令和4年度　学位申請論文

F9 mRNA splicing aberration due to a deep Intronic structural variation in a patient with moderate hemophilia B

(中等症血友病患者に認められるイントロン深部の構造的変異による *F9* mRNA のスプライシング異常)

名古屋大学大学院医学系研究科
総合保健学専攻

（指導：早川　文彦教授）

大平　晃也

主論文の要旨

_F9_ mRNA splicing aberration due to a deep Intronic structural variation in a patient with moderate hemophilia B

(中等症血友病患者に認められるイントロン深部の構造的変異による _F9_ mRNA のスプライシング異常)

大平　晃也

【緒言】

血友病 B とは血液凝固第 IX 因子 (FIX) 遺伝子 (*F9*) の異常に起因する FIX の量的・質的異常による遺伝性出血性疾患である。血友病 B は症例の血中 FIX 活性 (IU/dL) に基づき、重症 (<1 IU/dL), 中等症 (1-5 IU/dL), 軽症 (>5-<40 IU/dL)というように重症度分類がなされる。*F9* は X 染色体長腕末端側 (Xq27.1) に存在する 8 つのエクソンと 7 つのイントロンを持つ全長 33.5 kb の遺伝子であり、2.8 kb の mRNA に転写される。これまで血友病 B 症例では *F9* のエクソン領域において 1000 を超える遺伝子異常が報告されている。血友病 A やデュシェンヌ型筋ジストロフィといった遺伝子疾患の一部の症例には、各責任遺伝子のイントロン深部 (エクソン・イントロン境界から 100 塩基以上離れた領域) に複数の遺伝子異常が同定されており、それら遺伝子異常はスプライシング異常の原因となることが報告されている。その一方で、血友病 B 症例ではイントロン深部における遺伝子異常はほとんど報告されていない。本研究では血友病 B 中等症症例の *F9* イントロン 1 深部に新規

遺伝子異常を同定し、その遺伝子異常が *F9* mRNA のスプライシングに及ぼす影響を検討した。

【対象・方法】

名古屋大学医学部倫理審査委員会の承認のもと血友病 B 中等症 (血中 FIX 活性 ：3.0 IU/dL) と診断された男性よりインフォームドコンセントを得たのち、症例の末梢血白血球よりゲノム DNA を抽出、ダイレクトシーケンス解析や MLPA 解析、long-range PCR 解析などの遺伝子解析を実施した。同定した遺伝子異常が *F9* mRNA のスプライシングに及ぼす影響を調べるために、Exon-trap 解析を実施した。Exon-trap 解析では exon-trap cloning vector pET01 に野生型あるいは変異型の *F9* エクソン 1 からエクソン 3 までを組み込んだベクターを作製した (野生型 pET01, 変異型 pET01)。これらベクターを遺伝子導入した COS-7 細胞より total RNA を抽出し、ベクター上に組み込まれた *F9* エクソンのスプライシングパターンを RT-PCR やダイレクトシーケンスにて解析した。また正常なスプライシングを受けた mRNA の発現量を RT-qPCR で解析した。また遺伝子異常が FIX のタンパク質発現に及ぼす影響を調べるために、正常なスプライシングを受けた mRNA 量の減少をタンパク質発現量で評価することができる splicing-competent FIX 発現ベクターを使用した FIX 発現実験を実施した。pcDNA3.1 に野生型あるいは変異型 *F9* イントロン 1 を組み込んだ FIX 発現ベクターを作製した (野生型 FIX/pcDNA, 変異型 FIX/pcDNA)。これらベクターを遺伝子導入した HEK293 細胞より total RNA を抽出し、RT-qPCR にて *F9* mRNA のスプライシングパターンを解析した。また遺伝子導入細胞の

細胞溶解液、培養上清中の FIX タンパク質をイムノブロットや ELISA で検出、定量した。

【結果・考察】

F9 エクソン、エクソン・イントロン境界領域に対するダイレクトシーケンス解析や MLPA 法による *F9* エクソンの定量を実施したが、エクソン内の遺伝子異常や重複は認められなかった。続いて、long-range PCR を用いた *F9* 構造解析およびダイレクトシーケンス解析により症例の *F9* イントロン 1 深部に 28-bp の欠失と 476-bp の挿入を伴う新規構造的変異を同定した。挿入配列に対してホモロジー解析を行ったところ、挿入配列は 12 番染色体上の HNRNPA1 遺伝子エクソン 12 由来であることが明らかとなった。

我々は *F9* イントロン 1 深部で同定された構造的変異の *F9* スプライシングに対する影響を検討するため、Exon-trap 解析を実施した。その結果、野生型 pET01 を遺伝子導入した細胞では明瞭な 488-bp の増幅産物が得られた一方で、変異型 pET01 を遺伝子導入した細胞では 488-bp の増幅産物の著明な減少が認められた。また変異型 pET01 を遺伝子導入した細胞では 488-bp の増幅産物以外に新たに 553-bp と 729-bp の増幅産物も認められた。ダイレクトシーケンス解析により 488-bp の増幅産物は *F9* のエクソン 1、エクソン 2、エクソン 3 から構成される正常なスプライシングを受けた mRNA 由来である一方で 553-bp, 729-bp の増幅産物はそれぞれ *F9* エクソン 1 とエクソン 2 の間に挿入配列由来の偽エクソン 2 のみ、あるいは偽エクソン 2 と偽エクソン 3 を含む異常なスプライシングを受けた mRNA 由来であることが明らかとなった。これら異常スプライシング mRNA は偽エクソン 2 内に未成熟終止コドンが発生することから正常な FIX には翻訳されないと考えられた。さらに

変異型 pET01 より転写された正常スプライシング mRNA は野生型 pET01 より転写された正常スプライシング mRNA の 10%以下であった。

また HEK293 を使用した発現実験では、変異型 FIX/pcDNA より転写された正常スプライシング mRNA は野生型 FIX/pcDNA より転写された正常スプライシング mRNA の約 70%に減少していた。また細胞溶解液および培養上清中に対するイムノブロットで、細胞溶解液、培養上清中ともに変異型 FIX/pcDNA より翻訳された FIX タンパク質は減少していることが明らかとなった。さらに培養上清中の FIX タンパク質量を ELISA で定量したところ、変異型 FIX/pcDNA より翻訳・分泌された FIX 抗原量は野生型 FIX/pcDNA より翻訳・分泌された FIX 抗原量の約 60%に減少していた。以上の結果より症例で同定された新規構造的変異は *F9* mRNA のスプライシング異常を引き起こし、正常なスプライシングを受けた *F9* mRNA および FIX タンパク質を減少させていることが示唆された。

【結語】

本研究では血友病中等症患者の *F9* イントロン 1 深部に 28-bp の欠失を伴う 487-bp の新規構造的変異を同定した。このイントロン深部の新規構造的変異は *F9* mRNA のスプライシング異常を引き起こすことで、未熟終止コドンを持つ null な異常スプライシング mRNA を生じさせ、タンパク質に翻訳される *F9* mRNA を減少させていた。

主論文

Koya ODAIRA

Abstract

Introduction: Hemophilia B (HB) is a hereditary bleeding disorder caused by the genetic variation of the coagulation factor IX (FIX) gene (*F9*). Several *F9* structural abnormalities, including large deletion and/or insertion, have been observed to cause HB development. However, there is limited information available on *F9* deep intronic variations. In this study, we report about a novel large deletion/insertion observed in a deep region of *F9* intron 1 that causes mRNA splicing abnormalities.

Patient and methods: The patient was a Japanese male diagnosed with moderate HB (FIX:C = 3.0 IU/dL). The genomic DNA of the patient was isolated from peripheral blood leukocytes. DNA sequences of *F9* exons and splice donor/acceptor sites were analyzed via polymerase chain reaction and Sanger sequencing. Variant-affected *F9* mRNA aberration and FIX protein production, secretion, and coagulant activity were analyzed by cell-based exon trap and splicing-competent FIX expression vector systems.

