (2021) MIPP-Seq: ultra-sensitive rapid detection and validation of low-frequency mosaic mutations. BMC Med Genomics 14: 47. doi: 10.1186/s12920-021-00893-3

Du H, Jolly A, Grochowski CM, Yuan B, Dawood M, Jhangiani SN, Li H, Muzny D, Fatih JM, Coban-Akdemir Z, Carlin ME, Scheuerle AE, Witzl K, Posey JE, Pendleton M, Harrington E, Juul S, Hastings PJ, Bi W, Gibbs RA, Sedlazeck FJ, Lupski JR, Carvalho CMB, Liu P (2022) The multiple de novo copy number variant (MdnCNV) phenomenon presents with peri-zygotic DNA mutational signatures and multilocus pathogenic variation. Genome Med 14: 122. doi: 10.1186/s13073-022-01123-w

Duxin JP, Walter JC (2015) What is the DNA repair defect underlying Fanconi anemia? Curr Opin Cell Biol 37: 49-60. doi: 10.1016/j.ceb.2015.09.002

Gambin T, Liu Q, Karolak JA, Grochowski CM, Xie NG, Wu LR, Yan YH, Cao Y, Coban Akdemir ZH, Wilson TA, Jhangiani SN, Chen E, Eng CM, Muzny D, Posey JE, Yang Y, Zhang DY, Shaw C, Liu P, Lupski JR, Stankiewicz P (2020) Low-level parental somatic mosaic SNVs in exomes from a large cohort of trios with diverse suspected Mendelian conditions. Genet Med 22: 1768-1776. doi: 10.1038/s41436-020-0897-z

Goldmann JM, Wong WS, Pinelli M, Farrah T, Bodian D, Stittrich AB, Glusman G, Vissers LE, Hoischen A, Roach JC, Vockley JG, Veltman JA, Solomon BD, Gilissen C, Niederhuber JE (2016) Parent-of-origin-specific signatures of de novo mutations. Nat Genet 48: 935-9. doi: 10.1038/ng.3597

Jacquinet A, Brown L, Sawkins J, Liu P, Pugash D, Van Allen MI, Patel MS (2018) Expanding the FANCO/RAD51C associated phenotype: Cleft lip and palate and lobar holoprosencephaly, two rare findings in Fanconi anemia. Eur J Med Genet 61: 257-261. doi: 10.1016/j.ejmg.2017.12.011

Ju YS, Martincorena I, Gerstung M, Petljak M, Alexandrov LB, Rahbari R, Wedge DC, Davies HR, Ramakrishna M, Fullam A, Martin S, Alder C, Patel N, Gamble S, O'Meara S, Giri DD, Sauer T, Pinder SE, Purdie CA, Borg A, Stunnenberg H, van de Vijver M, Tan BK, Caldas C, Tutt A, Ueno NT, van 't Veer LJ, Martens JW, Sotiriou C, Knappskog S, Span PN, Lakhani SR, Eyfjord JE, Borresen-Dale AL, Richardson A, Thompson AM, Viari A, Hurles ME, Nik-Zainal S, Campbell PJ, Stratton MR (2017) Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. Nature 543: 714-718. doi: 10.1038/nature21703

Kaplanis J, Ide B, Sanghvi R, Neville M, Danecek P, Coorens T, Prigmore E, Short P, Gallone G, McRae J, Genomics England Research C, Carmichael J, Barnicoat A, Firth H, O'Brien P, Rahbari R, Hurles M (2022) Genetic and chemotherapeutic influences on germline hypermutation. Nature 605: 503-508. doi: 10.1038/s41586-022-04712-2

Kee Y, D'Andrea AD (2012) Molecular pathogenesis and clinical management of Fanconi anemia. J Clin Invest 122: 3799-806. doi: 10.1172/JCI58321

Kilpivaara O, Aaltonen LA (2013) Diagnostic cancer genome sequencing and the contribution of germline variants. Science 339: 1559-62. doi: 10.1126/science.1233899

