

# Identification of deep intronic individual variants in patients with hemophilia A by next-generation sequencing of the whole factor VIII gene

Hiroshi Inaba PhD<sup>1</sup> | Keiko Shinozawa PhD<sup>2</sup> | Kagehiro Amano MD, PhD<sup>1,2</sup> | Katsuyuki Fukutake MD, PhD<sup>1,2</sup>

<sup>1</sup>Department of Laboratory Medicine, Tokyo Medical University, Tokyo, Japan

<sup>2</sup>Department of Molecular Genetics of Coagulation Disorders, Tokyo Medical University, Tokyo, Japan

## Correspondence

Hiroshi Inaba, Department of Laboratory Medicine, Tokyo Medical University, Tokyo, Japan.

Email: hrsinb@tokyo-med.ac.jp

## Funding information

JSPS KAKENHI, Grant/Award Number: JP25461461; Japan Agency for Medical Research and Development

## Abstract

**Background:** No genetic defects are found in the coagulation factor VIII gene (*F8*) of approximately 2% of patients with hemophilia A. Recently, genomic variants causative of hemophilia A that were located deep within introns have been reported.

**Objectives:** We aimed to establish a comprehensive method of analysis of *F8* using next-generation sequencing (NGS) and investigate the variants located deep within the introns of *F8*.

**Patients/Methods:** Forty-five male patients with hemophilia A, including 31 with previously identified causative mutations, were investigated.

**Results:** Our NGS analysis allowed for the identification of genetic variants in roughly 99% of *F8*. We confirmed that our NGS analysis can detect the single nucleotide variants and small deletions with high accuracy. After filtering, a total of 27 rare and unique individual variants from 16 patients remained. Three of these variants, c.144-10810T>C, c.1010-365A>G, and c.5219+9065A>G, were predicted as deleterious with high expected accuracy by PredictSNP2 analysis. We also predicted the impact on splicing by *in silico* analysis using three different algorithms. Two patients with unknown causative mutations carried unique individual variants, c.144-10810T>C and c.6723+193G>A. We inferred that the c.144-10810T>C variant likely causes hemophilia, while the effect of the c.6723+193G>A variant remains unclear. Our analysis showed that the c.6429+14194T>C variant was significantly detected in patients carrying the intron 22 inversion.

**Conclusions:** Rare and unique individual variants located deep within the *F8* introns in patients with hemophilia A are not uncommon. Future studies are necessary to determine the function and effect of these variants on *F8* expression.

## KEYWORDS

factor VIII, hemophilia A, high-throughput nucleotide sequencing, intron, mutation

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2017 The Authors. *Research and Practice in Thrombosis and Haemostasis* published by Wiley Periodicals, Inc on behalf of International Society on Thrombosis and Haemostasis.

## Essentials

- Intronic variants of the factor VIII gene (*F8*) causing hemophilia A have been reported.
- We established an analysis method for whole *F8* and investigated the variants within its introns.
- Rare variants located within introns of *F8* in patients with hemophilia A are not uncommon.
- The c.6429+14194T>C variant was characteristically detected in patients with inversion.

## 1 | INTRODUCTION

Hemophilia A (MIM +306700) is the most common severe inherited bleeding disorder. It is the result of quantitative or qualitative abnormalities of blood coagulation factor VIII (FVIII), resulting from genetic defects in the coagulation factor VIII gene (*F8*).

Identification of genetic defects in patients with hemophilia A is essential for understanding the features of a given case of the disease and for providing more personalized treatment. Since *F8* was cloned in 1984<sup>1</sup>, various types of genetic mutations that cause hemophilia A have been identified in *F8*. Presently, approximately 3000 unique mutations have been identified and registered in a worldwide mutation database, the Factor VIII variant database (<http://www.factorviii-db.org/index.php>), and the CDC Hemophilia A Mutation Project (CHAMP) *F8* Mutation List (<http://www.cdc.gov/ncbddd/hemophilia/champs.html>). There are currently three standard methods that are applied to identify genetic defects in *F8*: (i) Direct sequencing of the *F8* coding region, promoter, 3'-UTR, and intron-exon boundaries either by polymerase chain reaction (PCR) and Sanger sequencing or by next-generation sequencing (NGS); (ii) Intron 22 inversion analysis by Southern blot, long-range PCR, or inverse PCR and Intron 1 analysis by PCR; and (iii) Copy number variant analysis by multiplex ligation-probe amplification analysis (MLPA) or array comparative genomic hybridization. However, recent studies have shown that no genetic mutations in *F8* can be found in approximately 2% of patients with hemophilia A.<sup>2-4</sup> In cases such as these, it is believed that mutations could be located deep within the introns of *F8*. Recently, we unexpectedly detected the c.1537+325A>G mutation within intron 10 by genomic DNA analysis.<sup>5</sup> We also detected the c.1443+602A>G mutation within intron 9 by mRNA analysis.<sup>6</sup> Both variants were predicted to cause a splicing abnormality by in silico analysis, and the abnormal transcripts were confirmed through mRNA analysis. Moreover, another study reported the existence of a causative variant located deep within an intron of the *F8*.<sup>7</sup>

The objective of the present study was to establish a method of whole-genetic analysis of the sequence of *F8* using NGS and investigate the variants located deep within its introns.

