Molecular insights into the mechanism of nonrecurrent F8 structural variants: Full breakpoint characterization and bioinformatics of DNA elements implicated in the upmost severe phenotype in hemophilia A.

Miguel Martín Abelleyro1 | Claudia Pamela Radic1 | Vanina Daniela Marchione1 | Karen Waisman1 | Tomas Tetzlaff2 | Daniela Neme3 | Liliana Carmen Rossetti1 | Carlos Daniel De Brasi1,4

1Instituto de Medicina Experimental (IMEX, CONICET-Academia Nacional de Medicina), Buenos Aires, Argentina
2Universidad Nacional de General Sarmiento, Los Polvorines, Argentina
3Fundación de la Hemofilia Alfredo Pavlovsky, Buenos Aires, Argentina
4Instituto de Investigaciones Hematológicas Mariano R Castex (Academia Nacional de Medicina), Buenos Aires, Argentina

Correspondence
Carlos Daniel De Brasi, Molecular Genetics of Haemophilia Laboratory, Academia Nacional de Medicina, Pacheco de Melo 3081, Buenos Aires (1425), Argentina.
Email: cdebrasi@hematologia.anm.edu.ar

Funding information
National Research Council (CONICET), Grant/Award Number: PIP 2014-00045; Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT), Grant/Award Number: PICT 2016-0899

Abstract
Hemophilia A (HA) provides excellent models to analyze genotype-phenotype relationships and mutational mechanisms. NhF8lds’s breakpoints were characterized using case-specific DNA-tags, direct- or inverse-polymerase chain reaction amplification, and Sanger sequencing. DNA-break’s stimulators (n = 46), interspersed repeats, non-B-DNA, and secondary structures were analyzed around breakpoints versus null hypotheses (E-values) based on computer simulations and base-frequency probabilities. Nine of 18 (50%) severe-HA patients with nhF8lds developed inhibitors, 1/8 affecting one exon and 8/10 (80%) affecting multi-exons. NhF8lds range: 2–165 kb. Five (45%) nhF8lds involve F8-extragenic regions including three affecting vicinal genes (SMIM9 and BRCC3) but none shows an extra-phenotype not related to severe-HA. The contingency analysis of recombinogenic motifs at nhF8ld breakpoints indicated a significant involvement of several DNA-break stimulator elements. Most nhF8ld’s breakpoint junctions showed microhomologies (1–7 bp). Three (27%) nhF8lds show complexities at the breakpoints: an 8-bp inverted-insertion, and the remnant two, inverted- and direct-insertions (46–68 bp) supporting replicative models microhomology-mediated break-induced replication/Fork Stalling and Template Switching. The remnant eight (73%) nhF8lds may support nonhomologous end joining/microhomology-mediated end joining models. Our study suggests the involvement of the retroposition machinery (e.g., Jurka-targets, Alu-elements, long interspersed nuclear elements, long terminal repeats), microhomologies, and secondary structures at breakpoints playing significant roles in the origin of the upmost severe phenotype in HA.

KEYWORDS
bioinformatics, F8, HEMA, large deletions, structural variants
1 | INTRODUCTION

Hemophilia A (HA; MIM##306700), the commonest X-linked coagulopathy with a frequency of 1:5,000 human males worldwide, is caused by pathogenic variants in the factor VIII gene (F8). Among these variants, HA is characterized by a prevalent involvement of an unusual type of structural variant (SV) unassociated with gain or loss of DNA sequences (copy number variations [CNVs]), the intron 22 inversion (Inv22; Lakich, Kazazian, Antonarakis, & Gitschier, 1993, Naylor, Brinke, Hassock, Green, & Giannelli, 1993), and the intron 1 inversion (Inv1: Bagnall, Waseem, Green, & Giannelli, 2002) causing 40–50% and 2–5% of severe cases worldwide, respectively (Antonarakis et al., 1995; Rossetti, Goodeve, Larripa, & De Brasi, 2004). The high frequencies of both Inv22 (and associated deletions, duplications) and Inv1 in severe-HA are fueled by the mechanism of nonallelic homologous recombination (NAHR), between 10 and 1 kb-long duplicons, int22h and int1h, respectively (Abelleyro et al., 2016; Bagnall et al., 2002; De Brasi & Bowen, 2008; Naylor et al., 1995).

The remaining cases of SVs causing 8–15% of severe-HA implicate nonrecurrent family-specific large deletions and complex rearrangements (noninversion F8-SVs) involving one or more F8-exons with or without upstream or downstream extragenic regions (Graw et al., 2005). The phenotypic features associated with these large F8-deletions, particularly those involving more than one exon, show the uppermost clinical severity including the highest predisposition to develop inhibitory antibodies against the therapeutic FVIII (odds ratio [OR; 95% confidence interval]: 7.07 [2.20–22.71], p < .0005; Marchione et al., 2017).

For experimental reasons, despite the growing involvement of massive parallel sequencing techniques in the clinical practice, data from SVs and their breakpoints still remain scarce as compared with other variant’s types. This relative deficit in clinical evidence weakens the investigation of the mechanism associated with the origin of SVs in the human genome. Hemophilia has been focused on as a model for understanding numerous characteristics of human genetics, such as genotype–phenotype relationship, mutational turnover and X-linked population genetics, and as a platform for developing genotyping methods for both, small variants and SVs.

Despite the limited number of fully characterized SVs in the human genome, several molecular mechanisms have been conceived to understand their origin. Among these models, there were described those associated with repair of DNA breaks, such as the mentioned NAHR, perhaps including homologous recombination (recombination between repeats homologs with lower nucleotide sequence identity [<90%], e.g., Rossetti et al., 2004), nonhomologous end joining (NHEJ; Moore & Haber, 1996), formerly designated as nonhomologous recombination (Woods-Samuels, Kazazian, & Antonarakis, 1991), and those associated with DNA synthesis, such as break-induced replication (BIR; Smith, Llorente, & Symington, 2007), microhomology-mediated BIR (MMBIR; Hastings, Ira, & Lupski, 2009) and Fork Stalling and Template Switching (FoSTeS; Lee, Carvalho, & Lupski, 2007). Albeit the cause that triggers the DNA repair mechanism are still under an active research, the literature has described an important number of DNA sequence motifs that stimulate the incidence of double strand breaks (DSBs) with potential to initiate an event of recombination (Abeyesinghe, Chuzhanova, Krawczak, Ball, & Cooper, 2003; Table ST1). Some molecular rearrangements in the F8 have been associated with SINE-type retrotransposons (short interspersed nuclear elements) such as Alu, and partial segments of long interspersed nuclear elements (LINE; Rossetti et al., 2004; Van de Water, Williams, Ockelford, & Browett, 1998; Vidal, Farssac, Tusell, Puig, & Gallardo, 2002).

Breakpoints’ characterization of SVs is crucial to gain insights into the possible molecular mechanism involved. To achieve this goal, we performed gap-polymerase chain reaction (PCR) by bridging the deletion breakpoints after chasing them by testing case-specific DNA sequence tags. Gap-PCR constitutes the gold standard approach for genotyping genomic deletions in hemizygous and heterozygous state representing the ideal route to perform carrier diagnosis in female relatives at risk in the affected family.