Kolinjivadi AM, Crismani W, Ngeow J (2020) Emerging functions of Fanconi anemia genes in replication fork protection pathways. Hum Mol Genet 29: R158-R164. doi: 10.1093/hmg/ddaa087

Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A, Jonasdottir A, Jonasdottir A, Wong WS, Sigurdsson G, Walters GB, Steinberg S, Helgason H, Thorleifsson G, Gudbjartsson DF, Helgason A, Magnusson OT, Thorsteinsdottir U, Stefansson K (2012) Rate of de novo mutations and the importance of father's age to disease risk. Nature 488: 471-5. doi: 10.1038/nature11396

Kottemann MC, Smogorzewska A (2013) Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. Nature 493: 356-63. doi: 10.1038/nature11863

Kutler DI, Patel KR, Auerbach AD, Kennedy J, Lach FP, Sanborn E, Cohen MA, Kuhel WI, Smogorzewska A (2016) Natural history and management of Fanconi anemia patients with head and neck cancer: A 10-year follow-up. Laryngoscope 126: 870-9. doi: 10.1002/lary.25726

Kutler DI, Singh B, Satagopan J, Batish SD, Berwick M, Giampietro PF, Hanenberg H, Auerbach AD (2003) A 20-year perspective on the International Fanconi Anemia Registry (IFAR). Blood 101: 1249-56. doi: 10.1182/blood-2002-07-2170

Kuznetsov S, Pellegrini M, Shuda K, Fernandez-Capetillo O, Liu Y, Martin BK, Burkett S, Southon E, Pati D, Tessarollo L, West SC, Donovan PJ, Nussenzweig A, Sharan SK (2007) RAD51C deficiency in mice results in early prophase I arrest in males and sister chromatid separation at metaphase II in females. J Cell Biol 176: 581-92. doi: 10.1083/jcb.200608130

Kuznetsov SG, Haines DC, Martin BK, Sharan SK (2009) Loss of Rad51c leads to embryonic lethality and modulation of Trp53-dependent tumorigenesis in mice. Cancer Res 69: 863-72. doi: 10.1158/0008-5472.CAN-08-3057

Lannoy N, Hermans C (2020) Genetic mosaicism in haemophilia: A practical review to help evaluate the risk of transmitting the disease. Haemophilia 26: 375-383. doi: 10.1111/hae.13975

Lee JA, Carvalho CM, Lupski JR (2007) A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell 131: 1235-47. doi: 10.1016/j.cell.2007.11.037

Liu P, Yuan B, Carvalho CMB, Wuster A, Walter K, Zhang L, Gambin T, Chong Z, Campbell IM, Coban Akdemir Z, Gelowani V, Writzl K, Bacino CA, Lindsay SJ, Withers M, Gonzaga-Jauregui C, Wiszniewska J, Scull J, Stankiewicz P, Jhangiani SN, Muzny DM, Zhang F, Chen K, Gibbs RA, Rautenstrauss B, Cheung SW, Smith J, Breman A, Shaw CA, Patel A, Hurles ME, Lupski JR (2017) An Organismal CNV Mutator Phenotype Restricted to Early Human Development. Cell 168: 830-842 e7. doi: 10.1016/j.cell.2017.01.037

Loeb LA (2001) A mutator phenotype in cancer. Cancer Res 61: 3230-9.

Lupski JR (2007) Genomic rearrangements and sporadic disease. Nat Genet 39: S43-7. doi: 10.1038/ng2084

Mamrak NE, Shimamura A, Howlett NG (2017) Recent discoveries in the molecular pathogenesis of the inherited bone marrow failure syndrome Fanconi anemia. Blood Rev 31: 93-99. doi: 10.1016/j.blre.2016.10.002

Manders F, Brandsma AM, de Kanter J, Verheul M, Oka R, van Roosmalen MJ, van der Roest B, van Hoeck A, Cuppen E, van Boxtel R (2022) MutationalPatterns: the one stop shop for the analysis of mutational processes. BMC Genomics 23: 134. doi: 10.1186/s12864-022-08357-3

Mehta P, Ebens C (2021) Fanconi Anemia.