## 2 | MATERIALS AND METHODS

### 2.1 | Patient samples

Genomic DNA was extracted from peripheral blood cells of patients. Forty-five Japanese male patients with hemophilia A were investigated. These patients were confirmed to not have any apparent

relatives with hemophilia. Thirty-one patients had been previously analyzed. Among them, no causative mutations were previously identified in two patients by conventional analysis including direct sequencing, intron 22 and 1 inversion analysis by long range (LR)-PCR, or MLPA. Moreover, von Willebrand disease type 2N was excluded by FVIII/von Willebrand factor binding assay. Fourteen of the 45 patients were analyzed for the first time in the present study.

The study was approved by the Ethics Committee of Tokyo Medical University. Written informed consent was obtained from all patients, and the study was carried out in accordance with the principles of the Declaration of Helsinki.

### 2.2 | Next-generation sequencing

The complete *F8* locus was amplified in 14 overlapping regions (5-23 kb) by LR-PCR using KOD FX neo (Toyobo Co., Ltd, Osaka, Japan). The primers used are shown in Table S1. Amplification was performed based on two-step touch-down PCR. Thermal cycling conditions are shown in Table S2. In total, approximately 197 kb (including the upstream and downstream regions of *F8*) were amplified. The PCR fragments were purified using an illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare UK Ltd, Little Chalfont, Buckinghamshire, UK), and mixed in equimolar amounts. The DNA library was prepared by fragmentation using a Nextera XT DNA sample preparation kit (Illumina Inc., San Diego, CA, USA). The paired-end adapter-ligated fragments of the pooled libraries were attached to the flow cell and sequenced using the amplicon sequencing application of the MiSeq software program (Illumina Inc.). The obtained nucleotide sequences were aligned to the GRCh37/hg19 coordinates of an *F8* reference sequence (ENSG00000185010) using the Burrows-Wheeler Aligner. The variants were detected using the Genome Analysis Toolkit and were annotated by the VariantStudio software program (Illumina Inc.).

### 2.3 | Inversion analysis

*F8* inversion was analyzed by the long-range PCR method described by Liu et al.,<sup>8</sup> with modifications. Briefly, the primers were designed more adequately and step-down amplification was adopted.

### 2.4 | Bioinformatic analyses

Two detection tools, BreakDancer<sup>9</sup> and Pindel,<sup>10</sup> were used to detect structural variants.

The Combined Annotation Dependent Depletion (CADD) score, which predicts the deleteriousness of single nucleotide variants as well as insertions/deletions in the human genome, was obtained from the CADD (version 1.3) website (<http://cadd.gs.washington.edu/home>).<sup>11</sup> PredictSNP2 analysis was also used for analysis of the prediction of disease-related mutations (<http://loschmidt.chemi.muni.cz/predictsnp2/>).<sup>12</sup>

Potential splice effects of variants were evaluated by Human Splicing Finder (<http://www.umd.be/HSF/>),<sup>13</sup> NNSPLICE at the Berkeley Drosophila Genome Project ([http://www.fruitfly.org/seq\\_tools/splice.html](http://www.fruitfly.org/seq_tools/splice.html)),<sup>14</sup> and the NetGene2 server (<http://www.cbs.dtu.dk/services/NetGene2/>).<sup>15</sup>

### 3 | RESULTS AND DISCUSSION

#### 3.1 | NGS data and analysis

Sequencing coverage was sufficiently high (>20 reads) to confirm the sequence, although it varied widely by region and analysis (Figure 1). However, a small part of intron 22, which differed in size (~1-2 kb) according to sample and analysis, showed very low coverage (0-20 reads). The low-coverage region correspond to *F8A1* (coagulation factor VIII-associated 1) gene and had a high GC content. Our NGS analysis therefore allowed for the identification of genetic variants within roughly 99% of *F8*. On average, 140 variants were detected in each patient. In the analysis of samples that were previously identified as having causative mutations, it was confirmed that the single nucleotide variants (such as point mutations) and small deletions could be detected with high accuracy and efficiency (Table 1). In contrast, structural variants (such as inversions and large duplications) could not be appropriately detected by bioinformatic analyses under the present conditions. *F8* appears to be susceptible to genetic rearrangements for the following reasons: (i) *F8* is very large and contains a large number of repetitive elements (e.g. Alu repeats and long interspersed elements); and (ii) *F8* is located on the tip of the X chromosome. Therefore, it becomes very effective in causative mutation analysis of hemophilia if we can detect not only single nucleotide substitution,

