# Investigation of new mechanisms causing hemophilia A

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

## Behnaz Pezeshkpoor

aus

Ahvaz /Iran

Bonn, April 2013

**Eidesstattliche Erklärung**

Hiermit versichere ich, dass die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne die Benutzung anderer als der angegebenen Quellen angefertigt wurde.

Teile dieser Arbeit wurden in Form eines wissenschaftlichen Artikels veröffentlicht:

- *Hämostaseologie, Volume: 30, Issue: 4A/2010, Supplement: 1, S158-S161*

  Long-range PCR screening for large rearrangements in the FVIII gene in patients without detectable mutations in the coding sequence
  **Pezeshkpoor B**., Nüsgen N., Dermer H., Vidovic N., Niemann B., Pavlova A., Oldenburg J., El-Maarri O.

- *Hämostaseologie, Volume: 30, Issue: 4A/2010, Supplement: 1, S162-S163*

  *F8a*, *F8b* and *F8c* expression and its association with 3 SNPs within the *F8* gene in healthy controls and HA patients with no mutation.
  O. El-Maarri; **B. Pezeshkpoor**; N. Nüsgen; J. Müller; J. Oldenburg

- *Journal of Thrombosis and Haemostasis*, 10: 1600–1608,*2012*

  Identification of a third rearrangement at Xq28 that causes severe hemophilia A as a result of homologous recombination between inverted repeats.
  **Pezeshkpoor B**., Rost, S., Oldenburg, J. and El-Maarri, O.

- *Journal of Thrombosis and Haemostasis*, 11: 1679–1687,*2013*

  Deep intronic 'mutations' cause hemophilia A: application of next generation sequencing in patients without detectable mutation in F8 cDNA
  **Pezeshkpoor B**.,Zimmer N, Marquardt N, Nanda I, Haaf T., Budde U., Oldenburg, J. and El-Maarri, O.

Bonn, 2014                                                     Behnaz Pezeshkpoor

"Any intelligent fool can make things bigger, more complex, and more violent. It takes a touch of genius -- and a lot of courage -- to move in the opposite direction."

Einstein

*For my parents*

The key to wisdom is knowing all the right questions.

John A. Simone

# Table of contents

- Abbreviations

| | |
|---|---|
| % | Percentage |
| °C | degree Celsius |
| µl | micro liter |
| A1, A2, A3, B, C1 and C2 | domains of FVIII |
| aa | amino acid |
| AD | Anno Domini |
| aPTT | activated partial thromboplastin time |
| ass | acceptor splice site |
| ATP | adenosine triphosphate |
| AUG | start codon (methionin) |
| bp | base pairs |
| BSA | bovine serum albumin |
| CCD | Charge-coupled Device |
| cDNA | complementary DNA |
| Cen. | Centromere |
| CGH | comparative genomic hybridization |
| $CO_2$ | carbonic acid |
| cm | centi meter |
| CMV | cytomegalovirus |
| CNV | copy number variation |
| *CTAG1A* | gene encoding cancer/testis antigen 1A |
| *CTAG1B* | gene encoding cancer/testis antigen 1B |
| dbSNP | single nucleotide polymorphism database |
| DDAVP | 1-Desamino-8-D-Arginin-Vasopressin, Desmopressin |
| deaza-dGTP | 7´-deaza-deoxyguanosine triphosphate |
| DMSO | dimethyl sulfoxide |
| DNA | deoxyribonucleic acid |
| DNase | deoxyribonuclease |
| dNTP | deoxynucleotide triphosphate |
| dl | deciliter |
| dsDNA | double-stranded DNA |
| dss | donor slice site |
| DTT | dithiothreitol |
| *E.coli* | *Escherichia coli* |
| EDTA | ethylenediaminetetraacetic acid |
| ER | endoplasmatic reticulum |
| ELISA | enzyme-linked immunosorbent assay |
| et al. | et altera |
| EtOH | ethanol |
| *FUNDC2* | gene encoding FUN14 domain containing 2 |
| g | gram |
| *F8* | gene encoding FVIII |
| FII | coagulation factor II |
| FISH | fluorescent in situ hybridization |
| FIX | coagulation factor IX |

| | |
|---|---|
| FV | coagulation factor V |
| FVIII | coagulation factor VIII |
| gDNA | genomic DNA |
| h | hours |
| HCl | Hydrochloride |
| hg | human genetics |
| HPLC | High-performance liquid chromatography |
| HPSF | High purity salt free |
| *IKBKG* | gene encoding inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase gamma |
| *IKBKGP* | gene encodine *IKBKG* pseudogene |
| IRES | internal ribosome entry site |
| kb | kilobase pairs |
| l | liter |
| LB | Luria-Bertani |
| LGT | low gelling temperature |
| LR-PCR | long range PCR |
| LMAN1 | lectin mannose-binding 1 |
| LRP | low density lipoprotein receptor-related protein |
| m | milli |
| M | molar |
| mA | milliampere |
| MCFD2 | multiple coagulation factor deficiency protein 2 |
| $MgCl_2$ | Magnesium chloride |
| min | minutes |
| MLPA | Multiplex ligation-dependent probe amplification |
| mRNA | messenger RNA |
| NaAc | natrium acetate |
| NCBI | National Center for Biotechnology Information |
| ng | nano gram |
| NGS | next generation sequencing |
| nm | nano meter |
| nt | nucleotide |
| OD | optical density |
| OMIM | Online Mendelian Inheritance in Man |
| ON | over night |
| pmol | Pico mole |
| ORF | open reading frame |
| P/S | penicillin/streptomycin |
| PBS | phosphate buffered saline |
| pH | potential of hydrogen |
| PCR | polymerase chain reaction |
| PT | Prothrombin time |
| qRT-PCR | quantitative real-time PCR |
| RNA | ribonucleic acid |
| rpm | revolutions per minute (U/min) |
| RTase | reverse transcriptase |
| RT-PCR | reverse transcriptase polymerase chain reaction |
| s | seconds |
| SDS | sodium dodecylsulfate |

| | |
|---|---|
| SOC | Super Optimal culture medium |
| SSC | saline-sodium citrate |
| SV40 | simian virus 40 |
| TAE | Tris-acetate-EDTA |
| Tel. | Telomere |
| TF | tissue factor |
| U | units |
| uAUG | upstream AUG |
| UV | ultra violet |
| UTR | untranslated region |
| V | volt |
| vWF | von Willebrand factor |
| VWD | von Willebrand disease |
| WT | wild type |
| w/v | weight per volume |
| X-Gal | 5-bromo-4-chloro-3-indolyl-$\beta$-D-galactoside |

# 1    Summary

The main aim of this work was identification of novel molecular mechanisms causing hemophilia A, with a special focus on patients without detectable mutations. In most of hemophilia A patients, mutations could be identified in the coding sequence of *F8*. However, despite tremendous improvements in mutation screening techniques, in a small number of patients no mutation is found. Therefore, the focus of this work was to analyze changes in regions of *F8* that are usually ignored in the routine screening. These variations involve regions such as introns that are not assessed by current routine molecular diagnostic techniques.

In this study a multistep experimental protocol was designed to screen the whole *F8* locus for novel genetic abnormalities and alternations. This protocol combined several mutation screening tools. Multiple ligation-dependent probe amplification and comparative genomic hybridization were used to detect copy number variations. Next generation sequencing (NGS) was applied in combination with long range (LR-) PCRs to re-sequence the whole genomic region of *F8*. The LR-PCR approach identified two novel breakpoints. Moreover, several unique deep intronic variations were found. Later, mRNA analysis was performed to prove the causality of these intronic variations.

Using this experimental strategy, causative mutations in most of the patients were found and characterized. Interestingly, all causal mutations were within the *F8* locus. The results of this work show once more the variety and complexity of mutations causing hemophilia A. Moreover, the data presented here show the association of deep intronic mutations with hemophilia A phenotype. To diminish the number of mutation negative patients, there is a need to apply several techniques. Based on the results of this work, an extended diagnosis flowchart is proposed. This flowchart demonstrates techniques for successful identification and confirmation of novel mutations in hemophilia A patients and could serve as a model for genetic analysis of other monogenic diseases, where no pathogenic mutation is found.

# 2    Introduction

Hemophilia A is a common X-linked recessive bleeding disorder (OMIM 306700). Typically, affected patients experience prolonged bleeding caused by lack or reduced activity of the coagulant Factor VIII (FVIII:C). Various types of mutations in the factor 8 gene (*F8*) are known to be responsible for the FVIII deficiency. Despite tremendous improvements in mutation screening methods, in a small group of hemophilia A patients no DNA change could be found. The presented work includes *F8* mutation analysis in hemophilia A patients beyond routine diagnosis to identify underlying mechanisms leading to FVIII deficiency in these patients.

## 2.1   Hemostasis and the blood coagulation cascade

Hemostasis is the protective physiological response to injury that results in exposure of blood to the subendothelial layers of the vessel wall. Hemostasis is a complicated and highly regulated process involving several coagulation factors and cofactors present in blood plasma. The generation of fibrin depends upon the exquisitely controlled interplay of the following cellular components: five proteases (Factor 7 (FVII), Factor 9 (FIX), Factor 10 (FX), protein C and prothrombin) interacting with five cofactors (tissue factor (TF; CD142)), thrombomodulin, protein S, Factor 5 (FV) and FVIII). The coagulation factors are designated by Roman numerals ranging from I to XIII with a lowercase "a" indicating the active form (reviewed in [1-3]). The role of each coagulation factor is best illustrated by the classical coagulation cascade (Figure 1), which consists of two distinct pathways; the contact pathway (also known as the intrinsic pathway), and the tissue factor pathway (also known as the extrinsic pathway), were first described 1964 by two independent groups [4, 5]). These pathways are assayed clinically using the activated partial thromboplastin time (aPTT) and the prothrombin time (PT), respectively. Factor 12 (FXII) and Factor VIIa (FVIIa) initiate the intrinsic and extrinsic pathways, respectively.
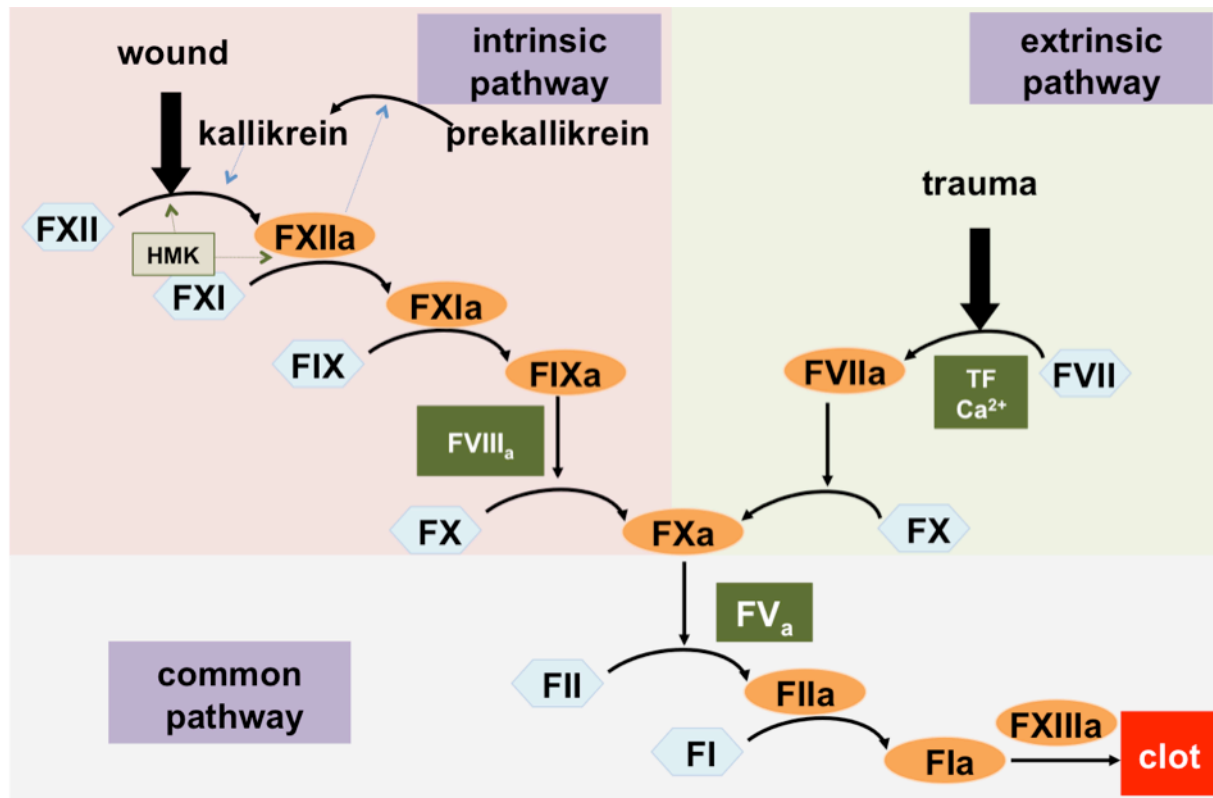
Figure 1: **Simplified scheme of the coagulation cascade and clot formation via intrinsic and extrinsic pathways.** The two pathways merge after the formation of FXa. The coagulation factors are abbreviated with capital F followed by their Roman numbers, HMK: high molecular weight kininogen, a: activated, $Ca^{2+}$= Calcium, TF=tissue factor.

After injury of a blood vessel, the coagulation cascade is initiated by the contact of blood with TF, a non-enzymatic lipoprotein present in the membrane of cells surrounding the vascular bed. Once in contact with blood the TF-dependent activation of coagulation, i.e. the extrinsic pathway, is initiated. When a vessel is injured, membrane bound TF is exposed. As a consequence the circulating FVII is activated, and binds to TF and results in formation of the FVIIa-TF complex. This complex acts as a cofactor and in converts circulating FIX and FX to the active serine proteases FIXa and FXa, respectively, which in turn activates TF-bound to FVII in a feedback mechanism (Figure 1). FXa works with its cofactor activated Factor V (FVa). A small amount of prothrombin activates thrombin, which works in a positive feedback loop itself. Thrombin can activate FV and FVIII [6], which activates FXI. FXI itself activates in turn FIX and allows activation of more thrombin. This will lead to releases of FVIII from being bound to von Willebrand Factor (vWF). Subsequently, after sufficient amount of thrombin is present in the circulation, fibrinogen is activated

and becomes fibrin. In the presence of activated factor 13 (FXIIIa), fibrin is polymerized and a fibrin clot is formed at the site of injury [7].

The described extrinsic pathway is responsible for trauma-initiated coagulation, whereas the physiological importance of the alternative "intrinsic" pathway is not fully understood. However, when activated, the procoagulant activities of the coagulation system are increased. The intrinsic pathway begins with formation of the primary complex on collagen which involves high molecular weight kininogen (HK)-dependent prekallikrein (PKK)/kallikrein (KK) regeneration cycle for FVIIa-dependent FXII activation (Figure 1). Through the sequential activation of several zymogens (FXII, FXI, FIX and FX), prothrombin is activated via the prothrombinase complex. FXII is activated and converts FXI into FXIa, following the activation of FIX which forms the tenase complex with its cofactor FVIIIa. At this point the intrinsic tenase shunts into the extrinsic pathway by generation of FXa [8].

## 2.2   Hemophilia A

Hemophilia A which is the most common bleeding disorder has a occurrence of 1 in every 5,000 males [9]. It is caused by lack or defective function of the coagulant FVIII protein. Endogenous FVIII is present in trace amounts of 100-200 ng/mL [10] and its blood-circulating form is bound to vWF. The severity of disease is related to the amount of residual FVIII. Bleeding occurs mainly in joints and muscles as well as internal organs. Although most of the factors involved in the coagulation cascade have been discovered in the 20$^{th}$ century, the earliest written references concerning hemophilia are attributed to Jewish writing of the 2$^{nd}$ century AD [11]. However, first in the 18$^{th}$ century, the three important characteristics of the disease where defined; bleeding, male gender and heredity [12]. Hemophilia has often been called the "Royal Disease" since Queen Victoria of England was carrier of hemophilia and passed the disease through her daughters to the Spanish, German, and Russian royal families [13].

The disease is classified as severe (FVIII:C< 1%), moderate (1%< FVIII:C< 5%), and mild (5%< FVIII:C< 35%) which represent about 40%, 10%, and 50% of patients with

hemophilia A, respectively [14, 15]. The severity of disease depends on the effect of the mutation on FVIII:C and protein levels. Although it is rare for women to suffer from hemophilia A, there are several reports describing molecular mechanisms behind the phenotype in females. This is caused by whether unbalanced X chromosome inactivation in [16-18] or *de novo F8* mutations [19-22] in obligate carriers of hemophilia A.

Broad spectrums of mutations cause a qualitative or quantitative defect leading to hemophilia A. Severely affected patients experience recurrent spontaneous excessive bleeding after trauma due to a lack of FVIII protein. The most common *F8* mutations leading to a severe phenotype are inversions. One in 30,000 males is born with an inversion in *F8* with a residual FVIII:C less than 1%. In about 90% of such severe hemophilia A cases, patients have an inversion involving intron 22. In the remaining 10%, an inversion of intron 1 is responsible for the severe phenotype [23, 24]. Rest of the patients carry a variety of missense, nonsense, and splice site mutations, as well as deletions and insertions [25]. Cases of moderate and mild phenotype are mainly due to missense mutations [26]. These patients are suffering from frequent bruising and bleeding into joints (mainly knee, ankle, hip and elbow).

Both, life expectancy and quality of life in hemophiliacs improved after therapy with fresh whole blood or fresh frozen plasma was available. After low success of the first trials with porcine and bovine plasma in the beginning of 1950s due to allergic reactions, human-derived plasma was successfully used to treat patients. However, as a consequence of receiving plasma from donors without an adequate donor screening and infective agent testing, a large number of hemophiliacs became infected with hepatitis C virus [27] and human immunodeficiency virus (HIV) [28]. At this time, AIDS became the major cause of death in hemophiliacs whereas until 1970s-1990s intracranial bleeding was the main cause. Later, additional virus inactivation steps (heating or solvent treatment) were included in the manufacture of plasma-derived FVIII products. Although plasma-derived FVIII replacement therapy is still used, an even safer option for the therapy became available after Gitschier et. al [29] and Toole et. al [30] independently cloned and characterized *F8* in 1984. Ten years later, recombinant FVIII (rFVIII) products were produced in cell culture, using **b**aby **h**amster **k**idney (BHK) or **c**hinese **h**amster **o**vary (CHO) cells. Since then rFVIII

products are widely used as a major source of the protein for the treatment. However, this therapy is extremely expensive due to the cost of *in vitro* production and purification. The variety of fractionation techniques used for the production of plasma-derived FVIII products make these also quite expensive. Because of the high costs and limited availability of FVIII products (both plasma-derived and recombinant), most of the patients worldwide are not under prophylaxis treatment.

Another alternative therapy method is gene therapy. Hemophilia A, as a monogenic disease, is a good candidate for gene therapy. This method has several advantages in comparison to the conventional substitution therapy. It is a safe and effective long-term treatment, which will avoid infection risks. Moreover, the disease can be corrected with as low as only 1-2% of FVIII levels. However, most of the clinical trials using different gene delivery systems [31, 32] do not guarantee to achieve the long-term availability of circulating FVIII in plasma, but for many patients even a temporary improvement of hemostasis would could be considered as a success. However, lack of long-term expression and the absence of a safe and effective viral delivery system are the limitations that still need to be improved.

## 2.3   Genomic organization of *F8*

In 1984, after cloning and characterization of the 'large' gene encoding FVIII, for the first time [29, 30], the precise and direct mutation screening of patients with hemophilia A became available. The genomic region of *F8* spans 186 kb and codes for a mRNA of 9.1 kb. It is mapped to the most distal band of the long arm of the X chromosome (Xq28) on the minus strand. The gene consists of 26 exons (Figure 2). Comparing all exons, which vary in size from 69 bp to 262 bp, exon 26 and exon 14 are remarkably huge with a size of 1959 bp and 3106 bp, respectively. Exon 26 mostly codes for the 3´ untranslated region (UTR), and exon 14 encodes the large B domain of FVIII protein. The introns vary in size between 200 bp and 32.5 kb (reviewed in [25]).
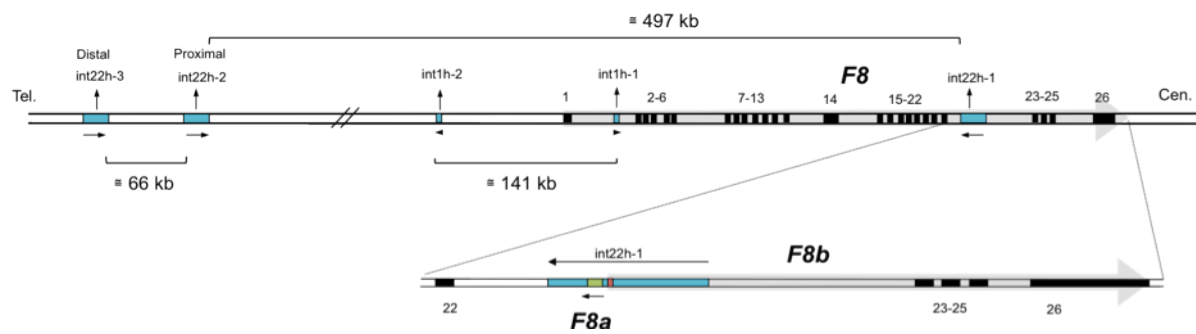
Figure 2: **Genomic organization of *F8* at Xq28.** The black arrows show the orientation of transcription of *F8a*, as well as the intron 1 and intron 22 homologous regions within and outside of *F8*. The gray arrows indicate the orientation of transcription of *F8* and *F8b*. The intron 1 repeats (int1h-1 and int1h-2) are involved in the intron 1 inversion. The two intragenic transcripts in the *F8a* and *F8b* are shown in a magnification. The first exon of each gene is shown with green and red blocks, respectively. Graph is not to scale. Modified from Oldenburg and El-Maarri [25].

Interestingly, the intron 22 of *F8* is not only very large (32 kb), but it also contains a CpG island that acts as a bidirectional promoter for two additional genes nested in the body of *F8*. The intronless *F8a* (OMIM 305423) and *F8b* (OMIM 305424). *F8a* is transcribed in the opposite and *F8b* in the same direction to *F8* [33, 34].

*F8a* codes for a protein called huntingtin-associated protein (HAP40), a 40 kD protein [35] that interacts with huntington protein. *F8b* is spliced to exons 23 through 26 by creating a transcript of 2.5 kb with the *F8* reading frame. The protein encoded by *F8b* transcript would include the C2 domain of FVIII, which is the phospholipid-binding region and essential for the coagulant activity. However, the biological function of *F8b* is obscure due to the fact that patients with a deletion comprising this part of the gene do not suffer from any other disease than hemophilia A. Transcripts of both genes have been detected in many tissues. However, the function of these two genes is still not well understood [33-35].

Of interest is also the location of these genes (*F8a* and *F8b*). Their promoter is located within a 9.5 kb region, about 5.78 kb downstream of exon 22 of *F8*, designated as intron 22 homologous region 1; int22h-1 [36] (Figure 2). This region is present in two other copies distal to 5´ end of *F8*: int22h-2 and int22h-3 that are located approximately 300 kb and 400 kb distal to *F8*, respectively. These homolog regions are hotspots of intrachromosomal recombination resulting in inversion of the intervening sequences, which interrupts *F8* at intron 22. This mutation is responsible

for 50% of severe hemophiliacs [24]. A similar pair of inverted repeats is present within intron 1 of *F8* about 15.26 kb downstream of exon 1 (annotated as int1h-1) and 125 kb upstream of exon 1 (annotated as int1h-2) [23] (Figure 2). This repeat is a ~1 kb element. Inversions involving this region are the causative mutation in 5 % of patients with severe hemophilia A. This inversion results in a short *F8* mRNA consisting of only exon 1 of *F8*. Both inversions can be diagnosed using PCR based mutation screening methods [37].