Milletti G, Strocchio L, Pagliara D, Girardi K, Carta R, Mastronuzzi A, Locatelli F, Nazio F (2020) Canonical and Noncanonical Roles of Fanconi Anemia Proteins: Implications in Cancer Predisposition. Cancers (Basel) 12: 2684-2706. doi: 10.3390/cancers12092684

Miyake T (1960) Mutator Factor in Salmonella Typhimurium. Genetics 45: 11-4. doi: 10.1093/genetics/45.1.11

Moldovan GL, D'Andrea AD (2009) How the fanconi anemia pathway guards the genome. Annu Rev Genet 43: 223-49. doi: 10.1146/annurev-genet-102108-134222

Moore L, Cagan A, Coorens THH, Neville MDC, Sanghvi R, Sanders MA, Oliver TRW, Leongamornlert D, Ellis P, Noorani A, Mitchell TJ, Butler TM, Hooks Y, Warren AY, Jorgensen M, Dawson KJ, Menzies A, O'Neill L, Latimer C, Teng M, van Boxtel R, Iacobuzio-Donahue CA, Martincorena I, Heer R, Campbell PJ, Fitzgerald RC, Stratton MR, Rahbari R (2021) The mutational landscape of human somatic and germline cells. Nature 597: 381-386. doi: 10.1038/s41586-021-03822-7

Moreno OM, Paredes AC, Suarez-Obando F, Rojas A (2021) An update on Fanconi anemia: Clinical, cytogenetic and molecular approaches (Review). Biomed Rep 15: 74. doi: 10.3892/br.2021.1450

Nalepa G, Clapp DW (2018) Fanconi anaemia and cancer: an intricate relationship. Nat Rev Cancer 18: 168-185. doi: 10.1038/nrc.2017.116

Nicolaides NC, Littman SJ, Modrich P, Kinzler KW, Vogelstein B (1998) A naturally occurring hPMS2 mutation can confer a dominant negative mutator phenotype. Mol Cell Biol 18: 1635-41. doi: 10.1128/MCB.18.3.1635

Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, Jones D, Hinton J, Marshall J, Stebbings LA, Menzies A, Martin S, Leung K, Chen L, Leroy C, Ramakrishna M, Rance R, Lau KW, Mudie LJ, Varela I, McBride DJ, Bignell GR, Cooke SL, Shlien A, Gamble J, Whitmore I, Maddison M, Tarpey PS, Davies HR, Papaemmanuil E, Stephens PJ,

McLaren S, Butler AP, Teague JW, Jonsson G, Garber JE, Silver D, Miron P, Fatima A, Boyault S, Langerod A, Tutt A, Martens JW, Aparicio SA, Borg A, Salomon AV, Thomas G, Borresen-Dale AL, Richardson AL, Neuberger MS, Futreal PA, Campbell PJ, Stratton MR, Breast Cancer Working Group of the International Cancer Genome C (2012) Mutational processes molding the genomes of 21 breast cancers. Cell 149: 979-93. doi: 10.1016/j.cell.2012.04.024

Normand EA, Braxton A, Nassef S, Ward PA, Vetrini F, He W, Patel V, Qu C, Westerfield LE, Stover S, Dharmadhikari AV, Muzny DM, Gibbs RA, Dai H, Meng L, Wang X, Xiao R, Liu P, Bi W, Xia F, Walkiewicz M, Van den Veyver IB, Eng CM, Yang Y (2018) Clinical exome sequencing for fetuses with ultrasound abnormalities and a suspected Mendelian disorder. Genome Med 10: 74. doi: 10.1186/s13073-018-0582-x

Oron E, Ivanova N (2012) Cell fate regulation in early mammalian development. Phys Biol 9: 045002. doi: 10.1088/1478-3975/9/4/045002

Pang PC, Chiu PC, Lee CL, Chang LY, Panico M, Morris HR, Haslam SM, Khoo KH, Clark GF, Yeung WS, Dell A (2011) Human sperm binding is mediated by the sialyl-Lewis(x) oligosaccharide on the zona pellucida. Science 333: 1761-4. doi: 10.1126/science.1207438