Perhaps due to the scarce literature addressing the contingency of DNA motifs and repetitive elements on intervals around the SV breakpoints as potential promoters for DSBs on statistical basis, the actual involvement of any of these elements into the molecular mechanism of gross rearrangements has been difficult to be assessed (Thomas et al., 2016; You et al., 2013).

Consequently, aimed to gain further insights into the mechanism generating nonhomologous/nonrecurrent SVs resulting in large deletions involving the F8, this study presents a full sequence characterization of a series of 11 novel SVs, and an extensive bioinformatics and statistical analysis of breakpoints measuring the likelihood to be involved of 46 DSB-stimulator motifs, repetitive elements, non-B-DNA, and secondary structure taking advantage of original null hypotheses based on unbiased probabilistic calculations and computer simulations.

2 | MATERIALS AND METHODS

2.1 | Patients and samples

This project was approved by the CEIANM (Ethics Committee of the Institutes of the National Academy of Medicine; approval date: 06/13/2012). Peripheral blood samples were collected from all participants, patients with severe HA (a comprehensive historical series of 336 families with 415 affected individuals, 18 of them showing SV and negative PCR signals for the inversions), and a written informed consent was obtained in all cases. High-quality genomic DNA samples (gDNA) from participants were extracted from 5 ml of frozen peripheral blood leukocytes, anticoagulated with ethylenediaminetetraacetic acid 5%, by the method of salting-out and alcoholic precipitation (Lahiri & Nurnberger, 1991).

2.2 | Primary detection of large F8 deletions

At first line, recurrent F8 inversions (i.e., Inv22 and Inv1) were genotyped by IS-PCR as it was modified (Abelleyro et al., 2016). For severe HA cases...
that tested negative for inversions, an amplification scheme of 38 PCR products, including all exonic coding sequences (26 exons), promoter regions and intronic sequences associated with splicing (Rossetti et al., 2013) was carried out to screen and detect the primary evidence of large F8 deletions. A large deletion was defined as a consistent absence of one or more contiguous amplification products in the F8.

2.3 | Chasing large F8 deletion breakpoints’ junctions

To chase, amplify, and characterize the breakpoints of large deletions, the strategy differed for the following two categories: deletions involving the upmost 5’ or 3’ amplicons of the F8 (a) and those limited by intragenic sequences not involving the extremes (b).

For deletions type (a) a bipartition approach was applied. This approach involves first, the interrogation of a specifically designed sequence tagged site (STS) at 100 kb upstream or downstream the F8 according to the SV. A positive PCR amplification of an extragenic STS allowed designing a second STS halfway (e.g., reducing from 100 to 50 kb the region of breakpoint uncertainty). Successive interrogation of halfway located STS narrows the region of uncertainty in the most effective manner. Finally, the molecular length of the region encompassing the breakpoint junctions permitted LD–PCR amplification (Table ST2).

Deletions type (b) were classified in deletions with positive F8 PCR products at both sides placed at a distance that can be directly amplified by long range–PCR (LR–PCR; Table ST3), or those that still need a bipartition approach because the estimated size of the PCR product largely exceeds those amplifiable by LR–PCR (Table ST2).

2.4 | Sequence tagged site amplification

STS PCR amplification analysis was carried out on 300 ng of patient’s gDNA applying standard reagents and thermal-cycling conditions. PCR product sizes, primers, and annealing temperatures are listed in Table ST2.

2.5 | Long-range polymerase chain reaction amplification

LR–PCR amplifications ranging 1.7–6.8 kb were performed in a volume of 25 μl with an input of 500 ng of gDNA using 0.75 U of KAPA LongRange DNA Polymerase (Kapa Biosystems) with primers located by the inferred 5’ and 3’ breakpoints (Table ST3) in a Thermal-Cycler BIOER (China) applying the conditions recommended by the enzyme manufacturer.

2.6 | Restriction analysis of LR–PCR products

To narrow the uncertainty of the SV’s breakpoints as much as possible, eight LR–PCR products were subjected to a restriction fragment length polymorphism analysis (Table ST3). Under conditions recommended by the manufacturer, 5 μl of LR–PCR products were digested with some restriction enzymes (TaqI, BclI, Swol, Smol, EcoRI, BamHI, HindIII, NcoI) in 15 μl of final volume. After estimating the molecular size of the restriction fragments associated with a specific LR–PCR product, the obtained fragment list was compared with a number of possible fragment lists generated by different hypotheses using the corresponding sequences on the reference genome.

2.7 | Standard PCR amplification of breakpoints

PCR amplifications ranging 400–1,559 bp were performed in a volume of 25 μl with a gDNA input of 300 ng using 1.5 U of Go Taq DNA Polymerase (Promega, Argentina) and primers designed to characterize the 5’ and 3’ breakpoints by Sanger sequencing (Table ST4) in a Thermal-Cycler Biometra (Germany) applying standard reagents and thermal-cycling conditions.

2.8 | Characterization of deletion’s breakpoints

The DNA sequence of each deletion breakpoint junction was characterized by bidirectional Sanger sequencing from the corresponding standard size PCR products bridging the 5’ and 3’ involved regions. DNA sequences of all 11 cases of F8 SV’s breakpoints were deposited on the GenBank (accession numbers: MH124155, MH124156, MH124157, MH124158, MH124159, MH124160, MH124161, MH124162, MH124163, MH124164, and MH124165).

2.9 | Bioinformatics, computer simulations, and statistical analysis

A comprehensive literature mining looking for DNA sequences associated with DSB stimulation resulted in 46 DNA motifs ranging 3–18 bp (Table ST1). The contingency analysis of these 46 DNA motifs on each deletion breakpoint was performed as follows. Each characterized breakpoint was mapped to the Human Reference Genome (GRCh38) on an interval of approximately 1.6 Mb on Xq28 including the F8 (NC_000023.11: 153364063-154955215 downloaded 07-Apr-2015 from the NCBI website [URL: www.ncbi.nlm.nih.gov/]). Centered in each breakpoint, two DNA segments per large deletion (5’, 3’), whose size (range: 10–24 bp) depending on the length of the specific motif searched (N: 3–18 bp) calculated as follows: 2×(N+2); except for Jurka sequences with 50-bp intervals (due to its specific mechanism of action in retrotransposition of Alu elements), were analyzed by aligning them with each DNA motif, as potential DNA break stimulators. In the case of complex rearrangements, cases #2 and #10, in which the large deletion combine with an ectopic inverted- and/or direct-insertion, six breakpoints (three 5’ and three 3’) were evaluated.