## 2.4  Molecular basis of hemophilia A

The spectrum of mutations causing hemophilia A includes a wide range of mutations scattered through the whole genomic region of *F8*. Most of the mutations are listed in the hemophilia A mutation database (HAMSTeRS) at http://hadb.org.uk/. In addition to the inversions described above, which account for gross rearrangements, various types of point mutations: missense (amino acid substitution), nonsense (premature termination of the peptide chain), and splice site mutations (altered mRNA), small and large deletions, insertions, and duplications have been reported. Another group of mutations include C>T and G>A substitutions at CpG dinucleotide hotspots. These mutations are caused by spontaneous methylation-induced deamination of the 5-methylcytosine [38, 39]. Hotspots for such mutations are codons 1689, 1941, 1966, 2147, 2150, 2159, and 2163. Small deletions and insertions caused by polymerase slippage errors in series of adenine nucleotides (A stretches; As) are another category of mutations. The main hotspot of these mutations are codon 1191 to 1194 (9 As), codon 1439 to 1441 (8 As), and 1588 to 1590 (6 As) [25]. Noteworthy, all of these mutations occur in exon 14. Furthermore, several deletions and duplications of *F8* locus have been reported. However, only few cases have been characterized in detail [40, 41]. Most of the mutations can be identified using the routine PCR-based analysis methods with a mutation detection rate of up to 97% [25]. In addition, mRNA analysis of the splicing variants in peripheral blood allows the detection of mutations at mRNA levels in patients with splicing defects [42, 43]. Although most of causative mutations leading to hemophilia A phenotype are in *F8*, mutations in interaction partners of FVIII considering the transport of the protein including lectin mannose-binding protein (LMAN1), multiple coagulation factor deficiency 2 protein (MCFD2),

and clearance of the protein such as low-density lipoprotein receptor-related protein (LRP) should be considered [44-46]. Moreover, mutations in vWF affecting its binding to FVIII, the so called vWF type 2N (Normandy) mutations, can reduce the half life of FVIII protein dramatically [47]. Interestingly, in about 2% of patients with hemophilia A, no mutation can be found at the genomic level [42, 48]. In addition mRNA analysis of most of these patients do not reveal any apparent mutation in coding sequences as well as promoter and 3´ UTR [42].

## 2.5 FVIII protein

The main sites of FVIII production in the liver are sinusoidal endothelial cells and hepatocytes [49]. Although several tissues express FVIII, the secretion of this protein is only limited to a number of cells [50, 51]. Endogenous FVIII is found in relatively low levels in blood (100-200 ng/ml corresponding to 0.35-0.7 nM). The primary *F8* cDNA translation product is a polypeptide of 2351 amino acid (aa) residues. The first 19 amino acids encode a hydrophobic signal peptide, which is removed during secretion and transport to endoplasmatic reticulum (ER). Finally, a mature glycoprotein of 2332 amino acids with a molecular weight of 267 kDa is synthesized. The mature protein consists of several main domains: three A domains, a large B domain, and two C domains (Figure 3). The largest domain is the B domain with 908 aa residues. The sizes of the A1, A2, and A3 domains are 336, 345, and 335 aa residues, respectively. The C1 and C2 domain are the smallest domains of FVIII protein with 155 and 152 aa residues, respectively. Three short acidic regions (a1 to a3) are rich in negatively charged residues. The a1 and a2 domains follow the A1 and A2 domains, respectively and connect these domains to the amino-terminal of B domain. The a3 domain is located before A3 domain and connects this domain to the carboxy-terminal end of B domain. Sequence analysis reveals a domain organization of A1-a1-A2-a2-B-a3-A3-C1-C2 (reviewed in [52]).
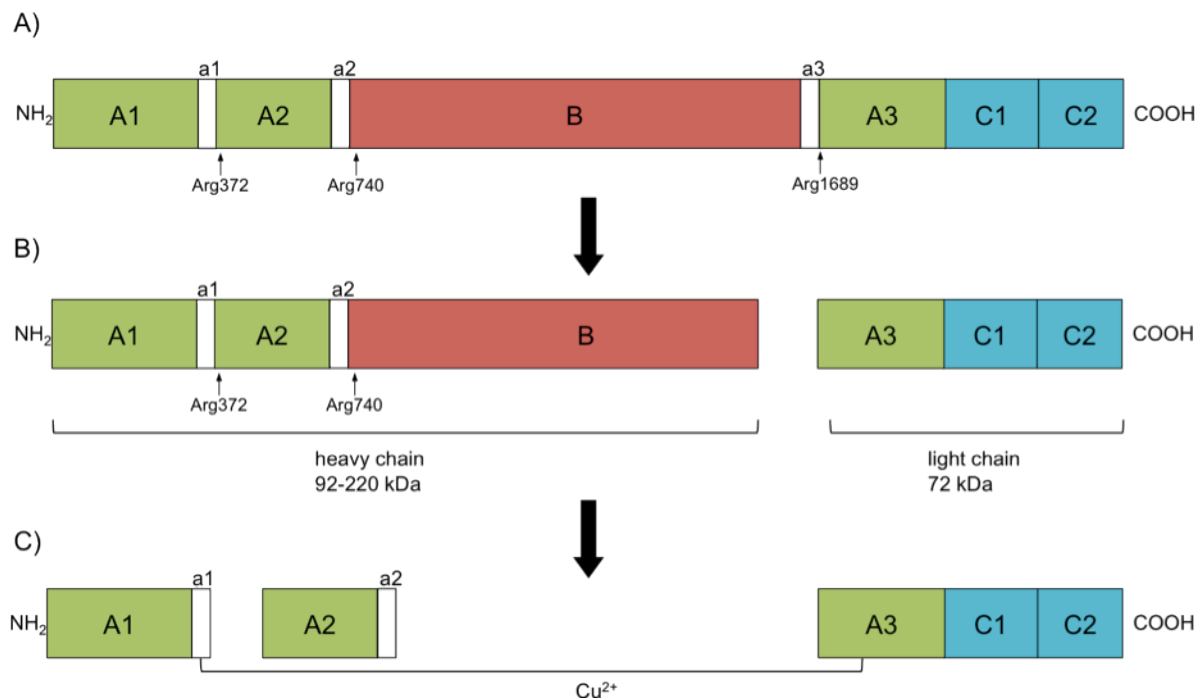
Figure 3: **Domain structure of FVIII protein showing the major interaction sites.** A) The precursor FVIII protein is 2351 aa (267 kDa) which is processed to a final product of 2332 aa that lacks 19 aa from its N terminus. B) This product will be cleaved by thrombin first at Arg 1689, which is the B-A3 junction. This cleavage results in a variably sized heavy chain (92-220 kDa) that consist of A1 and A2 domains and the partially proteolysed B domain. C) The A3 domain is as well released from the light chain (72 kDa) which consists of A3, C1 and C2. The final cleavage by thrombin results in complete removal of the B domain (Arg 740) and cleavage of the protein between A1 and A2 domains at position (Arg 372). The activated protein is held through a divalent metal ion-dependent interaction.

FVIII is highly homologous to FV [53]. Interestingly, the heavily glycosylated B domain is not homologous to FV B domain. In contrast, the three A domains share about 35 % of homology to each other and also to the A domains of ceruloplasmin [54]. Furthermore, the C domains share 35% identity to each other and 55% to the discoidin family of proteins, suggesting their interaction with phospholipids. The amino acid homology between FV and FVIII suggests that these genes originate from a duplication process and exon shuffling [55].

Prior to secretion into plasma, FVIII is intracellularly cleaved, at Arg1648 and Arg1313, within recognition motifs for the subtilisin-like protease PACE/furin [56]. FVIII is further cleaved within the B domain. This cleavage results in a heterodimer consisting of a heavy chain (a peptide of 92- 220 kDa consisting of A1-a1-A2-a2 domains with the B domain in variable extensions) and a light chain (a peptide of 72

kDa composed of the a3-A3-C1-C2 domains). Both chains are linked through a divalent metal ion-dependent interaction ($Cu^{2+}$ ions, the coordination with further ions such as $Ca^{2+}$ and $Mn^{2+}$ is also essential for maximal specific activity [57]) via A1 and A3 domains.

In plasma, FVIII heterodimers circulate noncovalently bound to vWF via the FVIII a3 domain. These heterodimers are activated by thrombin (FIIa) and cleaved at 3 arginine residues: Arg372, Arg740, and Arg1689. The active FVIII protein has a domain structure of A1-a1, A2-a2, and A3-C1-C2. This active FVIII is extremely unstable [58]. However, within the protein there is a stable metal-ion link within the light chain dimer, which is weakly associated with the A2 domain through electrostatic interactions. Inactivation of FVIII occurs via cleavage at two other arginine residues: Arg336 and Arg1721 by FXa, at Arg336 and Arg1719 by FIXa, or by activated protein C (APC) at Arg336 and Arg562. Interestingly, FIX has a dual role both in stabilization and inactivation of FVIII: beside its role in protection of FVIII against inactivation by APC, it can link the A2 and A3 domains due to its interaction with the active protein. On the other hand, FIX can cleave FVIII as well. Noteworthy is the protective role of vWF, which not only results in inhibiting phospholipids to interact with FVIII, but also helps by occupying the protease binding sites [59].

Two major proteins are known to play a role in clearance of FVIII: the low-density lipoprotein receptor (LDLR) and the hepatic low-density lipoprotein receptor-related protein (LRP). This clearance process is enhanced by heparan sulfate proteoglycan (HSGP) mediating the binding of FVIII to the cell surface [60]. The different sites for activation, inactivation and interaction of FVIII, highlights the complexity of regulation between hemostasis and thrombosis.

## 2.6 Determinants of plasma FVIII levels

Plasma FVIII activity (FVIII:C) is a highly variable quantitative trait. In healthy individuals several factors have been reported to be associated with elevated FVIII levels. While only one gene, the AB0 blood-group locus, has been shown to be associated with this broad variability [61-63], also vWF is a main determinant of

FVIII:C in plasma [62, 64]. The AB0 blood group account for variations in vWF levels as well. Evidence of the supporting role of factors such as age (an increase of 5-6 IU/dl per decade) [65, 66], body mass index (positively correlated with FVIII:C) [66], diabetes mellitus [67], sex ( FVIII:C is higher in women than men) [66, 68], race (FVIII:C is higher in blacks than whites) [69] oral contraception [70, 71], smoking [66, 72], lipid fractions [60, 67] exist. Moreover, several stimuli can cause a transient increase of FVIII:C such as exercise (as a possible result of adrenalin and $\beta_2$-adrenoreceptor stimulation) and vasopressin that can enhance FVIII levels directly or indirectly via the V2 receptor [73]. There are also reports of an increase in FVIII levels during pregnancy, surgery, liver disease and chronic inflammation. In most of these conditions, there is a simultaneous increase of FVIII and vWF levels [74]. Due to the homology between FV and FVIII and the fact that they share common biosynthetic pathways, the FV antigen (FV:Ag) levels correlate to some extent with FVIII:Ag levels. All in all, due to a multitude of variants that could influence FVIII levels in plasma, *F8* is not the only determinant that should be considered in cases with reduced FVIII:C levels.

## 2.7  Patients without mutation in *F8* gene

In about 98% of all hemophilia A patients, the molecular defect is within the *F8* coding sequence (cDNA). So far, more than 900 causal mutations in *F8* cDNA have been reported (http://hadb.org.uk). In addition, in about 10% of mild hemophilia A cases several point mutations have been described that have an effect on the splicing of *F8* mRNA. Such mutation could lead to the production of an aberrant mRNA, which will produce an abnormal FVIII protein or could reduce the protein synthesis. Also an effect on the secretion of the protein due to the production of a mis-folded protein can explain the hemophilia A phenotype. Despite tremendous improvements in mutation screening methods, in a minority of patients with reduced FVIII:C levels no mutation can be found. In such cases, application of more excessive methods in addition to the routine mutation screening methods is needed to identify the underlying molecular mechanisms. Therefore, mutations in non-coding areas of *F8* or even in other genes that might interact with FVIII protein should be considered to be responsible for the hemophilia A 'like' phenotype. Only few reports

have been published so far about these cases. El-Maarri et. al [42] performed detailed RNA analysis on a group of such patients, where mutations in the introns were excluded due to normal splicing pattern. One year later the same group reported absence of *F8* mRNA in a severe case, where no DNA change was detected [75]. Castaman et. al reported in 2010 several cases of patients where no candidate mutations where found after direct sequencing of the cDNA [43]. They found several splice site mutations located in or close to exon/intron boundaries. These mutations activated new intronic or exonic cryptic splice-sites. The same group reported deep intronic variations, which could be responsible for mild hemophilia A in patients where no other *F8* mutations have been identified. All these findings support the importance of further investigation of such cases to find the causal mutation responsible for reduced FVIII:C levels.

## 2.8  Study cohort

Patients referred to the hemophilia center in Bonn with a bleeding tendency and FVIII:C levels lower than 35% were diagnosed as hemophiliac. Patients were further analyzed using a simple diagnostic flowchart (Figure 4). All patients with FVIII:C less than 1% were initially tested for intron 1 and 22 inversions. All severe cases, which were negative for inversion PCRs together with mild to moderate patients (FVIII:C levels between 1-35%) were screened for mutations in the complete cDNA, intron/exon boundaries and promoter region of *F8.* When direct sequencing did not reveal any apparent mutations, these patients were designated as "**patients without mutation**".
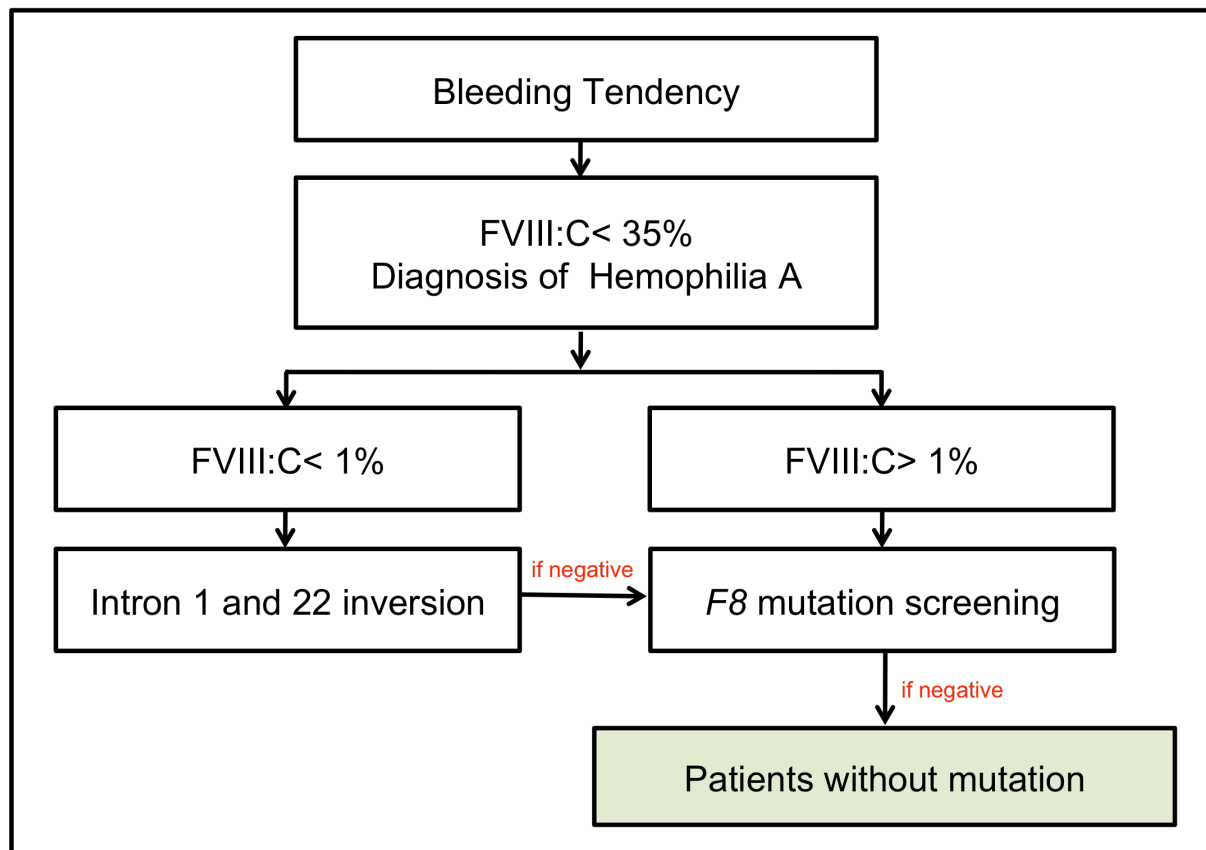
Figure 4: **Diagnostic workflow.** Algorithm of genetic testing approach of patients with reduced or no residual FVIII:C.

Our cohort of patient shares some important characteristics: 1) most of them have mild to moderate hemophilia A phenotype (FVIII:C levels of 5-35%), 2) none of the patients have experienced inhibitor development 3) and, also as reported by Castaman et al [76], these patients show a relatively poor response to desmopressin suggesting decreased or absent of FVIII protein. Desmopressin is a synthetic analogue of the hormone vasopressin that causes release of vWF from Weibel-Palade bodies in the vascular endothelium with a simultaneous rise in FVIII [77, 78]. Desmopressin causes two- to six-fold increase of FVIII in mild or moderate hemophilia A patients [79]. In the context of this work, a framework was established outlining a systematic approach to identify and characterize genetic determinants of FVIII deficiency in patients without mutation. This framework could serve as a rational practice guideline to analyze all genetic factors associated with altered FVIII:C levels.

## 2.9  Aim of the study and experimental strategy

The general aim of this study is to investigate new molecular mechanisms of FVIII deficiency. Diagnosis of hemophilia A is established on the basis of reduced or lack of FVIII:C. In a small group of clinically defined hemophilia A patients, even after applying various techniques, a 100% mutation detection rate yield is not achieved. A number of mutations are thus being missed, as some of these will involve sequence variations in intronic regions that are not being assessed by current techniques. To elucidate the molecular mechanism causing the hemophilia A phenotype in such patients with no detectable mutation in the *F8* cDNA, several mechanisms were proposed and investigated.

Therefore, a systematic screening of known genetic and non-genetic parameters associated with reduced FVIII:C levels was performed. For all the patients, hemophilia A was reconfirmed. This was done based on biochemical measurements of FVIII interacting factors in the coagulation cascade and comparing the clinical characteristics of the patients to mutation positive hemophilia A patients. Once the link between the disease and FVIII was verified, all known defects that mimic the a clinical hemophilia A phenotype caused by interacting partners of FVIII protein, including mutations in *LMAN1* and *MCFD2* affecting the transport of FVIII, type 2N von Willebrand deficiency (VWD) were excluded. The only factor accounting for the bleeding tendency was reduced amount of FVIII.

The analysis was then followed on *F8* at DNA and RNA level. Samples were analyzed for duplications and deletions in *F8* and when necessary for the whole X chromosome. Moreover, a targeted next generation sequencing (NGS) approach was applied, combining a universal long range PCR (LR-PCR) protocol and massive parallel sequencing, to identify sequence variants and rearrangements in the genomic region of *F8*. All variants were characterized. Moreover, qualitative and quantitative analysis of *F8* mRNA was done to investigate whether reduced expression or rapid degradation of mRNA can explain the hemophilia A phenotype.

# 3    Materials and Methods

## 3.1    Materials

### 3.1.1    Chemicals and reagents

All the chemicals used were purchased at *pro analysis* grade. Enzymes used in the molecular biology were obtained from Fermentas GmbH (Leon-Rot) or New England Biolabs GmbH (Frankfurt am Main). All cell culture media were purchased from Life Technologies, Gibco BRL (Eggstein). Primers were designed using the online software OlicoCalc and synthesized by Eurofins MWG Operon (Ebersberg) in desalted quality (HPSF).

| Chemical/ Reagent | Provider |
|---|---|
| 1 kb DNA ladder | Fermentas, St. Leon-Rot |
| 100 bp DNA ladder | Fermentas, St. Leon-Rot |
| Acetic acid | Merck KGaA, Darmstadt |
| Agar | Sigma-Aldrich Chemie GmbH, Steinheim |
| Agarose | Biozym Scientific GmbH, Oldendorf |
| BigDye terminator | Life Technologies, Applied Biosystems, Karlsruhe |
| Chloroform | Fluka, Neu Ulm |
| ddTTP | GE Healthcare Europe GmbH, Freiburg |
| DMSO | Merck KGaA, Darmstadt |
| DNA loading dye | Fermentas GmbH, Leon-Rot |
| dNTPs | Fermentas GmbH, Leon-Rot |
| DreamTaq DNA polymerase | Fermentas GmbH, Leon-Rot |
| EDTA | Sigma-Aldrich Chemie GmbH, Steinheim |
| Ethanol | Merck KGaA, Darmstadt |
| Ethidium bromide | Fluka, Neu Ulm |
| Expand long template PCR | Roche diagnostics, Mannheim |
| FIREPol® DNA polymerase | Solis Biodyne, Tartu, Estonia |

| Chemical/ Reagent | Provider |
|---|---|
| Fluorescein-12-dUTP | Roche Diagnostics, Mannheim |
| Glycerol | Merck KGaA, Darmstadt |
| HotStarTaq DNA polymerase | Qiagen, Hilden |
| HPLC –grade water | Merck KGaA, Darmstadt |
| Human FVIII-deficient plasma | Dade Behring, Marburg |
| iProof™ High-Fidelity DNA polymerase | Bio-Rad Laboratories, München |
| Isopropanol | Merck KGaA, Darmstadt |
| Kanamycin | Sigma-Aldrich Chemie GmbH, Steinheim |
| LGT agarose type VII | Sigma-Aldrich Chemie GmbH, Steinheim |
| Mineral oil | Sigma-Aldrich Chemie GmbH, Steinheim |
| *Paxgene* blood RNA tube | PreAnalytiX, Hombrechtikon, Switzerland |
| Polyclonal Rabbit anti human vWF | Dako Deutschland GmbH, Hamburg |
| Polyclonal goat anti rabbit IgG-HRP | Dako Deutschland GmbH, Hamburg |
| Rediprime™ II | Amersham, GE Healthcare Europe, Munich |
| Sodium metabisulfite | Merck KGaA, Darmstadt |
| SuperScript III™ | Life Technologies, Invitrogen, Darmstadt |
| T4 DNA Ligase | Life Technologies, Invitrogen, Darmstadt |
| Tetramethyl-rhodamine-5-dUTP | Roche Diagnostics, Mannheim |
| Trizol® | Life Technologies, Invitrogen, Karlsruhe |

### 3.1.2  Labware

| Labware | Provider |
|---|---|
| Centrifuge tubes (15 and 50 ml) | Greiner Bio-One GmbH, Solingen |
| CryoTube vials | VWR International GmbH, Langenfeld |
| Filtertips, 10, 200, 1000 µl | Sarstedt, Nürbrecht |
| Parafilm® | Pechiney Plastic Packing, Menasha, USA |
| Pasteur pipettes | Copan Innovation, Brescia, Italy |
| Petri dishes | Greiner Bio-One GmbH, Solingen |
| PCR 8-stripes | Thermo Fischer Scientific, Bonn |
| PyroMark Q24 Cartridge | Qiagen, Hilden |
| PyroMark Q24 Plate | Qiagen, Hilden |
| Serological pipettes | Sarstedt, Nürbrecht |
| Sephadex G-50 | Sigma- Aldrich Chemie GmbH, Steinheim |
| Reaction tubes (1.5 and 2 ml) | Eppendorf, Wesseling-Berzdorf |

### 3.1.3  Equipment

| Equipment | Provider |
|---|---|
| *3130xl* capillary sequencer | Life Technologies, Applied Biosystems, Karlsruhe |
| 7500 Fast Real-Time PCR | Life Technologies, Applied Biosystems, Karlsruhe |
| Centrifuge (5430R- X) | Eppendorf, Wesseling-Berzdorf |
| Centrifuge (3-16PK) | Sigma Laborzentrifugen GmbH, Osterode am Harz |
| Chemi doc (Gel Doc XR+) | Bio-Rad Laboratories, München |
| Fluorskan Ascent | Thermo Fisher Science, Bonn |
| Fluorochem® | Alpha Innotech Corp., San Leandro, USA |
| Genome Analyzer II | Illumina Inc., San Diego, CA, USA |
| Microscope | Carl Zeiss Microscopy GmbH, Göttingen |
| Mini Rocker | Peqlab Biotechnologie GmbH |
| Nano-Drop ND 1000 | Peqlab Biotechnologie GmbH |
| Power supply | Biometra, GE healthcare, München |
| PyroMark Q24 pyrosequencer | Qiagen, Hilden |
| PyroMark Q24 Vacuum Workstation (220 V) | Qiagen, Hilden |
| Thermomixer | Eppendorf, Wesseling-Berzdorf |
| Thermocycler (MJ Research) | Bio-Rad Laboratories, München |
| Thermocycler (T3000) | Biometra, GE healthcare, München |
| Vortexer | Janke & Kunkel IKA-Labortechnik |
| Water bath (Thermostat 2761) | Eppendorf, Wesseling-Berzdorf |

### 3.1.4  Databases and Programs

CGH array analysis:

- SignalMap (Roche NimbleGen, Mannheim)

Dissertation preparation:

- Office 2008 (Microsoft Corporation, Redmond, USA)
- Endnote X4 (Thomson Reuters, Carlsbad, CA, USA)

FISH analysis:

- FISHView EXPO 2.0 Applied Spectral Imaging, Neckerhausen, DE

MLPA analysis:

- Coffalyser (MRC-Holland, Amsterdam, The Netherlands)

Primer design:

- Olico calc (www.basci.northwestern.edu/biotools/oligocalc.html)

Sequence analysis:

- Geneious (Biomatters Ltd, Auckland, New Zealand)
- PyroMark Q24 software 2.0 (Qiagen, Hilden)
- GeneMapper® v.4.1 (Life Technologies, Applied Biosystems)
- BioEdit (IBIS Bioscience, USA) (http://www.mbio.ncsu.edu/bioedit/bioedit.html)

MicroRNA analysis:

- miRBase (http://www.mirbase.org/cgi-bin/blast.pl)
- RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi)

Sequence search:

- BLAST (Basic Local Alignment Search Tool) (www.ncbi.nlm.nih.gov/BLAST)
- BLAT (Basic Local Alignment Tool) (http://genome.ucsc.edu/cgi-bin/hgBlat)
- dbSNP (http://www.ncbi.nlm.nih.gov/SNP/)
- HapMap (http://hapmap.ncbi.nlm.nih.gov/)
- HEMApSTR (www.uenf.br/Uenf/Pages/CBB/LBT/HEMApSTR.html#F8A1)

Statistical analysis:

- GraphPad Prism (GraphPad Software Inc., La Jolla, CA, USA)

### 3.1.5 Statistical analysis

Statistical analysis was performed using GraphPad Prism software. P values were calculated using the Krustal-Wallis test for the autosomal loci and Mann Whitney test for the X-lined loci. A p-value below 0.05 shows a significant difference for all tests.