Park S, Mali NM, Kim R, Choi JW, Lee J, Lim J, Park JM, Park JW, Kim D, Kim T, Yi K, Choi JH, Kwon SG, Hong JH, Youk J, An Y, Kim SY, Oh SA, Kwon Y, Hong D, Kim M, Kim DS, Park JY, Oh JW, Ju YS (2021) Clonal dynamics in early human embryogenesis inferred from somatic mutation. Nature 597: 393-397. doi: 10.1038/s41586-021-03786-8

Pedersen BS, Quinlan AR (2018) Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics 34: 867-868. doi: 10.1093/bioinformatics/btx699

Phillips DH (2018) Mutational spectra and mutational signatures: Insights into cancer aetiology and mechanisms of DNA damage and repair. DNA Repair (Amst) 71: 6-11. doi: 10.1016/j.dnarep.2018.08.003

Plough HH (1941) Spontaneous Mutability in Drosophila. Cold Spring Harbor Symposia on Quantitative Biology 9: 127-137. doi: 10.1101/sqb.1941.009.01.016

Rageul J, Kim H (2020) Fanconi anemia and the underlying causes of genomic instability. Environ Mol Mutagen 61: 693-708. doi: 10.1002/em.22358

Rahbari R, Wuster A, Lindsay SJ, Hardwick RJ, Alexandrov LB, Turki SA, Dominiczak A, Morris A, Porteous D, Smith B, Stratton MR, Consortium UK, Hurles ME (2016) Timing, rates and spectra of human germline mutation. Nat Genet 48: 126-133. doi: 10.1038/ng.3469

Roach JC, Glusman G, Smit AF, Huff CD, Hubley R, Shannon PT, Rowen L, Pant KP, Goodman N, Bamshad M, Shendure J, Drmanac R, Jorde LB, Hood L, Galas DJ (2010) Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science 328: 636-9. doi: 10.1126/science.1186802

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. Nat Biotechnol 29: 24-6. doi: 10.1038/nbt.1754

Schulz KN, Harrison MM (2019) Mechanisms regulating zygotic genome activation. Nat Rev Genet 20: 221-234. doi: 10.1038/s41576-018-0087-x

Somyajit K SS, Babu S, Mishra A, Nagaraju G (2015) Mammalian RAD51 paralogs protect nascent DNA at stalled forks and mediate replication restart. Nucleic Acids Res 43: 9835-9855.

Somyajit K, Subramanya S, Nagaraju G (2012) Distinct roles of FANCO/RAD51C protein in DNA damage signaling and repair: implications for Fanconi anemia and breast cancer susceptibility. J Biol Chem 287: 3366-80. doi: 10.1074/jbc.M111.311241

Sopik V, Akbari MR, Narod SA (2015) Genetic testing for RAD51C mutations: in the clinic and community. Clin Genet 88: 303-12. doi: 10.1111/cge.12548

Supek F, Lehner B (2017) Clustered Mutation Signatures Reveal that Error-Prone DNA Repair Targets Mutations to Active Genes. Cell 170: 534-547 e23. doi: 10.1016/j.cell.2017.07.003

Thorvaldsdottir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14: 178-92. doi: 10.1093/bib/bbs017

Turner DJ, Miretti M, Rajan D, Fiegler H, Carter NP, Blayney ML, Beck S, Hurles ME (2008) Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. Nat Genet 40: 90-5. doi: 10.1038/ng.2007.40

Uchiyama Y, Nakashima M, Watanabe S, Miyajima M, Taguri M, Miyatake S, Miyake N, Saitsu H, Mishima H, Kinoshita A, Arai H, Yoshiura K, Matsumoto N (2016) Ultra-sensitive droplet digital PCR for detecting a low-prevalence somatic GNAQ mutation in Sturge-Weber syndrome. Sci Rep 6: 22985. doi: 10.1038/srep22985