Results: A 28-bp deletion/476-bp insertion was identified in the *F9* intron 1 of a patient with moderate HB. A DNA sequence identical to a part of the inverted *HNRNPA1* exon 12 was inserted. Cell-based transcript analysis revealed that this large intronic deletion/insertion disrupted *F9* mRNA splicing pattern, resulting in reduction of protein-coding *F9* mRNA.

Conclusion: A novel deep intronic *F9* rearrangement was identified in a Japanese patient with moderate HB. Abnormal *F9* mRNA splicing pattern due to this deep intronic structural variation

resulted in a reduction of protein-coding *F9* mRNA, which probably caused the moderate HB phenotype.

Introduction

Hemophilia B (HB) is a hereditary bleeding disorder caused by qualitative or quantitative abnormalities of coagulation factor IX (FIX) due to FIX gene (*F9*) variation. The frequency of HB is 1 in 25,000–30,000 male births, and based on FIX activity (FIX:C), the severity of HB is classified as severe (<1 IU/dL), moderate (1–5 IU/dL), or mild (>5–<40 IU/dL) [1].

F9 is located on the long arm of X chromosome (Xq27.1). It consists of eight exons and seven introns that range over 33.5 kb and is transcribed into the 2.8-kb mRNA of FIX [2]. To date, several kinds of *F9* variants have been identified in patients with HB. The CDC Hemophilia B Mutation Project (CHBMP) and EAHAD *F9* variant database have compiled >1,000 *F9* variations [3, 4]. In the CHBMP list, the frequency of *F9* variations is estimated as missense (58%); frameshift with small deletion, insertion, or duplication (16%); splice-site change (10%); nonsense (8%); large structural change (≥50 bp, 3%); small structural change (<50 bp, 2%); promoter (2%); and synonyms (1%).

Recent studies on genomic sequencing have indicated that genetic variation within the deep intron (>100 bases away from the exon–intron junction) can induce a disease-associated mRNA splicing aberration [5]. Deep intronic variation may trigger a pseudoexon inclusion by activating the noncanonical splice site or altering the splicing regulatory element. For hemophilia cases, several deep intronic variations in the coagulation factor VIII gene (*F8*) were identified as disease-associated variations that caused hemophilia A [6-10]. In the case of hemophilia B, very few reports have

discussed deep intronic variations in *F9* [11]; however, their disease-causing mechanisms have not been reported or characterized.

In this study, we present details about a novel complex rearrangement by deletions/insertions in deep *F9* intron 1 in a patient with moderate HB. Furthermore, this study demonstrates that this deep intron rearrangement disrupts the *F9* mRNA splicing pattern by creating pseudoexons, resulting in a reduction in protein-coding *F9* mRNA.

Materials and methods

Patient

This is the case of a Japanese male diagnosed with moderate HB (FIX:C = 3.0 IU/dL) and enrolled in the Japan Hemophilia Inhibitor Study 2 (J-HIS2) [12].

DNA sample

The genomic DNA of the patient was isolated from peripheral blood leukocytes using a previously described method [13] after obtaining a written informed consent. This study was approved by the ethics committee of the Nagoya University Graduate School of Medicine (Identification number: 2015-0391).

Polymerase chain reaction (PCR) and DNA sequence

All *F9* exons and exon–intron junctions were amplified by PCR using KOD FX polymerase (Toyobo, Osaka, Japan) and gene-specific primer sets (Supplementary Table 1). PCR amplicons were separated using electrophoresis with 1.5% agarose gel and were purified using the FastGene Gel / PCR Extraction Kit (Nippon Genetics Co. Ltd., Tokyo, Japan). Direct sequencing was performed using BigDye Terminator v1.1 Cycle Sequencing Kit and ABI Prism 310 or 3130 Genetic Analyzer (Applied Biosystems, Waltham, MA, USA). Genetic variation is described here based on the Human Genome Variation Society (HGVS) nomenclature [14].

Gene copy number variation analysis

To investigate exonic deletion or duplication in *F9*, multiplex ligation-dependent probe amplification (MLPA) was performed using SALSA MLPA Probemix (P207-C2 F9, MRC-Holland, Amsterdam, Netherlands) as per the manufacturer's protocol. Pooled gDNA derived from three normal males was used as a normal individual control.

Long-range PCR for *F9* introns

To investigate structural variations, long-range PCR analysis for *F9* introns was conducted. Variation-specific alteration in the PCR amplicon size was confirmed via agarose gel electrophoresis and densitometry analysis. Briefly, PCR amplicons were captured on 1.5% agarose gel electrophoresis and imported into the ImageJ program (National Institutes of Health). The signal intensity and mobility parameters of each amplicon were measured and presented graphically.

***In silico* analysis of mRNA splicing**

To predict the effect of *F9* variation on mRNA splicing, *in silico* simulation was performed using Splice Site Prediction [15] and NetGene2 [16].

Preparation of *F9* exon-trap vector and splicing-competent FIX expression vector

Cell-based splicing analysis of *F9* variation was conducted using the exon-trap cloning vector pET01 (MoBiTec GmbH, Göttingen, Germany). An exon-trap vector targeting a wild-type (WT) *F9* intron 1 was previously constructed by our group [17] and was named as pET01 F9 int 1 and 2 WT. An exon-trap vector targeting variant *F9* intron 1 was newly constructed by genetic engineering with pET01 F9 int 1 and 2 WT. Briefly, a variant-derived DNA alignment was generated by PCR with the patient's gDNA and specific primer set (Supplementary Table 1). pET01 was linearized via inverse PCR. The variant-derived DNA alignment and linearized pET01 were ligated using In-Fusion Premix (50˚C, 15 min; Takara Bio, Shiga, Japan), and the constructed product was subcloned by transformation into a DH5α competent cell.

Splicing-competent FIX expression vectors were constructed by rearranging the pcDNA3.1-based FIX expression vector that we previously generated [17]. Briefly, a variant-derived *F9* intron 1 alignment was amplified with the patient's gDNA and a specific primer set (Supplementary table 1). A reference *F9* intron 1 alignment was generated from normal male's gDNA. FIX expression vector was linearized by inverse PCR with a specific primer set (Supplementary table 1). The intron 1 alignment and linearized FIX expression vector were ligated by In-fusion premix (Takara Bio) and named as FIXwt int1(+) and FIXvariant int1(+), respectively.

Exon-trap analysis

The constructed vectors were transfected into COS-7 cells and HEK293 cells using the calcium phosphate transfection method, and the cells were cultured for 16 h [17]. They were cultured in Dulbecco's Modified Eagle Medium with high glucose (Wako, Osaka, Japan) supplemented with 10% fetal bovine serum (FBS; Sigma-Aldrich, St. Louis, MO, USA) and penicillin/streptomycin/amphotericin B (Wako) at 37˚C with 5% CO_2. Total RNA extraction method is described below.

RNA extraction and reverse-transcription

Total RNA from the transfected cells was isolated using the Reliaprep RNA Miniprep Systems (Promega, Madison, WI, USA) and was subsequently subjected to reverse transcription (RT)-PCR using PrimeScript II 1st Strand cDNA Synthesis Kit (Takara Bio) and specific primer sets (Supplementary Table. 1). RT-PCR products were analyzed via electrophoresis with 1.5% agarose gel and amplicon direct sequencing.

Reverse transcription-quantitative PCR (RT-qPCR)

RT-qPCR was performed using SYBR Premix Ex Taq II (Takara Bio) and Thermal Cycler Dice Real-Time System II (Takara Bio). The primer sets for RT-qPCR are presented in Supplementary Table 1. Glyceraldehyde-3-phosphate dehydrogenase gene (*GAPDH*) was simultaneously quantified as a reference gene.