### 3.1.6 Kits

- BAGene ABO-TYPE, BAG Healthcare, Lich
- BigDye Terminator v1.1, Life technologies, Applied Biosystems, Karlsruhe
- Large-Construct Kit, Qiagen, Hilden
- MLPA F8-P178 kit, MRC-Holland, Amsterdam, The Netherlands
- Omniscript® Reverse Transcription kit, Qiagen, Hilden
- PAXgene Blood RNA kit, Qiagen, Hilden
- PyroMark Gold Q24 Reagents, Qiagen, Hilden
- QIAGEN Plasmid Mini and Midi Kit, Qiagen, Hilden
- QIAquick Gel Extraction Kit, Qiagen, Hilden
- *VisuLize*TM FVIII Antigen Kit, Haemochrom Diagnostica GmbH, Essen

### 3.1.7 Buffers and media

- TAE buffer                              40mM Tris base
  (pH 8.5)                                1mM acetic acid
                                          1mM EDTA

- SOC medium                              2% tryptone
                                          0.5% yeast extract
                                          10mM sodium chloride
                                          2.5 mM potassium chloride
                                          10 mM magnesium chloride
                                          10 mM magnesium sulfate
                                          20 mM glucose

- Luria-Bertani (LB) medium               1% tryptone
                                          0.5% yeast extract
                                          0.5% NaCl

- SSC solution                            0.3 M NaCl
                                          30 mM sodium citrate dihydrate

- Church buffer                           1 mM EDTA,
  (pH 7.2)                                7% SDS
                                          0.5 M $Na_2HPO_4$

### 3.1.8  Plasmids

- *pCR*®-XL-TOPO® vector, Life Technologies, Invitrogen, Karlsruhe (Figure S 1)

### 3.1.9  Bacterial strains

Table 1: List of bacterial strains used for transformation of plasmids.

| DH5α | *end*A1 *hsd*R17(rk-mk+) *supE44 thi*-1 *recA1 gyrA relA1* Δ(*lacZYAargF*) U169 (φ80*lacZΔM15*) | (Hanahan, 1983) |
|---|---|---|
| TOP10 | F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(araleu) 7697 galU galK rpsL (StrR) endA1 nupG | Life Technologies, Invitrogen |

### 3.1.10  Biological samples

Blood samples from hemophilia A patients and healthy controls were obtained and collected in dipotassium EDTA tubes (1.75 mg/ml blood). For *FISH* analysis and karyotyping, heparinized blood was obtained from patients and submitted to the Department of Human Genetics of the University of Würzburg. All participants gave informed consent to participate in the study in accordance with the Declaration of Helsinki.
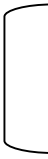
## 3.2  General methods

### 3.2.1  Polymerase chain reaction (PCR)

PCR was performed using FIREPol® DNA polymerase or DreamTaq DNA polymerase for DNA fragments up to 6 kb. Both polymerases are suitable for a TA cloning strategy. Annealing temperature was 2-6°C lower than primer melting temperature and elongation time was 1min/kb. Each PCR product was evaluated on 1% agarose gel (3.2.3). The PCR reaction mixture was as follows:

| Reaction mixture | Amount (µl) |
|---|---|
| 10x Buffer | 2.5 |
| dNTP Mix (5mM each) | 2.5 |
| MgCl$_2$ (25mM) | 2.5 |
| Gene-specific reverse primer (20µM) | 1 |
| Gene-specific forward primer (20µM) | 1 |
| Template DNA (100ng/µl) | 2 |
| DNA polymerase (5U/µl) | 0.5 |
| dH$_2$O | 13 |
| **Total** | **25** |

The cycler conditions for the PCR were as follows:

| Cycler protocol: | 94°C | 5 min | Initial Denaturation |
|---|---|---|---|
| | 94°C | 30-60 s | Denaturation |
| 35 cycles | 55-65°C | 30-60 s | Annealing |
| | 72°C | 1-5 min | Elongation |
| | 72°C | 10 min | Final elongation |

### 3.2.2  Site-directed mutagenesis

To analyze the effect of a mutation in a vector-based system, a site-directed mutagenesis approach was applied. Using this approach mutations were introduced at a specific residue inducing a desired amino acid substitution [80]. The mutation was introduced into the vector via a PCR reaction as described above using partial complementary primers incorporating the desired amino acid change and iProof™ High-Fidelity DNA Polymerase.

| Reaction mixture | Amount (µl) |
|---|---|
| 5x iProof™ HF/GC buffer | 10 |
| dNTP Mix (10 mM each) | 1 |
| Gene-specific reverse primer (20 µM) | 1 |
| Gene-specific forward primer (20 µM) | 1 |
| Template plasmid (100 ng) | 1 |
| iProof™ High-Fidelity DNA Polymerase (5U/µl) | 0.5 |
| dH$_2$O | 35.5 |
| **Total** | **50** |

PCR program:    98°C      30s      Initial Denaturation

                      98°C      10s      Denaturation

35 cycles    x*°C      30s      Annealing

                      72°C      12min    Elongation

                      72°C      10min    Final Extension

*= Annealing temperature was calculated according to the Tm of the primers used.

After the PCR reaction, a *DpnI* digestion was done to eliminate the methylated template plasmid. The reaction was incubated for 1 h at 37°C and inactivated at 80°C for 20 min. The PCR product was subsequently transformed into competent cell as described in section 3.2.7 and positive clones were used for expression studies after verification of the residue change by sequencing.

### 3.2.3 Agarose gel electrophoresis

1% (w/v) Agarose gels were prepared by dissolving 5 g agarose in 1x TAE buffer by heating. Ethidium bromide was added at a final concentration of 10 µg/ml to detect DNA. Samples were mixed with 5x loading dye and separated in 1x TAE at 5-10 mA/cm. A molecular weight size marker (1 kb or 100 bp ladder) was used depending on the size of the products. The gels were visualized and documented under UV light on a ChemiDoc system. For TA cloning, crystal violet was used instead of ethidium bromide to avoid DNA damage by UV light.

### 3.2.4  Gel extraction

PCR products obtained from different experiments (RT-PCRs (3.3.1.2), Long Range PCRs (3.3.3.2)) were gel purified using the QIAquick Gel Extraction Kit. Briefly, the products were resolved on agarose gels as described in section 3.2.3 and the fragments were excised subsequently. Gel slices were dissolved in a buffer containing a pH indicator and incubated for 10 minutes at 55°C. The mixture was then applied to the QIAquick spin column, and after several wash steps, the products were eluted in an appropriate high-salted buffer provided in kit or water.

### 3.2.5  Cloning of PCR products

Generally, a TA-cloning strategy was used to sequence low amounts of PCR products, multiple fragments obtained from RT-PCR and inverse PCR products. This was done using the TOPO® XL PCR cloning kit. The kit uses a linearized and topoisomerase I-activated *pCR*®-XL-TOPO® vector (3.1.8). The PCR products were generated using Expand Long Template PCR Enzyme mix, which is a unique enzyme mix containing *Tgo* DNA polymerase [81] a thermostable DNA polymerase with proof reading activity and Taq DNA polymerase. The later is essential for the cloning to ensure addition of 3´A-overhangs to the PCR products. Amplicons were gel purified after agarose gel electrophoresis (3.2.3) to increase the cloning efficiency. However, to avoid damage to PCR products by UV light, visualization and electrophoresis was done using crystal violet. PCR bands were excised and purified using the TOPO® XL Gel purification kit. The DNA was stored at -20°C or processed directly to TOPO® Cloning. Four µl of the gel-purified product was mixed with 1 µl of the *pCR*®-XL-TOPO® vector for 5 minutes up to 30 minutes at RT. The reaction was stopped by adding 1 µl of the 6x TOPO® Cloning Stop Solution. Subsequently, 2 µl of the TOPO® Cloning reaction was used to transform OneShot® TOP10 (Table 1) chemically competent cells according to the manufacturer's instructions as described in section 3.2.7. Selection of recombinants was done via disruption of the lethal E.coli gene, *ccdB* [82]. Ligation of a PCR product disrupts expression of the *ccdB* gene, which is fused to the C-terminus of the *LacZa* fragment, so that cells that contain the

non-recombinant vector are killed upon plating. Therefore, no blue/white screening is needed. Positive clones were selected and sequenced as described in section 3.2.6.1 using M13 Forward and M13 Reverse primers (M13F (-20): 5´-GTAAAACGACGGCCAG-3´; M13R: 5´-CAGGAAACAGCTATGAC-3´) provided from the TOPO® cloning kit. Alternatively, candidate clones were first selected on the basis of the insert size by direct amplification form bacterial colonies in a PCR reaction using the M13F and M13R primers. Bacterial stocks were prepared from positive clones as described in section 3.2.8.

### 3.2.6  Sequencing

Three different sequencing approaches were applied in this work: *ABI* Sanger sequencing, high throughput next generation sequencing and pyrosequencing.

#### 3.2.6.1  *ABI* Sanger sequencing

Sequencing was performed based on the chain termination method using BigDye terminators, which are labeled with dRhodamin acceptor dyes as follows:

| Terminator | Dye Label |
|:---:|:---|
| A | Dichloro (R6G) |
| C | Dichloro (ROX) |
| G | Dichloro (R110) |
| T | Dichloro (TAMRA) |

BigDye terminator contains a set of dye terminators labeled with novel, high-sensitivity dyes [83]. Each dye is fluorescing at a different wavelength. Sequencing is based on a mix of these fluorescence-labeled dideoxynucleotides. The excitation of dyes is done via a laser beam followed by detection of fluorescence by a CCD camera. Approximately 200 ng of plasmid or 2 µl of the enzyme-treated PCR products were used in the following sequencing mix:

| Reaction mixture | Amount (µl) |
|---|---|
| 5x BigDye sequencing buffer | 1.5 |
| Sequencing primer (3.2 pmol) | 0.5 |
| BigDye Terminator sequencing mix | 0.5 |
| Template | 0.5-2 |
| dH$_2$O | X |
| **Total** | **12** |

The PCR conditions for the sequencing PCR were as follows:

Cycler protocol:    96°C        1 min        Initial Denaturation

                        96°C        10 s         Denaturation

35 cycles       50°C         5 s          Annealing

                        60°C        4 min        Elongation

The products were subsequently purified by EtOH/NaAc precipitation (12 µl of the sequencing reaction product, 50 µl 96 % EtOH, 10 µl dH$_2$O and 2 µl of 3M NaAc) followed by centrifugation by 4°C at 4000 rpm for 45 minutes. The pellet was washed once with 70% EtOH for 10 minutes at the same centrifugation conditions and dried. Samples were resuspended in 15 µl HiDi-formaldehyde and loaded on the capillary sequencer.

### 3.2.6.2  Next generation sequencing (NGS)

NGS was performed as a service at GATC Biotech AG (Konstanz, Germany) on the *Illumina Genome Analyzer II (GA II)* platform using the standard protocols and the kits provided according to the manufacturers instructions. Briefly, the purified LR-PCR fragments (3.2.4) were equimolarly mixed and processed after fragmentation according to Illumina's Paired-End Sequencing library kits. The DNA samples are sheared to appropriate size (average 800 bp). After library construction each sample was tagged with a unique barcode. The tagged libraries were pooled and two unique

adapters were ligated to the fragments. The Paired-End adapter ligated fragments (200 bp ±50 bp) were isolated and attached to the flow cell. The flow cell is coated with single stranded oligonucleotides that correspond to the sequences of the adapters ligated during the library preparation. The adapter-ligated fragments are bound to the surface of the flow cell and amplified by bridge amplification via PCR (Figure 5).
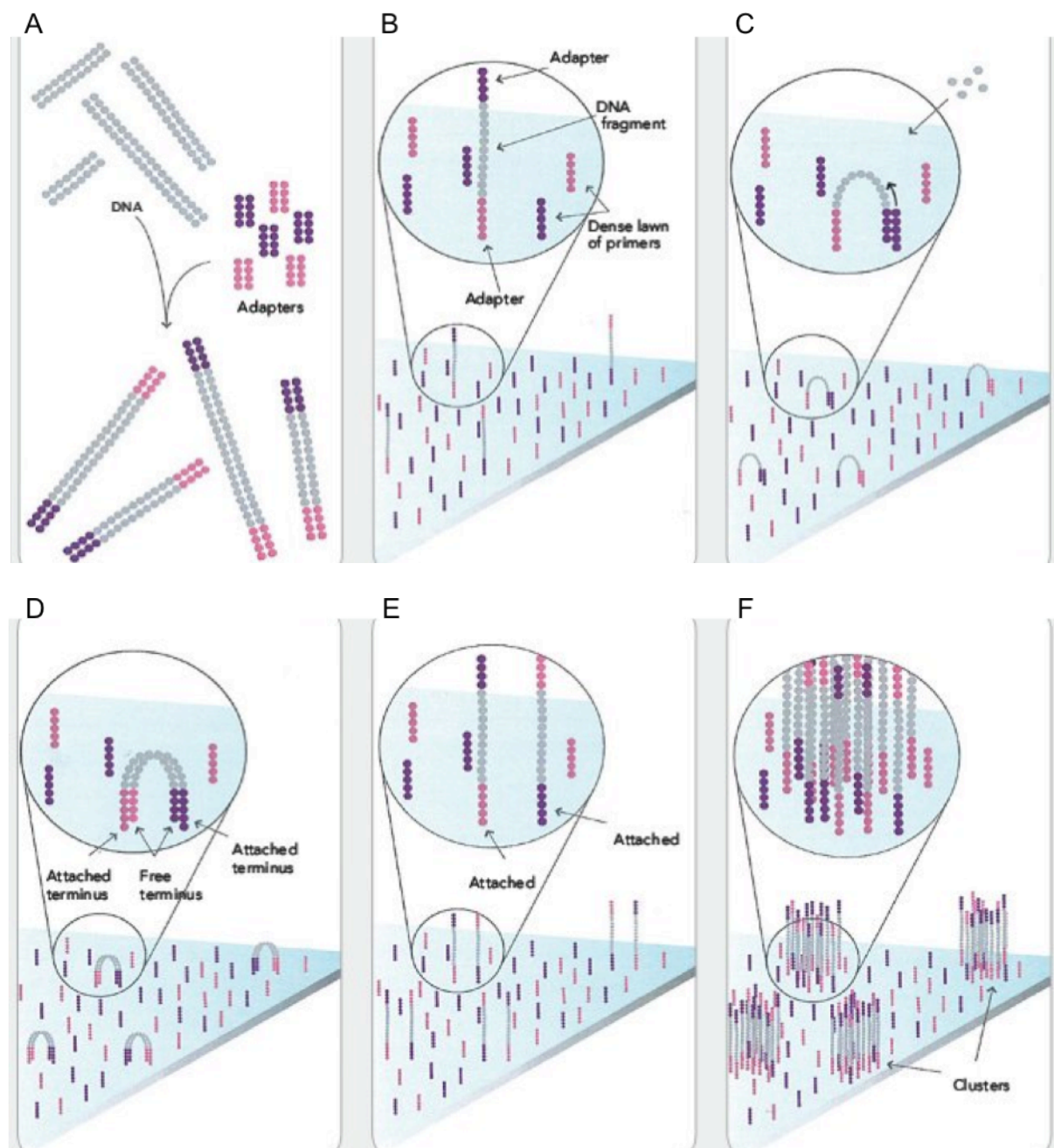
Figure 5: **Preparation of samples for NGS**. A) Preparing the DNA sample by random fragmentation and ligation of adapters to both ends of the fragments. B) Attachment of single-stranded fragments randomly to the flow cell channels. C) Bridge amplification on solid-phase after adding nucleotides and enzymes. D) Generation of double stranded bridges on the solid-phase. E) Denaturation of double stranded templates. F) Generation of several million dense clusters of double-stranded DNA in each channel of the flow cell (Pictures are modified based on data available on http://www.illumina.com/technology/sequencing_technology.ilmn).

The generated clusters on the flow cell are loaded into the sequencer for sequencing by synthesis approach. The sequences were analyzed using Geneious software. The reads were aligned onto the reference assembly (Figure 6).
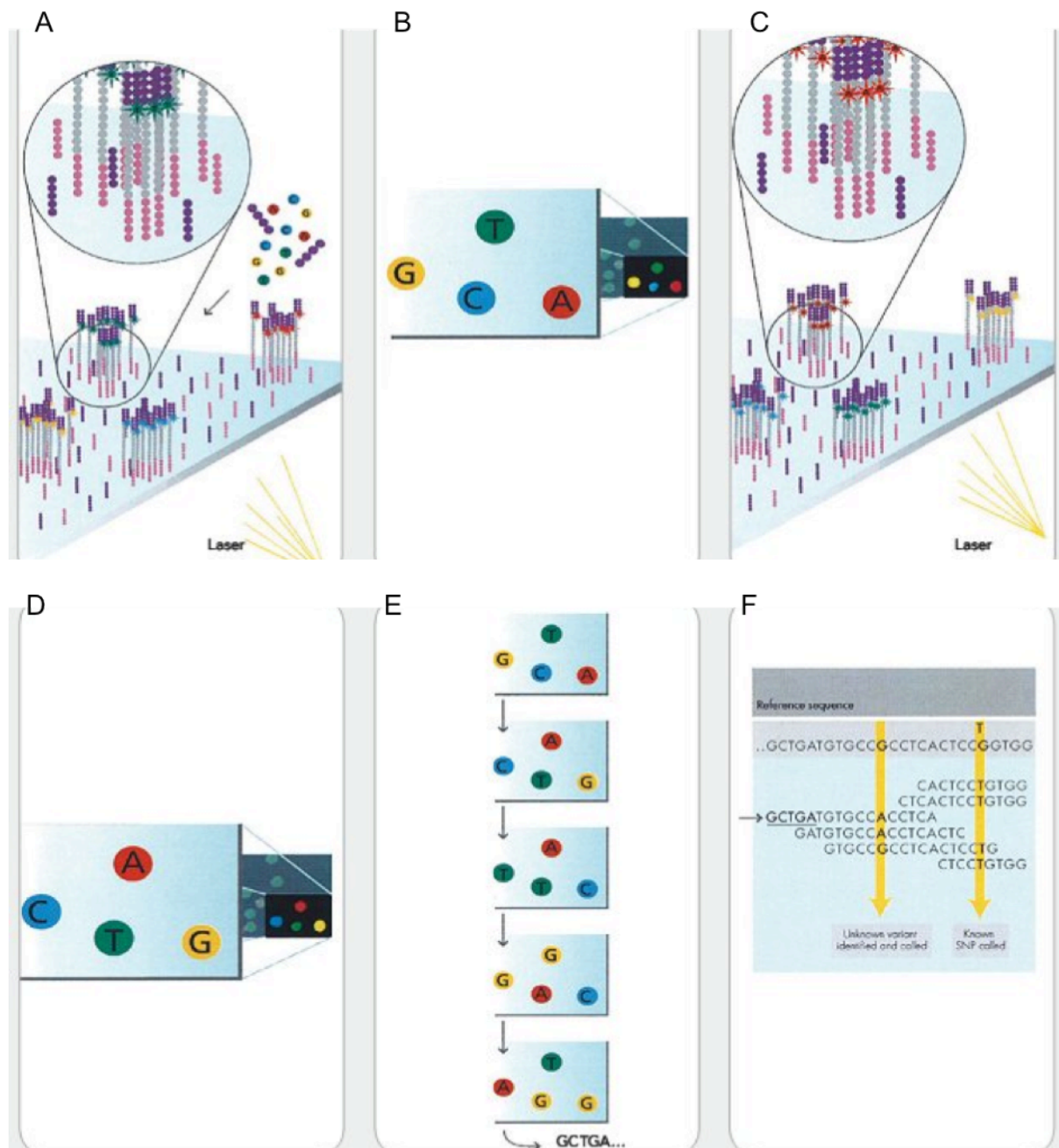
Figure 6: **Sequencing of clusters on the *GAII* Illumina sequencer.** A) All four labeled reversible terminator, primers, and DNA polymerase are added to the flow cell to initiate the first sequencing reaction. B) After laser excitation, the image of emitted fluorescence from the first base of each cluster is recorded. C) The second base is determined in the next sequencing cycle and the second base is recorded for each cluster. D) Cycles of sequencing are repeated to determine the sequence of bases in a given fragment. E) Aligning the data to a reference and identifying the differences (http://www.illumina.com/technology/sequencing_technology.ilmn).

### 3.2.6.3 Pyrosequencing

Pyrosequencing was carried out according to Tost et al. [84]. PCR fragments were generated using the primers listed in Table S 14, whereby either the reverse primer or the forward primer was biotin labeled. The sequencing primer was then hybridized to the labeled single-stranded PCR product that serves as template, and incubated with the enzymes, DNA polymerase, ATP sulfurylase, luciferase, and apyrase as well as the substrates, adenosine 5′phosphosulfate (APS) and luciferin. After the addition of the first dNTP to the reaction, the DNA polymerase catalyzes the incorporation of the dNTP into the DNA strand. If the first dNTP is complementary to the base in the DNA strand, pyrophosphate (PPi) is released in a quantity equimolar to the amount of the incorporated nucleotide. ATP sulfurylase converts PPi to ATP in the presence of APS. ATP drives the luciferase-mediated conversion of luciferin to oxyluciferin that generates visible light in amounts proportional to the amount of ATP (Figure 7 A-C).
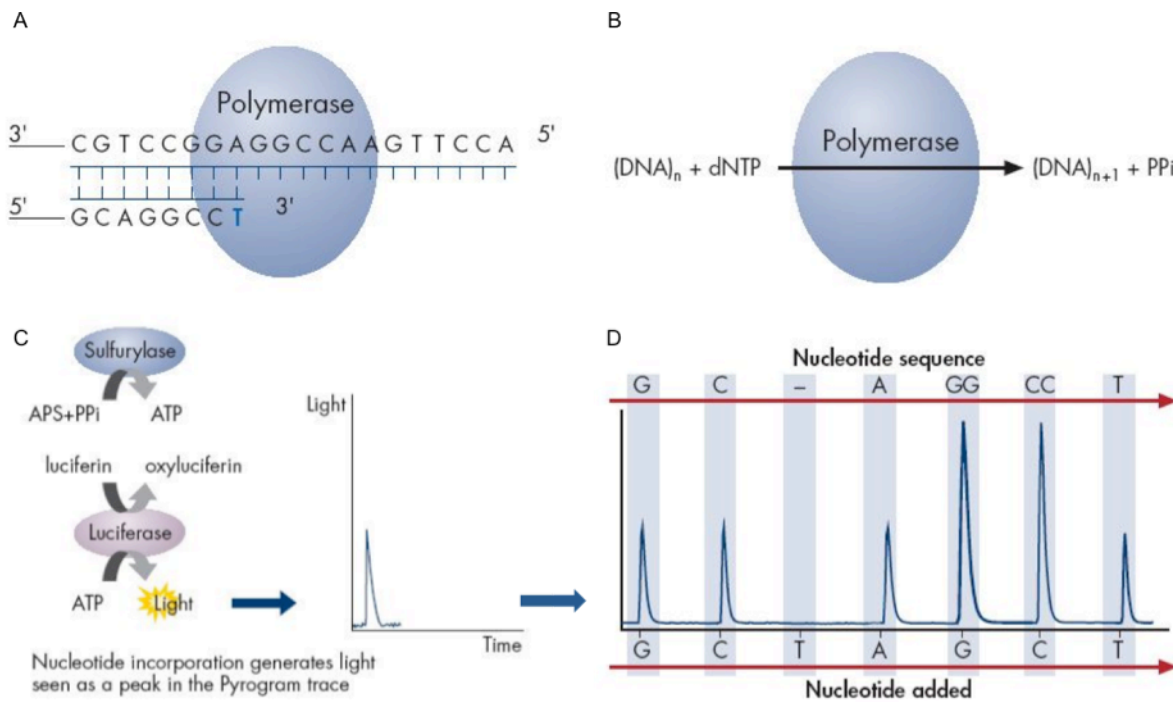
Figure 7: **Principle and procedure of pyrosequencing.** A) The sequencing primer is hybridized to a single-stranded PCR amplicon. B) The first dNTP is added to the reaction. If the dNTP is incorporated into the DNA strand, the reaction is accompanied by release of PPi, which is proportional to the amount of incorporated nucleotide. C) ATP sulfurylase converts PPi to ATP in the presence of APS. ATP drives the luciferase-mediated conversion of luciferin to oxyluciferin that generates light that is proportional to the amount of ATP. The light is detected by a CCD camera and is seen as a peak in the pyrogram. Unincorporated nucleotides and ATP is continuously degraded and another dNTP is added. D) Addition of dNTPs is performed sequentially until the complementary DNA strand is built up and the sequence is determined from the signal peaks in the pyrogram. The height of each peak is proportional to light signal and to the number of nucleotides incorporated.