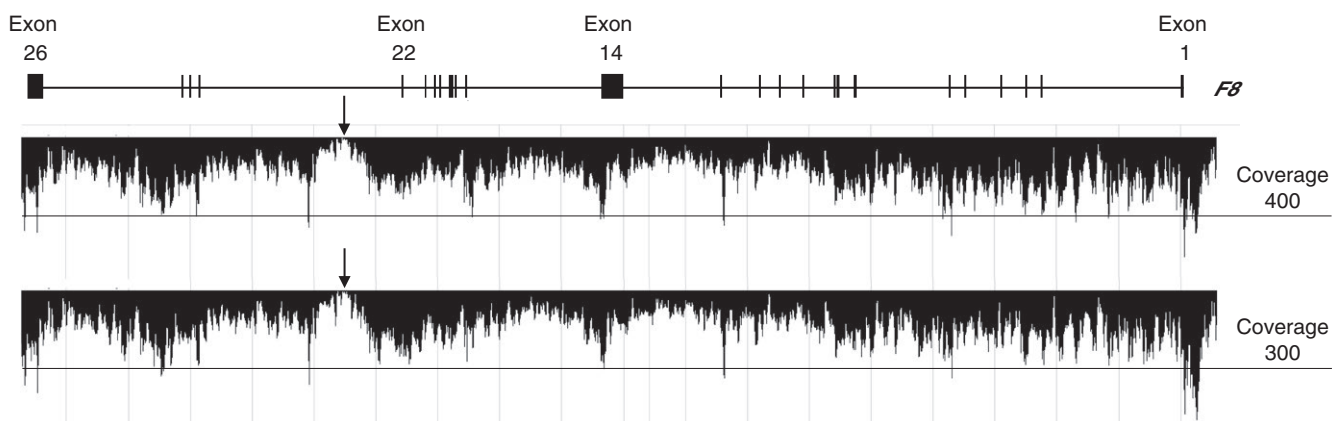
but also structural variant by NGS. Further studies are required to detect structural variants.

#### 3.2 | Variant analysis

To search for rare and causative variants located within the introns of *F8*, we narrowed down the variants. At first, in the VariantStudio software annotation, the variants were filtered by the following criteria: (i) "homozygote" (meaning hemizygote on the X chromosome of males) was applied to the category of "Genotype"; (ii) "PASS" (meaning all filters about the quality of variant call were passed in the VCF [Variant Call Format] file annotations) was applied to the category of "Filters"; and (iii) "no" was applied to the category of "Exonic" (meant intronic). We also ruled out variants registered in the Single Nucleotide Polymorphism Database (dbSNP), 1000 Genomes, COSMIC, and ClinVar databases. We further ruled out variants registered in the variant table in the *F8* transcript (*F8*-001 ENST00000360256.8 by GRCh38, *F8*-001 ENST00000360256.4 by GRCh37). Finally, we ruled out variants shared by more than one patient. After filtering, 27 variants remained from 16 patients (Table 2A). Although two duplication variants, c.5219+10174dupA and c.1903+2003dupT, passed through the VCF filters, the possibility of them being false positive variants cannot be excluded because they were located at a homopolymer sequence.

In the present study, we did not validate the detected variants. This is because we selected variants that were detected with high alternative variant frequency (approximately 100%) and recognized as homozygous. However, a concern of NGS for long-range amplified material is the appearance of false positive variants because of replication errors in PCR. Therefore, we believe that validation by Sanger sequencing is necessary when analyzing variants with low alternative variant frequency.

To predict the functional annotation of these variants, we first attempted to evaluate the variants by CADD analysis (Table 2). The C-scores of these variants obtained by the analysis ranged from 0.005 to 14.97. Four considerably high scores (over 10), were observed in four patients. According to these results, these four variants might



**FIGURE 1** Coverage of the whole factor VIII gene (*F8*). A typical coverage pattern obtained from two patients. Arrows indicate the position of the low-coverage region in intron 22

cause disease. Recently, however, it was reported that there is limited clinical validity for the identification of pathogenic variants in noncoding regions in a hereditary cancer panel.<sup>16</sup> Although CADD analysis can likely more accurately predict variants within the coding region of genes, it is believed to have limited accuracy for predicting variants in the noncoding regions. We therefore applied PredictSNP2, a unified platform for accurately evaluating the effects of SNPs by exploiting the different characteristics of variants in distinct genomic regions, in addition to CADD analysis, for more precise prediction. The results showed that three variants, c.144-10810T>C, c.1010-365A>G, and c.5219+9065A>G, were predicted as being deleterious with high expected accuracy of over 90%. The results also confirmed that the variants tested were not registered in the dbSNP, GenBank, Clinvar, OMIM, Regulome, or HaploReg databases. Together, these results indicated that these variants are considerably rare and may cause disease.