Transient expression of recombinant FIX protein

The splicing-competent FIX expression vectors were transfected into HEK293 cells using Lipofectamine 3000 (Thermo fisher scientific, Waltham, MA, USA) for 5 h, followed by medium change involving DMEM containing FBS and 5 μg/mL of vitamin K (Koa Isei Co., Itd, Yamagata, Japan). After 24 h following transfection, the cells were cultured in serum-free medium supplemented with 5 μg/mL of vitamin K for 24 h, and then the culture media and cells were collected.

Western blot analysis

Western blot analysis was performed as described previously [17]. In brief, after 24 h of culture under serum-free conditions, culture media or cell lysates of the transfected cells were collected. Culture media were centrifuged at 200 × g for 10 min and passed through 0.45-μm filter (Merck, Darmstadt, Germany) to remove cellular debris. Proteins in the culture media samples were concentrated via acetone precipitation method [18]. The precipitated pellet was dissolved using 1× sample buffer (50mM Tris-HCl, pH 6.8, 2% SDS, 10% Glycerol). To prepare cell lysate, cells were lysed with 1× sample buffer. Cell lysates or culture media precipitates were heated at 95°C for 5 min, and the protein concentration was determined using Bio-Rad DC™ Protein Assay Kit (Bio-Rad Laboratories, Hercules, CA, USA). Cell lysates or culture media precipitates were subjected to electrophoresis on 10% SDS- polyacrylamide gel with 850 mM 2-mercaptoethanol and 0.04%

bromophenol blue and then blotted onto Immobilon®-P Transfer Membrane (Merck). After blocking with 2.5% skim milk, the membranes were incubated with primary antibody against human FIX (1:1000 dilution, B-3, Santa Cruz Biotechnology Inc., Heidelberg, Germany) or a-tubulin (1:1000 dilution, AA13, Sigma-Aldrich) overnight at 4°C. Subsequently, the membranes were incubated with horseradish peroxidase conjugated secondary antibody (1:1000 dilution, Cell Signaling Technology, Danvers, MA, USA) at 25°C for 90 min. Signals were visualized with an ECL Select™ Western Blotting Detection Reagent (GE Healthcare UK Ltd., Little Chalfont, UK) and Light Capture II (Atto Corporation, Tokyo, Japan).

Measurement of FIX protein level

FIX protein levels in the culture media were measured using a Factor IX Human SimpleStep ELISA® Kit (Abcam, Cambridge, UK) according to the manufacturer's instructions. The antibodies used for capture and detection in this ELISA kit were anti-FIX polyclonal antibodies. The standard curve was created with plasma-derived FIX, which was included in the ELISA kit (Abcam).

Coagulant activity of recombinant FIX

The coagulant activity of recombinant FIX (rFIX) variant was measured by an activated partial thromboplastin time (APTT)-based one-stage clotting assay with C.K. Prest 2 (FUJIREBIO Inc., Tokyo, Japan). FIX-depleted plasma (SYSMEX CORPORATION, Hyogo, Japan) was

supplemented with rFIX-containing culture medium and subjected to the APTT-based assay. To create standard curve, coagulation times from serial diluted medium of rFIXwt were used. The specific coagulant activity was calculated as the ratio between the coagulant activity (% of rFIXwt) and the secreted protein level (% of rFIXwt).

Statistics analysis

All quantitative data are presented as standard error of the mean values. Two-group comparison was made using Mann-Whitney test. Statistical analyses were carried out using GraphPad Prism 5 (GraphPad Software).

Results and discussion

All *F9* exons and exon–intron junctions of the patient were sequenced, whereas any single-base substitutions or small deletion/insertions-causing HB were not detected. Moreover, MLPA analysis revealed that the patient had no *F9* exonic deletion or duplication (Fig. 1A). To investigate structural variations, long-range PCR analysis was conducted on each *F9* intron. The long-range PCR of *F9* intron 1 revealed a slight increase in size in the patient (Fig. 1B and 1C). PCR-restriction enzyme fragment analysis using *Pst*I or *Hpa*II further indicated that an approximately 500-bp fragment was inserted into the patient's *F9* intron 1 (Supplementary Fig. 1A: *Pst*I digestion, Supplementary Fig. 1B: *Hpa*II digestion). DNA sequencing of the variant amplicon revealed that the patient possessed a 28-bp deletion (NC_000023.11:g.139,534,834_139,534,861del) and a 476-bp insertion into the deep *F9* intron 1 (Fig. 1D). A homology search using NCBI basic local alignment search tool indicated that the insertion of 476 bp was the reverse sequence of heterogeneous nuclear ribonucleoprotein A1 gene (*HNRNPA1*) exon 12 that presented on chromosome 12 (NC_000012.12:g.54,285,853_54,286,328inv) (Supplementary Fig. 2). According to the HGVS nomenclature, this deep intron structural variation is described as follows: NC_000023.11:g.139,534,834_139,534,861delins[NC_000012.12:g.54,285.853_54,286,328inv].

Several genomic rearrangements in the deep intronic region have been reported to affect mRNA splicing regulation and are associated with monogenic diseases, including Duchenne muscular dystrophy [19-22]. To predict whether the identified deep intronic structural variation in *F9* intron 1

affected its mRNA splicing, the splicing pattern of the *F9* transcripts from the variant allele was simulated using Splice Site Prediction [23] and NetGene2 [24]. *In silico* simulations suggested that the deep intron deletions/insertions in *F9* intron 1 disrupted its transcriptional splicing pattern, producing three alternative splicing donor/acceptor sites (Table 1). In the present study, these variation-inducing segments are referred to as pseudointron 1, pseudoexon 2, pseudointron 2, pseudoexon 3, and pseudointron 3, respectively (Fig. 2A).

To confirm the *in silico* predictions, the variant transcripts were investigated by cell-based exon-trap analysis (Fig. 2B). The exon-trap vector is regulated by RSV LTR promoter and transfected into COS-7 cells and HEK293 cells. In the RT-PCR analysis, transcripts from WT exon-trap vector (pET01 F9 int 1 and 2 WT) revealed a predominant 448-bp amplicon, whereas those from variant exon-trap vector (pET01 F9 int 1 and 2 variant) demonstrated an obvious reduction of the 448-bp amplicon (Fig. 2C, Supplementary Fig. 3A). Moreover, extra amplicons (553 and 729 bp) were observed in the variant exon-trap vector. Direct sequence data revealed that the 448-bp amplicon (N1 transcript) consisted of *F9* exons 1, 2, and 3, which was normal *F9* mRNA (Fig. 2D). Conversely, the 553-bp amplicon (V1 transcript) consisted of *F9* exon 1, pseudoexon 2, exon 2, and exon 3, whereas the 729-bp amplicon (V2 transcript) consisted of *F9* exon 1, pseudoexon 2, exon 2, pseudoexon 3, and exon 3. Codon readings predicted that the V1 and V2 transcripts possessed a premature termination codon (PTC) in pseudoexon 2, indicating that the V1 and V2 transcripts were null phenotype transcripts. To further assess the impact of this deep intronic rearrangement, we quantitatively

evaluated the N1 transcription level of the transfectant with pET01 F9 int 1 and 2 variants. Quantification of N1 transcripts by RT-qPCR indicated that the levels of N1 transcripts expressed from the variant vector were <10% lower than those of WT vector (Fig. 2E). Same result was observed in an exon-trap analysis using HEK293 cells (Supplementary Fig. 3B). These observations suggest inverted *HNRNPA1* insertion into the deep *F9* intron 1 disrupts the *F9* transcript spicing pattern contributing to the reduction of *F9* N1 transcript.

