ORIGINAL ARTICLE

rpth
research & practice
in thrombosis & haemostasis

Detection of mosaics in hemophilia A by deep Ion Torrent sequencing and droplet digital PCR

Eric Manderstedt MSc[1] | Rosanna Nilsson MSc[1] | Rolf Ljung MD, PhD[2] | Christina Lind-Halldén PhD[1] | Jan Astermark MD, PhD[3] | Christer Halldén PhD[1]

[1]Department of Environmental Science and Bioscience, Kristianstad University, Kristianstad, Sweden

[2]Department of Clinical Sciences-Pediatrics, Lund and Malmö Center for Thrombosis and Hemostasis, Lund University, Skåne University Hospital, Malmö, Sweden

[3]Department for Hematology, Oncology and Radiation Physics, Center for Thrombosis and Hemostasis, Skåne University Hospital, Malmö, Sweden

Correspondence
Christer Halldén, Department of Environmental Science and Bioscience, Kristianstad University, S-291 88 Kristianstad, Sweden.
Email: christer.hallden@hkr.se

Funding information
Vetenskapsrådet, Grant/Award Number: 2015-02957

Handling Editor: Dr Pantep Angchaisuksiri

Abstract

Background: The occurrence of mosaicism in hemophilia A (HA) has been investigated in several studies using different detection methods.

Objectives: To characterize and compare the ability of AmpliSeq/Ion Torrent sequencing and droplet digital polymerase chain reaction (ddPCR) for mosaic detection in HA.

Methods: Ion Torrent sequencing and ddPCR were used to analyze 20 healthy males and 16 mothers of sporadic HA patients.

Results: An error-rate map over all coding positions and all positions reported as mutated in the *F8*-specific mutation database was produced. The sequencing produced a mean read depth of >1500X where >97% of positions were covered by >100 reads. Higher error frequencies were observed in positions with A or T as reference allele and in positions surrounded on both sides with C or G. Seventeen of 9319 positions had a mean substitution error frequency >1%. The ability to identify low-level mosaicism was determined primarily by read depth and error rate of each specific position. Limit of detection (LOD) was <1% for 97% of positions with substitutions and 90% of indel positions. The positions with LOD >1% require repeated testing and mononucleotide repeats with more than four repeat units need an alternative analysis strategy. Mosaicism was detected in 1 of 16 mothers and confirmed using ddPCR.

Conclusions: Deep sequencing using an AmpliSeq/Ion Torrent strategy allows for simultaneous identification of disease-causing mutations in patients and mosaicism in mothers. ddPCR has high sensitivity but is hampered by the need for mutation-specific design.

KEYWORDS
factor VIII, hemophilia A, high-throughput nucleotide sequencing, mosaicism, polymerase chain reaction

Essentials

- Detection of mosaics in hemophilia A (HA) has not previously used single-molecule techniques.
- Next generation sequencing and digital PCR (dPCR) were compared for mosaic detection in HA.
- Deep resequencing produced an error-rate map over *F8* and identified low-level mosaicism.
- dPCR has higher sensitivity than next generation sequencing but needs mutation-specific design.

1 | INTRODUCTION

Hemophilia A (HA) is an X-linked recessive disease caused by mutations in *F8*. The mutations cause deficiency or dysfunction of the factor VIII (FVIII) protein. *F8* has 26 exons that code for a 9-kb transcript.[1,2] More than 2000 mutations identified throughout *F8* have been associated with HA and are listed in a locus-specific variant database (www.factorviii-db.org). The most common type of mutation is an inversion involving repeated sequences in intron 22, detected in approximately 45% of all severe cases. Nonsense, deletion, and duplication mutations are also present more frequently in severe cases. In mild/moderate cases the dominant type of mutation is a missense variant.[1] *F8* is invariable with respect to common variation; only one nonsynonymous variant is present at a frequency >1% in Europeans (rs1800291, D1241E, with a minor allele frequency of 18%) and with an additional four nonsynonymous variants present in other populations.[3] Analysis of the genetic variation in *F8* using data from the 1000 Genomes Project revealed >3000 presumably benign rare variants and 18 variants previously associated with HA.[4] The "My Life, Our Future" initiative analyzed 3000 patients and discovered 924 unique variants, of which 285 were novel.[5] The authors detected novel variants continually throughout the project, indicating that additional variants most likely remain to be discovered. They also detected incidental variants unlikely to cause disease, including 11 variants previously associated with HA. Both reports clearly demonstrate the difficulty of associating specific variants with pathogenicity.

Mosaicism of somatic and germ cells is a condition in which cells within the same person have more than one genotype as a result of mutations acquired during cell development. The fraction of mutated cells depends on when during cell development the mutation occurs.[6] Somatic mosaicism has been described in >100 diseases (http://mosaicbase.cbi.pku.edu.cn/), but also in patients with cancer. Mosaic mutations can be difficult to detect as the frequency of the mutation can be very low.[7] Approximately 30%-50% of new cases of hemophilia are sporadic; that is, hemophilia is previously unknown in the family.[2] Accurate carrier diagnosis of a mother of a sporadic case is important and may have implications for a decision to perform prenatal diagnosis in future pregnancies. Conventional Sanger sequencing is insensitive and may not reveal mosaicism unless the rare variant (mutation) is present in >10%-15% of the cells. Thus, highly accurate and sensitive methods are of great importance.[8]

To date, only a small number of reports have investigated the existence of mosaicism in patients with HA. Early studies reported on single cases of mosaics for point mutations,[8-10] large deletions,[11-13]

and intron 22 inversions.[14] Later studies aimed to determine the frequency of mosaicism in different populations. Leuer et al[15] investigated a total of 61 families with sporadic severe HA.[15] They detected mosaicism in 8 of 32 families (25%) with substitutions, whereas no mosaics were observed in 13 families with small indels, nor in 16 families with intron 22 inversions. They used southern blot analysis of long-distance polymerase chain reaction (PCR) fragments for the inversion analysis and Sanger sequencing and allele-specific PCR for the analysis of point mutations. Tizzano et al[16] failed to detect mosaicism by southern blot analysis in 53 mothers of sons with intron 22 inversions.[16] In a study of HA in China, Lu et al[17] detected mosaic mothers in 3 of 10 families (30%) with intron 22 inversions and in 3 of 26 families (11%) with point mutations.[17] They used long-distance PCR in combination with the AccuCopy technique[18] for inversion analysis and SNaPshot analysis for point mutation analysis.