To investigate the effects of each variant on splicing, we performed *in silico* analysis using three types of prediction software (Table 2). However, a prediction by each algorithm did not accord in almost all specific variants, and did not lead to firm prediction results. This indicated that an alternative approach to verify the effect of splicing is necessary.

### 3.3 | Analysis of the patients without detectable mutation in F8

One of the two patients without detectable mutations in *F8* carried a unique individual variant, c.144-10810T>C, in intron 1 with a C-score of 13.56 (patient 1; Table 2A), which was considerably high. Furthermore, PredictSNP predicted the variant to be deleterious with a high expected accuracy of 97%. Although the interpretation by the human splicing finder predicted that the variant likely has no impact on splicing, considerable score changes suggesting the possibility of creating a new donor site were predicted in all splicing prediction algorithms. Taken together, these prediction results suggest that the c.144-10810T>C variant likely causes hemophilia A.

In the analysis of the other patient with unknown causative mutation, no variant remained after the aforementioned filtering. However, the patient carried a unique individual variant, c.6723+193G>A, in intron 24 with a C-score of 13.3 (patient 28; Table 2B). No significant splicing alteration was predicted by the *in silico* analysis. The variant was registered in the dbSNP as rs782551397, although the minor allele frequency and clinical significance were unavailable in the database. Therefore, it remains unclear whether 6723+193G>A is a causative mutation.

In the present study, we evaluated disease causality of each variant detected in the *F8* by several *in silico* analyses: CADD, PredictSNP2, and three types of splicing prediction software. Pezeshkpoor et al. reported an analysis of deep intronic mutations using NGS in patients without detectable mutations in *F8* cDNA.<sup>17</sup> Their methodology combined analysis by NGS and of mRNA. They identified two intronic variants (c.5998+530C>T and c.5998+941G>A) that create

new cryptic sites that lead to the insertion of intronic sequences in *F8* mRNA. In addition, they mentioned the necessity of verification of the splicing by experimental approaches, because the inconsistency between different algorithms in predicting the effect of specific variations on splicing was confirmed. Bach et al. also reported a study on deep intronic variants using NGS.<sup>18</sup> They identified deep intronic variants in 15 out of 15 patients with mild to moderate hemophilia A whose disease-causing mutations were not identified by conventional methods. Subsequently, the authors reported results confirming the impact of the variants on splicing using the mini-gene assay.<sup>19</sup> They reported that there were inconsistent results between *in silico* prediction and the mini-gene assay. Together, these reports indicated that predicting splicing by *in silico* analysis with complete reliability is difficult, and experimental verification is necessary. Further studies are necessary to determine the effects on splicing of the variants that we identified.

### 3.4 | Inversion analysis

A total of nine patients included in the present study carried the intron 22 inversion. Seven of these cases were previously detected by long-range PCR. The remaining two cases also were detected by long-range PCR in the present study. To identify the inversion by NGS analysis, we attempted *in silico* analysis using two software programs (BreakDancer and Pindel). Unfortunately, they did not predict the inversion precisely. A considerable false-positive detections and insufficient reproducibility were confirmed. However, we identified an interesting variant, c.6429+14194T>C, within the int22h-1 sequence, which is responsible for homologous recombination. This variant was detected in eight out of nine patients with inversion and was detected in one out of 37 patients without inversion. Therefore, the sensitivity, specificity, and positive predictive value for the prediction of inversion within the cohort by detection of this variant were 88.9%, 97.3%, and 88.9%, respectively.

## 4 | CONCLUSIONS

In the present study, we established a method of whole-genetic analysis of *F8* using NGS and investigated the variants located deep within *F8* introns. The application of NGS that can analyze deep intronic sequences can contribute to the clarification of etiology, and is expected to contribute to obtaining useful information on individual hemophilia patients. Our findings indicated that the existence of rare and unique individual variants located deep within the introns of *F8* of patients with hemophilia A is not uncommon. We believe that the majority of these variants are likely very rare and have no function. However, some of them are thought to have the possibility of being causative of hemophilia. Further studies are necessary to determine the actual functions and effects of these variants on *F8* expression. Comprehensive analysis using NGS will provide important information allowing for the personalized treatment of hemophilia.