To assess unbiased null hypotheses (Ho) to estimate the significance of a positive motif matching on the deletion, breakpoints’ intervals were
compared to motif matching on equal intervals of breakpoints selected at random in the same region of Xq28 (1.6 Mb). Associated with these Ho, expected values (E-values) were calculated by two different approaches. The first approach is based on computer simulations: each DNA motif was searched in motif-specific intervals around 240 artificial breakpoints selected at random on the 1.6 Mb Xq28 region. An E-value for each DNA motif was calculated according to the observed contingency at the random breakpoints. The randomness of the selection of those artificial breakpoints was guaranteed by use of an aleatory number generator accessed online (URL: www.random.org/) and the Xq28 genomic region limits indicated above.

A probabilistic approach consisted in a mathematical calculation of the expected number of matches in a random sequence of independently occurring bases taking into account the relative abundance of each base on the target interval on Xq28 (A and T, 28% each, and C and G, 22%) and the nucleotide ambiguity in some motifs (from unambiguous motif consensus, e.g., DNA polymerase alpha frameshift hotspot 1 "TCCCCC"; to extremely ambiguous, e.g., Vertebrate topoisomerase II consensus "RNYYNCCNNGYNGKTNYNY" associated with 4,194,304 different sequences; ST1). The bioinformatic analysis was mainly achieved using the SeqBuilder and MegAlign programs (LaserGene DNA Star), the ClustalW algorithm accessed online on the EMBL-EBI website (URL: www.ebi.ac.uk/Tools/msa/clustalw2/) and the BLAST algorithm accessed online from the NCBI website (URL: blast.ncbi.nlm.nih.gov/Blast.cgi/). The Repeat Masker algorithm and Dfam were accessed online (URL: www.dfam.org/) to identify interspersed repetitive elements (e.g., Alu, LINE L1, etc.). The modeling of secondary structure was depicted using the mfold Web Server online (URL: unafold.rna.albany.edu/?q=mfold), and the analysis of non B-DNA sequences (e.g., Z-DNA regions) was achieved by use of non-B-DNA motif search tool (nBMST) online (URL: nonb-abcc.ncbi.nih.gov/apps/nBMST/default/), and confirmed using RepeatAround (URL: portugene.com/repeataround.html/) and QGRS mapper (URL: bioinformatics.ramapo.edu/QGRS/analyze.php/).

The correlation of E-values calculated by both approaches (i.e., Simulations vs. Probabilistic) was analyzed by Spearman statistics. Contingency analyses of repetitive elements, non-B-DNA, and secondary structures were evaluated by estimating OR with 95% CI and the Fisher’s exact test.

3 | RESULTS

3.1 | F8 structural variants: Molecular and clinical characteristics

Among our comprehensive series of 336 unrelated families affected by severe HA, 163 show small F8 variants, 161 (48%) show recurrent SVs caused by NAHR (including 157 Inv22 and four Inv1), one nonrecurrent SV including a F8 exon 4–10 deletion mediated by AluSx homeologous recombination (Rossetti et al., 2004), and 11 (3%) F8 SVs, including large deletions and complex rearrangements not involving paired-recombination between homologs. The latter group represents the main focus of our analysis. All F8 large deletions represent paradigmatic null-variants establishing very strong evidence (PVS1) to classify them as pathogenic following the recommendations of the American College of Medical Genetics (ACMG; Richards et al., 2015). All 12 nonrecurrent SVs cases (i.e., 11 SV analyzed here and the one mediated by AluSx recombination previously reported; Rossetti et al., 2004) describe F8 partial deletions in which the FVIII:C activity levels (<1 IU/dl) closely correlate with a severe clinical phenotype. This group of nonrecurrent F8 partial deletions represents about 3.6% of all unrelated family probands affected by severe HA. A global figure of 50% of patients with F8 deletions (9/18) developed inhibitory antibodies against therapeutic FVIII (inhibitor). Among them, the subclass of F8 deletions involving more than an exon (multi-exon deletions) associated with 80% (8/10) of patients with inhibitor, whilst deletions involving a single exon (single-exon deletions) showed 12.5% (1/8) of patients developing inhibitor.

The DNA sequence from breakpoint junctions associated with the F8 SVs were sought using case-specific sets of STS designed by the bipartition principle, PCR amplified, and fully characterized at the sequence level (Table 1).

The molecular loss of DNA sequences associated with the presented F8 SVs range 2–165 kb (Table 1, Figure 1). Five out of 11 (45%) reported F8 partial deletions involved F8-extragenic regions either upstream (cases #1, #2, and #3; 27%) or downstream (cases #4 and #11; 18%) and six (58%) involved only intragenic F8 regions (cases #5, #6, #7, #8, #9, and #10; Figure 1). In all three groups of nonrecurrent SVs, no cases showed a clinically appreciable extra-phenotype beyond those related with severe HA to be mentioned, even though case #1 involved small integral membrane protein 9 (SMIM9) and cases #4 and #11, BRCA1/BRCA2-containing complex subunit 3 (BRCC3), which are Xq28 vicinal genes upstream and downstream the F8, respectively (Figure SF1).

Four (36%) out of all 11 characterized F8 SVs preserve the F8 open reading frame (ORF; Table 1). Case #3 Del(<e1;2), only affects one and five exons, respectively (Table 1). Beyond previous speculations, the frequency of inhibitor development in patients from the 12 families with nonrecurrent F8 SVs affected by in-frame large deletions (40%, 2/5) does not significantly differ from those affected by frameshift large deletions (54%, 7/13).

3.2 | Bioinformatics and statistical analysis of F8 SV breakpoints

Two out of the 11 (18%) described F8 SVs showed complex insertions at the gross deletion breakpoint junctions: case #2 (a 68-bp inverted insertion, a 6-bp insertion, and a 46-bp direct insertion) and case #10 (two inverted insertions of 61 and 67 bp), while case #7 showed an 8-bp insertion associated with an apparently unsuspected origin (Figure SF1). Two of the three juxtapositions of nonallelic sequences associated
| Case No. | Genomic DNA HGVS notationa (g.) | Coding DNA HGVS notationa (c.) | GenBank accession | Deletion size (bp)b | F8 ORF | References |
|----------|---------------------------------|-------------------------------|-------------------|-------------------|--------|------------|
| #1       | NC_000023.11:g.155011270_155089415del | NC_000023.11(NM_000132.3):c.-66863_143+11140del | MH124164         | 78,146           | No     | This study |
| #2       | NC_000023.11:g.[155001773_155025606delinsMH124158:g.55_165;155001755_155001753del] | NC_000023.11(NM_000132.3):c.-3055_144-2181delinsMH124158:g.55_165;144-2156_144.2154del | MH124158         | 23,834           | No     | This study |
| #3       | NC_000023.11:g.15502658_155024725del | NC_000023.11(NM_000132.3):c.-2174_107del | MH124163         | 2,068             | Conserved | Abellero et al, 2018 |
| #4       | NC_000023.11:g.154832958_154998257delinsAG | NC_000023.11(NM_000132.3):c.266-1163_*4636delinsAG | MH124156         | 165,300           | No     | This study |
| #5       | NC_000023.11:g.154989955_154968576del | NC_000023.11(NM_000132.3):c.602-2651_1009+754del | MH124162         | 21,380           | Conserved | This study |
| #6       | NC_000023.11:g.154982096_154986630del | NC_000023.11(NM_000132.3):c.670+606_787+2590del | MH124165         | 4,535            | Conserved | This study |
| #7       | NC_000023.11:g.154952229_154975239delinsTGTATCCCA | NC_000023.11(NM_000132.3):c.788-5684_1903+164delinsTGTATCCCA | MH124160         | 23,011           | Conserved | This study |
| #8       | NC_000023.11:g.154902624_154964956del | NC_000023.11(NM_000132.3):c.1443+1012_5999-457del | MH124157         | 62,333           | No     | This study |
| #9       | NC_000023.11:g.154959233_154962379del | NC_000023.11(NM_000132.3):c.1030_8_1537+1841del | MH124161         | 3,147            | No     | This study |
| #10      | NC_000023.11:g.154965646_154960991delinsMH124155:g.117_264 | NC_000023.11(NM_000132.3):c.1537+82_1752+499delinsMH124155:g.117_264 | MH124155         | 4,528            | No     | This study |
| #11      | NC_000023.11:g.154822704_154861819del | NC_000023.11(NM_000132.3):c.6620_*14892del | MH124159         | 39,116           | No     | This study |