Vanneste E, Voet T, Le Caignec C, Ampe M, Konings P, Melotte C, Debrock S, Amyere M, Vikkula M, Schuit F, Fryns JP, Verbeke G, D'Hooghe T, Moreau Y, Vermeesch JR (2009) Chromosome instability is common in human cleavage-stage embryos. Nat Med 15: 577-83. doi: 10.1038/nm.1924

Vaz F, Hanenberg H, Schuster B, Barker K, Wiek C, Erven V, Neveling K, Endt D, Kesterton I, Autore F, Fraternali F, Freund M, Hartmann L, Grimwade D, Roberts RG, Schaal H, Mohammed S, Rahman N, Schindler D, Mathew CG (2010) Mutation of the RAD51C gene in a Fanconi anemia-like disorder. Nat Genet 42: 406-9. doi: 10.1038/ng.570

Veltman JA, Brunner HG (2012) De novo mutations in human genetic disease. Nat Rev Genet 13: 565-75. doi: 10.1038/nrg3241

Wang AT, Kim T, Wagner JE, Conti BA, Lach FP, Huang AL, Molina H, Sanborn EM, Zierhut H, Cornes BK, Abhyankar A, Sougnez C, Gabriel SB, Auerbach AD, Kowalczykowski SC, Smogorzewska A (2015) A Dominant Mutation in Human RAD51 Reveals Its Function in DNA Interstrand Crosslink Repair Independent of Homologous Recombination. Mol Cell 59: 478-90. doi: 10.1016/j.molcel.2015.07.009

Webster ALH, Sanders MA, Patel K, Dietrich R, Noonan RJ, Lach FP, White RR, Goldfarb A, Hadi K, Edwards MM, Donovan FX, Hoogenboezem RM, Jung M, Sridhar S, Wiley TF, Fedrigo O, Tian H, Rosiene J, Heineman T, Kennedy JA, Bean L, Rosti RO, Tryon R, Gonzalez AM, Rosenberg A, Luo JD, Carroll TS, Shroff S, Beaumont M, Velleuer E, Rastatter JC, Wells SI, Surralles J, Bagby G, MacMillan ML, Wagner JE, Cancio M, Boulad F, Scognamiglio T, Vaughan R, Beaumont KG, Koren A, Imielinski M, Chandrasekharappa SC, Auerbach AD, Singh B, Kutler DI, Campbell PJ, Smogorzewska A (2022) Genomic signature of Fanconi anaemia DNA repair pathway deficiency in cancer. Nature 612: 495-502. doi: 10.1038/s41586-022-05253-4

Yang X, Xu X, Breuss MW, Antaki D, Ball LL, Chung C, Shen J, Li C, George RD, Wang Y, Bae T, Cheng Y, Abyzov A, Wei L, Alexandrov LB, Sebat JL, Network NBSM, Gleeson JG (2023) Control-independent mosaic single nucleotide variant detection with DeepMosaic. Nat Biotechnol Epub ahead of print. doi: 10.1038/s41587-022-01559-w

Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, Braxton A, Beuten J, Xia F, Niu Z, Hardison M, Person R, Bekheirnia MR, Leduc MS, Kirby A, Pham P, Scull J, Wang M, Ding Y, Plon SE, Lupski JR, Beaudet AL, Gibbs RA, Eng CM (2013) Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med 369: 1502-11. doi: 10.1056/NEJMoa1306555

Statements and Declarations

Funding

The somatic mosaicism research is funded by Baylor College of Medicine, by grant R01HD087292 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) to PS. RZ is supported by T32 GM07526-44 from the National Institute of General Medical Science (NIH/NIGMS). JRL is supported by the U.S. National Institute of Neurological Disorders and Stroke (NINDS) (R35NS105078). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Competing interests

JRL has stock ownership in 23andMe, is a paid consultant for Genomics International, and is a coinventor on multiple United States and European patents related to molecular diagnostics for inherited neuropathies, eye diseases, genomic disorders, and bacterial genomic fingerprinting. The Department of Molecular and Human Genetics at Baylor College of Medicine receives revenue from clinical genetic testing conducted at Baylor Genetics Laboratories.