**TABLE 1** Clinical data and information on the causative mutations in the patients studied

Patient #	FVIII:C (%)	Inhibitor	Causative mutation	Novel	dbSNP	C-score	Comment	Identified analysis	Alternative variant frequency in NGS analysis (%)
1	5.6		Unknown			n.a.			
2	<1	Positive	Intron 22 inversion			n.a.		previous (LR-PCR)	
3	1.6		Intron 22 inversion			n.a.		previous (LR-PCR)	
4	<1		c.6911G>A; p.Gly2304Glu			32		previous (DS), present (NGS)	100
5	10		c.601+3_601+4delAA	yes		12.9		previous (DS), present (NGS)	90.4
6	19		c.120C>A(p.Leu40=)	yes		2.686		previous (DS), present (NGS)	100
7	<1		c.5370_5372delCAT; p.Ile1790del	yes		15.34		previous (DS), present (NGS)	94
8	5		c.2120G>T; p.Trp707Leu			27.1		previous (DS), present (NGS)	100
9	<1	positive	c.1336C>T; p.Arg446Ter		rs137852372	36		previous (DS), present (NGS)	100
10	8.6		c.5347A>G; p.Arg1783Gly			26.3	also has rare polymorphism? c.3169G>A; p.Glu1057Lys	previous (DS), present (NGS)	100
11	<1	positive	Intron 22 inversion			n.a.	also has rare polymorphism? c.3169G>A; p.Glu1057Lys	previous (LR-PCR)	
12	3.4		c.1203G>T; p.Trp401Cys			33		previous (DS), present (NGS)	100
13	4		c.558C>G; p.Asp186Glu			23.3		previous (DS), present (NGS)	100
14	2.5		c.1470A>T; p.Arg490Ser			26.8		previous (DS), present (NGS)	100
15	4.5		c.6956C>T; p.Pro2319Leu		rs137852472	26.2		previous (DS), present (NGS)	100
16	<1		Intron 22 inversion			n.a.		previous (LR-PCR)	
17	1		c.3637delA; p.Ile1213PhefsTer5			13.81		previous (DS), present (NGS)	94.4
18	15-30		c.6547A>G; p.Met2183Val			25.5		previous (DS), present (NGS)	100
19	2.7		c.232T>C; p.Phe78Leu	yes		23.2		previous (DS), present (NGS)	100
20	6.6		c.6956C>T; p.Pro2319Leu		rs137852472	26.2		previous (DS), present (NGS)	100
21	10.6		c.1492G>A; p.Gly498Arg		rs137852414, rs28936969	34		previous (DS), present (NGS)	100
22	<1		Intron 22 inversion			n.a.		previous (LR-PCR)	

(Continues)

**TABLE 1** (Continued)

Patient #	FVIII:C (%)	Inhibitor	Causative mutation	Novel	dbSNP	C-score	Comment	Identified analysis	Alternative variant frequency in NGS analysis (%)
23	<1		c.6506G>A; p.Arg2169His		rs137852461	35		previous (DS), present (NGS)	100
24	<1		c.6464_6465delAA; p.Lys2155Thrfs*5		rs387906463	35		previous (DS), present (NGS)	90.8
25	<1		Intron 22 inversion			n.a.		previous (LR-PCR)	
26	1		c.1757T>A; p.Met586Lys			22.6		previous (DS), present (NGS)	96.7
27	30.1		c.6505C>T; p.Arg2169Cys			31		previous (DS), present (NGS)	100
28	2		Unknown			n.a.			
29	<1		Intron 22 inversion			n.a.		previous (LR-PCR)	
30	10.9		c.(787+1_788-1)_ (5998+1_5999-1)dup	yes		n.a.		previous (MLPA)	
31	6.4		c.4380delT; p.Asn1460Lysfs*5			28.3		previous (DS), present (NGS)	0 <sup>a</sup>
32	<1		c.2933- 2940delCATGGGGA; p.Ser978*fs	yes		29.6		present (NGS)	87
33	5.5		c.143+8C>T			6.315		present (NGS)	99.8
34	5.4		c.5879G>A; p.Arg1960Gln		rs28937294	33		present (NGS)	99.3
35	<1		c.6743G>A; p.Trp2248Ter			37		present (NGS)	100
36	2.6		c.326A>G; p.Asn109Ser	yes		24.5		present (NGS)	99.9
37	2.5		c.1226A>G; p.Glu409Gly		rs28933671	25.6		present (NGS)	100
38	<1		Intron 22 inversion			n.a.		present (LR-PCR)	
39	5.6		c.6977G>T; p.Arg2326Leu		rs137852360	27.2		present (NGS)	100
40	<1		Intron 22 inversion			n.a.		present (LR-PCR)	
41	29.2		c.5378C>A; p.Thr1793Asn			23.3		present (NGS)	100
42	36.6		c.923C>T; p.Ser308Leu		rs137852404, rs28937268	27.1		present (NGS)	100
43	3		c.142A>G; p.Arg48Gly	yes		10.77		present (NGS)	99
44	5.9		c.1475A>G; p.Tyr492Cys		rs137852412, rs28937275	26		present (NGS)	100
45	1.1	positive	c.322A>G; p.Lys108Glu	yes		27.3		present (NGS)	100

n.a., not available; DS, direct sequencing; LR-PCR, long-range PCR; MLPA, multiplex ligation-probe amplification analysis; NGS, next-generation sequencing.