The light is detected by a CCD chip and seen as a peak in the raw data output in the pyrogram (Figure 7 D). The height of the peak is proportional to the number of nucleotides incorporated. A nucleotide-degrading enzyme, apyrase, continuously degrades unincorporated nucleotides and ATP. After the degradation is completed, the next nucleotide is added. The dNTPs are added sequentially and a special form of dATP, the deoxyadenosine alfa-thio triphosphatase (dATP-S) is used in the reaction, which will be recognized by the luciferase but not by the DNA polymerase. The reactions were prepared using the PyroMark® Gold reagents, applied on a PyroMark Q24 system and analyzed subsequently on PyroMark Q24 software 2.0.

### 3.2.7  Transformation of bacteria

Plasmid vectors were introduced into bacterial cells, competent *E.coli* (DH5α or OneShot® TOP10, 3.1.9) according to manufacturer's protocol. Briefly, competent cells were thawed on ice.  Five µl of each plasmid was added to the cells, mixed gently, and incubated for 30 minutes on ice. Followed by a heat-shock for 30 seconds in a 42°C water bath, cells were chilled on ice for 5 minutes. Afterwards, 250 µl SOC medium was added to the cells, followed by an incubation for 1h at 37°C at ~200 rpm. Cells were plated on LB-agar plates containing the appropriate antibiotic (50 µg/ml end concentration). Plates were incubated in overnight a 37°C incubator.

### 3.2.8  Isolation and purification of bacterial plasmid DNA

Colonies were picked from LB-Plates and cultured overnight in LB-medium supplied with the appropriate antibiotic. Bacterial stocks were prepared from each clone by mixing 350 µl glycerol with 650 µl of the overnight LB-culture and stored at -80°C.

For plasmid purification, cells were grown overnight in 5 ml LB-medium. Cells were harvested by centrifugation for 15 minutes and subsequently lysed using the lysate buffer from Qiagen Plasmid Mini kit. Bacterial lysates were cleared by centrifugation and loaded onto the anion-exchange column. Plasmid DNA selectively binds to the column and is cleared from impurities. After several wash steps plasmid DNA was eluted in high-salt buffer and the DNA was collected by centrifugation. Plasmids were subsequently sequenced (3.2.6.1) to verify the insert and/or the mutation introduced in the vector. A size validation was done on an agarose gel as described in section 3.2.3. For plasmid preparation in higher amounts, 50 ml of LB-medium containing the appropriate antibiotic was inoculated and grown overnight. Plasmid isolation was done using Qiagen Plasmid Midi kit with the same principle as Qiagen Plasmid Mini Kit. DNA concentration was assessed (3.2.9) and plasmids were stored at -20°C until use.

Purification of bacterial artificial chromosome (BAC) was done using the Qiagen Large-Construct Kit. This method ensures DNA of significantly greater purity. The

efficient removal of genomic DNA is achieved by an integrated treatment with ATP-dependent exonucleases performed after the lysis of a 500 ml culture of bacterial cells. The sample was then loaded onto the column, where all RNA, protein, metabolites, and other low-molecular-weight impurities are removed after a wash step. The ultrapure DNA is eluted in high-salt buffer, concentrated and desalted by isopropanol precipitation and collected by centrifugation.

### 3.2.9  Measurements of DNA and RNA concentrations

DNA and RNA concentration was determined by measuring by photometry (1 $OD260_{nm}$ = 50µg dsDNA/ml, 1 $OD260_{nm}$ = 40µg RNA/ml) on a Nanodrop® machine.

## 3.3  Mutation analysis

### 3.3.1  RNA analysis

#### 3.3.1.1  RNA extraction from blood

For RNA analysis, 2.5 ml of human whole blood was collected in a *PAXgene* Blood RNA tube. To ensure complete lysis of blood cells, the tubes were incubated at least for 2h at room temperature after blood collection. If the *PAXgene* tubes were stored at -70°C, the samples were equilibrated to RT before starting the RNA extraction procedure. The *PAXgene* Blood RNA tube contains a reagent composition that protects RNA from degradation by RNases and minimizes *ex vivo* changes in gene expression.

Total RNA extraction was performed according to the manufacturer's procedure. Briefly, the tubes were centrifuged for 10 minutes at 5000 xg. After one wash step with RNase-free water (RNFW), the pellet was resuspended in an optimized buffer included in the kit supplemented with proteinase K and incubated for 10 minutes at 55°C. Subsequently, the lysate was applied to a *PAXgene* Shredder spin column to homogenize and remove the cell debris. The supernatant of the flow-through was transferred to a microcentrifuge tube and mixed with ethanol to adjust binding

conditions. The sample was then applied to a *PAXgene* RNA spin column, where the RNA was selectively bound to the *PAXgene* silica membrane. The RNA was treated with DNase I using the RNase-Free DNase set to remove traces of genomic DNA. DNase I was diluted in DNA digestion buffer (RDD buffer) and incubated for 15 minutes at RT. After several wash steps, RNA was eluted in elution buffer and heat-denatured at 65°C for 5 minutes. RNA was stored at -70°C until used.

### 3.3.1.2 Semi-Quantitative analysis of mRNA via Reverse Transcription PCR (RT-PCR)

Total RNA was reverse transcribed to complementary DNA (cDNA) using the Omniscript® Reverse Transcription kit. Briefly, template RNA was thawed on ice and reverse-transcription reaction was prepared as follows and incubated for 60 minutes at 37°C.

| Reaction mixture | Amount (µl) |
|---|---|
| 10x Buffer RT | 2 |
| dNTP Mix (5 mM each) | 2 |
| Gene-specific primer (10 µM) | 2 |
| RNase inhibitor (10 U/µl) | 1 |
| Omniscript® Reverse Transcriptase | 1 |
| Template RNA (~ 50 ng) | 5 |
| RNFW | 7 |
| **Total** | **20** |

The cDNA of *F8* was synthesized using 4 different gene-specific primers and amplified in eight overlapping fragments using a nested PCR approach as described by El-Maarri et al. [42]. RT-PCR products were resolved on a 1% (w/v) agarose gel as described in section 3.2.3. Aberrant mRNA products and fragments with low amounts of PCR product were cloned using a TA-cloning strategy (3.2.5) and subsequently sequenced (3.2.6.1).

### 3.3.1.3  Splice site prediction analysis

Wildtype and mutant sequences were submitted for splice site prediction to the Berkeley Drosophila Genome Project (BDGP) (http://fruitfly.org/seqtools/splice.html), to the NetGene2 server (http://www.dtu.dk/services/NetGene2/) and to the Human Splicing Finder (http://www.umd.be/ HSF/) servers.

### 3.3.1.4 Quantitative analysis of mRNA expression via real time PCR (qRT-PCR)

We performed a one-step qRT-PCR using probes to analyze *F8* expression in blood. Total RNA was reverse transcribed using SuperScript® III Reverse Transcriptase and the cDNA was amplified with HotStarTaq DNA Polymerase using the manufacture's protocol except that the SuperScript® III was diluted 1:10 in the buffer supplemented in the kit and mixed with the reaction mixture prepared as follows:

| Reaction mixture | Amount (µl) |
|---|---|
| 10x Buffer RT | 5 |
| MgCl (25nM) | 5 |
| dNTP Mix (5mM each) | 0.4 |
| Gene-specific forward primer (10µM) | 1 |
| Gene-specific reverse primer (10µM) | 1 |
| Gene-specific probe (5µM) | 1 |
| Rox | 0.03 |
| SuperScript® III | 0.4 |
| HotStarTaq polymerase | 0.25 |
| Template RNA (~ 50ng) | 2 |
| RNFW | 33.92 |
| **Total** | **50** |

The PCR conditions for the qRT-PCR were as follows:

| Cycler protocol: | 50°C | 20 min | cDNA synthesis |
|---|---|---|---|
| | 95°C | 15 min | RTase inactivation/ |
| | | | HotStarTaq Polymerase activation |
| 40 cycles | 94°C | 30 s | |
| | 60°C | 60 s | |

The experiment was done on a *7500* Real-Time PCR System. An internal control from the human porphobilinogen deaminase gene *(PBGD;* NM_000190) was included and the *F8* mRNA levels were normalized to their respective *PBGD* mRNA expression. The sequences of qRT-PCR primers are listed in Table S16.

### 3.3.2  DNA analysis: general methods

#### 3.3.2.1  DNA extraction from blood

Ten ml of peripheral blood was collected from patients and control groups. Additionally, blood samples from extended family members (when available from both affected and not affected individuals) were collected. Genomic DNA was extracted using a standard salt out precipitation procedure [85].

#### 3.3.2.2  PCR amplification of exons for investigated genes

PCR amplification from genomic DNA for investigated genes: *F8* (NG_011403.1); *vWF (*NC_000012); *LMAN1* (NG_012097) and *MCFD2* (NG_016428.1). The reaction was performed using standard PCR conditions (3.2.1). All PCR products were resolved on a 1% agarose gel (3.2.3) and were sequenced on an *ABI* genomic analyzer (3.2.6.1). Sequences were analyzed using the Geneious software.

### 3.3.3  DNA analysis: methods for gross rearrangement analysis

### 3.3.3.1  Southern blot

Restriction enzymes (100 U each) in combination with the appropriate enzyme buffers were used in separate reactions for cleavage of 10 µg genomic DNA. All reactions were incubated at 37°C for 2h and subsequently separated on a 0.7% (w/v) agarose gel in TAE buffer for 48 hours at low voltage (30 V). After depurination in 0.25 M HCl solution for 10 minutes and denaturation in 1.5 M NaCl and 0.5 M NaOH for additional 10 minutes. DNA fragments were transferred from gel to a positively charged nylon membrane overnight by capillary blotting. Subsequently, the membrane was neutralized in 2x SSC solution and pre-hybridized in Church buffer at 65°C while rotating. The probe was amplified by standard PCR conditions using healthy control DNA with primers described in Table 2. Later, the probe was labeled with α-$^{32}$P-dCTP using Rediprime$^{TM}$ II (Random Prime Labeling System) according to the manufacturer's recommendation. After incubation at 37°C for 20 minutes, the probe was purified by centrifugation through a Sephadex G50 spin-column. The probe was denatured at 95°C for 5 minutes and immediately added to the Church buffer in which the membrane was pre-incubating. Hybridization of the membrane was performed overnight at 65°C in a rotating incubator. The next day, the membrane was serially washed three times in washing solutions with different concentrations of SSC (2x, 0.5x and 0.1x) supplemented with 0.1% (w/v) SDS in order to prevent non-specific binding. Subsequently, the membrane was wrapped in plastic wrap and exposed to X-ray film. Band patterns were visualized by autoradiography.

Table 2: Primers used for amplification of the southern blot probe.

| Name | Sequence (5´-->3´) | Tm | Position on chrX: hg18 | Length |
|------|--------------------|----|------------------------|--------|
| Probe-F | AAG AAC AGT TCA TGT CAA GGA CAA | 64 | 153891795 | 502 |
| Probe-R | TGG AAG ATC AGA GGG AAC A | | 153891293 | |

### 3.3.3.2 Long range (LR)- PCR across the *F8* genomic region

LR-PCR was performed on genomic DNA to amplify the complete genomic region of *F8* by amplification in 28 overlapping fragments using the Expand Long Template PCR System according to the manufacturer's instructions. Briefly, 200 ng DNA was used for each reaction. Thermal cycling conditions were as follows: 2 min initial denaturation at 92°C, 10 cycles at 92°C for 10 s, 55-65°C (depending on the primers used) for 30 s and 68°C for 1 min/kb (depending on the size of the amplicons), 20 cycles at 92°C for 10 s, 55-65°C for 30 s and 68°C for 1 min/kb + 20 s/cycle, followed by 7 min at 68°C. PCR products were analyzed by gel electrophoresis (Figure 16, 3.2.3). All PCR fragments were purified by gel extraction (3.2.4) and were stored at -20°C until use. For each patient, the purified LR-PCR fragments were equimolarly mixed and were applied to a NGS approach on an Illumina sequencer (3.2.6.2).

### 3.3.3.3 Inverse PCR

One µg of genomic DNA was digested with 10U of an appropriate restriction enzyme for 4 h at 37°C in a reaction volume of 50 µl. After a phenol:chloroform (1:1) extraction, DNA was precipitated with 2 volumes of ethanol, 0.1 volume of sodium acetate (pH) 4,6 for 30 minutes at -80°C. The precipitate was recovered by centrifugation at 14.000 x *g* for 15 minutes at 4°C, followed by a final wash with 70% ethanol (v/v) and a final centrifugation at 14.000 x *g* for 15 minutes at 4°C. The precipitate was resuspended in 50 µl of $dH_2O$. Digested DNA was ligated using 15U of T4 DNA Ligase and 120 µl of 5x Ligase buffer and was incubated overnight at 15°C. After extraction with phenol:chloroform (1:1), the ligated circular products were precipitated and resuspended in 30 µl of $dH_2O$. The circular products were used as templates for a LR-PCR reaction using the Expand long template PCR System. The PCR conditions were as follows: 2 minutes at 92°C, 10 cycles of 92°C for 10 s, 58°C for 30 s and 68°C for 5 minutes, followed by 25 cycles of 92°C for 15 s, 58°C for 30 s, 68°C for 5 minutes plus 20 s for each successive cycle and a final step for 7 minutes at 68°C.

### 3.3.3.4  Multiple Ligation-Dependent Probe Amplification (MLPA)

MLPA was carried out using F8-P178 kit according to the manufacturer's instructions. The technique comprises 4 steps: denaturation, hybridization, ligation, and amplification (Figure 8). Briefly, 250 ng of genomic DNA was denaturized and hybridized to the probe mix for 16h at 60°C. The probes were then ligated by treatment with ligase65 at 54°C for 15 minutes. Subsequent PCR reactions are performed using a 6-carboxyfluorescein (FAM)-labeled primer for each exon. The amplification product of each probe has a unique length. PCR products were resolved according to the size on an ABI *3130xl* capillary sequencer. Results were visualized using the GeneMapper® software and analyzed using the Coffalyser™ software (MRC-Holland). Samples were compared to a control sample. The differences in relative peak height indicate a copy number change of the target exon.
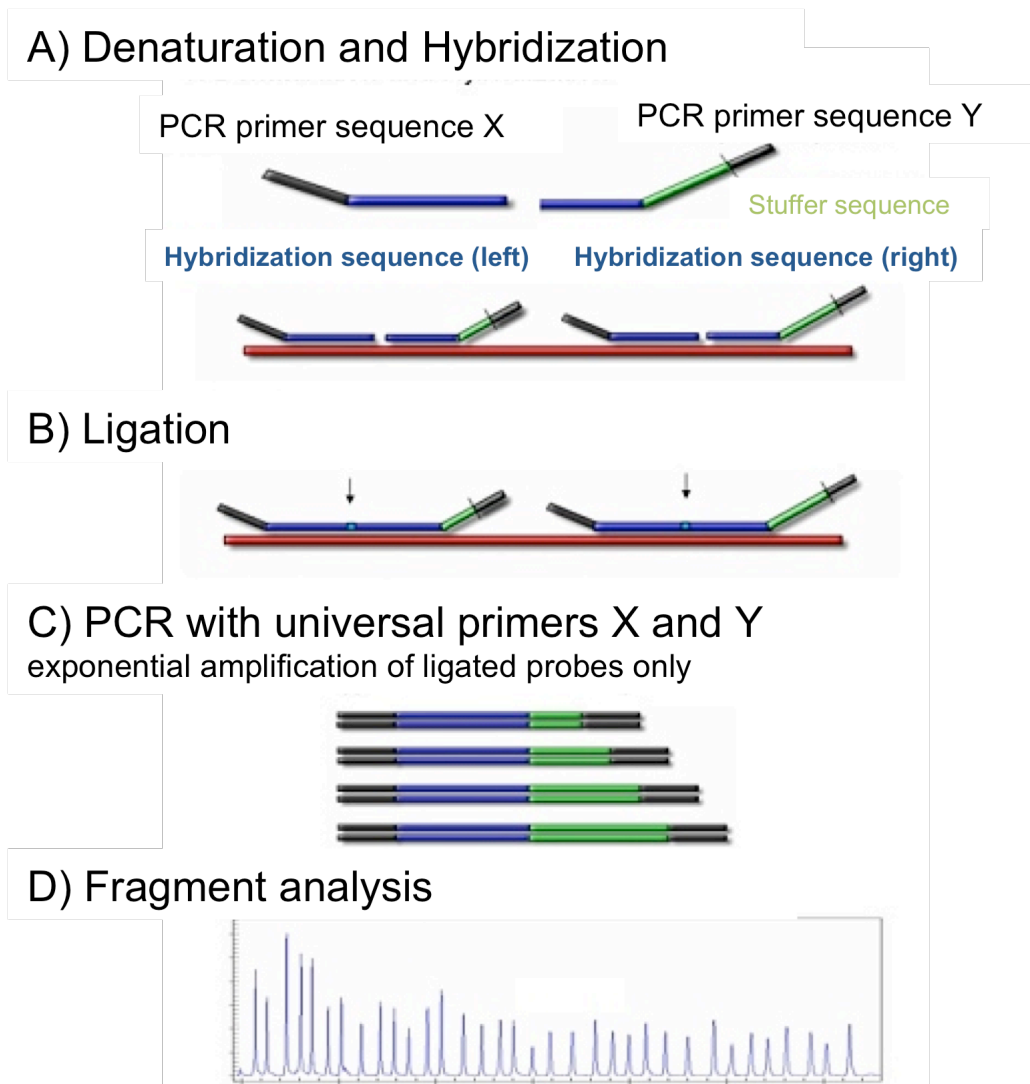
Figure 8: **Principle and procedure of MLPA.** A) Genomic DNA is denaturized and hybridized to the probe mix. B) After a ligation step, C) a PCR reaction is performed using universal primers D) and the products are resolved on a capillary sequencer. Fragments are visualized using the GeneMapper® software. The black arrows indicate the ligation site of the probes to each other (http://www.mlpa.com/WebForms/WebFormMain.aspx).

### 3.3.3.5 Fluorescence *in situ* hybridization (*FISH*) analysis

From each index patient, 10 ml blood was taken in heparin-blood collecting tubes. A bacterial artificial chromosome (BAC) clone containing genomic region of *F8*, clone number: IMGSB737D062145D (Imagenes GmbH, Berlin, Germany), was isolated as described in section 3.2.8 and used as a probe for the hybridization procedure. *FISH* analysis was performed in collaboration with Dr. Indrajit Nanda at the Institute of Human Genetics, University of Würzburg. DNA from BAC was directly labeled with either fluorescein-12-dUTP or tetramethyl-rhodamine-5-dUTP by standard nick

translation. For each *FISH* experiment approximately 200 ng of labeled DNA was hybridized to patient's metaphase spreads. Hybridization of repetitive sequences within the labeled probe was suppressed by adding excess unlabelled human Cot-1 DNA (10x probe concentration) in the hybridization mixture. Following hybridization, chromosomes were counterstained in 1.5 µg/ml DAPI (4',6-diamidino-2-phenylindole) in Vectashield anifade. Slides were examined under a Zeiss epifluorescence microscope (Imager, A.1) equipped with a CCD camera and appropriate filter sets. Selected metaphases displaying DAPI counterstaining and rhodamine fluorescence were photographed and images were merged using the software FISHView EXPO 2.0 (Figure 9).
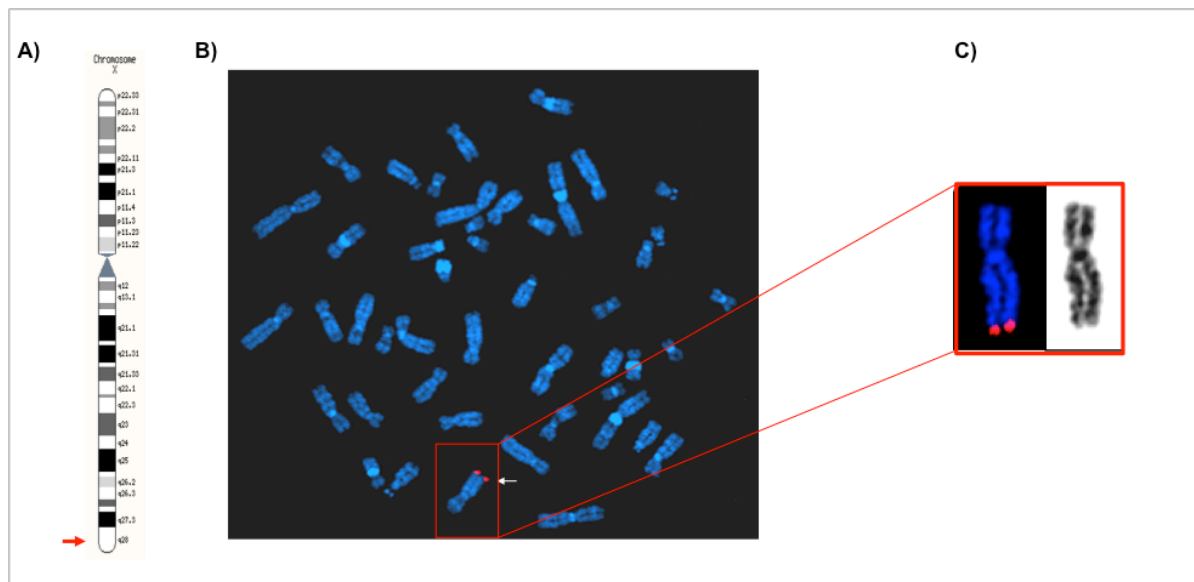


Figure 9: *FISH* **analysis of metaphase cells.** A) Schematic representation of the X chromosome. The arrow indicates the position of *F8* at Xq28 where the probe binds. B) Hybridization of the *F8* specific probe (red signal) and counterstained with DAPI (blue). C) Magnification of X Chromosome cut outs with and without hybridization.

### 3.3.3.6  Comparative Genomic Hybridization (CGH) Array

CGH-Arrays were done as a service at Source BioScience imaGenes GmbH in Berlin. Two µg of DNA from the index patient and standard control male DNA was labeled with dUTP-coupled cyanine Cy3 and Cy5, respectively. Both samples were co-hybridized on a NimbleGen Human CGH 385K chromosome X Tiling Array (Build: hg18, NCBI36). Data was analyzed and visualized using the SignalMap software.

### 3.3.3.7  Haplotype analysis

For haplotype analysis, 15 extragenic and intragenic SNPs and Short Tandem Repeats (STRs) were analyzed using polymorphic markers selected from HapMap and HEMApSR [86] database. For each marker, the PCR reaction mixture consisted of 0.2 µM of each primer, 200 µM of each dNTPs, and 50 ng of genomic DNA in a 25 µl total volume. The amplicons were sequenced as described above. Microsatellite repeat analyses were done using a fluorescent-labeled reverse primer in the PCR reaction. Subsequently, the amplicons were loaded on an ABI *3100xl* capillary sequencer and distinguished by size using GeneMapper® software v4.1.

### 3.3.3.8  Likelihood of carriership

Likelihood of being carrier was determined using FVIII:C values after correction for the errors/deviations introduced by age and blood group. Transformed value ($T_{FVIII:C}$) was calculated according to the formula $T_{FVIII:C} = \ln(FVIII:C) - age^2 \times 0.00011 - bg \times 0.18017 - 3.61700$ as described by Oldenburg et al [87], where bg=0 if the subject has blood group 0 or A2 and bg=1 if the subject has blood group A1, B or AB. The T values are than used to extrapolate on the density curves of likelihood of being carrier or non-carrier.