Abbreviations: ORF, open reading frame

aHGVS notation: Variant nomenclature recommended by the Human Genome Variation Society (den Dunnen et al., 2016)
bDeletion size represents the larger deletion of the rearrangement, do not show insertion or small deletions which could be included in the corresponding nomenclature
with complex cases #2 and #10 showed microhomologies (range: 2–7 bp). Among the remaining eight cases, large deletions without insertions at the breakpoints, five (63%) showed junction microhomologies (range: 1–5 bp; Figure SF1, Table 3).

To gain insights into the origin of the 11 characterized F8 SVs, an exhaustive research of DNA motifs that may act as stimulators or drivers for recombination were performed. In particular, the search focused on elements within motif-specific intervals (range: 10–50 bp) around breakpoints, which were reported as promoters or stimulators for DNA breaks triggering an incorrect DNA repair/DNA replication leading to nonallelic recombination. These DNA motifs (n = 46), mined from the literature, constitute a heterogeneous group of sequences with different lengths (range: 3–18 bp) and positions with ambiguities that represent from a single unambiguous sequence (e.g., X-element Escherichia coli; 8 bp-long) to more than 4 million sequences (i.e., Vertebrate topoisomerase II consensus, 18 bp; Table ST1). To estimate the significance of at least a positive match on a breakpoint interval, each recombinogenic motif was associated with two specific hypotheses: one based on computer Simulations (Sim) involving artificial breakpoints randomly selected on the F8 ± 700 kb region on Xq28 and other named Probabilistic (Prob) calculated by extending the independent base count frequency, on the same Xq28 region, through all base positions on the motif length (Section 2).

Table 2 shows the contingency of DNA motifs found within each motif-specific interval around the breakpoints from the analyzed 11 cases with F8 nonrecurrent SVs. Among all 22 DNA motifs showing at least a positive match at the breakpoint intervals, null-hypothesis E-values estimated from Sim vs. Prob approaches showed a highly significant correlation (Spearman’s r coefficient 0.88, p < .0001; Figure SF2). This correlation recognize three outliers—Consensus SAR 2, SAR 4, and Murine Parvovirus recombination hotspot—showing Sim E-values > 100, > 30, and > 20 fold higher than expected by Prob, respectively (Figure SF2). This evidence may reflect an adaptive role of these DNA elements in the genome dynamics or at least on the target region Xq28.

Twenty-two out of 46 (48%) searched DNA motifs were detected with at least a positive match around the breakpoints of F8 SVs, which is somehow expected for DNA motifs associated with Sim E-values higher than 0.1 (n = 12), rare for DNA motifs with Sim E-values ranging 0.1–0.05 (n = 2) and it is significantly unexpected for the remnant 8 DNA motifs (17%) with Sim E-values ranging 0.0013–0.046 (Table 2). Six out of the 11 (54%) F8 SV cases showed positive motif matches within at least a breakpoint interval in the latter group of significantly unexpected DNA motifs (cases #2, #3, #4, #7, #9, #10) and this figure extended to 64% (adding case #1) when Sim E-values up to 0.071 were considered (Table 2).

In addition, all 11 F8 SVs showed the involvement of repetitive elements (i.e., long terminal repeats [LTRs], n = 6); long and short interspersed nuclear elements, such as LINE L1 [n = 7], L2 [n = 4], and SINE such as Alu subfamilies [n = 6] on at least one 50-bp interval centered on the breakpoint (Table 3). Among all 30 breakpoints analyzed in the 11 cases, 23 (76.7%) presented repetitive elements, as compared with 52.9% obtained on 240 random Xq28 breakpoints in an experiment in silico. These figures indicate a significant increase in the risk of being associated with a SV breakpoint with an OR (95% CI) of 2.92 (1.21–7.07; p = .018), when a repetitive...
TABLE 2  Contingency of recombinogenic sequences/motifs in clinical cases compared with null-hypotheses designed by computer simulations and probabilistic estimations