Next, we sought to investigate the pathogenic impact of the inverted *HNRNPA1* insertion into the deep *F9* intron 1 at FIX protein level (Supplementary Fig. 4). We conducted a FIX expression experiment using splicing-competent FIX expression vectors, FIXwt int1(+) and FIXvariant int1(+), following the design of Scalet et al. (Supplementary Fig. 4A) [25]. In this study, the splicing-competent FIX expression vector was constructed using pcDNA3.1 (+) vector with CMV promoter and transfected into HEK293 cells. RT-PCR showed that the *F9* transcript pattern derived from FIXvariant int1(+) was disrupted (Supplementary Fig. 4B). RT-qPCR analysis showed that the N1 transcript level of the transfectant with FIXvariant int1(+) was reduced by approximately 30% as compared with that of FIXwt int1(+) (Supplementary Fig. 4C). FIX protein level was reduced in the cell lysate and media of FIXvariant int1(+)-transfectants, and secreted FIX of FIXvariant int1(+)-transfectant was decreased by approximately 40% as compared with that of FIXwt int1(+)-transfectant (Supplementary Fig. 4D and E). Besides, APTT-based one stage clotting assay denoted that the specific coagulant activity of rFIXvariant was 1.67 ± 0.23 as compared with that of rFIXwt with no statistical significance

(Supplementary Fig. 4F). This data was a rational result because the N1 transcript was WT FIX-coding mRNA. The splicing-competent FIX expression suggests that the deep intronic structural variation reduces FIX protein level, whereas the synthesized FIX has no alteration on the FIX coagulant activity.

In this study, we have shown the molecular pathogenesis of the *F9* deep intronic structural variation. We observed that the N1 transcript level was different between the exon-trap and the splicing-competent FIX expression assays (Supplementary Fig. 3 and 4). This inconsistence could be explained by the difference of vector's promoter system (RSV LTR promoter in the exon-trap assay, CMV promoter in the splicing-competent FIX expression assay). Kornblihtt et al reported that the difference of promoter could affect the transcript splicing pattern [26]. One of the severe limitations of this study is to lack plasma FIX antigen / protein-level data of proband and his relatives. In this study, we genetically analyzed genomic DNA derived from samples of the patient who was enrolled in J-HIS2 study [12]. However, we were not able to obtain further clinical samples (e.g., citrated plasma or whole blood sample containing leukocytes) of proband and his relatives. These clinical samples would be key materials used to perform *in vivo* and/or *ex vivo* research. To gain further insights about the disease-associated impact of *F9* deep intronic structural variation(s), researchers should set up a system to preserve clinical samples including whole blood.

Conclusion

In this study, a novel *F9* deep intron 1 rearrangement, a 28-bp deletion/476-bp insertion, was identified in a Japanese patient with moderate HB. To the best of our knowledge, this is the first report to demonstrate a HB-causing mechanism of a *F9* deep intronic structural variation. This deep intronic structural variation induces *F9* mRNA splicing aberration, which results in the generation of null phenotype transcripts with PTC and a reduction in the levels of the protein encoding *F9* mRNA.

Acknowledgments

We thank Mrs. Wakamatsu for her excellent technical assistance. This study was supported in part by grants-in-aid provided from the Japanese Ministry of Education, Culture, Sports, Science, and Technology (Grant No. 17H05073: S. Tamura, Grant No. 19K08853: A. Katsumi), the Japanese Ministry of Health, Labor, and Welfare (Grant No. 2017-012: T. Kojima, Grant No. JP17fk0410306: K. Nogami), the Japan Agency for Medical Research and Development (AMED, Grant No. JP20fk0410017: K. Nogami), Aichi Health Promotion Foundation (S. Tamura), and the National Center for Geriatrics and Gerontology (the Research Funding for Longevity Sciences, Grant No. 30-11: A. Katsumi). We would like to thank Enago for English language editing.

Authorship contributions

K. O., F. K., and S. T. designed and performed the research, analyzed the data, and drafted the manuscript. N. S. designed the project and collected and analyzed the clinical data. M. T. and Y. H. conducted the research and analyzed the data. A.S., T. Kanematsu., S. O., A.T., A.K., M. S., T. M., T. Kojima., and F. H. developed the project and collected and analyzed the clinical data.

Conflicts of interest

The authors have no conflicts of interest to declare associated with this manuscript.

References

[1] F. Peyvandi, I. Garagiola, G. Young, The past and future of haemophilia: diagnosis, treatments, and its complications, Lancet 388(10040) (2016) 187-97.

[2] S. Yoshitake, B.G. Schach, D.C. Foster, E.W. Davie, K. Kurachi, Nucleotide sequence of the gene for human factor IX (antihemophilic factor B), Biochemistry 24(14) (1985) 3736-50.

[3] CDC Hemophilia Mutation Project (CHAMP & CHBMP). https://www.cdc.gov/ncbddd/hemophilia/champs.html.

[4] EAHD Factor IX Gene (F9) Variant Database. https://f9-db.eahad.org/index.php.

[5] R. Vaz-Drago, N. Custodio, M. Carmo-Fonseca, Deep intronic mutations and human disease, Hum Genet 136(9) (2017) 1093-1111.

[6] R.D. Bagnall, N.H. Waseem, P.M. Green, B. Colvin, C. Lee, F. Giannelli, Creation of a novel donor splice site in intron 1 of the factor VIII gene leads to activation of a 191 bp cryptic exon in two haemophilia A patients, Br J Haematol 107(4) (1999) 766-71.

[7] G. Castaman, S.H. Giacomelli, M.E. Mancuso, G. D'Andrea, R. Santacroce, S. Sanna, E. Santagostino, P.M. Mannucci, A. Goodeve, F. Rodeghiero, Deep intronic variations may cause mild hemophilia A, J Thromb Haemost 9(8) (2011) 1541-8.

[8] H. Inaba, T. Koyama, K. Shinozawa, K. Amano, K. Fukutake, 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 19(1) (2013) 100-5.

[9] B. Pezeshkpoor, N. Zimmer, N. Marquardt, I. Nanda, T. Haaf, U. Budde, J. Oldenburg, O. El-Maarri, Deep intronic 'mutations' cause hemophilia A: application of next generation sequencing in patients without detectable mutation in F8 cDNA, J Thromb Haemost 11(9) (2013) 1679-87.

[10] C.Y. Chang, C.L. Perng, S.N. Cheng, S.H. Hu, T.Y. Wu, S.Y. Lin, Y.C. Chen, Deep intronic variant c.5999-277G>A of F8 gene may be a hot spot mutation for mild hemophilia A patients without mutation in exonic DNA, Eur J Haematol 103(1) (2019) 47-55.

[11] J. Feng, Q. Liu, J. Drost, S.S. Sommer, Deep intronic mutations are rarely a cause of hemophilia B, Hum Mutat 14(3) (1999) 267-8.

[12] K. Shinozawa, K. Yada, T. Kojima, K. Nogami, M. Taki, K. Fukutake, A. Yoshioka, A. Shirahata, M. Shima, J.H.I.S. study group on, Spectrum of F8 Genotype and Genetic Impact on Inhibitor Development in Patients with Hemophilia A from Multicenter Cohort Studies (J-HIS) in Japan, Thromb Haemost (2020).

[13] T. Kojima, M. Tanimoto, T. Kamiya, Y. Obata, T. Takahashi, R. Ohno, K. Kurachi, H. Saito, Possible absence of common polymorphisms in coagulation factor IX gene in Japanese subjects, Blood 69(1) (1987) 349-52.

[14] Human Genome Variation Society. http://www.hgvs.org/mutnomen/.

[15] Splicing site Site prediction Prediction. https://www.fruitfly.org/seq_tools/splice.html.

[16] Netgene2. http://www.cbs.dtu.dk/services/NetGene2.

[17] K. Odaira, S. Tamura, N. Suzuki, M. Kakihara, Y. Hattori, M. Tokoro, S. Suzuki, A. Takagi, A. Katsumi, F. Hayakawa, S. Okamoto, A. Suzuki, T. Kanematsu, T. Matsushita, T. Kojima, Apparent synonymous mutation F9 c.87A>G causes secretion failure by in-frame mutation with aberrant splicing, Thromb Res 179 (2019) 95-103.

[18] Acetone precipitation of proteins, Tech Tip #49. https://tools.thermofisher.com/content/sfs/brochures/TR0049-Acetone-precipitation.pdf?CID=Social_LAB.