### 3.3.3.9  X-Chromosome inactivation

X-chromosome inactivation test was performed using the human androgen receptor (*HUMARA*) assay as described by Allen et al [23]. Two µl of the PCR product was mixed with 12 µl of deionized formamide and 2 µl of GeneScan 500 ROX size standard. The mixture was denaturized and loaded on an ABI *3130xl* capillary sequencer. GeneMapper® software v4.1 was used for quantification and interpretation of the data. Peak heights of the fluorescence intensity were used to calculate ratios. The percentage of inactivation of each X chromosome in the possible/obligate carriers was calculated according to the formula: A/(A+1)*100 where A=((Peak 1 / Peak 2) after digestion) / ((Peak 1 / Peak 2) before digestion).

### 3.3.4   Coagulation assays and biochemical methods

For coagulation factor assays, 5 ml blood was collected in 3.1% sodium citrate blood. The plasma was isolated by centrifugation at 2000x g, 4°C for 15 minutes. The plasma was used directly or was stored at -80°C until use.

#### 3.3.4.1   FVIII activity assays

FVIII:C was measured using a two-stage chromogenic assay and a one stage clotting assay.

##### 3.3.4.1.1  Clotting assay

The clotting assay is an aPTT-based assay [89] done on a Behring Coagulation System. The samples were diluted in FVIII deficient plasma. After incubation with aPTT reagent, $CaCl_2$ was added to the mixture and the time of the clot formation is determined. The machine measures the change in light scattering. The time when the maximum scattering is reached reflects aPTT measurement. For each measurement, standard plasma and medium from untransfected cells (mock cells) were included as positive and negative controls, respectively.

##### 3.3.4.1.2  Chromogenic assay

The chromogenic assay [90] was performed also on a Behring Coagulation System using reagents from Siemens Healthcare Diagnostics. It is a two-step test: in the first step, FVIII is activated in the plasma with thrombin, in the next step the activated FVIII, accelerates the conversion of FX to FXa in the presence of FIXa, phospholipids, and $Ca^{2+}$ ions. The amount of activated FX is measured at 405 nm by hydrolysis of a p-nitroanalide substrate. The amount of p-nitroanalide is proportional to the generated FXa, which is dependent on the amount of FVIII in the sample. Samples are diluted in 0.9% sodium chloride. For each measurement, two control plasmas were included: Control N (normal range) and Control Plasma P (pathogenic range).

### 3.3.4.2  FVIII antigen assay

FVIII antigen was measured using *VisuLize™* FVIII antigen kit, a commercial enzyme-linked immunosorbent assay (ELISA) assay. The wells are pre-coated with an anti-human FVIII polyclonal antibody produced in sheep, which can bind to the FVIII antigen present in the sample. The samples were diluted 1:8 and 1:16 in sample dilution buffer supplied in the kit and the assay was performed according to manufacturers instructions. Briefly, the samples were pipetted to the pre-coated stripes and incubated for 1h at room temperature. The wells were washed with ELISA washing buffer to remove unbound material. The secondary antibody, an anti-sheep peroxidase-labeled antibody, was applied to the wells and the reaction was incubated for 45 minutes at room temperature. The wells were again washed and a solution of the peroxidase substrate tetramethylbenzidine (TMB) was applied for exactly 10 minutes. During this reaction the wells will develop a blue color. Adding an acidic stop solution, which turns the blue color to yellow, stopped the assay. The intensity of the formed color, was measured spectrophotometrically in a microplate reader at 450 nm. The absorbance is directly proportional to the FVIII antigen amount captured in each well. The reference curve was a log-log plot of the mean absorbance values versus the FVIII antigen concentration. A dilution factor of 1, 2, or 4 was calculated for the 1:4 and, 1:8, and 1:16, respectively.

### 3.3.4.3  vWF activity measurement

VWF activity was measured using an automated latex enhanced immunoassay (*HemosIL™*) on IL Coagulation System. The vWF activity kit is a latex particle enhanced immunoturbidimetric assay. The activity of vWF is determined by measuring the increase of the turbidity produced by the agglutination of the latex reagent. An anti-vWF monoclonal antibody, directed against the platelet-binding site of vWF, is adsorbed onto the latex reagent and will react with the vWF of the test plasma. The degree of agglutination is directly proportional to the vWF activity in the sample and is determined by measuring the decrease of transmitted light caused by the aggregates.

### 3.3.4.4  vWF antigen measurement

VWF antigen was measured using an Enzyme-Linked Fluorescent Assay (ELFA). VIDAS vWF was applied to perform a quantitative measurement of vWF antigen in the plasma samples of patients. The test is a two-step enzyme immunoassay sandwich method combined with a fluorescent detection. The samples were diluted and cycled in and out of the solid phase receptacle (SPR) to enable the antigen to bind to the immunoglobulin adsorbed on the inner side or the SPR. A washing step was included to wash the unbound components. The conjugate is then cycled in and out of the SPR to bound to the antigen retained during the first step. A second washing step is applied to remove unbound material. The substrate, 4-methyl-umbelliferyl phosphate, is then cycled in and out of the SPR and the conjugate enzyme catalyzes the hydrolysis of the substrate into a fluorescent product, which is measured at 450 nm. The fluorescence intensity is proportional to the concentration of the antigen present in the sample.

### 3.3.4.5  vWF collagen- and FVIII binding-assay

For both, vWF collagen- and FVIII-binding assays, frozen plasma from the patients was measured in collaboration with Prof. Budde at Medilys Laborgesellschaft GmbH in Hamburg. The collagen binding assay (vWF:CBA) was essentially done according to Thomas et al. [91], while the vWF:FVIII binding assay was done according to Schneppenheim et al. [47].

### 3.3.4.6  vWF multimer analysis

SDS-agarose discontinuous gel electrophoresis was carried out essentially as previously described [92, 93] in collaboration with Prof. Budde at Medilys Laborgesellschaft GmbH in Hamburg. In brief, low (1.2% w/v) and medium (1.6% w/v) resolution gels were prepared. Samples were diluted according to their vWF content to approximately 0.5 µg/ml. After electrophoresis, the multimers were transferred to a nitrocellulose membrane. The blots were incubated with mouse anti human vWF antibody as primary antibody. As secondary antibody, an anti-mouse

IgG antibody was used. Detection was carried out by luminescence using the video detection system Fluorchem™. In addition to visual inspection of the gels, quantitative evaluation was performed to improve the reproducibility and comparability of multimer analysis. For densitometric evaluation, small, intermediate and large multimers were defined as oligomers 1-5, 6-10 and >10, respectively.

### 3.3.5 AB0 blood group determination

AB0 blood group was determined using the BAGene AB0-TYPE or routine methods from genomic DNA or EDTA anticoagulated blood samples, respectively.

### 3.3.6 C-reactive protein determination

C-reactive protein (CRP) was run on a clinical chemistry analyzer COBAS® INTEGRA using the Roche CRP Latex immunoassay.

# 4   Results

All patients with a bleeding tendency were referred to the hemophilia center in Bonn and collaborating laboratories. Patients with FVIII:C levels lower than 35% were diagnosed as hemophiliacs. The molecular genetic analysis was done to detect the causal mutation in *F8*. Patients were analyzed using a routine diagnostic flowchart (Figure 4). Briefly, all severe patients were tested for intron 1 and intron 22 inversions. All patients who were negative for inversion PCRs together with mild to moderate patients (FVIII:C levels between 1-35%) were screened for mutations in the complete *F8* cDNA as well as the exon/intron boundaries and promoter region of *F8.* A total of 14 patients were included in this study. Among these patients 3 had severe, eight had mild and two had moderate phenotype (Table 3). The study was approved at the ethic committee of the university of Bonn (approval number 183/07). All participants received oral and written information concerning the study and signed the informed consent form. All patients were negative for mutations in *F8* cDNA except patient HA#14, where the amplification of exon 18 failed. The investigation of the molecular mechanism behind the severe hemophilia A phenotype in this patient was performed as described in section 4.2.6.

## 4.1   Biochemical     and     clinical    reconfirmation    of hemophilia A phenotype

The first step was to reconfirm hemophilia A in these patients. This was done based on biochemical measurements of all factors in the coagulation cascade associated with a bleeding phenotype and later by comparing the clinical characteristics of the patients without detectable mutation in *F8* to mutation positive hemophilia A patients. Table 3 shows the clinical characteristics of the patients included in this study.

Table 3: Characteristics of the patients included in this study.

(*nd*= no data available, Neg.=negative)

| Patient number | | Normal range | HA #1 | HA #2 | HA #3 | HA #4 | HA #5 | HA #6 | HA #7 | HA #8 | HA #9 | HA #10 | HA #11 | HA #12 | HA #13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FVIII | Activity | 57-245% | <1% | 21 | 5.0 | 5.6 | 3.0 | 5.3 | <1% | 9.9 | 29 | 14 | 35.0 | 7.4 | 12.3 |
| | Antigen | 50-183% | *nd* | 34.0 | 7.4 | 7.6 | 4.0 | 7.3 | *nd* | 14.5 | *nd* | *nd* | 9.9 | 22.1 | 18.8 |
| | FVIII:C/FVIII:Ag | 0.67-2.05 | *nd* | 0.62 | 0.70 | 0.74 | 0.75 | 0.73 | *nd* | 0.68 | *nd* | *nd* | 3.53 | 0.34 | 0.65 |
| | mRNA expression | 0.029-1.121 | *nd* | 0.263 | 0.067 | 0.098 | 0.066 | 0.063 | *nd* | 0.038 | 0.122 | 0.038 | 0.056 | 0.046 | 0.033 |
| vWF | Activity | 44-276% | 76 | 53 | 52 | 96 | 76 | 80 | *nd* | 168 | 86 | 110 | 60 | 132 | 80 |
| | Antigen | 40-250% | 86 | 70 | 66 | 88 | 90 | 108 | 159 | 164 | 79 | 71 | 59 | 170 | 89 |
| | vWF activity/vWF:Ag | 0.406-1.844 | 0.88 | 0.76 | 0.79 | 1.09 | 0.84 | 0.74 | *nd* | 1.02 | 1.09 | 1.55 | 1.02 | 0.78 | 0.90 |
| | Collagen binding | 50-160% | 77 | 63 | 79 | 56 | 61 | 87 | 129 | 97 | *nd* | *nd* | 46 | 118 | 87 |
| | FVIII binding | 60-170% | 88 | 103 | 80 | 102 | 95 | 96 | 92 | 100 | *nd* | *nd* | 98 | 126 | 106 |
| FII | | 87-127% | 80 | 101 | 105 | 94 | 105 | 97 | 107 | 88 | *nd* | *nd* | *nd* | 95 | 103 |
| FV | | 67-145% | 80 | 134 | 134 | 88 | 86 | 122 | 125 | 119 | *nd* | *nd* | *nd* | 111 | 123 |
| FVII | | 77-154% | 74 | 119 | 134 | 63 | 110 | 115 | 128 | 137 | *nd* | *nd* | *nd* | 88 | 55 |
| FX | | 80-140% | 79 | 108 | 99 | 95 | 133 | 125 | 133 | 157 | *nd* | *nd* | *nd* | 92 | *nd* |
| aPTT | | 25-36 sec. | 60.5 | 34.3 | 33.7 | 46.1 | 46.6 | 43.6 | 48.5 | 42.9 | *nd* | *nd* | *nd* | 44.5 | 32.6 |
| Antithrombin activity | | 85-130% | 92 | 99 | 87 | 100 | 98 | 96 | 96 | 92 | *nd* | *nd* | *nd* | 99 | 87 |
| FVIII-Inhibitor | | Negative | Neg. | Neg. | Neg. | Neg. | Neg. | Neg. | Neg. | Neg. | Neg. | Neg. | Neg. | Neg. | Neg. |
| Blood group | | AB0 | A1 | 0 | 0 | 0 | A | B | A | B | A2 | 0 | 0 | A1 | 0 |
| Clinical severity | | | severe | mild | moderate | mild | moderate | mild | severe | mild | mild | mild | mild | mild | mild |

The activity levels of investigated factors (FII, FV, FVII, FX, and vWF) were in the normal range, thus excluding a quantitative deficiency of these factors. Moreover, the only reduced parameters among the patients were FVIII:C and FVIII:Ag levels. The clinical phenotype of patients without mutation do not differ from hemophilia A patients that harbor known mutations in the *F8* cDNA. Both groups receive comparable amounts of administrated FVIII concentrate to abolish the bleeding symptoms. The residual FVIII:C measured 24h and 48h after infusion is also comparable in both groups as well . The bleeding tendency of all patients is efficiently corrected with substitution of FVIII concentrates indicating a normal clearance of FVIII molecule in these patients (Table 4).

FVIII:Ag levels from 8 patients with mild to moderate hemophilia A phenotype that do not undergo a prophylaxis treatment show a very clear and strong linear correlation to FVIII:C and the severity of the disease. There is a significant reduction in both FVIII:C and FVIII:Ag levels when comparing the patients to a cohort of healthy individuals (Table S 9, Figure 10). The ratio of FVIII:C to FVIII:Ag in patients is within the normal range (0.74 - 1.55). This result clearly associates the deficiency of FVIII:C in these patients to a quantitative deficiency of FVIII protein in blood rather than its dysfunction or problems in its structure.
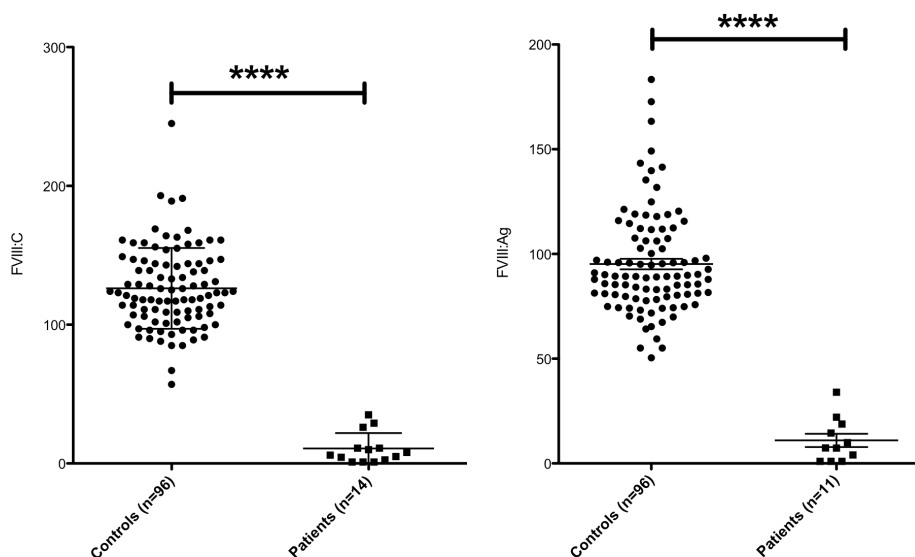


Figure 10: **Comparison of FVIII:C and FVIII:Ag levels in normal healthy male individuals and patients without mutation.** Patients show a significant reduction of FVIII:C and FVIII:Ag levels compared to that of healthy controls. FVIII:C and FVIII:Ag levels are shown as % to the normal controls. FVIII:Ag levels were not measured for severe patients due to their prophylaxis treatment (Mann Whitney test **** = $p < 0.0001$).

Table 4: Clinical characteristics of patients without mutation.

Rec: Recombinant FVIII concentrate, Plasma: Plasma-derived FVIII concentrate, ave.: average, nd= no data available, * vWF containing concentrate, [a] compared to hemophilia A patients with a known mutation and same severity of disease. DDAVP= Desmopressin

| Patient number | | | HA#1 | HA#2 | HA#3 | HA#4 | HA#5 | HA#6 | HA#7 | HA#8 | HA#12 | HA#13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Current type of FVIII substitution medicament | | | Rec. | Rec. | Rec. | Rec. | Rec. | Rec. | Rec. | Rec. | Plasma* | Rec. |
| FVIII:C after | 24 h | Infusion units | nd | 3000 | 4000 | nd | 2000 | 3000 | nd | 3000 | 3000 | 3000 |
| | | FVIII:C | nd | 24% | 16% | nd | 29% | 17% | nd | 21% | 40% | 26% |
| | 48 h | Infusion units | nd | 2000 | 4000 | 2000 | nd | nd | nd | 2000 | nd | 3000 |
| | | FVIII:C | nd | 16.5% | 3.0% | 7.8% | nd | nd | nd | 13.5% | nd | 18.5% |
| Bleeding tendency [a] | | | nd | normal | normal | normal | normal | normal | nd | normal | normal | normal |
| Amount of factors used [a] | | | nd | below ave. | above ave. | below ave. | below ave. | below ave. | nd | below ave. | above ave. | below ave. |
| Response to DDAVP: fold increase in FVIII:C after 30 sec. Mild:0.8-16   Mod.:0.8-156 | | | nd | 4.17 | 10 | 2.10 | nd | 0.98 | nd | nd | nd | 6.50 |
| Clinical severity | | | severe | mild | moderate | moderate | moderate | mild | severe | mild | mild | mild |

### 4.1.1  Investigating known determinants of FVIII:C

Once the link between the disease and FVIII was verified, all known defects in parameters that mimic a clinical hemophilia A phenotype were analyzed. This includes all FVIII interacting proteins, such as: *LMAN1* and *MCFD2,* both affecting the intracellular transport of FVIII protein, and *vWF* causing VWD type 2N that affects the stability of FVIII protein. Furthermore the recently described SNPs associated with reduced FVIII:C levels were screened for patients without detectable mutation [94].

### 4.1.1.1  Searching for mutations in *LMAN1* and *MCFD2*

Mutations in *LMAN1* and *MCFD2* cause a combined FV/FVIII deficiency through interfering with the intracellular transport of FVIII and FV protein [44, 45]. Although all patients have normal FV activity, the possibility of putative rare variations was investigated. The hypothesis was that these mutations could result in less optimal functionality of these cargo proteins by specifically affecting the trafficking of FVIII protein rather than FV. The complete coding regions of both genes were screened for mutations (exons 1-13 for *LMAN1* and exons 2-4 for *MCFD2*). No DNA alterations were found. Therefore, a link between these proteins and reduced FVIII:C levels in patients was excluded.

### 4.1.1.2  Searching for a qualitative defect in vWF

After the release of FVIII protein in the circulation, vWF plays an important role as carrier and protector of FVIII protein [95-98]. For the patients, the FVIII:C/vWF:Ag ratio ranged from 0.03-0.58 (mean=0.14) comparing to 0.88-2.44 (mean=1.31) in the control group. Such a reduction is similar to what is observed in vWD-2N [99]. Even so, not only the FVIII binding domains of vWF protein (which are described to harbor mutations causing vWD-2N) but all 52 exons of *vWF* were analyzed for mutations and/or polymorphisms that might affect binding of vWF to FVIII protein. Several known SNPs were found but no causal mutations associated with VWD (Table S 1) Moreover, to exclude a functional defect in the vWF protein, collagen-binding activity,

FVIII binding assay, and multimer pattern of vWF protein were excessively analyzed (except for 2 patients -HA#9 and HA#10- where no plasma was available). Our vWF:FVIII and vWF:CBA analysis showed normal values in the range of 80-126% and 46-129%, respectively (Table 3). In all patients multimer analysis revealed a normal pattern of both, high- and low molecular weight multimers (Figure 11). As expected, vWF activity levels were largely attributable to vWF:Ag levels and a good correlation was observed between vWF activity and antigen levels (Table 3). Based on the above analysis, a defect in vWF protein was excluded (whether quantitative or qualitative) that could explain reduction of FVIII:C in these patients.
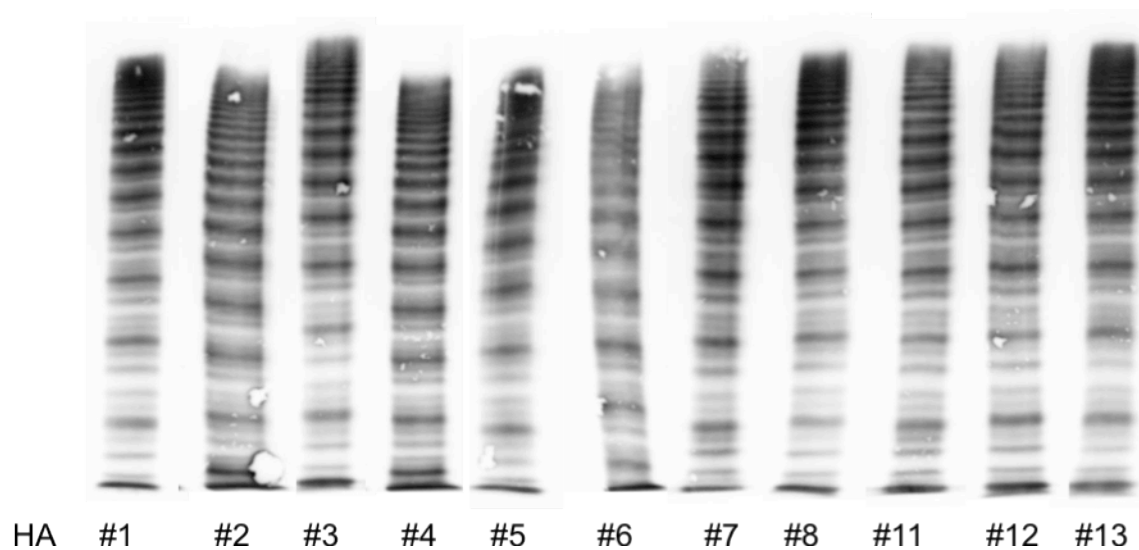


Figure 11: **Multimer analysis of hemophilia A patients without mutation.** All analyzed patients show normal vWF multimer patterns (No plasma was available from HA#9 and HA#10 to perform this analysis).

### 4.1.1.3 Searching for associations between FVIII:C and genetic polymorphisms

Recently a genome wide association study (GWAS) screening reported five loci (*STAB2, STXB5, SCARA5, vWF* and *AB0*), to be associated with FVIII:C; these loci included one SNP in *AB0* (rs687289) and one SNP in *vWF* (rs1063856). Therefore, the possibility was investigated whether these patients harbor several 'un-favored' polymorphisms that are associated with lower FVIII:C levels. All 5 polymorphisms were investigated in patients as well as in controls (Table S 3 and S 5).

In addition, three polymorphisms in *F8* itself (rs7058826, rs1800291 and rs1800292), previously suggested to show association with FVIII activity were also studied [69, 100]. In healthy controls a significant association was only found between FVIII:C levels and the intronic polymorphism in AB0 (Kruskal-Wallis ANOVA test, p=0.0004), and 4 out of 9 patients carry the GG genotype, while three are heterozygous. The G genotype is reported to be associated with lower FVIII:C levels as also confirmed in healthy cohort [100]. The occurrence of the G allele in some of the patients could contribute to further lowering the FVIII:C levels. A borderline significance was observed with the exonic S1269S (rs1800292) SNP in exon 14 of *F8* (Mann Whitney test, p=0.0627); however, all the patients carry the AA genotype that is associated with higher FVIII:C levels (Figure 12, Table S 5).
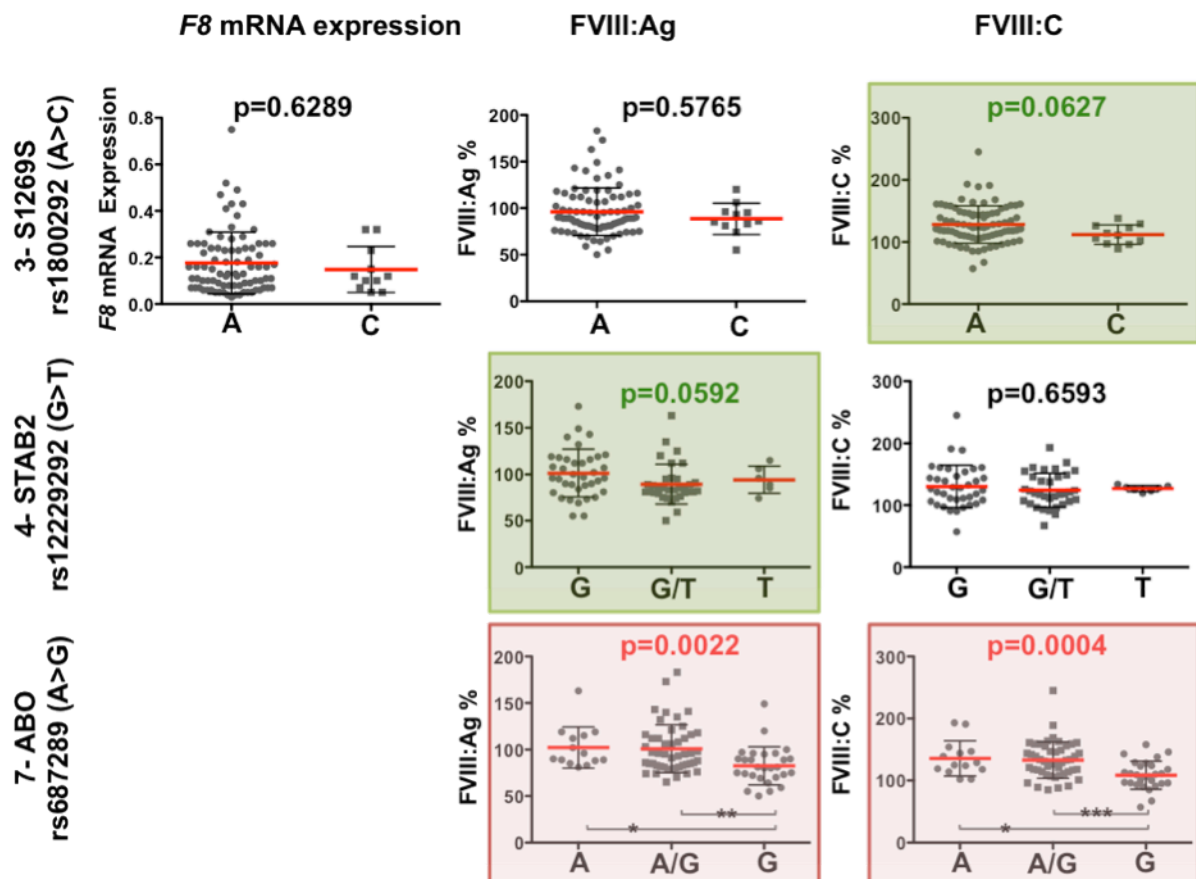


Figure 12: **Association between SNPs and the FVIII:Ag and FVIII:C levels in healthy population.** Only the data of the SNPs that show significant or borderline association with previously known associated SNPs is presented. Graphs that are boxed in green and red correspond to statistically borderline significant or significant associations, respectively. FVIII:C and FVIII:Ag levels are presented as % to the normal range (P values are calculated using the Krustal-Wallis test for the autosomal loci and Mann Whitney test for the X-lined loci, *=p<0.05, **= p<0.01, ***=p<0.001).