Authors' contributions

RZ, HD, TG and PS conceived the project and designed the experiments. RZ, HD, TG, JRL, PL and PS performed the experiments and analyzed the data. RZ, HD, TG, JRL and PS wrote the manuscript. All authors read and approved the submitted version.

Availability of data and materials

The dataset supporting the conclusions of this article is included within the article (and its additional files).

Ethics approval

Our study was approved by Baylor College of Medicine Institutional Review Board (H-46683, H-41191).

Consent to participate

The deceased child's parents signed an informed consent for trio genome analysis as part of the research protocol. This research conformed with the principles of the Declaration of Helsinki.

Table 1. Distribution of VAFs among the potential *de novo* heterozygous and mosaic SNVs/InDels

a. Customized variant filtering pipeline

VAF range	Fanconi SNVs	Fanconi InDels	Fanconi All	Control SNVs	Control InDels	Control All
9.55% - 18.69%	384	8	392	8	3	11
18.7% - 37.41%	446	13	459	3	0	3
37.41 – 62.59%	45	5	50	34	6	40

b. DeepMosaic filtering pipeline

VAF range	Fanconi SNVs	Fanconi InDels	Fanconi All	Control SNVs	Control InDels	Control All
9.55% - 18.69%	413	9	422	35	7	42
18.7% - 37.4%	437	16	453	3	1	4
37.41 – 62.59%	44	5	49	34	6	40

SNV, single nucleotide variant; VAF, variant allele frequency

Table 2. Mutational signature analysis for *de novo* substitutions: Breakdown of the germline and somatic variants in the FA proband and control

		Fanconi Anemia patient				Control			
	Mutation type	**Transition (Ti)**	**Transversion (Tv)**	**Total**	**Ti/Tv**	**Transition (Ti)**	**Transversion (Tv)**	**Total**	**Ti/Tv**
Germline	SBS	27	40	67	0.68	22	13	35	1.69
	DBS	0	0	0	NA	0	0	0	NA
	CBS	2	0	2	NA	0	0	0	NA
	BBS	NA	NA	0	NA	NA	NA	0	NA
	Total	29	40	69	0.73	22	13	35	1.69
Somatic	SBS	278	521	799	0.53	4	3	7	1.33
	DBS	6	12	18	0.50	0	0	0	NA
	CBS	1	23	24	0.04	1	2	3	0.50
	BBS	NA	NA	3	NA	NA	NA	0	NA
	Total	285	556	844	0.51	5	5	10	1.00

BBS, Biallelic base substitution; CBS, Clustered-base substitution; DBS, Doublet-base substitution; SBS, Single-base substitution; NA, not applicable

Figure legends

Fig. 1. *De novo* variant allele frequency distribution

VAF density plots in the FA proband (red histogram) with two components of the bimodal distribution (red curve line) compared with the control subject (blue histogram and line) generated using the home-brew pipelines: (i) with manual IGV curation step (A), and (ii) with increased sensitivity (ploidy = 50 in GATK HC) and automated filtration using DeepMosaic (B). The lines denote the potential postzygotic mutational events (VAF <37.4%) and the potential germline cell events (VAF 37.41-62.6%).

Fig. 2. Genome-wide distribution of *de novo* substitution and examples of *de novo* substitution pattern in the FA patient

a. Rainfall plot of *de novo* substitutions.

b-e. Integrative Genomics Viewer (IGV) visualization of the proband (top) and parents (bottom) short read sequencing data at four loci, illustrating four different substitution patterns: b) Single-base substitution (SBS), c) double-base substitution (DBS), d) clustered-base substitution (CBS), e) biallelic-base substitution (BBS)

Fig. 3. Mutational pattern of *de novo* single base substitution.

a) 96 tri-nucleotide profile

b) Pan-nucleotide river plot with the extended context (2 bp)

Figure 4. Mutational single base substitution signature.

a) The top 20 known single base substitution (SBS) signatures ranked by the cosine similarity (from top to bottom).

b) The absolute contribution of the fitted SBS signatures.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Suppfig1.pdf
- Supplement2.7.23.docx
- SupplementaryTable2.csv