Next-generation sequencing (NGS) is based on the simultaneous sequencing of many single molecules analyzed in parallel. The use of targeted resequencing allows for deep resequencing, presenting data on the exact number of sequenced molecules and the exact percentages of variants. This makes it possible to detect mosaicism even at low frequencies. Ion Torrent (Thermo Fisher Scientific, Waltham, MA, USA) is an NGS platform based on semiconductor chip technology.[19,20] Using the AmpliSeq strategy, targeted resequencing is performed through the simultaneous amplification of many amplicons in a multiplex PCR. Recently, a number of NGS-based studies have resequenced gene panels of different sizes associated with inherited bleeding disorders.[21-23] Depending on the panel size and the selected read depth, this type of mutation screening can also identify the presence of mosaic mutations.

Droplet digital PCR (ddPCR) is an alternative to NGS that also analyzes many single molecules in parallel. ddPCR is a quantitative method based on water-in-oil emulsion droplet technology and uses TaqMan genotyping assays with limiting dilution in end-point PCR. Poisson statistics are subsequently used to make an absolute quantification. By amplifying single molecules using end-point PCR, the signal in each droplet is derived from one or a low number of molecules. As a result, even rare alleles are present at a high frequency in each specific reaction and can be detected against a background of wild-type alleles.[24]

The aims of the present study were to produce an error rate map over all coding positions and all positions reported as mutated in the *F8*-specific database and to investigate the properties of Ion Torrent sequencing and ddPCR for mosaic detection in HA. For this purpose, 16 selected noncarrier mothers with substitutions and indels from a Swedish HA population were analyzed. The study used and

compared AmpliSeq-based Ion Torrent sequencing and ddPCR for the detection of mosaics.

2 | MATERIALS AND METHODS

2.1 | Study populations

The study population consisted of 20 healthy male individuals who were used in the basic characterization of reference and alternative alleles. Mosaic analysis was performed on 16 unrelated Swedish families, each represented by a sporadic HA index case and a noncarrier mother according to previous mutation analysis.[25,26] These families were selected based on the mutation type of the index patient, eight with substitution mutations and eight with indel mutations. The families were selected from the families described by Mårtensson et al,[26] and their details were available to us at Malmö Hemophilia Center. Genomic DNA was isolated from blood collected in EDTA using the QIAamp DNA Blood kit (Qiagen, Hilden, Germany), and DNA concentrations were initially determined by PicoGreen fluorometry (Molecular Probes, Eugene, OR, USA). An *RNase P* TaqMan assay in combination with ddPCR was subsequently used for a more careful determination of concentration. This study was approved by the Regional Ethical Review Board in Lund, and written informed consent was obtained from all subjects.

2.2 | Ion Torrent sequencing

The primer set used in this study was obtained from Ion AmpliSeq Designer (http://www.ampliseq.com). A gene panel encompassing 123 primer sets distributed into two pools and targeting the exonic positions of *F8*, *F9*, and *VWF* was designed and used to extract data on reference and alternative allele frequencies for *F8* and was also used for mosaic detection (Table S1). In total, 43 577 bases were covered by the panel. PCR amplification was performed using the Ion AmpliSeq Library Kit 2.0 according to the manufacturer (Life Technologies, Rockville, MD, USA). All samples were barcoded individually using Ion Xpress barcodes, and clean-up was performed using Agencourt Ampure XP beads (Beckman Coulter, Indianapolis, IN, USA). Amplicon concentration and quality were determined by using capillary electrophoresis and High Sensitivity NGS Fragment Analysis Kit 1-6000 bp (Advanced Analytical Technologies, Orangeburg, NY, USA) on a Fragment Analyzer (Advanced Analytical Technologies). Emulsion PCR was then performed on the Ion OneTouch 2 System (Life Technologies) according to the manufacturer's instructions. After emulsion PCR, the template-positive ion sphere particles were recovered using Dynabeads MyOne Streptavidin C1 beads (Life Technologies). The sequencing was 400-base reads with a total of 850 flows using the default flow order and was performed with a 316v2 chip on an Ion PGM sequencer with Ion PGM Hi-Q sequencing kit (Life Technologies). A more detailed description of the sequencing procedure can be found in the paper by Manderstedt et al.[27] Artificial mixtures containing approximately 1% mutant DNA against a wild-type background (in total 30 ng of DNA) were prepared for three different substitution mutations.

2.3 | Sequencing data analysis

The sequences were aligned against the hg19 human reference sequence using the Torrent Suite Software 5.0.2 (Thermo Fisher Scientific). Variant calling was performed using the Torrent Variant Caller version 5.0.2 with the recommended parameters for AmpliSeq libraries. Identified variants were annotated using the Variant Effect Predictor.[28] The BAM files were used in downstream analysis by investigating all positions for each sample using mpileup from SAMtools. Base calls and read depth from both strands were extracted from the mpileup output. The reference and alternative allele/indel frequencies were determined for each position. The number of reads in forward and reverse strand and their quotient (strand bias) were also determined for each position. The number of unique genome equivalents sequenced for the different disease-causing mutations was estimated using random sampling with replacement. Data from all positions with variants associated with HA were extracted from the *F8*-specific mutation database (www.factorviii-db. org). Unique control groups were generated for each mutation by combining data from all mutation-negative samples. To detect mosaic mutations, Fisher's exact test was used to compare the mutation frequencies of the noncarrier mother and the control group. The limit of detection (LOD) for each mutated position was determined by calculating the average of the alternative allele frequency in the control group and adding three standard deviations.