## 4.2 Mutation analysis on DNA and RNA level in hemophilia A patients

### 4.2.1 Mutation analysis: DNA level

#### 4.2.1.1 Exclusion of gross chromosomal recombination via FISH analysis

Up to this point, all results showed a link of the hemophilia A phenotype to *F8* gene itself. Therefore cytogenetic analysis was performed to exclude chromosomal rearrangements in all patients using a probe for *F8* gene. At least chromosomes of twenty metaphase nuclei were hybridized with a *F8* gene specific probe synthesized from a BAC clone (3.3.3.5). The analysis was done for patients where we could provide fresh blood material. In 7 patients (for 6 of them the analysis was done in this study and for HA#1 it was previously reported) no gross chromosomal rearrangements could be found. The probe was hybridized to the tip of the long arm of the X chromosome (Xq28) and no translocation of *F8* gene was identified (Figure 13).
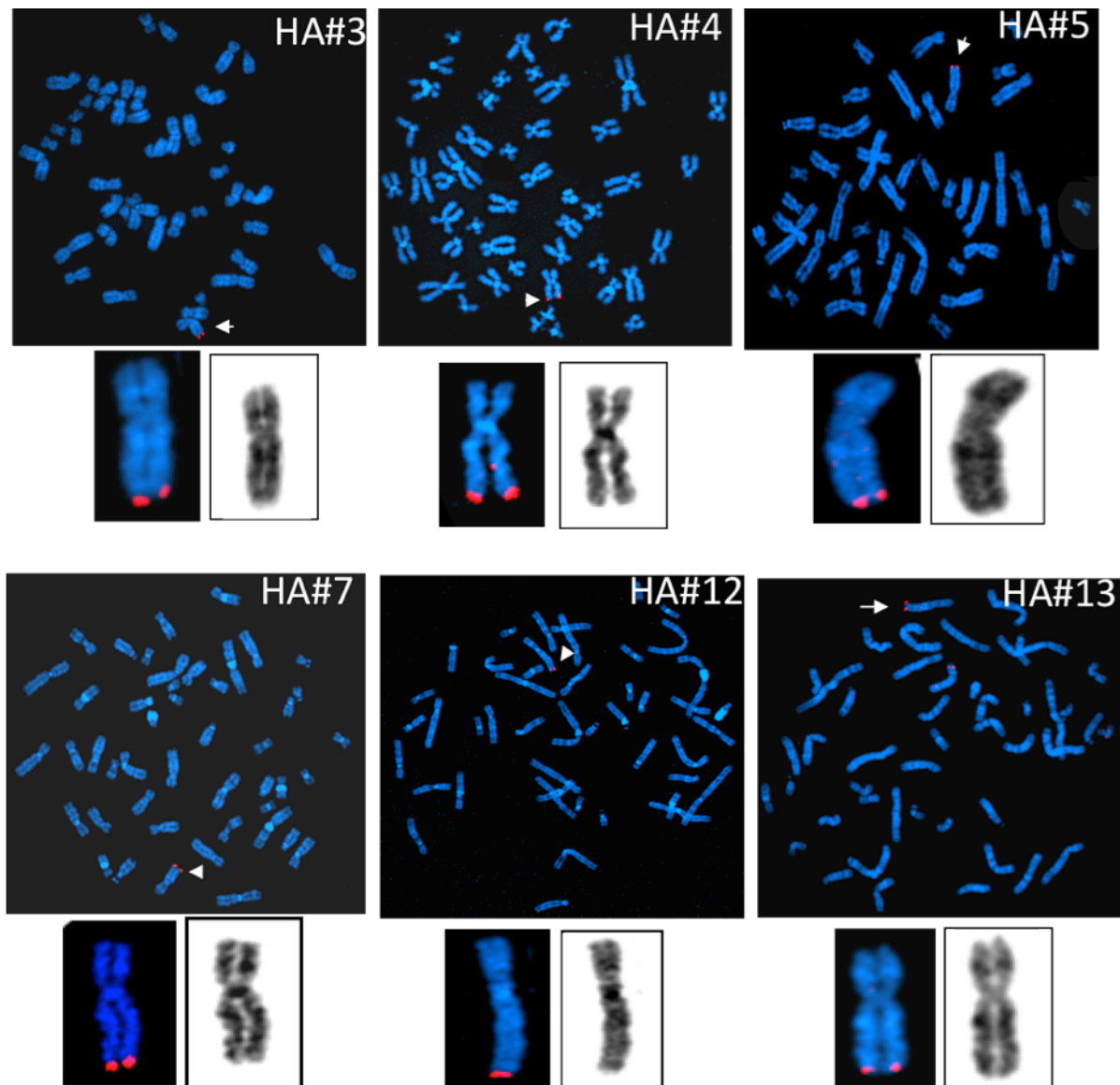
Figure 13: **FISH analysis of metaphase cells for 6 patients.** Hybridization of the *F8* specific probe (red signal) and counterstained with DAPI (blue) is shown for each analyzed patient. The magnification of cut outs of X Chromosome with and without hybridization, is shown below. The white arrow indicates the site of hybridization.

### 4.2.1.2 Exclusion of duplications via MLPA analysis

After exclusion of gross chromosomal rearrangements, a copy number variation (CNV) in the body of *F8* gene was investigated. Owing to limitations in resolution of FISH analysis failing to detect small microdeletions and microduplications, it is important to perform a complementary method. To detect such variations, MLPA was carried out for all patients. Except for patient HA#7 all patients showed normal MLPA

pattern (Figure 14). For patient HA#7, MLPA analysis revealed a novel large duplication of all exons of *F8* except exon 26 and partial triplication of exons 5-22. The investigation of the molecular mechanism of the severe hemophilia A phenotype was then continued as described in section 4.2.5.
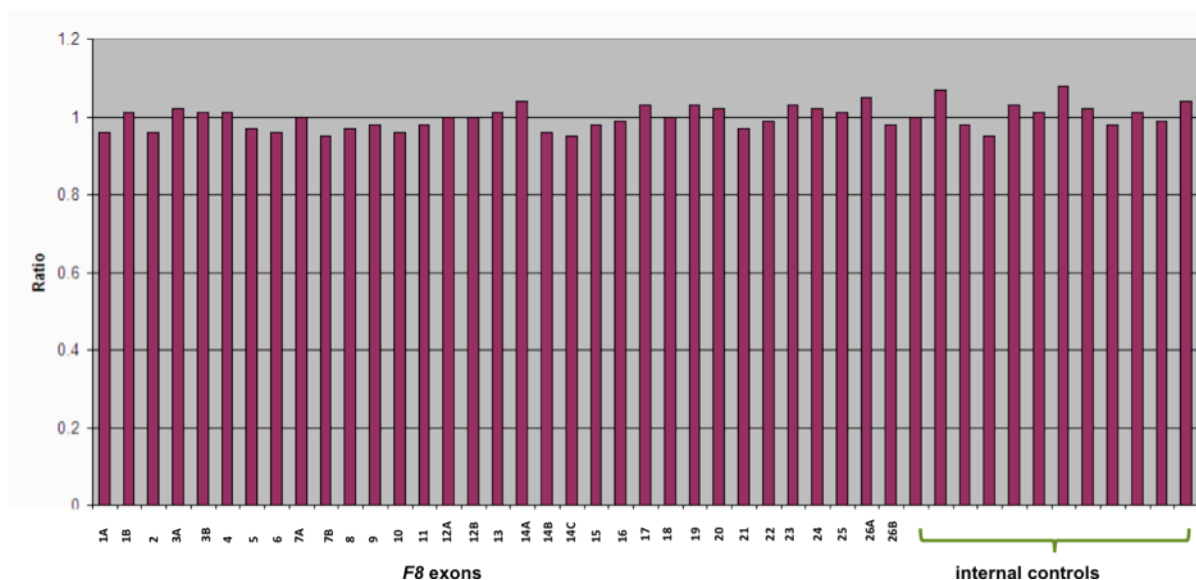


Figure 14: **Representive picture showing result of MLPA of *F8* for the analyzed patients.** All the patients showed no CNVs in *F8* gene expect for patient HA#7. On the X-axis exons of *F8* are shown. For some exons more than one probe is included. On the Y-axis the copy number of each exon is shown as a ratio.

### 4.2.1.3  Investigation of integrity of *F8* locus via LR-PCR

The whole genomic region of *F8* (~190 kb) was analyzed using a LR-PCR approach. After excluding the small and large rearrangements at chromosomal and exonic level, the intronic regions of *F8* were in focus of analysis. Therefore, the complete *F8* locus was analyzed using primers listed in Table S 10 by amplification of 28 overlapping PCR fragments. In all the patients except for patient HA#1, the amplification of all LR-PCRs was successful (Figure 15).
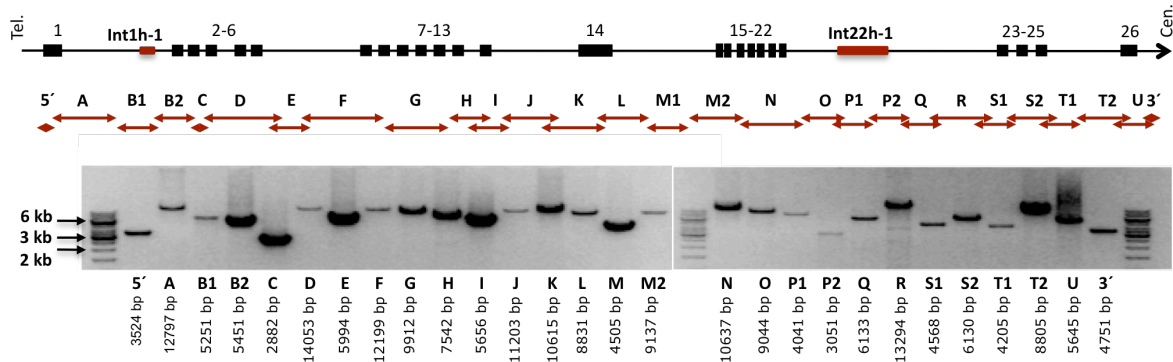
Figure 15: **Long Range PCRs covering the *F8* locus.** A schematic diagram of the genomic *F8* locus is shown together with the relative locations of individual LR-PCRs with the gel amplification picture and the expected product size. Amplification of all regions was successful for all patients except for HA#1 and HA#14.

This data indicates that the *F8* locus in all the other patients is intact. For patient HA#1 the amplification of region B1 (an amplicon of 5.25 kb region in intron 1 of F8, Figure 30 A) failed. The characterization of the molecular mechanism behind the hemophilia A phenotype in this patient is described in section 4.2.4. A size based analysis of the LR-PCR fragments failed to demonstrate any abnormalities in either of the other patients. Thereby, inversions and gross deletions and insertions were excluded.

### 4.2.1.4  Identification of deep intronic mutations using a NGS sequencing

In the next step, all 28 amplicons of each patient were sequenced using a NGS approach (3.2.6.2). The first step after the sequencing was to align the NGS reads, each with a length of 36 bp, to the reference genome sequence. After analysis of the differences between the patient's sequences and the reference sequence a list with variants was generated (Table S 2). The significance of these variants was determined by sorting out false positive calls. Validation of the variant calls was done by verification of each variant with frequency calls above 50%. Five SNPs corresponding to frequency of 50% calls and above were sequenced using ABI Sanger sequencing. A threshold of 90% call was found to be the false positive border and a minimum coverage of 3 reads as a cutoff corresponding to a real variation. Forty-four intronic variations were found to fulfill these detection threshold; 36 of them were reported in dbSNP database. Nine novel intronic variations were identified that are not reported in dbSNP (Table 5).

Table 5: List of unique intronic variations found in patients after NGS

(c.= coding sequence;*= stop codon; ins= insertion, a= patient HA#4 and HA#5 were not included in the NGS approach because another index patient from the same family was included).

| | SNP Nr. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| | NG_11403.1 g. | 88055 | 94667 | 123107 | 124347 | 124759 | 130722 | 138108 | 160166 | 190183 |
| | Location in *F8* gene | Intron 13 | Intron 13 | Intron 16 | Intron 18 | Intron 18 | Intron 21 | Intron 22 | Intron 22 | UTR |
| | Nucleotide change | C>G | G_ins | C>T | C>T | G>A | A>C | G>T | A>G | G>T |
| | GRCH37.p5 37.3 | c.1903+8034 | c.1903+11641 | c.5586+194 | c.5998+530 | c.5998+941 | c.6273+2864 | c.6429+6461 | c.6429+28519 | c.7055*+56 |
| Patients | HA#1 | This patient was not included in the NGS approach (for more details see section 4.2.4). | | | | | | | | |
| | HA#2 | C | **G_ins** | C | C | G | A | T | A | G |
| | HA#3 | C | - | C | C | G | A | T | A | **T** |
| | HA#4[a] | C | - | C | C | G | A | T | A | **T** |
| | HA#5[a] | C | - | C | C | G | A | T | A | G |
| | HA#6 | C | - | C | C | G | A | T | A | G |
| | HA#7 | This patient was not included in the NGS approach (for more details see section 4.2.5). | | | | | | | | |
| | HA#8 | **G** | - | C | C | **A** | A | T | A | G |
| | HA#9 | C | **G_ins** | C | C | G | A | T | **G** | G |
| | HA#10 | C | - | **T** | C | G | A | T | A | G |
| | HA#11 | C | - | **T** | C | G | A | T | A | G |
| | HA#12 | C | - | C | **T** | G | **C** | **G** | A | G |
| | HA#13 | C | - | **T** | C | G | A | T | A | G |
| | HA#14 | This patient was not included in the NGS approach (for more details see section 4.2.6). | | | | | | | | |

### 4.2.1.5  Association of SNPs found after NGS with *F8* mRNA expression, FVIII:Ag and FVIII:C levels in patients without mutation and healthy controls

In order to investigate, if the detected polymorphisms after NGS in *F8* could have an effect on the *F8* mRNA expression, FVIII:Ag or FVIII:C levels, an association study for all 36 detected SNPs was performed (Table S 6). A significant association between *F8* mRNA expression was found with rs28857481 (Mann Whitney test, p=0.0326) and rs6643622 (Mann Whitney test, p=0.0071); however no significant association was found between the later two SNPs and FVIII:Ag or FVIII:C levels. Among all the 36 SNPs, only rs34552198 was significantly associated with FVIII:C (Mann Whitney test, p=0.0490). Two SNPs showed border significance (rs78327897 (p=0.0611) and rs1050705 (p=0.0535)). From this data, one can conclude that none of the SNPs, which are present in healthy individuals alone have an effect on reduced FVIII:C levels in these patients (Figure 16 and Table S 7).
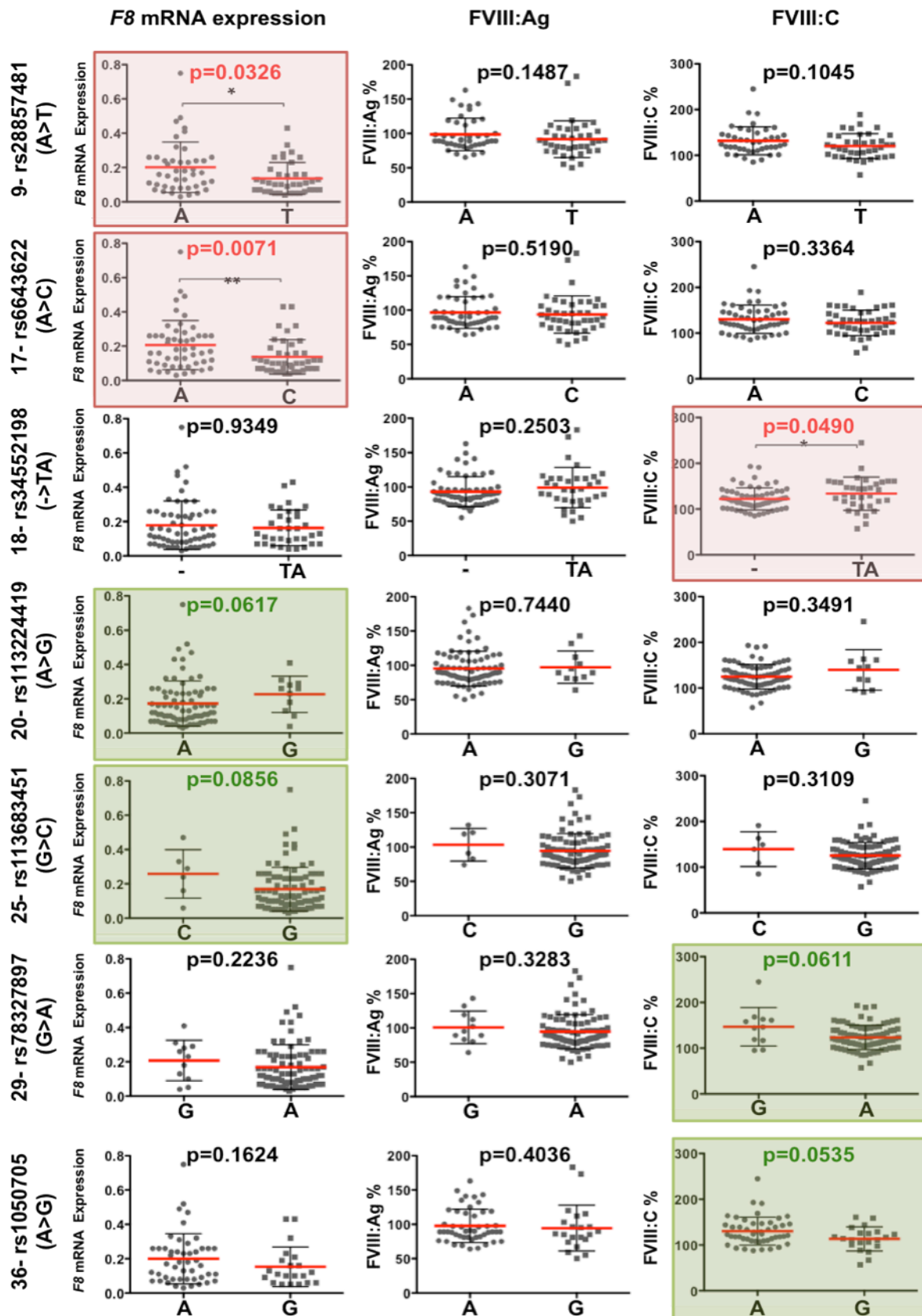
Figure 16: **Association between SNPs and the *F8* mRNA expression, FVIII antigen and activity levels in healthy population.** Only the data of the SNPs that show a significant or borderline association is shown. Graphs that are highlighted with red and green transparent boxes correspond to statistically significant or borderline significant association, respectively. P values are calculated using Mann Whitney test for the X-lined loci.

Therefore, all four significant (rs34552198) or border significant SNPs (rs1800292; rs78327897; rs1050705), found to be associated with FVIII:C levels, where used to construct haplotypes in both healthy controls and patients without mutations. In the control group, we could detect 6 haplotypes (Figure 17). When stratified for the AB0 polymorphisms the group of individuals with AA and AG genotypes at the rs687289 shows significant differences for the FVIII:C levels among different haplotypes (ANOVA Krustal-Wallis test p=0.0048). Two haplotypes, H4 and H5, were associated with significantly lower FVIII:C levels in comparison to H1, H2 and H3 haplotypes. Interestingly, all patients have haplotypes associated with higher FVIII:C levels (HA#2 and HA#3 have H3 haplotype while the rest of the patients have H1 haplotype). This analysis clearly excludes the effect of common polymorphisms in *F8* on lowering the FVIII:C in the studied patients (Table S 8).
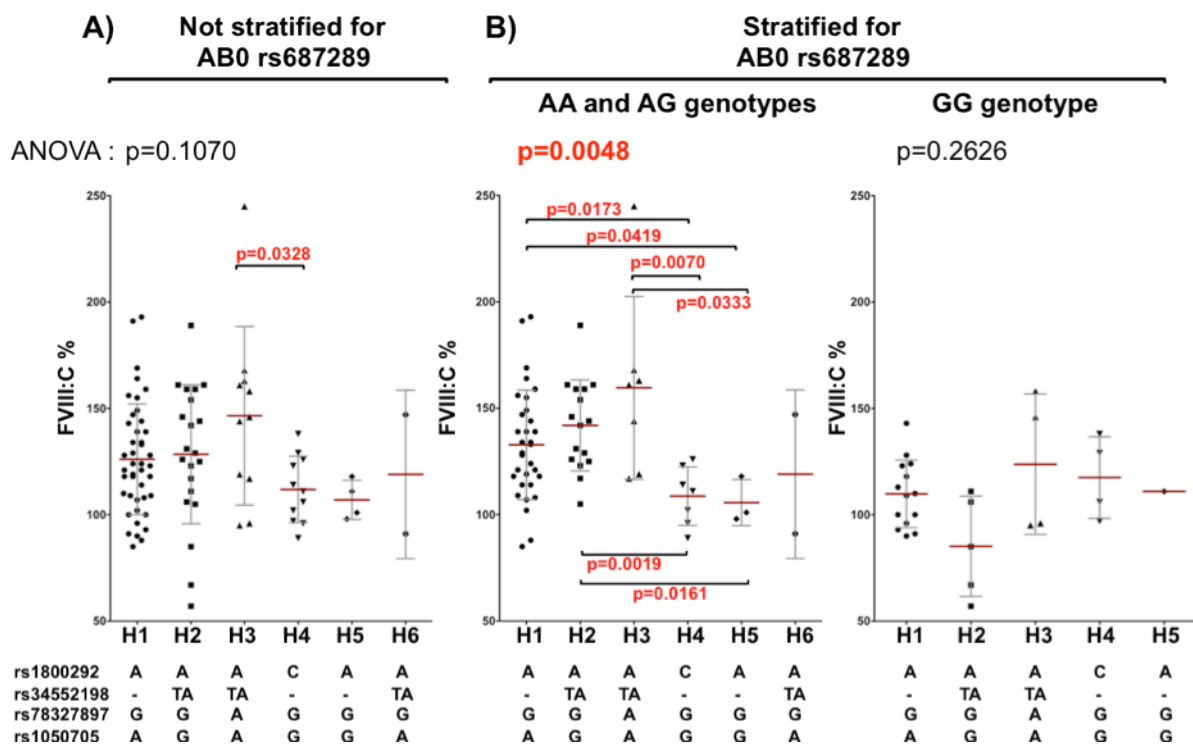


Figure 17: **Haplotypes analysis based on four SNPs showing the highest significance with FVIII:C.** FVIII:C is plotted for 6 constructed haplotypes when the haplotypes where A) not stratified and B) stratified for the AB0 SNP (rs687289). Significant differences between haplotypes are shown with their p-values. P values are calculated using the Krustal-Wallis test for the autosomal loci.