| Recombinogenic sequence \ Case no. | #1 | #2 | #3 | #4 | #5 | #6 | #7 | #8 | #9 | #10 | #11 | E-value\* | E-value\* | Motif sequence\* (interval window bp\*) |
|----------------------------------|----|----|----|----|----|----|----|----|----|----|----|-----|-----|---------------------------------|
| DNA polymerase alpha/beta frameshift hotspot 2 | 1 | - | - | - | - | - | - | - | - | - | - | 0.013 | 0.009 | TGGNGT (16) |
| Ig heavy chain class switch repeat 4 | - | - | - | 1 | - | - | - | - | - | - | - | 0.017 | 0.007 | TGGGG (14) |
| Ig heavy chain class switch repeat 2 | - | - | - | 1 | - | - | - | - | - | - | - | 0.021 | 0.007 |GGGCT (14) |
| Jurk a Sequence 5 | - | - | - | - | 1 | - | - | - | - | - | - | 0.042 | 0.017 |TAAAG (50) |
| Jurk a Sequence 6 | - | - | - | - | - | - | 3 | 1 | - | - | - | 0.042 | 0.017 |CTAAAA (50) |
| Consensus SAR 2 | - | - | - | - | - | - | 3 | - | - | - | - | 0.042 | 0.000 |TTWTTWTTWT (24) |
| Jurk a Sequence 4 | - | - | - | - | - | - | 1 | - | - | - | - | 0.046 | 0.017 |TTGA (50) |
| Murine parvo virus recombination hotspot | - | - | - | - | - | - | - | - | - | - | - | 0.002 | 0.002 |CTWTTY (16) |
| Jurk a Sequence 9 | 2 | - | - | - | - | - | - | - | - | - | - | 0.071 | 0.022 |TTTAA (50) |
| Vaccinia topoisomerase I consensus | 1 | - | - | - | - | 1 | - | - | - | - | - | 0.088 | 0.019 |YCCCV (14) |
| DNA polymerase arrest site | 1 | 1 | - | 1 | 1 | - | - | - | - | - | - | 0.104 | 0.017 | WGGAG (14) |
| Consensus SAR 4 | - | 1 | - | - | - | - | - | 8 | - | - | - | 0.133 | 0.004 |TWWTDTWAMP (24) |
| Ig heavy chain class switch repeat 5 | - | - | 1 | - | - | - | - | 1 | - | - | - | 0.188 | 0.034 | TGA (12) |
| Deletion hotspot consensus | 1 | - | - | 1 | - | 1 | - | 2 | - | - | - | 0.204 | 0.042 | TGRKRKM (16) |
| Murine MHC deletion hotspot | 1 | 2 | - | 1 | 2 | 2 | - | - | - | 2 | - | 0.417 | 0.061 |CAGR (12) |
| DNA polymerase alpha pause site core sequence 1 | 3 | - | - | 1 | 2 | 1 | 1 | - | - | - | - | 0.475 | 0.108 | CAT (10) |
| DNA polymerase alpha pause site core sequence 3 | - | 3 | 1 | 2 | 1 | 1 | - | - | - | - | - | 0.525 | 0.170 | GTY (10) |
| DNA polymerase alpha pause site core sequence 2 | - | 1 | - | - | - | - | - | - | - | - | - | 0.535 | 0.108 | GYT (10) |
| Vertebrate/plant topoisomerase I consensus cleavage | - | - | 1 | 2 | 1 | - | 1 | 2 | 1 | 1 | 1 | 0.583 | 0.138 | CAT (10) |
| Vertebrate/plant topoisomerase I consensus cleavage | - | 1 | 1 | 2 | 2 | - | 2 | 1 | 2 | 1 | 1 | 0.742 | 0.246 | GYT (10) |
| Vertebrate/plant topoisomerase I consensus cleavage | - | 1 | - | 3 | 3 | 3 | 2 | 1 | 2 | 5 | 2 | 4 | 0.904 | 0.314 | RAT (10) |
| Vertebrate/plant topoisomerase I consensus cleavage | 4 | - | 2 | - | 1 | 4 | 4 | 1 | - | 1 | - | 1.013 | 0.246 | CAY (10) |

\*Estimated in silico based on 240 simulations random sites on Xq28 and motif-specific intervals (10–50 bp)

\*Theoretical or Probabilistic estimations based on probabilities of base counts on Xq28 and motif-specific intervals

\*W=A o G; Y=C o T; K=G o T; M=A o C; S=G o C; D=A o G o T; B=G o T o C; H=A o C o T; V=C o G o A; N=G o T o A o C

\*Motif-specific intervals was computed according the motif-specific length (N) bp as (2 x (N + 2)), except for Jurk a hexanucleotides with intervals of 50 bp due to its mechanism of action (Jurk a, 1997, Proceedings of the National Academy of Sciences of the United States of America, 94: 1872–7)

Element (LINE, SINE, or LTR), strictly a part of it, is detected within the 50-bp breakpoint interval. Disaggregated analysis of SINEs indicated that Alu elements associated with a nonsignificant OR of 2.4 (0.87–6.37), while LINES (L1/L2) participated even less than expected showing a nonsignificant OR of 0.82 (0.38–1.81). On the other hand, LTR showed a highly significant OR of 9.75 (2.92–32.61; p = .0007) when its contingency in nonhomologous F8 SVs was compared with the Ho (240 random breakpoints).

The analysis of non-B-DNA structures in our series showed an uneven distribution characterized by a breakpoint with three matches (case #1, 5’ breakpoint; involving cruciform DNA, a hairpin loop, and a segment of Z-DNA) and four breakpoints with either a hairpin, or a cruciform or a triplex DNA, indicating a global involvement of 16.7% (5/30; Table 3). Comparison of these figures with the Sim Ho, showing a proportion of 5.8% (14/240), indicated that vicinal non-B-DNA structures imply a significant risk for suffering F8 SV breakpoints with an OR of 3.23 (1.07–9.72; p = .042).

In contrast, the detection of secondary DNA structures based on ΔG analysis by the Mfold algorithm resulted in a nonsignificant association with SV breakpoints (76.7%, n = 30) as compared with the Ho (67.9%, n = 240), OR 1.55 (0.64–3.77; Table 3, Figure SF3). The ΔG values of the most stable secondary structure on the breakpoints’ 50-bp-interval resulted heterogeneous showing an average of ~3.4 kcal but a wide range from ~11.35, in case #1 5’ breakpoint,
to −0.30 in case #9 3’ breakpoint (Figure SF3). Interestingly, most of the 30 breakpoints were found near or on the limits between single and double strand DNA structure (Figure SF3).

### 4 | DISCUSSION

The F8 genotype of a series of severe HA patients with large deletions has been fully characterized.

The inhibitor development in all patients with HA showing nonrecurrent SV (n = 18; large deletions ranging 1–24 F8 exons) has been estimated in 50% but this figure does not represent the significantly higher inhibitor risk associated with multi-exon deletions (80%) versus those associated with single-exon deletions (12.5%). These observations from our population closely agree with the inhibitor risks estimated from the literature (Garagiola, Palla, & McVey & Lee, 2008; Moore & Haber, 1996; Szostak, Orr-Weaver, Rothstein, & Stahl, 1983), and those associated with DNA synthesis, that is, BIR (McEachern & Haber, 2006) and its derivative MMBIR and FoSTeS (Hastings et al., 2009; Lee et al., 2007).

Aimed to estimate the origins of the molecular mechanism involved in our series of F8 SVs, we classified the mechanisms according to the scar leftover at the breakpoint junctions. First, we classified two groups, recurrent and nonrecurrent. In recurrent SVs, HR plays the central role involving two homologous sequences (>95% identical; formerly described ranging from a minimum of 15–400 bp; Perez et al., 2005; Szostak et al., 1983) in which a specific mechanism can be recognized by the finding of a crossover. Molecular scars at the breakpoint junctions representing an intrinsic obstacle to assign one of them properly (e.g., MMEJ–MMBIR and MMBIR–FoSTeS).

In this group of F8 recurrent SVs in our historical series of patients with severe HA, large F8 inversions (48%), including the prevalent Inv22 and the Inv1 (1.2%), show the highest prevalence in severe HA in close agreement with the literature and F8 pathogenic variants databases (Gouw et al., 2012).

On the other hand, nonrecurrent SVs involves different molecular models in which some of them could leave the same or similar molecular scars at the breakpoint junctions representing an intrinsic obstacle to assign one of them properly (e.g., MMEJ–MMBIR or MMBIR–FoSTeS).