[19] A. Ferlini, N. Galie, L. Merlini, C. Sewry, A. Branzi, F. Muntoni, A novel Alu-like element rearranged in the dystrophin gene causes a splicing mutation in a family with X-linked dilated cardiomyopathy, Am J Hum Genet 63(2) (1998) 436-46.

[20] H.R. Madden, S. Fletcher, M.R. Davis, S.D. Wilton, Characterization of a complex Duchenne muscular dystrophy-causing dystrophin gene inversion and restoration of the reading frame by induced exon skipping, Hum Mutat 30(1) (2009) 22-8.

[21] B. Baskin, W.T. Gibson, P.N. Ray, Duchenne muscular dystrophy caused by a complex rearrangement between intron 43 of the DMD gene and chromosome 4, Neuromuscul Disord 21(3) (2011) 178-82.

[22] M.M. Khelifi, A. Ishmukhametova, P. Khau Van Kien, D. Thorel, D. Mechin, S. Perelman, J. Pouget, M. Claustres, S. Tuffery-Giraud, Pure intronic rearrangements leading to aberrant pseudoexon inclusion in dystrophinopathy: a new class of mutations?, Hum Mutat 32(4) (2011) 467-75.

[23] M.G. Reese, F.H. Eeckman, D. Kulp, D. Haussler, Improved splice site detection in Genie, J Comput Biol 4(3) (1997) 311-23.

[24] S. Brunak, J. Engelbrecht, S. Knudsen, Prediction of human mRNA donor and acceptor sites from the DNA sequence, J Mol Biol 220(1) (1991) 49-65.

[25] D. Scalet, I. Maestri, A. Branchini, F. Bernardi, M. Pinotti, D. Balestra, Disease-causing variants of the conserved +2T of 5' splice sites can be rescued by engineered U1snRNAs, Hum Mutat 40(1) (2019) 48-52.

[26] A.R. Kornblihtt, Promoter usage and alternative splicing, Curr Opin Cell Biol 17(3) (2005) 262-8.

Figure Legends

Fig 1. Identification of *F9* intronic structural variation.

A) *F9* exon copy number analysis by MLPA. Data were normalized using pooled gDNA of healthy male subjects. B) Genomic long-range PCR with primer set, F9_Ex1-3_PCR_Fw vs. Rv. N: Pooled normal male gDNA, P: patient gDNA. C) Densitometry analysis of genomic long-range PCR amplicon. To examine the size alteration of PCR amplicon derived from the patient's gDNA, gel mobility of each amplicon was measured by densitometry analysis of the image presented in panel B. D) Breakpoint junction DNA sequencing. The patient carried a 28-bp deletion (from NC_000023.11: g.139,534,834 to g.139,534,861) and a 476-bp insertion (NC_000012.12: g.54,285,853_54,286,328inv) into *F9* intron 1. The 476-bp insertion was an identical sequence of a part of inverted *HNRNPA1* exon 12.

Fig 2. Cell-based transcript analysis of the *F9* intronic structural variation.

A) Diagram of the predicted anomalous splicing pattern in variant *F9* intron 1. This diagram was designed based on the splice-site prediction *in silico* using Splice Site Prediction and NetGene2 (Table 1). B) Design of variant *F9* exon-trap vector, pET01 F9 int 1 and 2 variant. The PCR-amplified DNA alignment of variant *F9* was inserted into a previously constructed pET01 F9 int 1 and 2 WT (15). C) Result of RT-PCR in COS-7 cells transfected with pET01 F9 int 1 and 2 WT or pET01 F9 int 1 and 2 variant. Abnormal amplicons of 553 bp (V1 transcript) and 729 bp (V2 transcript) were detected in cells transfected with the variant, and 448-bp amplicon (N1 transcript) was significantly reduced. D) Direct sequencing of RT-PCR amplicon. The N1 transcript (448-bp amplicon) was a normal-type (protein-coding) transcript composed of *F9* exons 1, 2, and 3. The V1 (553 bp) and V2 (729 bp) transcripts were aberrant transcripts containing pseudoexon(s). DNA sequences of the boundary region between exons or pseudoexons were displayed. The asterisk indicates a PTC within pseudoexon 2. pEx indicates a pseudoexon. E) Quantification of N1 transcript (normal *F9* mRNA) in COS-7 cells transfected with the variant exon-trap vector. The transcription diagram shows the design of the N1 transcript-specific primer set. The amount of N1 transcript was normalized with respect to the *GAPDH* expression level. n = 5 (independently transfected samples).

Fig 1. Identification of *F9* intronic structural variation.

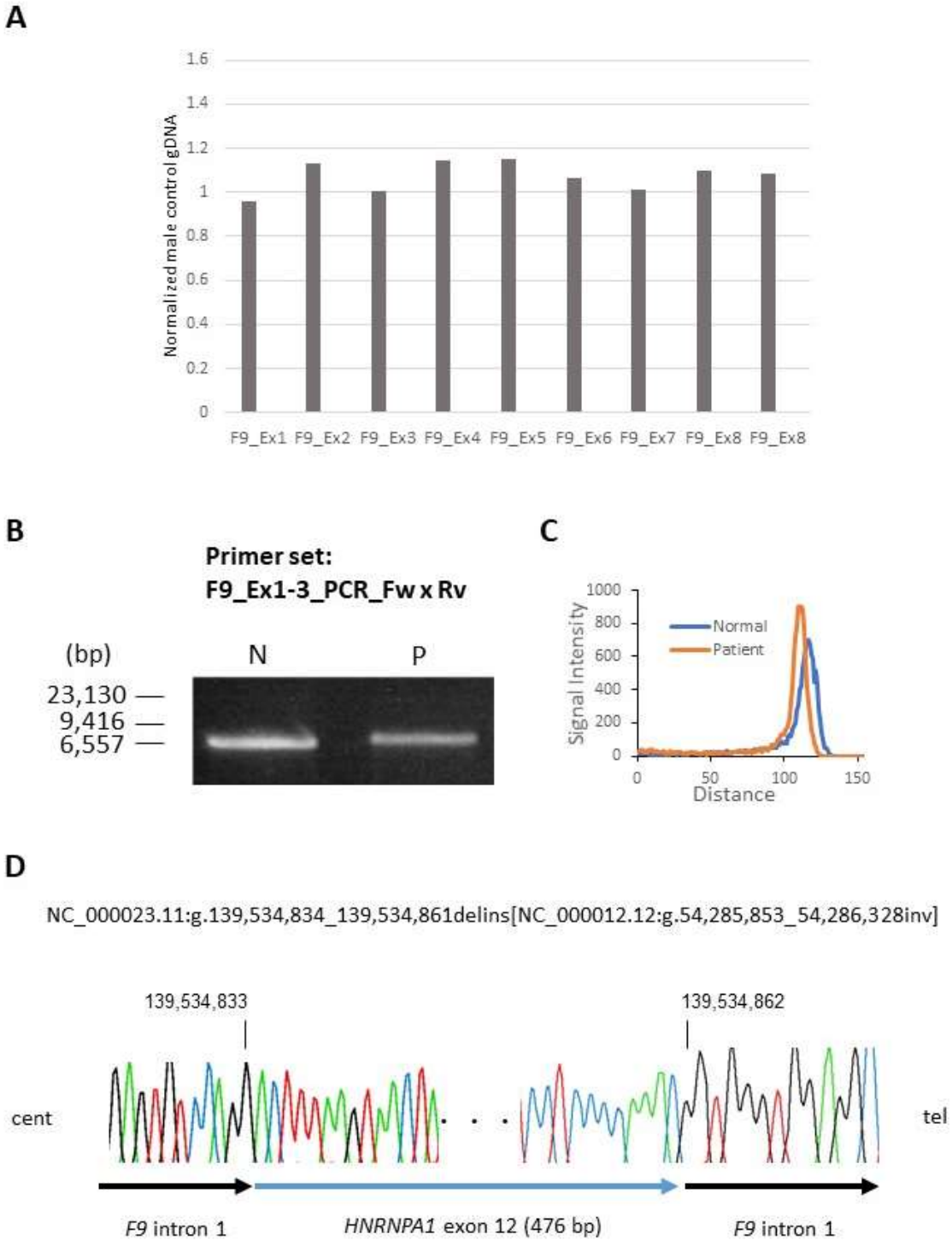


Fig 2. Cell-based transcript analysis of the *F9* intronic structural variation.