### 4.2.2  Mutation analysis: RNA level

#### 4.2.2.1 *In silico* analysis predicts new donor and acceptor sites for the majority of the unique NGS variants

The next step was to perform *in silico* analysis for all nine unique variants to study the possible effect of these transitions in generation of new donor splice site (dss) and acceptor splice site (ass) at mRNA level. Three different algorithms were used for this analysis. All SNPs were predicted to create new dss and/or ass sites at least by one of the prediction softwares, except SNP1 (c.1903+8034 C>G) and SNP8 (c.6429+28519 A>G). For two SNPs (SNP5 (c.5998+941 C>T) and SNP6 (c.6273+2864 A>C)), the softwares predict generation of new ass and dss in intron 18 and intron 21, respectively. However the predicted strength of the new splice site, according to Human Splicing Finder (HSF), were lower than that for the wild type sequence (Table 6).

For all the other SNPs, the splice site prediction servers predicted new ass and dss sites higher than the wild type sequence. The inconsistencies between different algorithms to predict the effect of specific variation on splicing necessitate the verification by experimental approaches. Therefore, two independent approaches were used to further investigate the effect of these variations on mRNA splicing pattern.

Table 6: *In silico* analysis of novel deep intronic SNPs found in patients after NGS.

(c.= coding sequence;*= stop codon; ins= insertion; dss= donor splice site; ass= acceptor splice site; mut= mutation; wt= wild type).

| SNP | Patient | Position | Intron | Nucleotide change | FRUITFLY (score 0-1) | Netgene2 (score 0-1) | HSF (score 0-100) |
|---|---|---|---|---|---|---|---|
| 1 | HA#8 | c.1903+8034 | 13 | C >G | not predicted | new ass: mut= 0.15 | not predicted |
| 2 | HA#2 HA#9 | c.1903+11641 | 13 | G_ins | not predicted | not predicted | dss: wt= 44.80,  mut= 72.71 ass: wt= 19.10,  mut= 72.76 |
| 3 | HA#10 HA#11 HA#13 | c.5586+194 | 16 | C>T | new dss: mut= 0.88 | new dss: mut= 0.56 | dss: wt= 63.06,  mut= 89.90 ass: wt= 65.56,  mut= 66.41 |
| 4 | HA#12 | c.5998+530 | 18 | C>T | new dss: mut= 0.99 | new dss: mut= 0.35 | dss: wt= 59.30,  mut= 86.13 |
| 5 | HA#8 | c.5998+941 | 18 | G>A | not predicted | not predicted | ass: wt= 75.72,  mut= 46.77 |
| 6 | HA#12 | c.6273+2864 | 21 | A>C | not predicted | new dss: mut= 0.45 | dss: wt= 65.68,  mut= 60.93 |
| 7 | HA#12 | c.6429+6461 | 22 | G>T | not predicted | not predicted | ass: wt= 51.30,  mut= 80.25 |
| 8 | HA#9 | c.6429+28519 | 22 | A>G | not predicted | not predicted | not predicted |
| 9 | HA#3 HA#4 | c.7055*+56 | UTR | G>T | new dss: mut= 0.41 new ass: mut= 0.14 | new dss: mut= 0.96 | dss: wt= 68.14,  mut= 94.98 ass: wt= 89.89,  mut= 90.49 |

### 4.2.2.2 RT-PCR analysis reveals aberrant splicing associated with two deep intronic variations found in two patients

To investigate the association of the identified variations with the hemophilia A phenotype, the corresponding molecular alterations were analyzed. To this end, *F8* mRNA was amplified for the patients harboring an intronic variation as previously described by El-Maarri et al. [42]. This approach would show the effect of the predicted new dss and ass of the variants on mRNA splicing. Indeed, in two of the patients with unique variants (SNP4; c.5998+530 in patient HA#12 and SNP5; c.5998+941 in patient HA#8; table 4 and 5), in region C2, an aberrant mRNA transcript was detected (exons 17-21). Both aberrant mRNA transcripts were bigger than normal expected mRNA transcript (Figure 18 C and Figure 19 C).
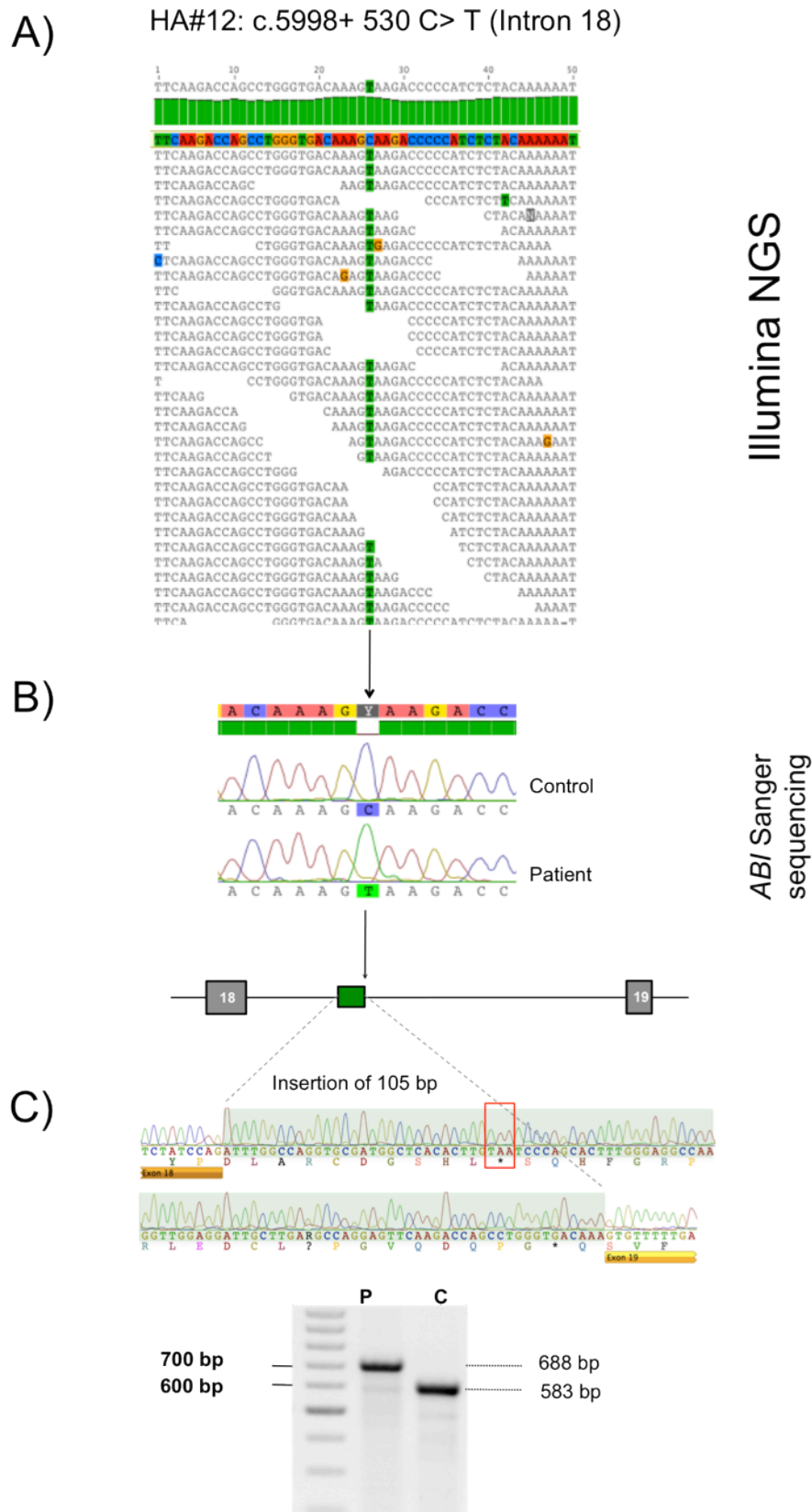
Figure 18: **: Identification of a novel intronic mutation in intron 18 in patient HA#12.** A) The intronic variations was identified in intron 18 using a NGS approach B) and verified using an *ABI* sequencing method. C) An aberrant mRNA band was observed when amplifying fragments spanning exons 17-21 (normal mRNA 583 bp; aberrant mRNA 688 bp in patient HA#12). Sequencing of the aberrant mRNA transcript revealed the introduction of a 105 of intron 18 into the mRNA sequence. The 100 bp ladder is used (C=control, P=patient).
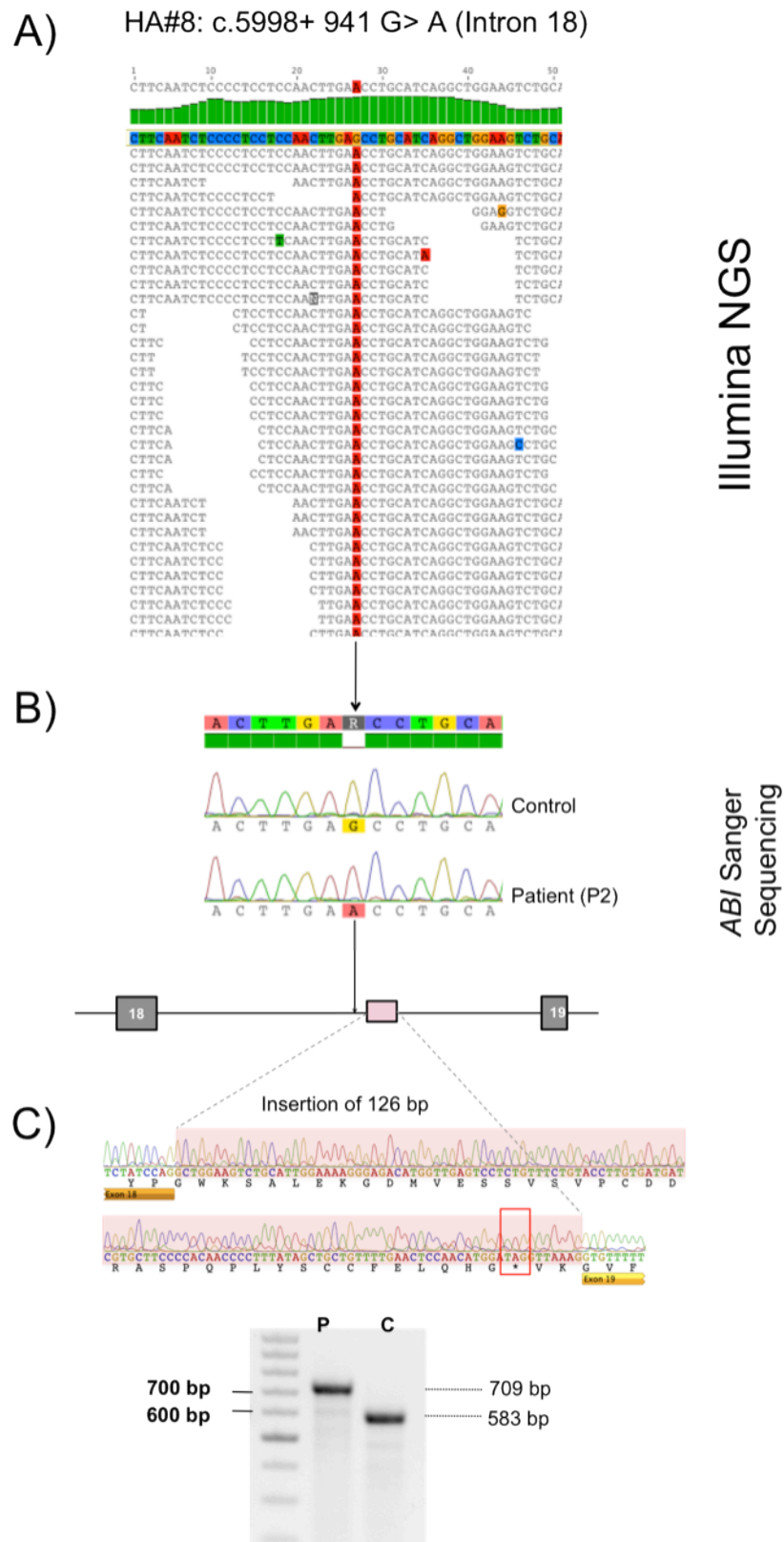
Figure 19: : **Identification of a novel intronic mutation in intron 18 in patient HA#8.** A) The intronic variations were identified in intron 18 using a NGS approach B) and verified using an *ABI* sequencing method. C) An aberrant mRNA band was observed when amplifying fragments spanning exons 17-21 (normal mRNA 583 bp; aberrant mRNA 709 bp in patient HA#8). Sequencing revealed the introduction of a 126 bp of intron 18 into the mRNA sequence. The 100bp ladder is used. (C=control, P=patient).

The PCR products of aberrant mRNAs were sequenced and the origin of the inserted sequences were determined using a Blast analysis (Figure 18 B and Figure 19 B). The analysis revealed an alignment with sequences of intron 18 (105 and 126 bp sequence insertion for patient HA#12 and HA#8, respectively). Analysis of the flanking sequence of these insertions at genomic DNA level revealed and confirmed the presence of the previously detected two intronic variations by the NGS analysis, the C>T transition at position c.5998 +530 in patient HA#12 and the G>A transition at position c.5998 +941 in patient HA#12. The new dss for the intronic mutation for patient HA#12 was highly predicted by two of the predictions algorithms.

Interestingly, for patient HA#8, the intronic variant appears to create a false donor splice site in intron 18 of *F8*; AG→AA at –950 bases downstream exon 18, which was not predicted using the splice site prediction programs. Sequencing results show that this transition activates another splice site in the reference sequence +10 bp downstream of the mutation. In accordance with the RT-PCR result and the fact that the inserted sequence was absent in controls and other patients, one can consider these variations as causal mutation. Both intronic mutations introduce stop codons that lead to truncated protein. Therefore, it is expected that the normally spliced product of *F8* should also be present in order to be compatible with a mild phenotype as observed in the patients.

Moreover, an unsteady result in region A2 (exons 4-8) was observed for patients HA#5 and HA#6 that belong to the same family. Additional products, smaller and larger in size, were amplified along with the normal band in HA#5 and HA#6, respectively. However, no intronic variations were found close to the insertion and deletion sites. This could result from activation of a pseudo-exon due to a mutational event reinforcing the insertion.

In general, the success rate of the RT-PCR was lower for patients than for healthy controls. All the other patients showed normal splicing patterns (Figure 20). The PCR products of all regions for each patient were sequenced. No alterations were found in the sequence of *F8* mRNA in the remaining patients.
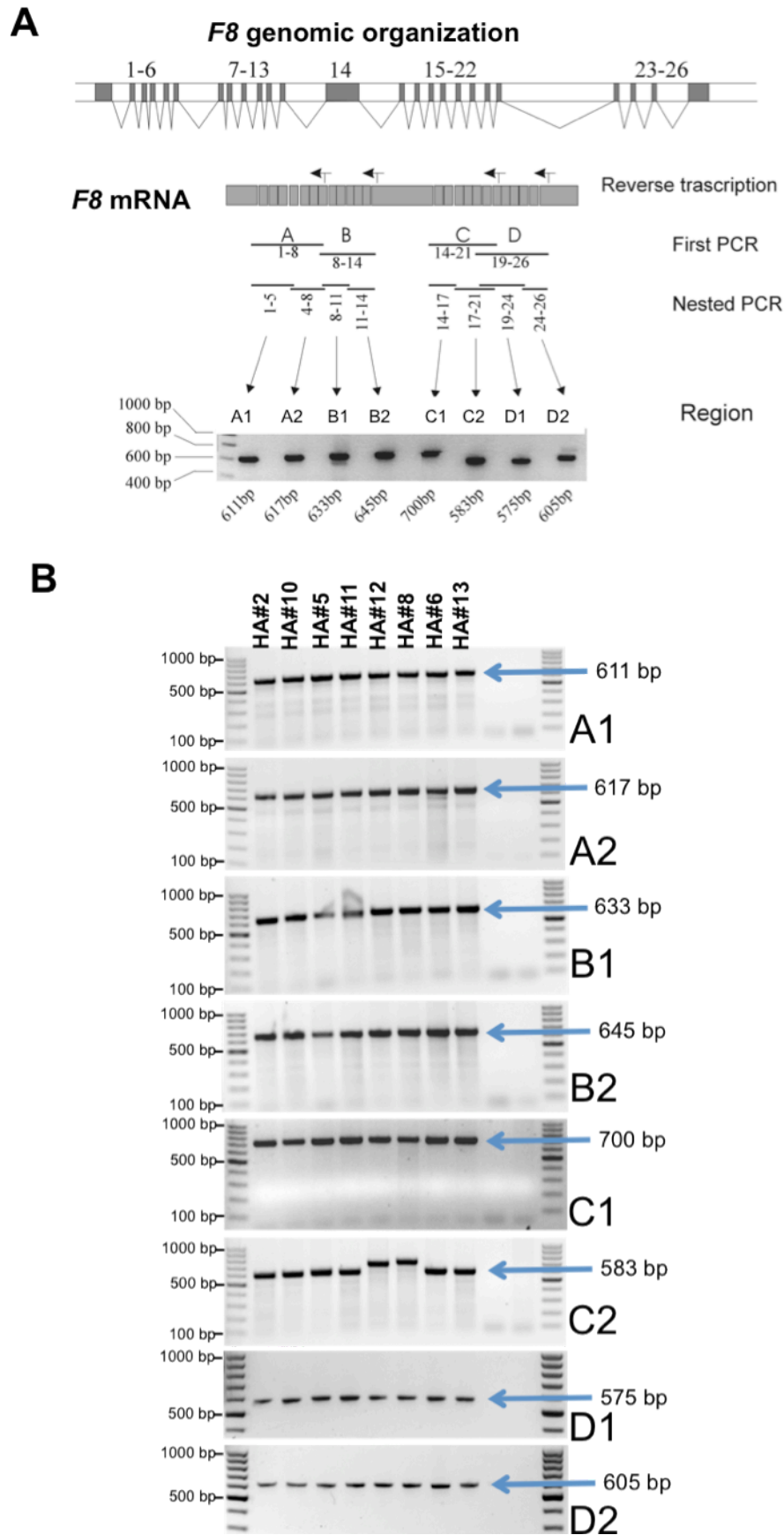
Figure 20: **RT-PCR approach of *F8* mRNA.** A) Schematic diagrams representing different RT-PCR fragments (according to [42]) with the B) corresponding experimental gel pictures of analyzed patients. The 100bp ladder wasa used.

### 4.2.2.3 Quantitative mRNA analysis reveals reduction of mRNA level in exons flanking the intronic mutations

Since the RT-PCR analysis did not reveal aberrant mRNA splicing in all patients with unique intronic variations, a Taqman-based real-time PCR assay was designed. Using this assay, the amount of *F8* mRNA was measured at exon boundaries, where an alternative splicing was expected. The amount of *F8* mRNA was also measured at a region of *F8* that is not subject to alternative splicing. As an internal reference control, a probe was used, which was designed for exon 1-2 (Hs00240767_m1, Applied Biosystems). This probe was upstream to all the variations found after NGS approach. The second probe spans the two exons flanking the intronic variations. The ratio of the amount of *F8* mRNA at internal control was calculated to the second probe and compared between a given patient and healthy controls. A reduced ratio in comparison to the controls indicates the presence of alternative splicing at the investigated region. Primers are listed in Table S 17.

For patients HA#12 and HA#13 with aberrant mRNA splicing in intron 18 (after RT-PCRs), the amount of normal *F8* mRNA was significantly reduced compared to the controls (Figure 21). The inserted sequences in both patients led to a frameshift generating a premature termination codon. This reduction can explain the mild hemophilia A phenotype by decreasing the amount of normal *F8* mRNA transcript in both patients (Mann Whitney test, p= 0.0005 for c.5998+530 C>T and p=0.0004 for c.5998+941 G>A).
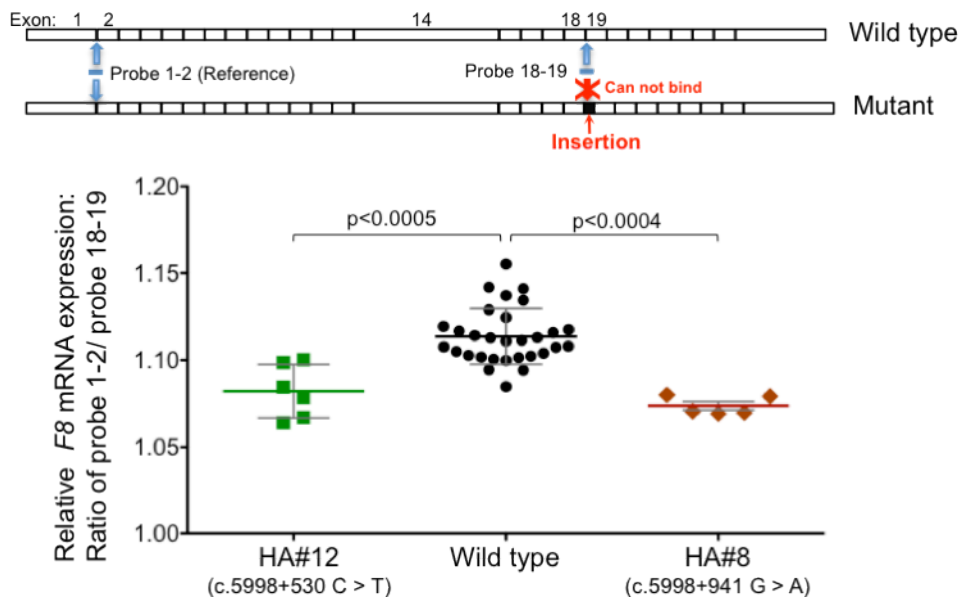
Figure 21: **Quantitative mRNA analysis for two intronic variants in intron 18.** The positions of the Taq-Man probes are shown on the upper panel. The ratio of control probe to the tested probe is shown at the lower panel. A significant reduction of wild type F8 mRNA was observed for both patients (p=0.0005 for patient HA#12 and p=0.0004 for patient HA#8).

For patient HA#11 with a unique intronic variation in intron 16 (SNP3 c.5586+194 C>T), a reduction of *F8* mRNA expression was detected in the exons flanking the intron 16 (Mann Whitney test p= 0.019). The same SNP was also detected in two other patients (HA#10 and HA#13). However, no mRNA was available from these patients to perform the qRT-PCR assay. This data suggestive for the presence of an aberrant mRNA for patient HA#11. This aberrant transcript is most probably caused by the intronic variation. However, this is missed using the nested RT-PCR approach applied in this study (Figure 22 A).

On the other hand, for SNP2 (c.1903+11641 (-) > G in intron 13), which is present in patients HA#2 and HA#9, quantitative RT-PCR approach did not show a significant difference in amount of wild type *F8* mRNA level for the patient HA#2 when compared to healthy controls (Figure 22 B). No RNA was available from patient HA#9 to perform the qRT-PCRs. The later result suggests that even if this variation is causative for the disease, it is not through inducing abnormal splicing. But rather through another still unexplained mechanism. Interestingly, SNP2 was present in all affected members of the index patient and was absent in all healthy individuals (Figure 27).
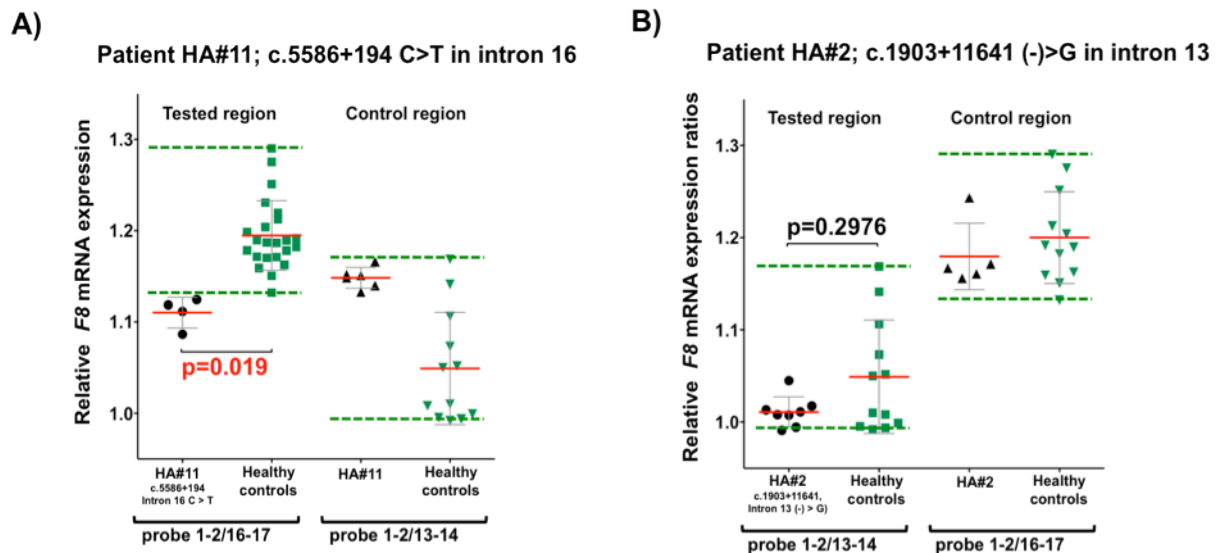
Figure 22: **Quantitative RNA analysis for two intronic variants.** A) The c.5586+194 C>T in intron 16 and B) the c.1903+11641 (-)>G in intron 13 were analyzed using a qRT-PCR approach. In each case, the tested region is shown for the patient (black) and healthy controls (green). In addition, a control region is shown at the right side of each graph. In this region no differences between the patient and healthy controls was expected. The upper and lower normal ranges are represented by horizontal dashed green lines. Different data points represent independent repetitive measurement for the patients, while different data pints of the healthy controls, represent independent measurements of different individuals. P-values are based on a non-parametric Mann Whitney t-test.