2.4 | Droplet digital PCR

ddPCR used a QX100 ddPCR system from Bio-Rad (Bio-Rad, Hercules, CA, USA) and TaqMan genotyping to quantify the mutant alleles. TaqMan systems (Table S2) were designed using RealTimeDesign Software, (BioSearch Technology; https://www.biosearchtech.com/display.aspx?pageid=54), checked for cross-hybridization using NCBI Primer–BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) and ordered from DNA Technology A/S (Risskov, Denmark). The 20-μL ddPCR reaction mixture contained template DNA, 1× Supermix (Bio-Rad), 900 nM of each primer and 250 nM of each probe. The reaction mixture was mixed with 70 μL of droplet generation oil in a DG8 disposable droplet generator cartridge using a QX100 Droplet Generator (Bio-Rad). The generated droplets were transferred to a 96-well PCR plate and heat sealed using pierceable foil. PCR amplification was performed using the following conditions: incubation at 95°C for 10 minutes followed by 40 cycles of incubation at 94°C for 30 seconds and the optimal annealing temperatures for 60 seconds, with a final incubation at 98°C for 10 minutes; ramp rate was 2.5°C per second. Droplets were then counted in a QX100 droplet reader, and the data were analyzed using

QuantaSoft software (Bio-Rad). Artificial mixtures containing varying concentrations of mutant and wild-type DNA, as well as nontemplate controls, were prepared for LOD determination. The mixtures contained a decreasing frequency of the mutant allele (100%, 50%, 10%, 1%, 0.1%, 0.01%, and 0%) against a wild-type background (in total 30 ng of DNA) and were analyzed in replicates of five. To obtain an adequate number of positive droplets, an additional 16 replicates of the 0.01% and 0% artificial mixtures were analyzed.

3 | RESULTS

3.1 | Characteristics of Ion Torrent sequencing

F8 was analyzed for variants in 20 male individuals using AmpliSeq-based Ion Torrent sequencing. The system presented an average read depth of 1515X for the complete coding sequence, with considerable variation in read depth between both individuals and amplicons. Only a few positions showed a strand bias of >95%. To describe the ability of the Ion Torrent system to discriminate between low-level alternative (mosaic) alleles and sequencing errors, an error rate map was produced by calculating reference and alternative allele/indel frequencies for each position. All bases in the complete coding sequence (9059 bases; GRch37.p13) adding five nucleotides to all exon boundaries (260 bases) were evaluated for their alternative allele frequencies. These data were then summarized according to the reference base of the interrogated positions. Figure 1 shows box plots of the alternative allele frequencies for all positions, subdivided into positions where the reference alleles were either C or G or A or T. The 3880 positions with C or G as reference alleles showed a median error frequency of 0.00030 with an interquartile range (IQR) of 0.00033. There was a total of 166 outliers where five positions had mean error frequencies >1%. The 5439 positions with A or T as reference alleles showed a median error frequency of 0.00075 with an IQR of 0.00077. There was a total of 126 outliers where 12 positions had mean error frequencies >1%. Thus, most positions with A or T as reference alleles had slightly higher error frequencies compared to positions with C or G (Figure 1B).

When the interrogated positions were instead evaluated based on their surrounding bases, a different pattern emerged. Figure 2 shows the mean error frequencies in four different situations. The most common situation was when A or T preceded the interrogated position, which was then followed by A or T. This occurred for 3185 positions, which showed a median error frequency of 0.00029 and an IQR of 0.00038. The least common situation was when C or G preceded the interrogated position, which was then followed by C or G. This occurred for 1626 positions, which showed a median error frequency of 0.00094 and an IQR of 0.00080. The two other situations were intermediate both regarding their actual numbers and regarding their median error frequencies. The 17 positions with a mean error frequency >1% were highly enriched in positions that were surrounded on both sides with C or G (Figure S1). To further investigate these 17 erroneous positions, they were plotted against
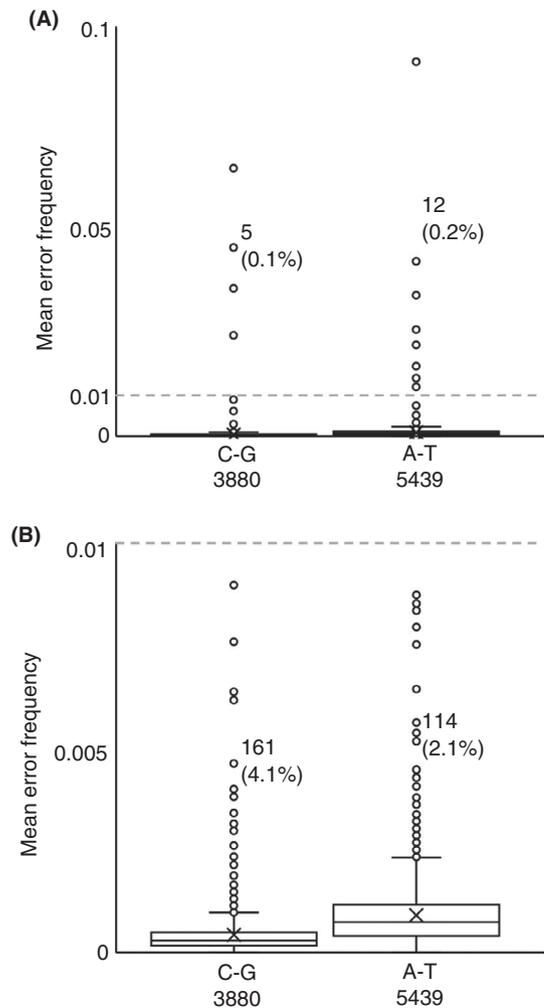
FIGURE 1 Mean alternative allele frequencies for the interrogated positions. Reference base is given as either C or G or A or T and their mean error frequencies are given using two different scales: A. <0.1 mean error frequency and B. <0.01 mean error frequency

the read depth and strand bias of all positions (Figure 3). No obvious correlation to read depth or strand bias was found. However, 4 of the 17 positions were found in regions with overlapping amplicons. To investigate the indel error frequencies the complete sequence was analyzed for the number of mononucleotide repeats of different sizes and then their mean error frequencies were determined (Figure 4). Although the actual numbers of repeats with large numbers of repeat units were low, the longer repeats were very error prone, giving rise to high indel error frequencies.