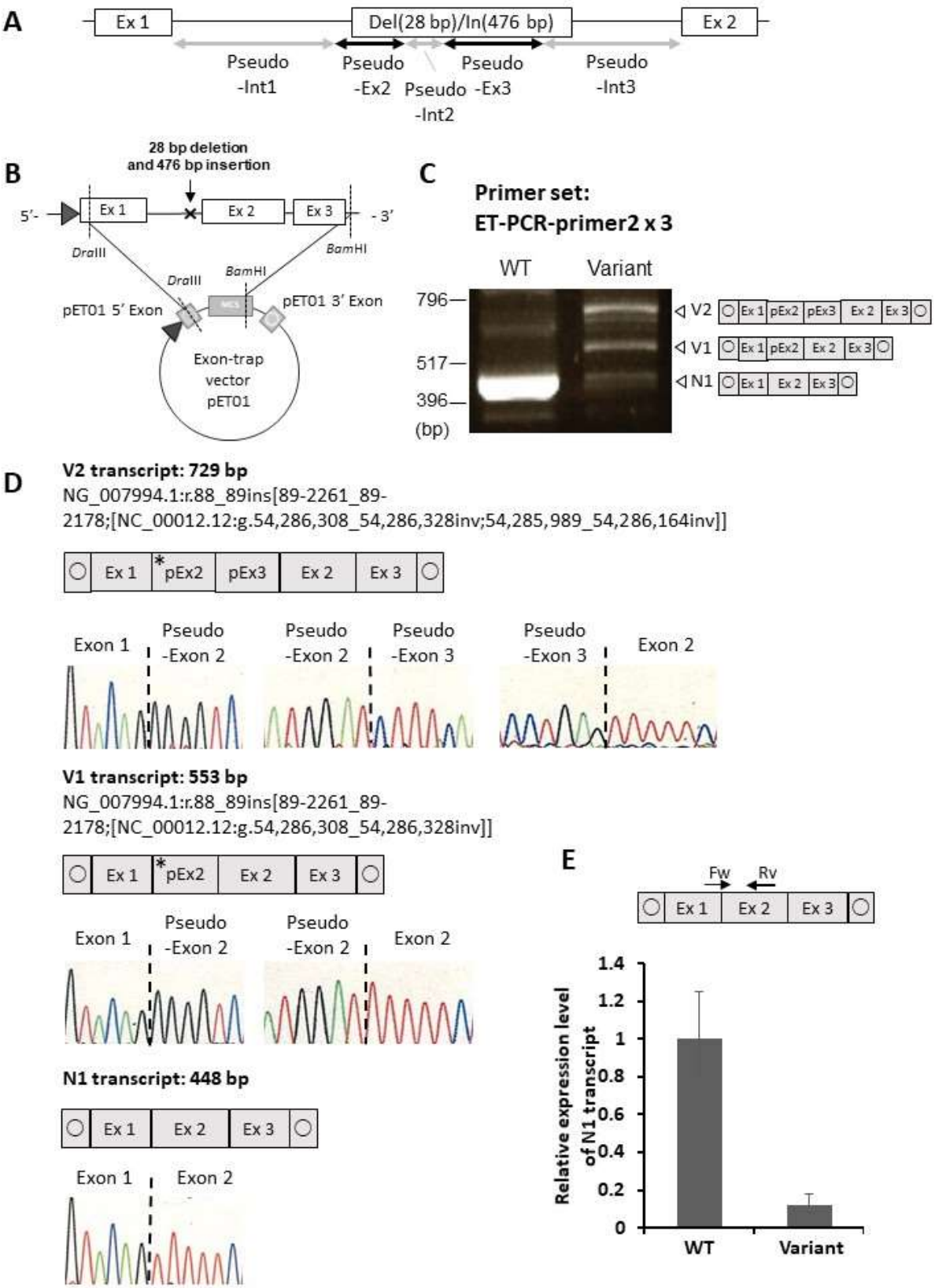


Table 1. *In silico* splicing simulation of *F9* structural variation

		pseudo -Int 1	pseudo-Int 2	pseudo -Int 3
Splice Site Prediction	Donor Site Score	0.83	0.96	0.60
	Acceptor Site Score	n.d	0.98	0.94
NetGene2	Donor Site Score	0.7	0.39	0.60
	Acceptor Site Score	0.16	0.56	n.d

Score ranged from 0.4 to 1.0 for Splice Site Prediction and from 0.0 to 1.0 for NetGene2. Higher scores indicate more potential splice sites.

n.d: not detected.

Koya ODAIRA

Supplementary data

Table of contents

Supplementary Figure Legends

Supplementary Fig. 1. Restriction enzyme fragment analysis.

A) Long-range genomic PCR amplicons ranging from *F9* 5′-untranslated region to intron 3 were digested with *Pst*I. Upper panel indicates an image of agarose gel electrophoresis analysis. Abnormal fragments (magenta arrowhead) were detected in the PCR amplicons derived from the patient's gDNA. In the control gDNA pooled with healthy male subjects, the *Pst*I-digensted fragments were detected as 1857 bp, 2280 bp, and 3012 bp. In the gDNA of the patient, the *Pst*I-digensted fragments were detected as 1857 bp, 2280 bp, and approximately 3500 bp. Lower panel is an illustration to explain the result of *Pst*I-digensted fragments analysis. (B) Long-range genomic PCR amplicons ranging from F9 5′-untranslated region to intron 3 were digested with *Hpa*II. Upper panel indicates an image of agarose gel electrophoresis analysis. Abnormal fragments (magenta arrowhead) were detected in the PCR amplicons derived from the patient's gDNA. In the control gDNA pooled with healthy male subjects, the *Hpa*II -digensted fragments were detected as 1479 bp, 2326 bp, and 3344 bp. In the gDNA of the patient, the *Hpa*II -digensted fragments were detected as approximately 2100 bp, 2326 bp, and 3344 bp. Lower panel is an illustration to explain the result of *Hpa*II -digensted fragments analysis. These observations indicated that an approximately 500 bp DNA alignment was inserted into the central part of the *F9* intron 1. N: Control gDNA pooled with healthy male subjects. P: patient's gDNA.

Supplementary Fig. 2. Genomic DNA sequence of breakpoint junction in the patient *F9*.

The sequences displayed are the DNA alignments of the patient's gDNA, the reference *F9* intron 1 (NC_000023.11:g.139534803_139534882), and the inserted sequence of *HNRNPA1* on chromosome 12 (NC_000012.12:g.54285853_54286328inv). Microhomology sequences are indicated by underlined characters.