Some of these nonrecurrent rearrangements are driven by HR-like between homologs associated with 70–90% of sequence identity (homologous recombination), often involving repetitive elements such as SINEs or parts of LINEs building a genomic architecture that catalyzed the rearrangement (Mezárd, Pompon, & Nicolas, 1992; Rossetti et al., 2004; Stankiewicz et al., 2003). This latter mechanism can be recognized by the finding of a crossover (CO) with or without an associated event of gene conversion (GC).

| Case no. | SV nickname | Junction microhomology | Interspersed repeats | Non-B-DNA structures | Secondary structure |
|----------|-------------|------------------------|----------------------|----------------------|--------------------|
|          |             | 5’ to 3’                | 5’                   | 3’                   |                    |
| #1       | Del(<e1;78)| TCT                    | LINE L2              | LINE L1              | Cruciform, Hairpin, Z-DNA | Yes Yes |
| #2       | Del/Ins(>e1;24)| – (+7)⁴, TG, CG | AluY, -,  AluSx, AluJb | –, –, –, –, –, –, Hairpin | Yes, Yes, Yes, Yes, Yes, Yes |
| #3       | Del(<e1;2) | T                      | LINE L2              | –                    | Cruciform          | Yes No |
| #4       | Del(e3;>165)| – (+2)⁴               | -                    | LINE L1              | –                  | Yes No |
| #5       | Del(e5;e7;21) | GA                    | LINE L1              | LTR                  | –                  | Yes Yes |
| #6       | Del(e6;5) | GA                     | -                    | AluSg                | –                  | Mirror repeat Yes Yes |
| #7       | Del/Ins (e7e12;23) | – (+9)⁴          | LINE L1              | –                    | –                  | – Yes |
| #8       | Del(e10e18;62) | TG                   | LINE L1              | LTR                  | Mirror repeat      | – No |
| #9       | Del(e10;3) | AATT                   | LINE L2              | AluJb                | –                  | – Yes |
| #10      | Del/Ins(e11;5) | TG, – (+0)⁴, ATAAATT | LTR, LINE L1, LTR  | LINE L1, LTR        | –, –, –, –         | Yes, Yes, No, No, Yes |
| #11      | Del(e24;>39)| – (+0)⁴               | -                    | LINE L2              | –                  | – Yes |

Abbreviations: LINE, long interspersed nuclear element; LTR, long terminal repeat; SV, structural variant

*Parenthesis (+N) indicates the length of an eventual insertion at the junction in bp (range: 1–9 bp), blunt ends without microhomology is indicated as – (+0)
inside two paired repetitive elements (e.g., CO and GC between AluSx within F8 intron 3 and intron 10 in a large deletion of exons 4–10, Rossetti et al., 2004). BIR is a similar mechanism requiring at least 15 bp of perfect sequence identity but it is associated with DNA synthesis (Liskay, Letsou, & Stachelek, 1987; McEachern & Haber, 2006), and due to the involvement of the replicative-repair polymerase Pol ι, BIR is error prone (Deem et al., 2011).

In contrast, NHEJ mechanism is easily recognized by the juxtaposition of nonhomologous DNA at junctions. NHEJ model involves the rejoining of two somewhat distant DSBs with loss of the encompassed genomic DNA, ranging from less than 1 kb to more than 100 kb, and the eventual involvement of repetitive elements could help in stabilizing the interaction of the breakpoint regions termed ectopic synapsis (Liu, Carvalho, Hastings, & Lupski, 2012; Woods-Samuels et al., 1991). MMEJ and MMBIR mechanisms have similar molecular scars at junctions, a microhomology of 2–10 bp, but MMBIR involves new DNA synthesis, perhaps encompassing short sequence tracks not associated with the references 5’ or 3’ regions, whilst MMEJ does not (Hastings et al., 2009; McVey & Lee, 2008; Woods-Samuels et al., 1991).

FoSTeS mechanism associates with stalls or pauses at the replication fork, phenomenon that tends to occurs at regions of genomic instability related to a complex genomic architecture, and it was initially proposed to explain complex rearrangements including deletions, duplications and direct or inverted insertions at junctions, which are difficult to be explained by other models (Lee et al., 2007). Similar to MMBIR, FoSTeS may associate with microhomology bites between the sequences involved making difficult the assignment of a single model only based on the sequence scars leftover and the analysis of their reference sequence counterparts. Although some authors refer FoSTeS/MMBIR as an unresolved couple (Zhang et al., 2009), we would rather classify MMBIR to describe rearrangements with only one template change and FoSTeS when more than one template switching are found.

Even though the molecular scar provides hints to choose the potential involved molecular mechanism, it does not provide information about the origin of the DNA breakpoint. It is clear that breakpoints may cluster close to DNA elements, motifs or structural features at the corresponding targets on the reference sequences. For example, retrotransposable DNA elements such as LINE and Alu generate ruptures to initiate transposition into elsewhere on the genome, besides its potential participation on the recombination event via pairing by sequence identity (Brouha et al., 2003; Deininger & Batzer, 1999; Kazazian 2004; Rüdiger, Gregersen, & Kielland-Brand, 1995). Interestingly, repetitive elements were found specifically enriched at deletion breakpoints but not at duplication breakpoints (Vissers et al., 2009).

Non-B-DNA, such as tetruplex, cruciform, bent DNA, Z-DNA, and secondary DNA structures, represent another source of DNA breakpoint stimulators as they were detected surrounding breakpoints in different genomic rearrangements (Bacolla & Wells, 2004; Wang & Vasquez, 2014). The literature showed a number of DNA motifs associated with DNA breakpoints, for example, cleavage sites of Topoisomerase I or II enzymes (Been, Burgess, & Champoux, 1984; Spitzner & Muller, 1988), DNA polymerase arrest sites (Weaver & DePamphilis, 1982), Alu/LINE specific retrotransposition target sequences (Jurka, 1997), and many others, all listed in Table ST1 (n = 46) highlighting the 22 motifs involved in our set of SV breakpoints (Table 2).

Taking into account the above introduced literature evidence; we attempted to classify our set of 11 nonrecurrent F8 SVs identifying the most likely mechanism according to the molecular features found at the breakpoint junctions and the related reference sequences. Accordingly, cases #3 and #11 do not left a molecular scar at the breakpoint junction to permit its association with a replication mechanism model, and neither microhomology, characteristics that allow clustering them into classical NHEJ. Furthermore, necessary DNA breaks at both deletion breakpoints to support the choice of NHEJ model may be stimulated, in case #3, by Jurka sequence 6 (on the 5′ break) and Murine parvovirus recombination hotspot (on the 3’ break) or/and by LINE L2, non-B-DNA and a highly stable secondary structure at the 5′ breakpoint site. In case #11, even though there is no evidence of the involvement of significantly unexpected recombinogenic motifs, it shows a LINE L2 on the 3′ breakpoint and a secondary structure at the 5′ rupture site as possible elements involved in generating DNA breaks.