To investigate if the mutated positions showed the same pattern with respect to error rates as all of the coding positions, data were extracted from all positions reported as mutated in the *F8*-specific mutation database (www.factorviii-db.org). These positions were then investigated for alternative alleles and the LOD for each position was calculated by adding three SDs to the average alternative allele/indel frequencies. Alternative alleles and indels were analyzed separately. This analysis used a mean read depth of 1606X and a coverage corresponding to > 30X for 99.7% and > 100X for 98.5% of positions. The
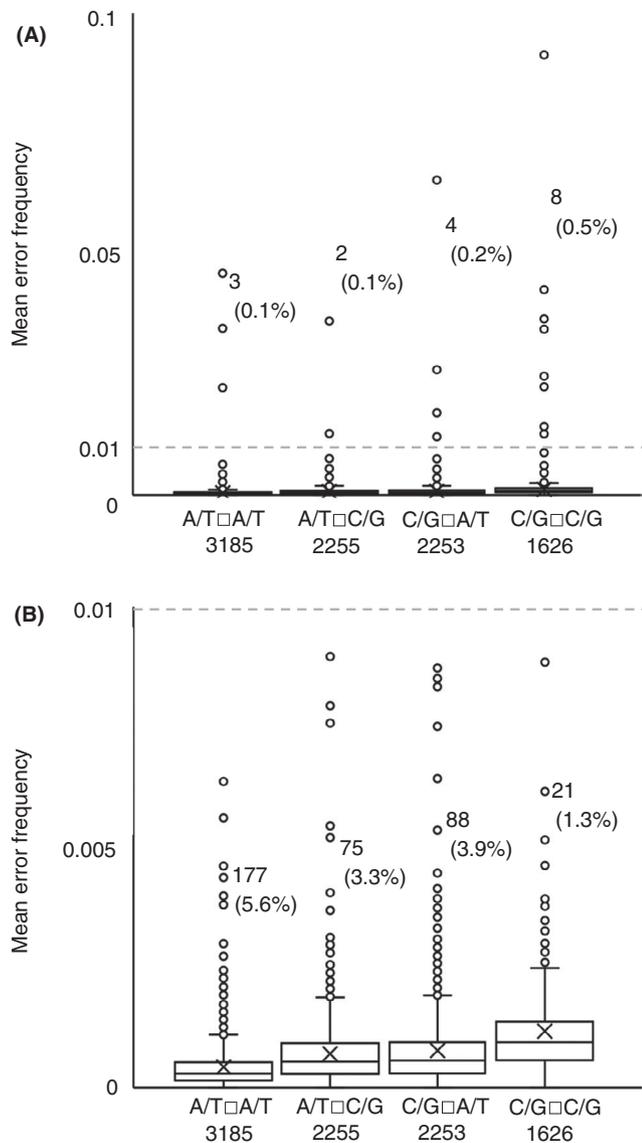
FIGURE 2 Mean alternative allele frequencies for the interrogated positions according to their surrounding bases. Mean error frequencies are given for four combinations of bases A/T or C/G before and after the interrogated positions. A. <0.1 mean error frequency and B. <0.01 mean error frequency

LOD was calculated for all 1125 unique positions with substitutions (Figure 5A and 5C). Most positions showed low LOD: 925 positions with a LOD < 0.5% (82% of positions), 1096 positions with LOD < 1% (97% of positions). Of the remaining 29 positions, all but one had a LOD < 2.5%. A single position had a LOD of 6.6%. Positions with a higher LOD showed a weak tendency to cluster along the chromosome. Positions with indels were analyzed in the same way, with a mean read depth of 1443X and a coverage of > 30X for 99.6% and > 100X for 97.2% of positions. The LOD was calculated for the 450 positions (Figure 5B and 5D). The majority of all positions showed a low LOD: 368 positions with LOD < 0.5% (82% of positions) and 404 positions with LOD < 1% (90% of positions). Of the remaining 46 positions, 29 had a LOD < 5%. The remaining 17 positions had an average LOD of 17% and varied between 6% and 72%.

Discrimination between low-level mosaic alleles and errors is determined primarily by read depth and error rate of each specific position. To describe the influence of varying read depth on the LOD, these two parameters were compared (Figure S2A and B). The read depth varied widely, showing a weak correlation to increased LOD for both alternative alleles and indels. As strand bias is not independent to the read depth but can add further complexity to the error estimates, this factor was also compared with the LOD (Figure S2C and D). The strand bias varied widely, showing a weak correlation to increased LOD for indels. These correlations were noted in the overall data pattern and were not derived from formal testing.

The errors observed in each specific position during Ion Torrent sequencing may be both random and systematic in nature. To investigate the relationship between random and systematic errors, two replicates were made of several samples. After sequencing and determination of the error frequencies for all positions the results of the two data sets were plotted against each other (Figure 6). As expected, substitutions generally showed a lower error level and most of the variation in the error level seemed to be random in nature. There was a total of 29 positions that were more error prone with a LOD > 1% representing approximately 3% of all positions (marked in red in Figure 6A). This tendency was more pronounced for indels, but also in this case a major part of the error variation seemed random in nature. For indels there was a total of 10 positions with a LOD > 9.7% and a total of 46 positions had a LOD > 1%. These represented approximately 10% of all positions and are marked in red in Figure 6B.

3.2 | Mosaic detection using Ion Torrent sequencing

To confirm the ability of the AmpliSeq/Ion Torrent system to detect low-level mosaic mutations in *F8*, 1% artificial mixtures were made by mixing DNA from each of three patients with DNA from a wild-type control. The samples containing approximately 1% artificial mixtures of the mutations c.1834C > T, c.5878C > T and c.5393C > T were then compared with a control group consisting of 20 wild-type samples. The number of reads of the mixtures and the wild-type controls were compared using Fisher's exact test to determine if there was a significant difference between the frequencies of the mutant alleles in patients and controls. All three ~ 1% mixtures were detected at highly significant levels (Table 1).

F8 was analyzed for the presence of mosaicism in a total of 16 noncarrier mothers or grandmothers of patients with sporadic HA (Table 2). All DNA samples were represented by between 8000 and 16 000 genome equivalents (GEs) in the initial PCR amplification. Since the sequenced molecules were sampled randomly from the pools of PCR products, the total number of unique GEs sequenced could be calculated from the read depths of the mutated positions. This estimate was made by random sampling with replacement and was in all but one case > 2000 unique GEs. Fisher's exact test was then used to determine if the differences observed between the mutation frequencies for the respective mother and the control groups were significant. One mother showed a significant difference for a

(A)