Patients 1-31 were previously identified with causative mutations. In 21 of these patients, the mutations were confirmed by the present NGS analysis. Patients 32-45 were analyzed for the *F8* gene for the first time in this study.

<sup>a</sup>An accurate frequency value could not be calculated because of program error.

**TABLE 2** Rare and unique individual variants detected in introns of the factor VIII gene from patients with hemophilia A. The variants that passed all filtering criteria (A) and the variant detected in a patient with unknown causative mutation (B)

(A)								Human splicing finder <sup>a</sup>
Patient #	Individual variant	Coordinate of X-chr. (GRCh37)	Intron	Coverage Depth	dbSNP	C-score	PredictSNP2	Predicted signal
1	c.144-10810T>C	154238685	1	348		13.56	97%	No significant splicing motif alteration detected.
2	c.788-364_788-356delTTGGAGTTCC	154198182	6	40		6.305	n.a.	Alteration of an intronic ESS site.
3	c.602-1484G>A	154217064	4	53		1.241	88%	No significant splicing motif alteration detected.
	c.6430-3498T>C	154095000	22	73		1.219	77%	Creation of an intronic ESE site.
15	c.5219+10174dupA	154146671	14	86		0.402	n.a.	Alteration of an intronic ESS site.
	c.5220-10889A>G	154145737	14	98		2.244	88%	Alteration of an intronic ESS site.Creation of an intronic ESE site.
16	c.787+1870C>T	154211092	6	62		0.713	88%	No significant splicing motif alteration detected.
18	c.2113+3832C>T	154172141	13	93		2.889	77%	No significant splicing motif alteration detected.
	c.5373+301T>C	154134394	15	85		0.985	88%	No significant splicing motif alteration detected.
	c.6900+4491C>T	154084216	25	38		0.356	88%	Alteration of an intronic ESS site.Creation of an intronic ESE site.
22	c.6430-14725T>G	154106227	22	154		0.644	88%	Alteration of an intronic ESS site.Creation of an intronic ESE site.
24	c.143+6775A>G	154243910	1	403		0.213	88%	Creation of an intronic ESE site.
27	c.1753-535A>G	154182852	11	250		12.21	73%	Alteration of an intronic ESS site.Creation of an intronic ESE site.
31	c.144-7336G>A	154235211	1	474		4.566	88%	Creation of an intronic ESE site.
	c.601+169T>C	154221042	4	586		2.172	88%	Alteration of an intronic ESS site.
	c.6429+14259G>A	154110093	22	18		2.502	88%	No significant splicing motif alteration detected.
	c.6901-1476A>G	154067503	25	539		2.751	88%	Alteration of an intronic ESS site.
34	c.6901-1650C>T	154067677	25	84		0.005	74%	Creation of an intronic ESE site.
36	c.1010-365A>G	154195327	7	257		11.76	97%	Creation of an intronic ESE site.

(Continues)

Donor by HSF matrices (0-100)	Acceptor by HSF matrices (0-100)	5' motif by MaxEntScan (-20-20)		NNSPLICE		NetGene2	
		3' motif by MaxEntScan (-20-20)	Donor (0-1)	Acceptor (0-1)	Donor (0-1)	Acceptor (0-1)	
58.51>85.35	78.59>81.72	0.37>8.55	n.p.>0.9			n.p.>0.81	
	86.55>19.81, 43.4>82.49, 79.14>26.43						
38.98>65.82	73.03>74.28						
	47.71>76.66						
	84.51>36, 27.96>84.51, 85.09>27.96, 47.52>85.09	4.73>-12.87,-12.08>5.45		0.61>0.59, 0.86>0.72, 0.61>0.78		0.27>0.25, 0.31>0.15, 0.26>0.25	
	71.69>71.63						
66.12>66.52, 43.26>70.09						0.34>0.31, 0.34>0.31, 0.20>0.19	
70.3>59.72	46.82>75.77, 71.41>71.34						
		4.93>5.97		n.p.>0.52		n.p.>0.17	
	68.76>69.41						
	69.05>69.94					0.41>0.39	
		6.46>6.47		0.83>0.81			
	76.68>78.53						
	84.53>84.67	7.02>6.49					
	89.22>89.29	3.61>4.42		n.p.>0.43			
	72.64>70.8						