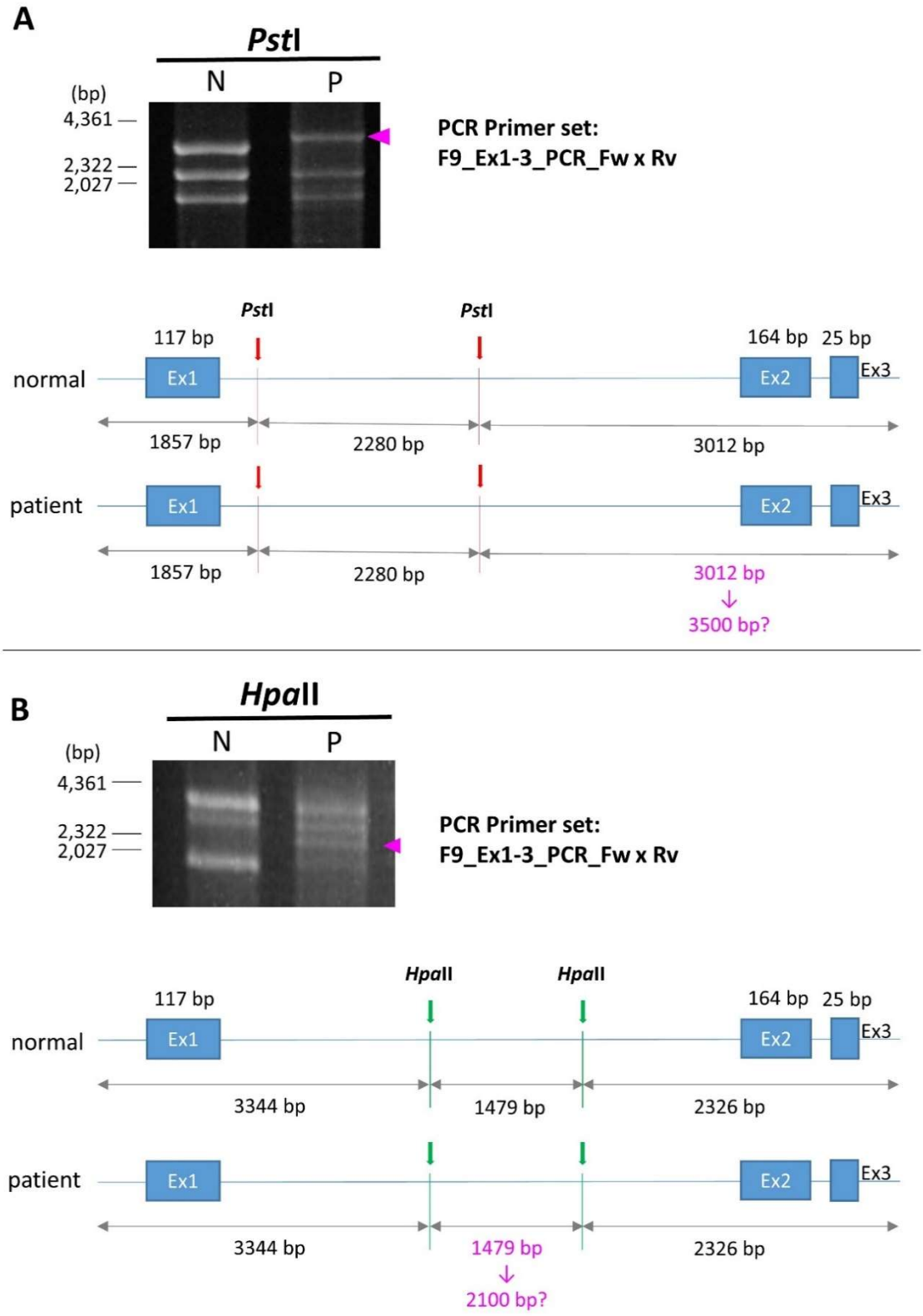
Supplementary Fig. 3. Result of cell-based transcript analysis using HEK293 cells.

A) Result of RT-PCR in HEK293 cells transfected with pET01 F9 int 1 and 2 WT or pET01 F9 int 1 and 2 variant. Abnormal amplicons were detected in cells transfected with the variant, and 448 bp amplicon (N1 transcript) was significantly reduced. B) Quantification of N1 transcript (normal *F9* mRNA) in HEK293 cells transfected with the variant exon-trap vector. The amount of N1 transcript was normalized with respect to the *GAPDH* expression level. n = 5 (independently transfected samples).

Supplementary Fig. 4. FIX protein expression analysis using splicing-competent vector with *F9* intron 1 variation.

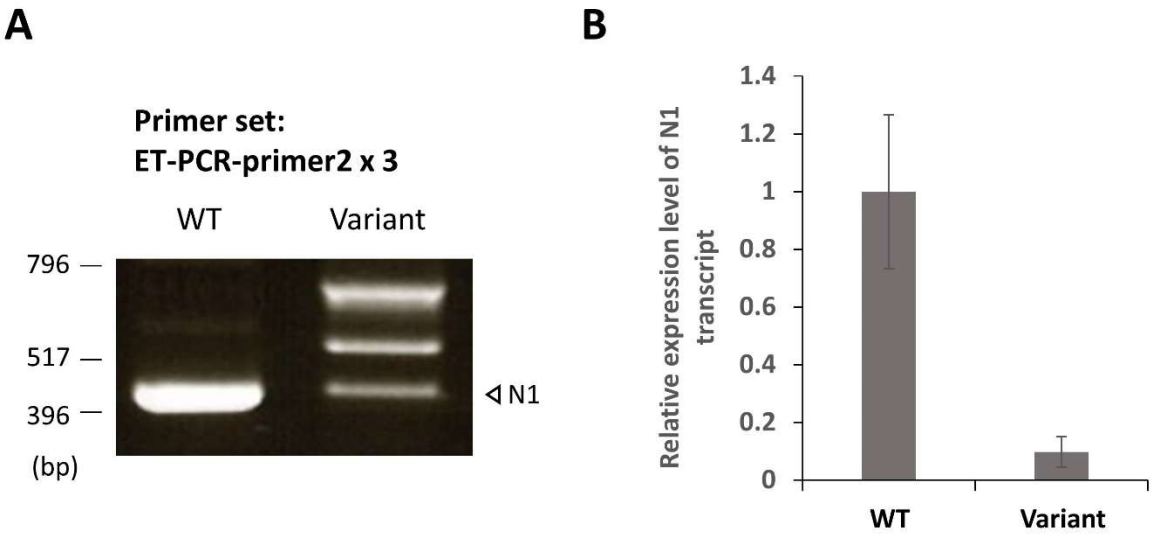
A) Design of a splicing-competent FIX expression vector. The PCR-amplified DNA alignment of normal or variant *F9* intron 1 was inserted into a pcDNA3.1-based FIX expression vector that we previously constructed [17]. B) Result of RT-PCR in HEK293 cells transfected with FIXwt int1(+) (WT) or FIXvariant int1(+) (Variant). Normal amplicon is indicated as N1 transcript. C) Quantification of N1 transcript in HEK293 cells transfected with FIXvariant int1(+). The amount of N1 transcript was normalized with respect to the *GAPDH* expression level. n = 4 (independently transfected samples). D) Western blot analysis of rFIX produced by HEK293 cells transfected with pcDNA3.1 (Mock), FIXwt int1(+) or FIXvariant int1(+). Here, 10 μg protein in cell lysate and 5 μg protein in culture media precipitation were loaded. a-tubulin was used as a loading control. E) Quantification of secreted rFIX from HEK293 cells transfected with FIXwt int1(+) or FIXvariant int1(+). n=4 (independently transfected samples). F) The specific coagulant activity of rFIXvariant. The specific activity was calculated as the ratio between the coagulant activity and the secreted protein level. The dotted line indicates the specific activity of rFIX-WT.

Supplementary Fig. 1. Restriction enzyme fragment analysis.