(B)
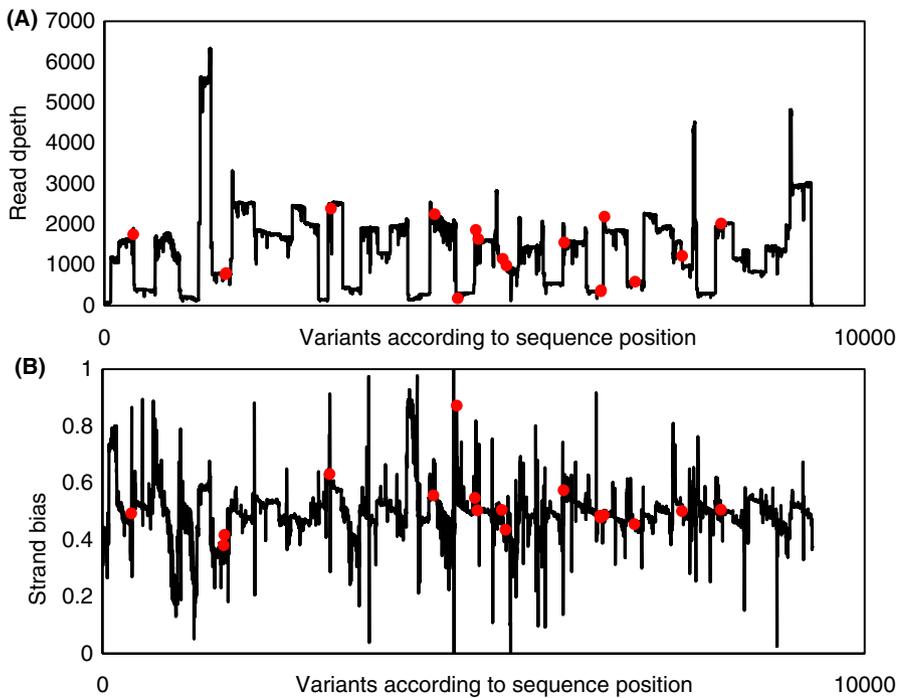
FIGURE 3 Read depth and strand bias of all interrogated positions. The 17 positions with an error frequency > 1% are marked in red
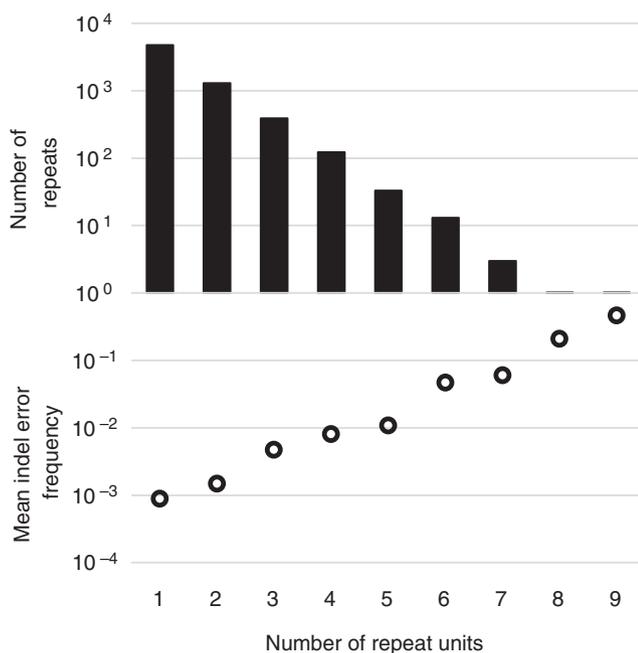
FIGURE 4 Number of repeat units and mean indel error frequency as a function of mononucleotide repeat size

c.805A > T mutation (8% frequency of the mutant allele; odds ratio, 42; *P* value ≤3.10^{-16}). This mutation showed limited frequency variation between the individuals in the control group.

3.3 | Mosaic detection and validation using droplet digital PCR

To investigate the properties of ddPCR for rare allele (mosaic) detection, artificial mixtures were produced by mixing mutant and wild-type DNA samples in varying proportions. Samples with 100%, 50%, 10%, 1%, 0.1%, 0.01%, and 0% mutant DNA against a background of wild-type DNA were prepared for seven of the mutations analyzed above: c.805A > T, c.1834C > T, c.5393C > T, c.5878C > T, c.209delTTGT, c.6469delAA, and c.2738insT (Table 2). All mixtures were analyzed in replicates of five, except for samples containing 0.01% and 0% mutant DNA, which were analyzed in replicates of 21. The LOD varied slightly for the different systems but was in all cases < 0.1% and even lower for some systems. Given a mutation such as c.209delTTGT, where the TaqMan probe can identify the mutant allele with high specificity and total absence of cross hybridization to the wild-type allele, the detection level would be limited almost exclusively by the number of fragments analyzed. Assuming the use of 20 000 GE/reaction and the analysis of 10 reactions, this would result in a detection level of < 0.01%. Other mutations showed low levels of cross hybridization but despite this could still detect the rare allele at a level of < 0.1%. The level of mosaicism detected by ddPCR was similar to the level detected by Ion Torrent sequencing for the c.805A > T mutation, NGS = 7.8%; ddPCR = 7.2% (Figure S3). The generally much lower LOD of ddPCR failed to detect additional cases of mosaicism in the remaining mothers who were analyzed.

4 | DISCUSSION

The present study was performed to characterize the ability of the AmpliSeq/Ion Torrent sequencing strategy to detect low-level mutations for the detection of mosaics, as well as to compare Ion Torrent sequencing and ddPCR for mosaic detection.

Ion Torrent sequencing and ddPCR have a fundamental similarity regarding frequency determination of rare alleles. Both systems rely on emulsion PCR-based analysis of single molecules:

FIGURE 5 Limit of detection (LOD) for all variants annotated as mutations in the *F8*-specific mutation database (www. factorviii-db.org). Cumulative distribution of LOD values for alternative alleles (A) and indels (B). The means of the variant frequencies underlying the LOD values are given in gray. LOD values as a function of *F8* sequence position for alternative alleles (C) and indels (D). LOD = 1% is marked by a dashed line
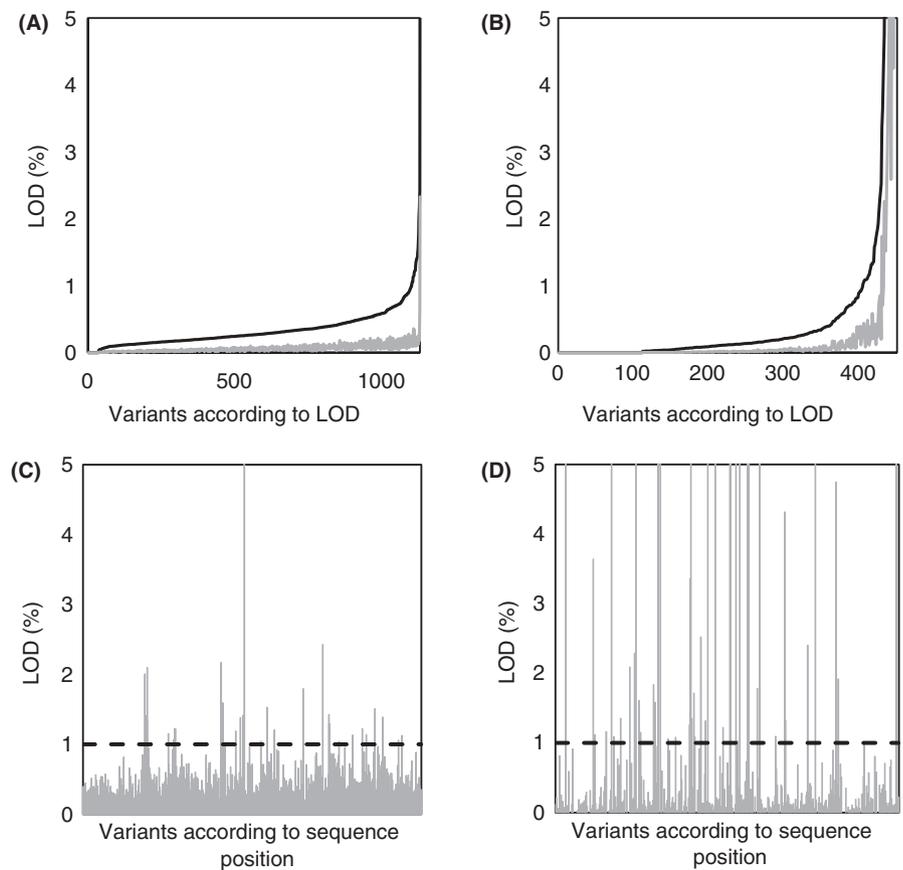
FIGURE 6 Correlation between error frequencies for replicates. (A) alternative alleles and (B) indels. Positions with a LOD > 1% are marked in red
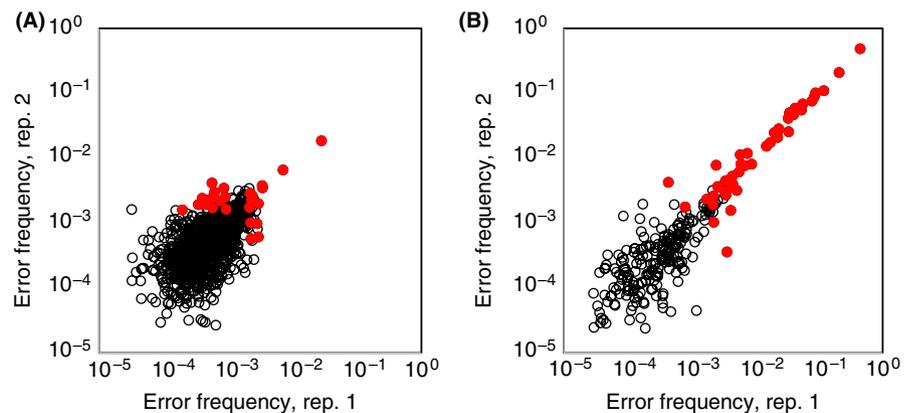
TABLE 1 Detection of mosaicism in ~ 1% artificial mixtures of mutant and wild type DNA

	Artificial mixture		Control group[a]			
Mutation	Reads	Frequency	Reads	Frequency	P value	LOD
c.5878C > T	5893	0.0058	77 600	0.0006	$2.8 \cdot 10^{-18}$	0.0014
c.5393C > T	3368	0.0161	56 110	0.0005	$4.0 \cdot 10^{-47}$	0.0016
c.1834C > T	5831	0.0132	90 269	0.0006	$8.3 \cdot 10^{-60}$	0.0014

Abbreviation: LOD, limit of detection.

[a]Control group consisting of 20 wild-type samples.

Ion Torrent by sequencing and ddPCR by TaqMan genotyping. Ion Torrent sequencing is a universal analysis strategy that has a great advantage for the analysis of the large spectrum of different mutations causing HA, while the TaqMan system design depends on initial identification of the disease-causing mutations and subsequent design and optimization of mutation-specific probe systems. In Ion Torrent sequencing analysis, there is a basic noise level inherent to the system, whereas in ddPCR it is possible to

TABLE 2 Detection of mosaics among presumed noncarrier mothers

Family	Mutation	Noncarrrier mother[a]				Control group[a]				
		Total[b]	Nonref[c]	Bias[d]	Nonref/ Total (%)	Total[b]	Non-ref[c]	Bias[d]	Nonref/ Total (%)	P value[e]
Substitutions										
631	c.266G > A	3013	3	0.63	0.1	56 557	21	0.67	0.0	.12
733	c.805A > T[f]	8428	672	0.61	8.0	84 599	160	0.72	0.2	$<3\cdot10^{-16}$
717	c.902G > A	4369	8	0.54	0.2	88 582	105	0.55	0.1	.16
653	c.1834C > T[f]	9264	8	0.52	0.1	146 451	106	0.54	0.1	.37
459	c.5393C > T[f]	2382	3	0.58	0.1	95 487	68	0.61	0.1	.25
417	c.5878C > T[f]	11 767	5	0.42	0.0	137 436	71	0.47	0.1	.73
677	c.6230C > G	3677	3	0.49	0.1	74 104	51	0.48	0.1	.47
667	c.6682C > T	2912	5	0.39	0.2	69 941	46	0.40	0.1	.05
Indels										
628	c.173delC	3305	2	0.64	0.1	60 993	25	0.63	0.0	.41
310	c.209delTTGT[f]	3722	0	0.57	0.0	55 123	0	0.58	0.0	1
371	c.209delTTGT[f]	3467	0	0.60	0.0	55 123	0	0.58	0.0	1
703	c.1861delC	7394	0	0.50	0.0	150 372	4	0.50	0.0	1
607	c.4694delTTCT	11 402	0	0.57	0.0	186 634	0	0.49	0.0	1
714	c.6469delAA[f]	2237	1	0.68	0.0	47 696	5	0.71	0.0	.24
614	c.6565delGA	3426	0	0.56	0.0	44 777	0	0.54	0.0	1
701	c.2738insT[f]	3317	21	0.50	0.6	80 860	718	0.54	0.9	.95

[a]Number of reads of each noncarrier mother was compared with the sum of reads of all remaining mothers who did not have the disease-causing mutation (control group).