TABLE 2 (Continued)

(A)								Human splicing finder <sup>a</sup>
Patient #	Individual variant	Coordinate of X-chr. (GRCh37)	Intron	Coverage Depth	dbSNP	C-score	PredictSNP2	Predicted signal
37	c.787+2302G>A	154210660	6	82		0.165	88%	No significant splicing motif alteration detected.
	c.5219+9065A>G	154147781	14	121		14.97	91%	No significant splicing motif alteration detected.
	c.6901-7339G>A	154073366	25	119		2.36	77%	No significant splicing motif alteration detected.
38	c.1444-2189A>G	154191632	9	135		6.234	73%	Creation of an intronic ESE site.
44	c.787+3098T>C	154209864	6	222		4.466	77%	No significant splicing motif alteration detected.
	c.1903+2003dupT	154180163	12	216		0.349	n.a.	Creation of an intronic ESE site.
	c.6429+14453G>T	154109899	22	126		6.766	77%	Creation of an intronic ESE site.
45	c.2113+3105T>G	154172868	13	160		1.501	88%	Alteration of an intronic ESS site. Creation of an intronic ESE site.
(B)								Human splicing finder
Patient #	Individual variant	Coordinate of X-chr. (GRCh37)	Intron	Coverage Depth	dbSNP	C-score	PredictSNP2	Interpretation
28	c.6723+193G>A	154089800	24	79	rs782551397	13.3	73%	No significant splicing motif alteration detected.

n.a., not available; ESE, Exonic Splicing Enhancers; ESS, Exonic Splicing Silencers; n.p., not predicted.

Gray background in the PredictSNP2 column indicate "deleterious" and the remaining indicate "neutral". A percentage indicates the normalized confidence which corresponds to the observed accuracy measured for similar score on the actual data.

<sup>a</sup>In the prediction using Human Splicing Finder analysis, it was interpreted that all variants likely have no impact on splicing.

## AUTHOR CONTRIBUTIONS

H. Inaba performed conception and design, the experiment, data analysis and interpretation, and conducted drafting and revising of the manuscript. K. Shinozawa provided expert technical assistance. K. Amano supervised the study and performed interpretation and revising the manuscript. K. Fukutake supervised the study and performed data interpretation, revising the manuscript and final approval of the manuscript.

## ACKNOWLEDGMENTS

We thank H. Takedani and A. Nagao for providing one patient's blood sample. We thank J.H. Ohyashiki and the members of her laboratory for their support in the NGS analysis.

**Funding:** This work was partially supported by JSPS KAKENHI Grant Number JP25461461. This study was partially supported by the Research Program on HIV/AIDS from the Japan Agency for Medical Research and Development (AMED).

## RELATIONSHIP DISCLOSURE

H. Inaba has received honoraria from Biogen and Bayer outside of the submitted work. K. Shinozawa is an endowed assistant professor funded by Baxalta and has received honoraria from Baxalta, Bayer, and Novo Nordisk, outside of the submitted work. K. Amano hold concurrent posts as professors for the department of Molecular Genetics of Coagulation Disorders without additional salary; is a board member of the Factor Eight Inhibitor Bypass Activity Post Marketing Surveillance Study Board in Japan organized by

				NNSPLICE		NetGene2	
Donor by HSF matrices (0-100)	Acceptor by HSF matrices (0-100)	5' motif by MaxEntScan (-20-20)	3' motif by MaxEntScan (-20-20)	Donor (0-1)	Acceptor (0-1)	Donor (0-1)	Acceptor (0-1)
	80.47>79.82, 69.58>70.29						
	88.69>87.98		8.81>8.78		0.90>0.86		0.33>0.23
	77.59>76.74		4.37>3.38		0.48>n.p.		
	89.88>90.05				0.41>0.50		
			4.76>-11.38,-11.55>4.76				
	65.83>64.87						
	71.04>73.53						
				NNSPLICE		NetGene2	
Donor by HSF matrices (0-100)	Acceptor by HSF matrices (0-100)	5' motif by MaxEntScan (-20-20)	3' motif by MaxEntScan (-20-20)	Donor (0-1)	Acceptor (0-1)	Donor (0-1)	Acceptor (0-1)
	97.26>96.6, 84.2>83.79		4.74>4.91				