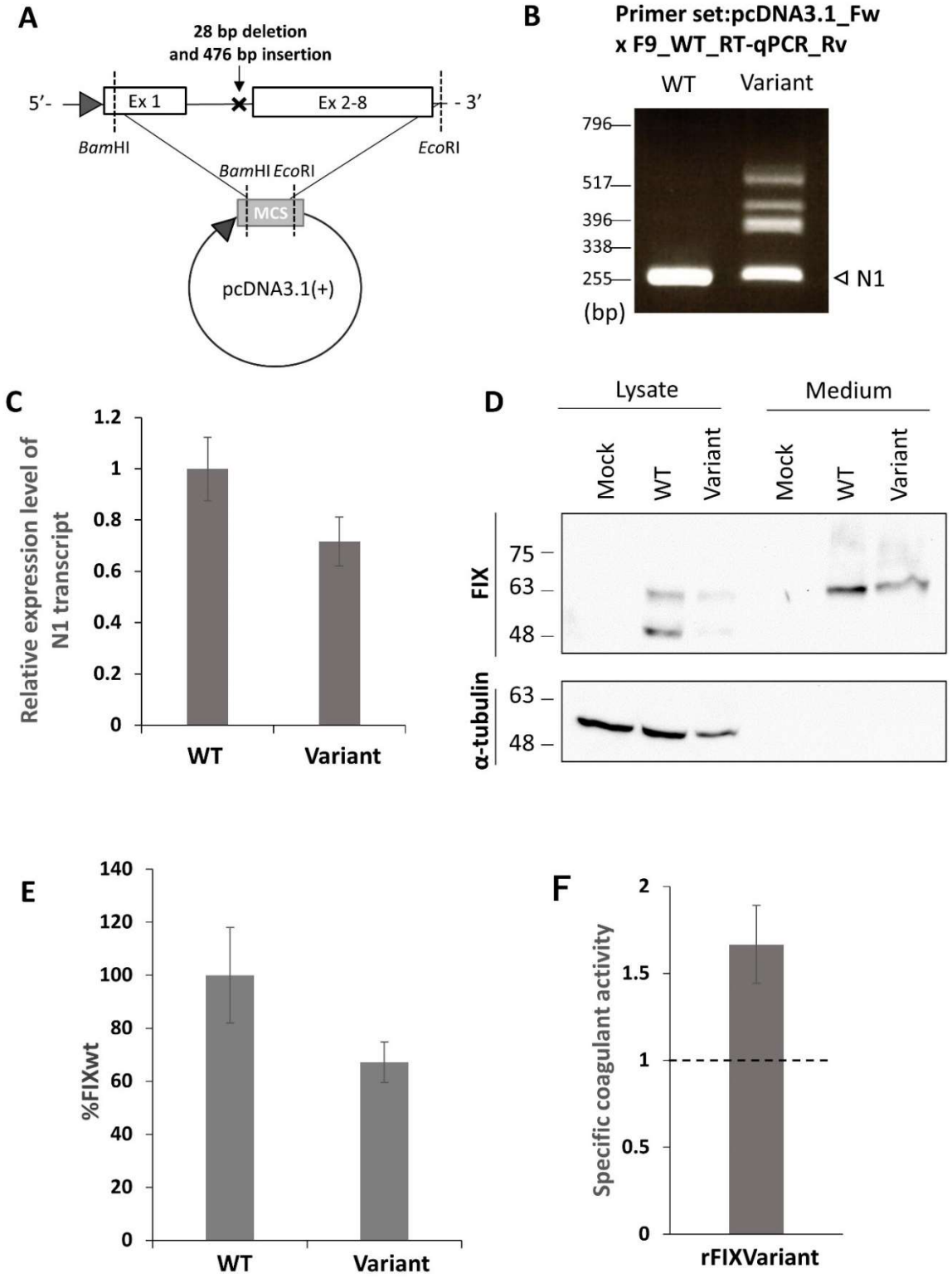


Supplementary Fig. 2. Genomic DNA sequence of breakpoint junction in the patient *F9*.

Supplementary Fig. 3. Result of cell-based transcript analysis using HEK293 cells.



Supplementary Fig. 4. FIX protein expression analysis using splicing-competent vector with *F9* intron 1 variation.



Supplementary table. Oligonucleotide primers.

Name of oligonucleotide	Sequence (5'-3')	Application
F9_ex1_Fw	AATCAGACTAACTGGACCAC	*F9* exon sequence
F9_ex1_Rv	TATCTAAAAGGCAAGCATAC	*F9* exon sequence
F9_ex2.3_Fw	ATGATGTTTTCTTTTTTGCT	*F9* exon sequence
F9_ex2.3_Rv	GGTTGGACTGATCTTTCTG	*F9* exon sequence
F9_ex4_Fw	TTCTAAGCAGTTTACGTGCC	*F9* exon sequence
F9_ex4_Rv	GTAGCTTCTTGAACTCATATCC	*F9* exon sequence
F9_ex5_Fw	CCCCCAATGTATATTTGACC	*F9* exon sequence
F9_ex5_Rv	CCGTCCTTATACTAGAAGCC	*F9* exon sequence
F9_ex6_Fw	AATACTGATGGGCCTGCT	*F9* exon sequence
F9_ex6_Rv	AACTTGCCTAAATACTTCTCAC	*F9* exon sequence
F9_ex7_Fw	CCAATATTTTGCCTATTCCT	*F9* exon sequence
F9_ex7_Rv	CTTCTGGTATGGAAATGGCT	*F9* exon sequence
F9_ex8-1_Fw	TGTGTATGTGAAATACTGTTTG	*F9* exon sequence
F9_ex8-1_Rv	TTATAGATGGTGAACTTTGTAGA	*F9* exon sequence
F9_ex8-2_Fw	TTGGATCTGGCTATGTAAGT	*F9* exon sequence
F9_ex8-2_Rv	AGTTAGTGAGAGGCCCTG	*F9* exon sequence
F9_Ex1-3_PCR_Fw	AATCAGACTAACTGGACCAC	Variant specific PCR in Fig. 1B
F9_Ex1-3_PCR_Rv	GGTTGGACTGATCTTTCTG	Variant specific PCR in Fig. 1B
Int1_fusion_Fw	ATAGTAAGCCATTTTTATATCGGAG	Construction of exon-trap vector in Fig. 2B
Int1_fusion_Rv	TTCACTCGTTTGCAATGCT	Construction of exon-trap vector in Fig. 2B
pET01_Int1wt_inversePCR_Fw	ATTGCAAACGAGTGAAGGAAATTGAGAAATATGG	Construction of exon-trap vector in Fig. 2B
pET01_Int1wt_inversePCR_Rv	AAAATGGCTTACTATTTGCATATAACCTAAACACATCCTC	Construction of exon-trap vector in Fig. 2B
FIX_int1_infusion_Fw	GTGCTGAATGTACAGGTTTGTTTCCTTTTTTAAAATACATTG	Construction of splicing-competent FIX expression vector in supplementary Fig.4A
FIX_int1_infusion_Rv	CATGATCAAGAAAAACTGAAATGTAAAAGAATAATTCTTTAGTTT	Construction of splicing-competent FIX expression vector in supplementary Fig.4A
FIX_ex1_inv_Rv	CTGTACATTCAGCACTGAGTAGATATCCTAAAAG	Construction of splicing-competent FIX expression vector in supplementary Fig.4A
FIX_ex2_inv_Fw	TTTTTCTTGATCATGAAAACGCCAACAAAA	Construction of splicing-competent FIX expression vector in supplementary Fig.4A
ET-PCR-primer2	GATCGATCTGCTTCCTGGCCC	RT-PCR in Fig. 2C, supplementary Fig.3A
ET-PCR-primer3	CTGCCGGGCCACCTCCAGTGCC	RT-PCR in Fig. 2C, supplementary Fig.3A
pcDNA3.1_Fw	AGTGCTTACTGGCTTATCGAAAT	RT-PCR in supplementary Fig. 4B
GAPDH_qRT_Fw	GGCTGCTTTTAACTCTGGTA	RT-qPCR in Fig. 2E, supplementary Fig. 3B, supplementary Fig. 4C
GAPDH_qRT_Rv	CATGGGTGGAATCATATTGG	RT-qPCR in Fig. 2E, supplementary Fig. 3B, supplementary Fig. 4C
F9_WT_RT-qPCR_Fw	TACTCAGTGCTGAATGTACAGTTTT	RT-qPCR in Fig. 2E
F9_WT_RT-qPCR_Fw2	CTCAGTGCTGAATGTACAGTTTTTC	RT-qPCR in supplementary Fig. 3B, supplementary Fig. 4C
F9_WT_RT-qPCR_Rv	TGAACAAACTCTTCCAATTTACCTG	RT-PCR in supplementary Fig. 4B, RT-qPCR in Fig. 2E, supplementary Fig. 3B, supplementary Fig. 4C