[b]Total number of reads.

[c]Number of reads with nonreference alleles.

[d]Forward strand bias.

[e]One-sided Fisher's exact test.

[f]Mutations also investigated with droplet digital polymerase chain reaction.

design and optimize the TaqMan-systems to very high specificities, effectively reducing noise. This, together with the analysis of a larger number of GE, increases the sensitivity of this method in comparison with NGS-based methods such as Ion Torrent sequencing. Both techniques work well for substitution mutations and for indels that are not part of mononucleotide repeats. For indels that are part of mononucleotide repeats, both systems show poor performance. Standard TaqMan systems are very difficult to design and optimize for this type of target sequences and the analysis of mononucleotide repeats is a well-known Achilles' heel of Ion Torrent sequencing. Repeated analysis and the analysis of artificial mixtures of mutant and wild-type DNA can solve the problem in some cases; however, low-frequency mononucleotide alleles where the LOD in many cases is higher than the frequency of the signal remain problematic.

Sequence analysis of the coding sequence of *F8* revealed a few larger mononucleotide repeats, with one 9-mer, one 8-mer, three 7-mers, thirteen 6-mers, and thirty-three 5-mers. Together, these repeats make up approximately 0.5% of the total sequence and will need alternative analysis strategies. The difficulties in the analysis of such sequences could probably be greatly reduced by using Illumina-based sequencing that does not depend upon resolving the signal differences from long and differently sized mononucleotide repeats, but instead uses reversible terminator chemistry to sequence a single base at a time regardless of the local sequence context.

Previously, several other studies have investigated the existence of mosaicism among noncarrier mothers of boys with HA using different assay methods resulting in the detection of mosaicism among some of the mothers. One group used Sanger sequencing and allele-specific PCR for the analysis of point mutations,[15] whereas another group relied upon Sanger sequencing in combination with SNaPshot analysis.[17] Both allele-specific PCR and SNaPshot analysis are mutation-specific and semiquantitative techniques that require substantial investments in system design, production, and optimization. To produce reliable data, they require comparisons with serial dilutions of each investigated variant.

In the present study, mosaicism was detected in one mother with a c.805A > T substitution mutation. Comparison between the noncarrier mother and the control group showed a considerably higher mutation frequency in the mosaic mother. The noncarrier mother mosaic for the c.805A > T mutation had an 8.0% mutation

frequency and the control group had 0.2%. ddPCR confirmed the mosaicism in the mother with the c.805A > T mutation. Thus, a mosaic mother was detected in 1 of 16 investigated families (6%). In the study by Leuer et al,[15] a total of 45 families with substitutions and indels were investigated and eight were found to be mosaic (18%),[15] whereas Lu et al[17] detected mosaic mothers in 3 of 26 families (11%).[17]

It is interesting to note that the majority of the observed mosaic mutations had frequencies in the interval 5%–25% (present study, 7.8%; Lu et al[17] study, 10%, 15%, and 20%; Leuer et al[15] study, 0.2%, 0.5%, 2%, 5%, 5%, 10%, 20%, and 25%).[14,16] Only 3 of 12 had frequencies < 5%, and none had a mutation frequency > 25%. A possible explanation for the apparent lack of higher frequency mutations may be due to misclassification of carrier status in mothers with high-frequency mosaicism. This would underestimate the true frequency of mosaics. This would be unproblematic from a clinical perspective as these mothers were already regarded as carriers. At the other end of the frequency spectrum, there may be a detection bias caused by low sensitivity of the methods used. Leuer et al,[15] for example, estimated the sensitivity of the allele-specific PCR for point mutations to 0.1% in a wild-type background, but it decreased to 2%-5% for small deletions/insertions.[15] Even when supported by dilution experiments there is a risk of false-positive results when using PCR-based techniques for the analysis of indels.

Another interesting observation is that the four sources from which DNA was extracted in the study by Lu et al[17] seemed to detect similar frequencies of the mosaic mutations in all investigated sources of DNA (blood, oral mucosa, hair follicle, and urine).[17] This may indicate the introduction of the mutation early during ontogenesis. We speculate that the use of nongermline cells such as whole blood may bias the detection of mosaics to a certain frequency interval. This observation was also made by Leuer et al[15] who noted that, in a somatic mosaic individual, the number of mutant alleles in the germ cells may not decrease below a certain proportion due to the early appearance of the mutation during development.[15]

NGS is a truly universal and quantitative technique with the ability to both identify the disease-causing mutation and detect the occurrence of low-frequency mutations. The testing involves estimating the frequency of the mutation in the individual investigated for mosaicism and comparing it to the frequency observed in a control group. Statistical testing for frequency differences is performed and the mutation frequencies of significant cases are then compared with the LOD estimated from the control group to ascertain the result. The determination of the LOD depends on both the number of sequenced unique GEs and the noise level for the specific position of the disease-causing mutation. The number of unique GEs analyzed in the present study was, in all but one case, >2000 unique GEs. The LOD for this study varied from one disease-causing mutation to another but was in most cases fairly low: for substitutions, 925 positions reported as mutated in the mutation database (www.factorviii-db.org) had a LOD of < 0.5% (82% of positions). The corresponding number for indels was 368 positions representing 82% of positions. This means that an absolute majority of the mutated positions in *F8*

can be investigated for the exact mutation in the patient while the mother is investigated simultaneously for mosaicism at the 1% level.

Given a scenario where the initial mutation detection in the patient is performed using a NGS technique such as Ion Torrent sequencing, it is convenient and sufficient to include the mother in the initial analysis to investigate her carrier status in parallel to determination of the mutation in the patient. This will in most cases be sufficient to detect an absolute majority of all cases of mosaicism as reported in the literature. If confirmation of a mosaicism is desired ddPCR is a good alternative that will also in most cases allow a lower level of detection in special cases such as for noncarrier mothers with more than one child with the same mutation. Thus, deep Ion Torrent sequencing is suitable for both mutation detection in patients and carrier mothers[27] and evaluation of mosaic status.