Baxalta; has received payment for lectures from Baxalta, Bayer, Biogen, Kaketsuken, Novo Nordisk, and Pfizer; has received payment for consultancy meetings with Baxalta, Bayer, CSL Behring, Kaketsuken, Novo Nordisk, and Pfizer; and has received unrestricted grants supporting research from Pfizer, outside the submitted work. K. Fukutake hold concurrent posts as professors for the department of Molecular Genetics of Coagulation Disorders without additional salary; is an investigator of Hemophilia Research Study Update organized by Baxalta, a board member of the Advate Safety Board in Japan organized by Baxalta, and a board member of the Benefix Post Marketing Surveillance Study Board in Japan, organized by Pfizer; has received payment for consultancy meetings with Baxalta, Pfizer, Biogen, Bayer, CSL Behring, Kaketsuken, SRL, LSI Medience, and Novo Nordisk; has received unrestricted grants supporting research from Baxalta, Pfizer, Bayer, Kaketsuken, Japan

Blood Products Organization, Ortho Clinical Diagnostics, and CSL Behring; has received payment for lectures from Baxalta, Bayer, Pfizer, Novo Nordisk, CSL Behring, Roche Diagnostics, Fujirebio Inc., Torii pharmaceuticals, Siemens, Abbott, Octapharma, and Sekisui Medical; and has received fee for post marketing survey from Cimic, outside the submitted work.

## REFERENCES

- Gitschier J, Wood W, Goralka T, et al. Characterization of the human factor VIII gene. *Nature*. 1984;312:326–30.
- Oldenburg J, El-Maarri O. New insight into the molecular basis of hemophilia A. *Int J Hematol*. 2006;83:96–102.
- Klopp N, Oldenburg J, Uen C, Schneppenheim R, Graw J. 11 hemophilia A patients without mutations in the factor VIII encoding gene. *Thromb Haemost*. 2002;88:357–60.

4. Vidal F, Farssac E, Altisent C, Puig L, Gallardo D. Rapid hemophilia A molecular diagnosis by a simple DNA sequencing procedure: identification of 14 novel mutations. *Thromb Haemost.* 2001;85:580–3.
5. Inaba H, Koyama T, Shinozawa K, Amano K, Fukutake K. Identification and characterization of an adenine to guanine transition within intron 10 of the factor VIII gene as a causative mutation in a patient with mild haemophilia A. *Haemophilia.* 2013;19:100–5.
6. Inaba H, Shinozawa K, Hagiwara T, Amano K, Fukutake K. The etiology of hemophilia hiding deep inside the F8 intronic sequence. *Blood.* 2011;118:1223.
7. Castaman G, Giacomelli SH, Mancuso ME, et al. Deep intronic variations may cause mild hemophilia A. *J Thromb Haemost.* 2011;9:1541–8.
8. Liu Q, Nozari G, Sommer SS. Single-tube polymerase chain reaction for rapid diagnosis of the inversion hotspot of mutation in hemophilia A. *Blood.* 1998;92:1458–9.
9. Chen K, Wallis JW, McLellan MD, et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods.* 2009;6:677–81.
10. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics.* 2009;25:2865–71.
11. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 2014;46:310–5.
12. Bendl J, Musil M, Stourac J, Zendluka J, Damborsky J, Brezovsky J. PredictSNP2: a unified platform for accurately evaluating SNP effects by exploiting the different characteristics of variants in distinct genomic regions. *PLoS Comput Biol.* 2016;12:e1004962.
13. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* 2009;37:e67.
14. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol.* 1997;4:311–23.
15. Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S. Splice site prediction in *Arabidopsis thaliana* pre-mRNA by combining local and global sequence information. *Nucleic Acids Res.* 1996;24:3439–52.
16. Mather CA, Mooney SD, Salipante SJ, et al. CADD score has limited clinical validity for the identification of pathogenic variants in noncoding regions in a hereditary cancer panel. *Genet Med.* 2016;18:1269–75.
17. Pezeshkpoor B, Zimmer N, Marquardt N, et al. Deep intronic ‘mutations’ cause hemophilia A: application of next generation sequencing in patients without detectable mutation in F8 cDNA. *J Thromb Haemost.* 2013;11:1679–87.
18. Bach JE, Wolf B, Oldenburg J, Muller CR, Rost S. Identification of deep intronic variants in 15 haemophilia A patients by next generation sequencing of the whole factor VIII gene. *Thromb Haemost.* 2015;114:757–67.
19. Bach JE, Muller CR, Rost S. Mini-gene assays confirm the splicing effect of deep intronic variants in the factor VIII gene. *Thromb Haemost.* 2016;115:222–4.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Inaba H, Shinozawa K, Amano K, Fukutake K. Identification of deep intronic individual variants in patients with hemophilia A by next-generation sequencing of the whole factor VIII gene. *Res Pract Thromb Haemost.* 2017;1:264–274. <https://doi.org/10.1002/rth2.12031>