Similar to the model of NHEJ in which we cannot find evidence of newly synthesized DNA but presenting 2–5 bp of microhomology, cases #1, #5, #6, #8, and #9 may be classified in the MMEJ repair model. Case #1 was chosen as a paradigm of the models of mechanisms of recombination not involving DNA synthesis and, consequently, to show a typical graphic representation of the group (NHEJ/MMEJ; Figure 2a). The research of DNA breaks stimulator motifs has indicated recombinogenic sequences at both sides in cases #1 and #9 including some Jurka and SAR sequences, while cases #1, #6, and #8 show non-B-DNA in one breakpoint and all MMEJ cases show repetitive elements (Alus, LINEs and LTRs) or secondary structures at least in one 5′ and 3′ breakpoint site.

On the other hand, all four remaining cases (#2, #4, #7, and #10) show traces of DNA synthesis (scars) at breakpoints and consequently allowed us to classify them in the groups associated with DNA replication but lacking relatively long/intermediate homologous sequences (MMBIR/FoSTeS). The couple of complex rearrangements cases (#2 and #10) are clearly associated to a FoSTeS repair mechanism as they involve more than one direct or inverted insertion associated with the deletion as well as scars on their boundaries. The required collapses of the DNA synthesis or/and replication fork stalling could be associated with recombinogenic motifs, repetitive elements and, notably, with secondary structures observed in each region where the new strand that is being synthesized invades other region of genomic DNA.

It is interesting to realize that case #7 may be adequately explained either by MMBIR or FoSTeS models both involving DNA synthesis of a short tract of eight bases representing an inverse complementary insertion defining the junction scar (Figure 2). This SV shows three bases of microhomology and the collapse that may explain both mechanisms could be associated with Jurka and Vaccinia topoisoasme I consensus sequences at the 5′ and 3′ breakpoints, as
FIGURE 2  Schematic representation of likely molecular mechanisms originating two representative rearrangements of our series. Two specific examples are shown: case #1 (a) and #7 (b and c). (a) The model of microhomology-mediated end joining provides the most parsimonious mechanism supporting most cases in our SV series including case #1. Breaks defining the 5′ and 3′ edges of a structurally simple large deletion trigger the classical route of DNA reparation where a microhomology, in this case #1 "TCT", may stabilize the gap repair. With similar characteristics, in some cases that present one bp (e.g., case #3) or no microhomology (case #11), the model of nonhomologous end joining provides a simple explanation to understand the events of DNA breaks occurrence and gap repair. (b and c) Although models of microhomology-mediated break-induced replication (MMBIR) (b) and Fork Stalling and Template Switching (FoSTeS) (c) both provide relatively simple explanations for the common occurrence of large deletions associated with direct or inverted insertions at the breakpoint junctions, such as case #7 discussed in detail, they differ in the mechanistic events involved (e.g., DNA breaks, replication fork stalling and collapse). (b) MMBIR. Breakpoints (black triangles) start the recombinogenic events and the secondary structure formed on 5′ (stem-loop structure shown in Figure SF3, case #7) and a microhomology of three bases "GTA" facilitates the reverse pairing on the same strand used as a template to synthesize a new DNA segment (dashed line) of eight bases, which create an inverted insertion ("TCCCATAA") at the breakpoint junction in case 7. The collapse of the new strand and the subsequent invasion of the 3′ breakpoint mediated by a microhomology of "TAA" and a final repair may complete the event. (c) FoSTeS. The collapse of the replication fork (black stars) stops the DNA synthesis on the leading strand allowing the invasion of the lagging strand by the three-nucleotide "GTA" and the synthesis of the eight-base inverted insertion. A new collapse stops the synthesis allowing a second template switching, 23 kb downstream on F8 intron 12 (case #7). SV, structural variant.
well as the presence of repetitive elements and secondary structure. In particular, considering the involvement of MMBIR model in case #7, the secondary structure detected on the 5’ extremity of the leading to the lagging strand, followed by the subsequent involvement of MMBIR model in case #7, may be generated by a first template switching, on F8 intron 6, from the leading to the lagging strand, followed by the subsequent inversion of the F8 intron 12, 23 kb apart, representing the second template switching (Figure 2c).

Our results about the molecular characteristics of the large F8 deletion breakpoints perfectly align with those formerly described in the literature about no-recurrent rearrangements, which have been explained by NHJEJ or MEMEJ model as the molecular mechanism involved in their origin (Woods-Samuels et al., 1991). Likewise, our results agree with other authors reporting a high participation of repetitive elements, either directly involved in the deletion mechanism, or indirectly, surrounding one or both breakpoint regions (Thomas et al., 2016; Van de Water et al., 1998; Vidal et al., 2002). Furthermore, the literature showed a number of fully characterized large deletions, whose molecular characteristics are very similar to what we have shown in this study that after being reanalyzed in view of the new corpus of theoretical and bioinformatics knowledge, it will surely result in similar conclusions about the molecular mechanisms possibly involved (You et al., 2013).

To the best of our knowledge, this study represents the first effort to integrate and systematize the molecular analysis of clinically relevant nonrecurrent SVs with a main focus posed on the origin event. The present study of nonrecurrent SVs affecting the F8 provides a number of molecular lessons including the prevalent participation of various elements of the retrotransposition machinery (e.g., Jurka sequences, Alu elements, LINEs, LTRs), common microhomologies at the breakpoint junctions and the involvement of relatively stable secondary structures exposing short tracks of potentially feeble single-stranded DNA by each event site.

Our findings point out the utility to extend the presented combination of bioinformatic and statistical analysis to gain molecular insights into the mechanism of nonhomologous/nonrecurrent SVs involved in other genetic disorders in humans.

ACKNOWLEDGMENTS

This study was supported by grants from the National Research Council (CONICET), National Agency for Scientific and Technological Promotion (ANPCyT), and the World Federation of Hemophilia. The authors thank Miguel de Tezanos Pinto (National Academy of Medicine and National Hemophilia Foundation) and Laura Primiani (National Hemophilia Foundation) for their help in different phases of the work.

ORCID

Miguel Martín Abellero [http://orcid.org/0000-0002-9961-9414]
Liliana Carmen Rossetti [http://orcid.org/0000-0002-4683-7899]

REFERENCES

Abellero, M. M., Rossetti, L. C., Curto, Mde. L., Radic, C. P., Marchione, V. D., & De Brasi, C. D. (2016). F8 intron 22 inversions and SNP rs73563631 in unrelated families with severe haemophilia A: Clinical features and gene testing implications. Thrombosis and Haemostasis, 115, 678–681. https://doi.org/10.1160/TH15-08-0643

Abyesinghe, S. S., Chuzhanova, N., Krawczak, M., Ball, E. V., & Cooper, D. N. (2003). Translocation and gross deletion breakpoints in human inherited disease and cancer I: Nucleotide composition and recombination-associated motifs. Human Mutation, 3, 229–244.

Antonarakis, S. E., Rossiter, J. P., Young, M., Horst, J., de Moerloose, P., Sommer, S. S., & Lnaba, H. (1995). FVIII gene inversions in severe haemophilia A: Results of an international consortium study. Blood, 86, 2206–2212.

Bacolla, A., & Wells, R. D. (2004). Non-B DNA conformations, genomic rearrangements, and human disease. Journal of Biological Chemistry, 279, 47411–47414.