RELATIONSHIP DISCLOSURE

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

EM, RN, and CLH performed experiments. EM, RN, and CH performed data analysis and wrote the manuscript. All authors contributed to the final manuscript. RL and JA contributed patient material. CH supervised the study.

REFERENCES

1. Swystun LL, James PD. Genetic diagnosis in hemophilia and von Willebrand disease. Blood Rev. 2017;31(1):47–56.
2. Hoffman R, Benz EJ, Silberstein LE, et al. Hematology: Basic Principles and Practice. Philadelphia, PA: Elsevier; 2018. ISBN 9780323357623.
3. Viel KR, Machiah DK, Warren DM, Khachidze M, Buil A, Fernstrom K, et al. A sequence variation scan of the coagulation factor VIII (FVIII) structural gene and associations with plasma FVIII activity levels. Blood. 2007;109(9):3713–24.
4. Li JN, Carrero IG, Dong JF, Yu FL. Complexity and diversity of F8 genetic variations in the 1000 genomes. J Thromb Haemost. 2015;13(11):2031–40.
5. Johnsen JM, Fletcher SN, Huston H, Roberge S, Martin BK, Kircher M, et al. Novel approach to genetic analysis and results in 3000 hemophilia patients enrolled in the My Life. Our Future initiative. Blood Adv. 2017;1(13):824–34.
6. Kasper CK, Buzin CH. Mosaics and haemophilia. Haemophilia. 2009;15(6):1181–6.
7. Biesecker LG, Spinner NB. A genomic view of mosaicism and human disease. Nat Rev Genet. 2013;14(5):307–20.
8. Costa C, Frances AM, Letourneau S, Girodon-Boulandet E, Goossens M. Mosaicism in men in hemophilia: is it exceptional? Impact on genetic counselling. J Thromb Haemost. 2009;7(2):367–9.
9. Casey GJ, Rodgers SE, Hall JR, Rudzki Z, Lloyd JV. Grandpaternal mosaicism in a family with isolated haemophilia A. Br J Haematol. 1999;107(3):560–2.
10. Costa JM, Vidaud D, Laurendeau I, Vidaud M, Fressinaud E, Moisan J-P, et al. Somatic mosaicism and compound heterozygosity in female hemophilia B. Blood. 2000;96(4):1585–7.

11. Higuchi M, Kochhan L, Olek K. A somatic mosaic for haemophilia A detected at the DNA level. Mol Biol Med. 1988;5(1):23–7.

12. Brocker-Vriends AH, Briet E, Dreesen JC, Bakker B, Reitsma P, Pannekoek H, et al. Somatic origin of inherited haemophilia A. Hum Genet. 1990;85(3):288–92.

13. Levinson B, Lehesjoki AE, de la Chapelle A, Gitschier J. Molecular analysis of hemophilia A mutations in the Finnish population. Am J Hum Genet. 1990;46(1):53–62.

14. Oldenburg J, Rost S, El-Maarri O, Leuer M, Olek K, Müller CR, et al. De novo factor VIII gene intron 22 inversion in a female carrier presents as a somatic mosaicism. Blood. 2000;96(8):2905–6.

15. Leuer M, Oldenburg J, Lavergne JM, Ludwig M, Fregin A, Eigel A, et al. Somatic mosaicism in hemophilia A: a fairly common event. Am J Hum Genet. 2001;69(1):75–87.

16. Tizzano EF, Cornet M, Domenech M, Baiget M. Exclusion of mosaicism in Spanish haemophilia A families with inversion of intron 22. Haemophilia. 2003;9(5):584–7.

17. Lu Y, Xin Y, Dai J, Wu X, You G, Ding Q, et al. Spectrum and origin of mutations in sporadic cases of haemophilia A in China. Haemophilia. 2018;24(2):291–8.

18. Ding Q, Wu X, Lu Y, Chen C, Shen R, Zhang X, et al. AccuCopy quantification combined with pre-amplification of long-distance PCR for fast analysis of intron 22 inversion in haemophilia A. Clin Chim Acta. 2016;458:78–83.

19. Merriman B, Ion Torrent R, Team D, Rothberg JM. Progress in ion torrent semiconductor chip based sequencing. Electrophoresis. 2012;33(23):3397–417.

20. Morey M, Fernandez-Marmiesse A, Castineiras D, Fraga JM, Couce ML, Cocho JA. A glimpse into past, present, and future DNA sequencing. Mol Genet Metab. 2013;110(1–2):3–24.

21. Bastida JM, Del Rey M, Lozano ML, Sarasquete ME, Benito R, Fontecha ME, et al. Design and application of a 23-gene panel by next-generation sequencing for inherited coagulation bleeding disorders. Haemophilia. 2016;22(4):590–7.

22. Bastida JM, Gonzalez-Porras JR, Jimenez C, Benito R, Ordoñez GR, Álvarez-Román M, et al. Application of a molecular diagnostic algorithm for haemophilia A and B using next-generation sequencing of entire F8, F9 and VWF genes. Thromb Haemost. 2017;117(1):66–74.

23. Simeoni I, Stephens JC, Hu F, Deevi SVV, Megy K, Bariana TK, et al. A high-throughput sequencing test for diagnosing inherited bleeding, thrombotic, and platelet disorders. Blood. 2016;127(23):2791–803.

24. Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, Makarewicz AJ, et al. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. Anal Chem. 2011;83(22):8604–10.

25. Hallden C, Nilsson D, Sall T, Lind-Hallden C, Liden AC, Ljung R. Origin of Swedish hemophilia A mutations. J Thromb Haemost. 2012;10(12):2503–11.

26. Martensson A, Ivarsson S, Letelier A, Manderstedt E, Hallden C, Ljung R. Origin of mutation in sporadic cases of severe haemophilia A in Sweden. Clin Genet. 2016;90(1):63–8.

27. Manderstedt E, Nilsson R, Lind-Hallden C, Ljung R, Astermark J, Hallden C. Targeted re-sequencing of F8, F9 and VWF: Characterization of Ion Torrent data and clinical implications for mutation screening. PLoS One. 2019;14(4):e0216179.

28. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17(1):122.

SUPPORTING INFORMATION

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