Bagnall, R. D., Waseem, N., Green, P. M., & Giannelli, F. (2002). Recurrent inversion breaking intron 1 of the factor VIII gene is a frequent cause of severe Hemophilia A. Blood, 99, 168–174.

Been, M. D., Burgess, R. R., & Champoux, J. J. (1994). Nucleotide sequence preference at rat liver and wheat germ type 1 DNA topoisomerase breakage sites in duplex SV40 DNA. Nucleic Acids Research, 12, 3097–3114.

Brouha, B., Schustak, J., Badge, R. M., Lutz-Puggle, S., Farley, A. H., Moran, J. V., & Kazazian, H. H., Jr. (2003). Hot L1s account for the bulk of retrotransposition in the human population. Proceedings of the National Academy of Sciences of the United States of America, 100, 5280–5285.

De Brasi, C. D., & Bowen, D. J. (2008). Molecular characteristics of the intron 22 homologs of the coagulation factor VIII gene: An update. Journal of Thrombosis and Haemostasis, 6, 1822–1824. https://doi.org/10.1111/j.1538-7836.2008.03094.x

Deem, A., Keshthelyi, A., Blackgrove, T., Vayl, A., Coffey, B., Mathur, R., & Malkova, A. (2011). Break-induced replication is highly inaccurate. PLoS Biology, 9(2):e1000594. https://doi.org/10.1371/journal.pbio.1000594

Deininger, P. L., & Batzer, M. A. (1999). Alu repeats and human disease. Molecular Genetics and Metabolism, 67, 183–193.

den Dunnen, J. T., Dalgleish, R., Maglott, D. R., Hart, R. K., Greenblatt, M. S., McGowan-Jordan, J., & Taschner, P. E. (2016). HGVS recommendations for the description of sequence variants: 2016 Update. Human Mutation, 37, 564–569. https://doi.org/10.1002/humu.22981

Garagioi, I., Palla, R., & Pevyandi, F. (2018). Risk factors for inhibitor development in severe hemophilia a. Thrombosis Research, 168, 20–27. https://doi.org/10.1016/j.thromres.2018.05.027

Gouw, S. C., van den Berg, H. M., Oldenburg, J., Astermark, J., de Groot, P. G., Margaglione, M., & van der Bom, J. G. (2012). F8 gene mutation type and inhibitor development in patients with severe hemophilia A: Systematic review and meta-analysis. Blood, 119, 2922–2934. https://doi.org/10.1182/blood-2011-09-379453

Graw, J., Brackmann, H. H., Oldenburg, J., Schneppenheim, R., Spannagl, M., & Schwaab, R. (2005). Haemophilia A: From mutation analysis to new therapies. Nature Reviews Genetics, 6, 488–501.

Hastings, P. J., Ira, G., & Lupski, J. R. (2009). A microhomology-mediated break-induced replication model for the origin of human copy number variation. PLoS Genetics, 5(1), 1–9. https://doi.org/10.1371/journal. pgen.1000327. e1000327

Jurka, J. (1997). Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proceedings of the National Academy of Sciences of the United States of America, 94, 1872–1877.

Kazazian, H. H., Jr. (2004). Mobile elements: Drivers of genome evolution. Science, 303, 1626–1632.

Lahiri, D. K., & Nurnberger, J. I., Jr. (1991). A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. Nucleic Acids Research, 19, 5444.
Lakich, D., Kazazian, H. H., Antonarakis, S. E., & Gitschier, J. (1993). Inversions disrupting the factor VIII gene are common cause of severe Haemophilia A. Nature Genetics, 5, 236–241.

Lee, J. A., Carvalho, C. M., & Lupski, J. R. (2007). A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell, 131, 1225–1247.

Liskay, R. M., Letsou, A., & Stachelk, J. L. (1987). Homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells. Genetics, 115, 161–167.

Liu, P., Carvalho, C. M., Hastings, P. J., & Lupski, J. R. (2012). Mechanisms for recurrent and complex genomic rearrangements. Current Opinion in Genetics & Development, 22, 211–220. https://doi.org/10.1016/j.gde.2012.02.012

Marchione, V. D., Zuccoli, J. R., Abelleyro, M. M., Radic, C. P., Neme, D., Candela, M., & Rossetti, L. C. (2017). A prevalent CTLA4 missense variant significantly associates with inhibitor development in Argentine patients with severe haemophilia A. Haemophilia, 23, e166–e169. https://doi.org/10.1111/hae.13194

McEachern, M. J., & Haber, J. E. (2006). Break-induced replication and recombinational telomere elongation in yeast. Annual Review of Biochemistry, 75, 111–135.

McVey, M., & Lee, S. E. (2008). MMEJ repair of double-strand breaks (director's cut): Deleted sequences and alternative endings. Trends in Genetics, 24, 529–538. https://doi.org/10.1016/j.tig.2008.08.007

Mezard, C., Pompon, D., & Nicolas, A. (1992). Recombination between similar but not identical DNA sequences during yeast transformation occurs within short stretches of identity. Cell, 70, 659–670.

Moore, J. K., & Haber, J. E. (1996). Cell cycle and genetic requirements of two pathways of non-homologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. Molecular and Cellular Biology, 16, 2164–2173.

Naylor, J., Brinne, A., Hassock, S., Green, P. M., & Giannelli, F. (1993). Characteristic mRNA abnormality found in half the patients with severe Haemophilia A is due to large inversions. Human Molecular Genetics, 2, 1773–1778.

Naylor, J., Buck, D., Green, P. M., Williamson, H., Bentley, D., & Giannelli, F. (1995). Investigation of the factor VIII intron 22 repeated region (int22h) and the associated inversion junctions. Human Molecular Genetics, 4, 1217–1224.

Perez, C., Guyot, V., Cabaniols, J. P., Gouble, A., Micheaux, B., Smith, J., & Duchateau, P. (2005). Factors affecting double-strand break-induced homologous recombination in mammalian cells. Biotechniques, 39, 109–115.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., & Gastier-Foster, J. ACMG Laboratory Quality Assurance Committee. (2015). Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in Medicine, 17, 405–424. https://doi.org/10.1038/gim.2015.30

Rossetti, L. C., Goodeve, A., Larripa, I. B., & De Brasi, C. D. (2004). Homologous recombination between AluSx-sequences as a cause of hemophilia. Human Mutation, 24, 440.

Rossetti, L. C., Szurkalo, I., Radic, C. P., Abelleyro, M. M., Primiani, L., Neme, D., & De Brasi, C. D. (2013). Factor VIII genotype characterization of haemophilia A affected patients with transient and permanent inhibitors: A comprehensive Argentine study of inhibitor risks. Haemophilia, 19, 511–518. https://doi.org/10.1111/hae.12105

Rüdiger, N. S., Gregersen, N., & Kielland-Brand, M. C. (1995). One short well conserved region of Alu-sequences is involved in human gene rearrangements and has homology with procaryotic chi. Nucleic Acids Research, 23, 256–260.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.