### 4.2.2.4   Investigating the effect of SNP9 on *F8* mRNA

To evaluate, if SNP9 could have an effect on *F8* mRNA, a nested RT-PCR approach was designed to amplify the UTR of *F8* using the patients mRNA. The cDNA synthesis was done using a primer in 3´ UTR (chrX: 154,064,856- 154,064, 877;hg 19). The nested PCR approach revealed two bands, one having the expected size (740 bp) and one smaller in size (581 bp, Figure 23 A). Sequencing of the aberrant band revealed a deletion of 159 bp in UTR located 2 bp downstream of SNP9. The variation results in alternative splicing that results in an aberrant mRNA with a shorter 3´ UTR, which could influence the stability of *F8* mRNA (Figure 23 B).
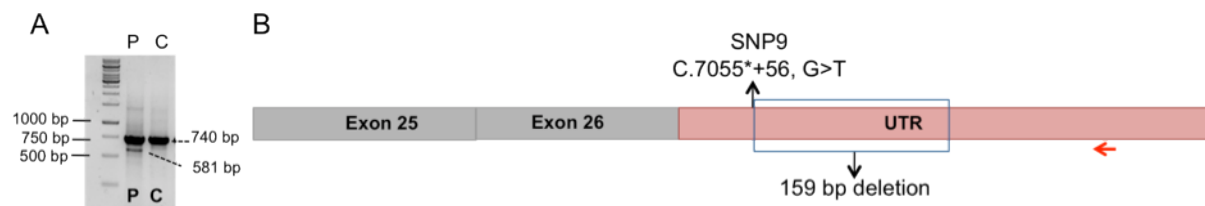


Figure 23: **Nested RT-PCR approach of SNP9 for patient HA#3.** A) RT-PCR using a gene specific primer in UTR. An aberrant mRNA transcript was detected in patient HA#3. B) Sequencing of the aberrant mRNA revealed deletion of 159 bp (boxed) within the UTR of *F8* mRNA. The arrow indicates the position of the primer used for the reverse transcription.

Notably, the same variation was found in the obligate carrier, which shows a significant reduction of FVIII:C levels (Figure 24).
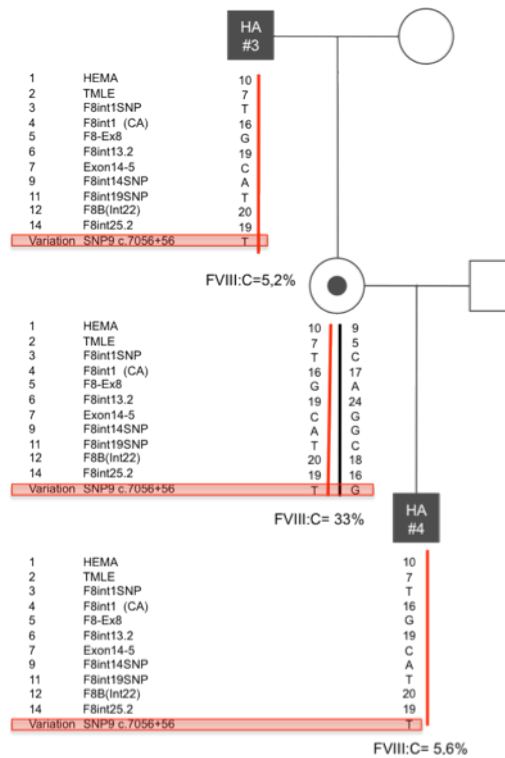


Figure 24: **Haplotype segregation analysis**. Both affected family members (HA#3 and HA#4) and the obligate carrier with significant reduced FVIII:C levels have SNP9.

This mutation results in changes in alternative splicing of *F8* mRNA. The deletion in UTR in the aberrant band might affect the structure and accessibility of *F8* mRNA to its interacting partners. To explore this possibility, the wild type and mutant sequence were screened for their microRNA-binding pattern, using the miRBase microRNA prediction server. No difference was observed in binding of mature microRNAs. However, miRBase results showed differences in 4 stem-loop microRNAs. The mutant has lost it ability to bind to three microRNAs (hsa-mir-1470, hsa-mir-493 and hsa-mir-607). On the other hand because of the variation a new microRNA can bind to UTR (hsa-mir-4754) (Figure 25).
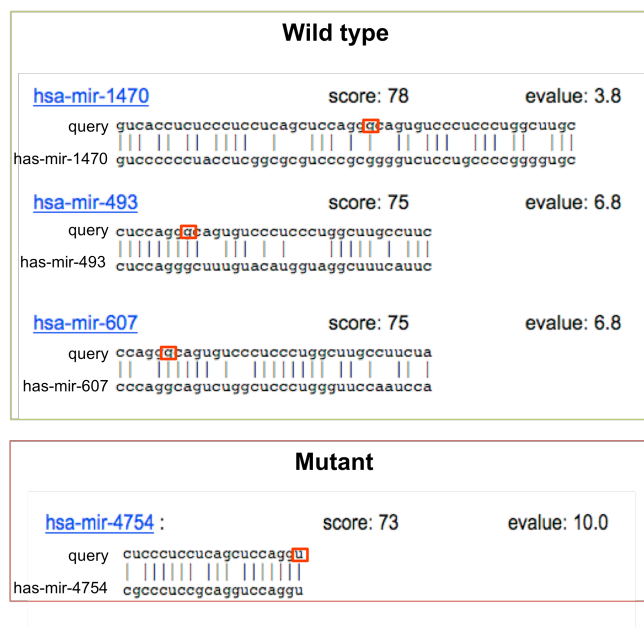
Figure 25: **Results of microRNA binding prediction using miRBase server.** The analysis revealed differences between the stem-loop microRNAs binding to the 3´ UTR of *F8* mRNA in presence and absence of the mutation. The base change is shown in red box. The wild type and mutant sequences are shown in positive and negative strands, respectively.

The secondary structure of the RNA was analyzed using the RNAFold program. The G>T transition shifts the base pairing in the stem loop. The free energy of the thermodynamic ensemble is -4.44 and -1.53 kcal/mol for wild type and mutant sequence, respectively. This could be a hint that through the structural change in the mRNA the accessibility of mRNA for its interacting protein, which is essential for the translation, is disrupted (Figure 27). One guanine–cytosine pair is not present in the mutant RNA structure, when compared to the wild type. Longer and stronger paired zones are more stable (guanine-cytosine have three hydrogen bonds and adenine-uracil only two). This could prevent the ribosome from breaking the pairing and proceeding translation. The G>T transition is not in disfavor of RNA in regard to translation processing.
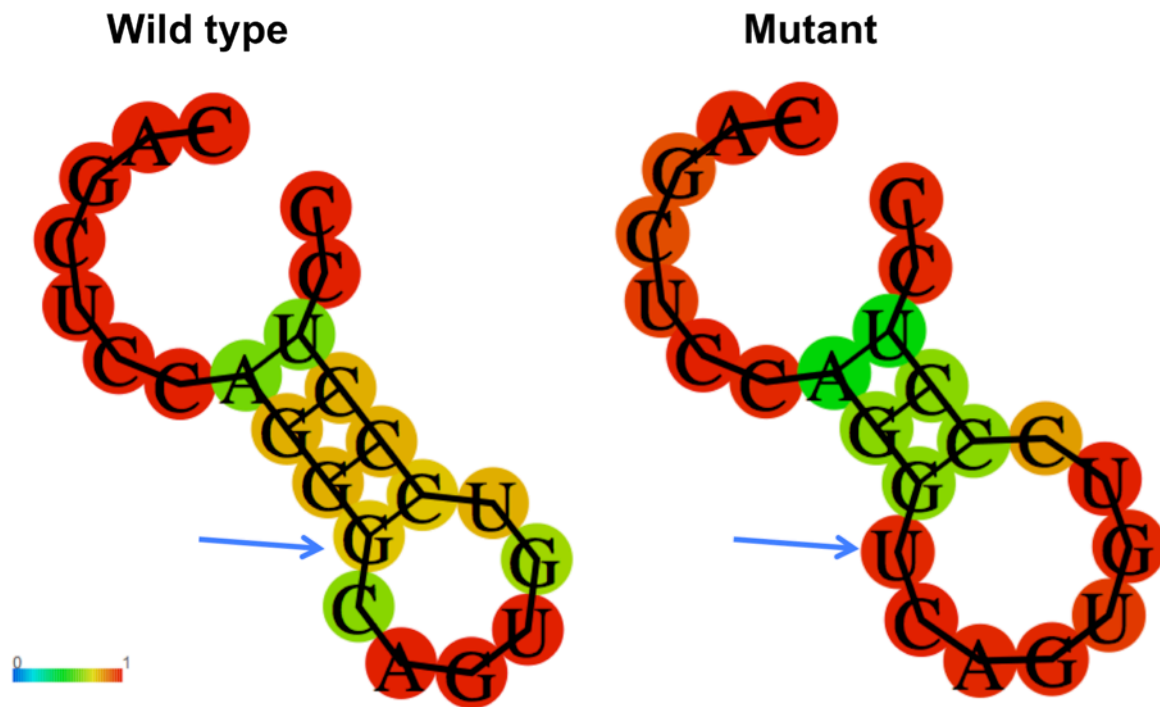
Figure 26: **Minimal free energy RNA structure of wild type and mutant RNA predicted with the RNAfold program.** The structures are colored by base-pairing probabilities. The arrow indicates the location of the base change in wild type and mutant sequence. The color indicates the probability of being unpaired from 0 to 1.

### 4.2.3  Within one family the same *F8* allele is segregating with all affected members

In the remaining patients where the causal mutation was not found, a haplotype analysis was done to elucidate if the same allele is shared between the index patients and the affected family members in a given family. To address this, pedigrees of 2 families were constructed. Fifteen intragenic and extragenic polymorphic markers on X chromosome in the *F8* locus and its vicinity were selected and analyzed using two SNP databases (dbSNP and HapMap) for the most informative polymorphic markers. Using these markers it was possible to follow the segregation of the *F8* allele in the index patient and the family members. Thus, it was possible to identify whether all individuals with reduced FVIII:C levels in a family share the *F8* allele genotype with the index patient.

*Family 1 (HA#2):* This family includes only one patient with mild hemophilia A phenotype of about 21% FVIII:C and 34% FVIII:Ag levels (Figure 27). In this pedigree the *F8* allele of the index patient was also present in 5 potential carriers (I-2, II-4, II-5, III-1, III-6) of which only one show significant reduction of FVIII:C (III-6: 35%). In all cases the XIR pattern was analyzed to explain the reason for different FVIII:C levels despite the presence of the index patient allele. The *HUMARA* assay in cases with normal FVIII:C levels (I-2, II-4, II-5 and III-1) showed inactivation of index patient's allele in 82%, 100%, 95% and 69% of the cells, respectively. This could explain their relative normal FVIII:C levels.
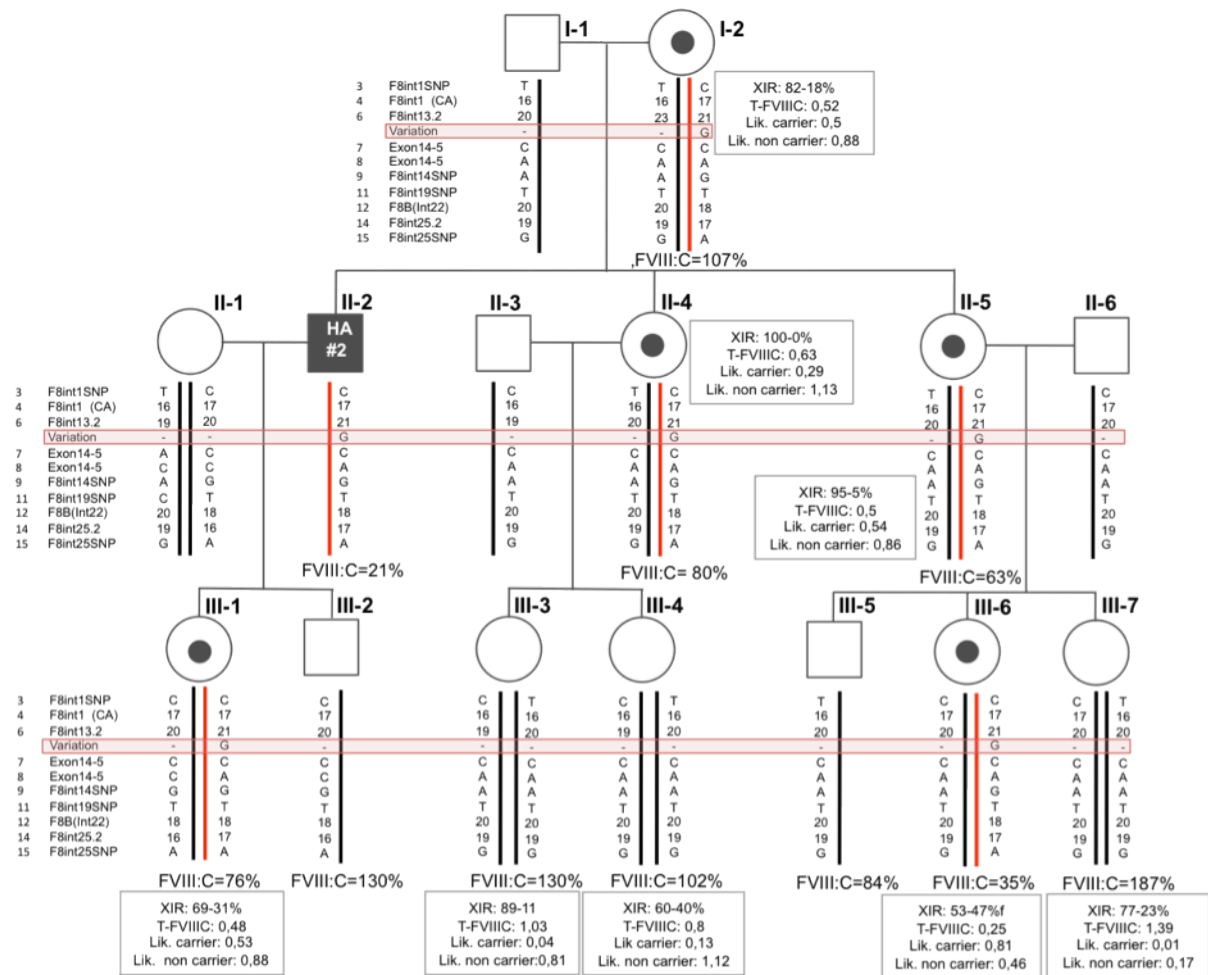


Figure 27: **Haplotype segregation analysis in family 1.** The X-chromosome inactivation (XIR) ratio is shown for the potential carrier females as well as the likelihood of being carrier and non-carrier. A horizontal transparent pink box highlights the unique variation of G insertion (SNP2).

Moreover, for III-6 about 50% of cells were expressing the patient's allele, explaining the reduction of FVIIII:C levels. Based on an informative marker in exon 14 of *F8*, RNA analysis was performed for member III-1 (daughter of the index patient). Using this assay we could determine that only the non index-patient allele at RNA level was detectable. This result supports the normal FVIII:C levels expected based on the XIR pattern (Figure 28).
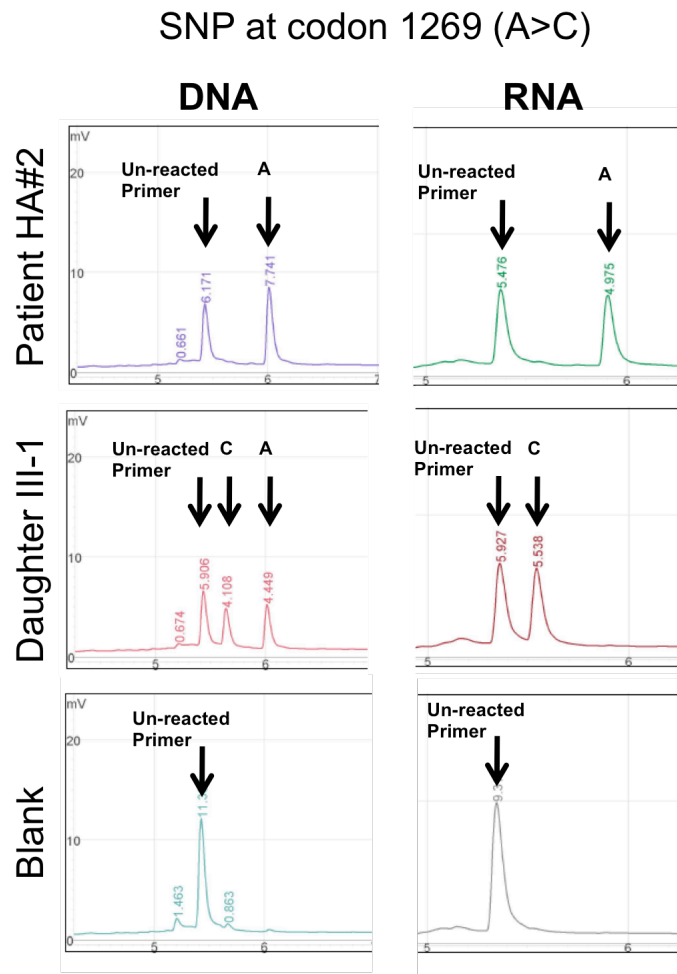


Figure 28: **Allelic analysis of the DNA and RNA of C to A SNP at codon 1269 of *F8* in patient HA#2 and his daughter**. Two alleles are present in the DNA of the daughter, while only the maternally inherited allele is present in the RNA, which indicates the skewed inactivation of the paternal allele.

***Family 2 (HA#5 and HA#6):*** This family includes two patients and three obligate carriers. Patient HA#6 has a moderate to mild phenotype with FVIII:C of 5.3% and FVIII:Ag level of 7.3%. The second patient (HA#5), appears to be more affected and shows a moderate phenotype with FVIII:C of 3%. FVIII:Ag level was also reduced to 4%. The slight difference in severity of phenotype as well as activity and antigen level could be attributed to modifying environmental, genetic or epigenetic factors. Based on the haplotype analysis, both patients share the same *F8* haplotype. Moreover, all three obligate carriers share the same haplotype with the index patients (Figure 29). Three of them showed also a decrease in FVIII:C that corresponds quite well to the XIR pattern. Markers in exon 14 of *F8* were not informative in this family thus it was not possible to follow the allelic presence on the mRNA levels.



Figure 29: **Haplotype segregation analysis in family 2.** The X-chromosome inactivation (XIR) ratio is shown for the potential carrier females as well as the likelihood of being carrier and non-carrier.
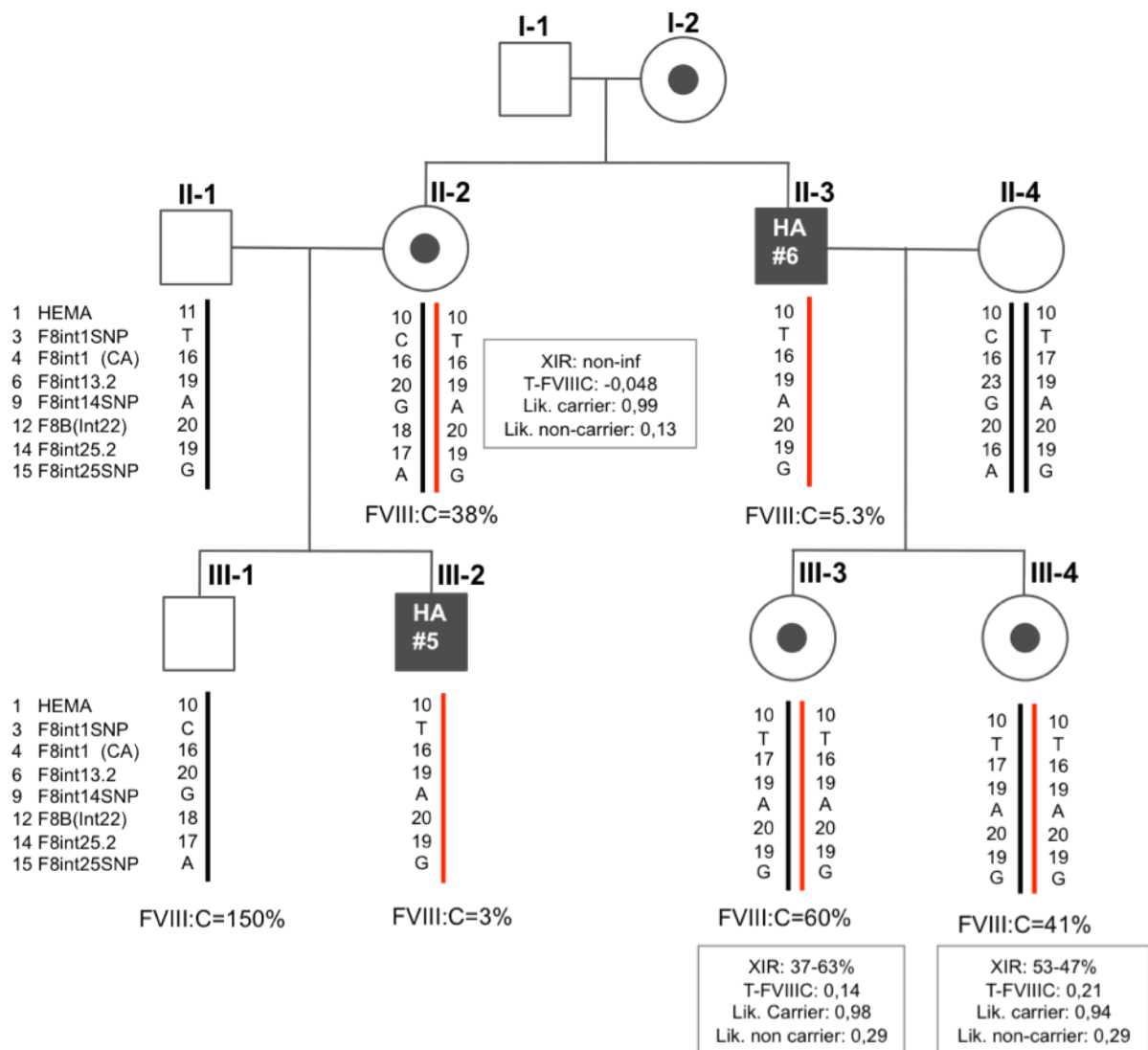
### 4.2.4 Identification of a rearrangement within the intron 1 of *F8* in patient HA#1

The index patient (HA#1) is a male hemophilia A patient with severe phenotype (FVIII:C<1%) who suffers from no other known health problems. No inhibitors have been detected. MLPA analysis showed normal allele copy number for *F8*. No *F8* mRNA transcript was detectable in his total peripheral blood mononuclear cells [75]. Both, intron 1 and 22 inversions, were excluded. Both suggestive carriers of the family show reduced levels of FVIII:C (33% and 42% for his mother and sister, respectively). Skewed X-chromosome inactivation was excluded in both carriers using HUMARA assay.

#### 4.2.4.1 LR-PCRs reveals a breakpoint in *F8* gene

To elucidate the molecular mechanism behind the lack of mRNA synthesis for patient HA#1, LR-PCRs were performed. This approach was designed to analyze the integrity of *F8* locus by amplifying 28 overlapping fragments (3.3.3.2). Inability to amplify region B1, a 5.25 kb fragment located in intron 1 of *F8* from the genomic DNA of the index patient indicated the presence of a genomic breakpoint in this region (Figure 30 A). Mapping PCRs using 5 different primer pairs were accordingly designed to further narrow the borders of the non-amplifiable region (Table S 11). Using this approach, the extent of the non-amplifiable region was further localized to a 942 bp region (Figure 30 B). The putative break was localized in intron 1 about 1 kb upstream of Int1h-1 repeat (Int1h-1 on chrX: 153,887,576- 153,888,616; hg18), which is involved in the recurrent intron 1 inversion.
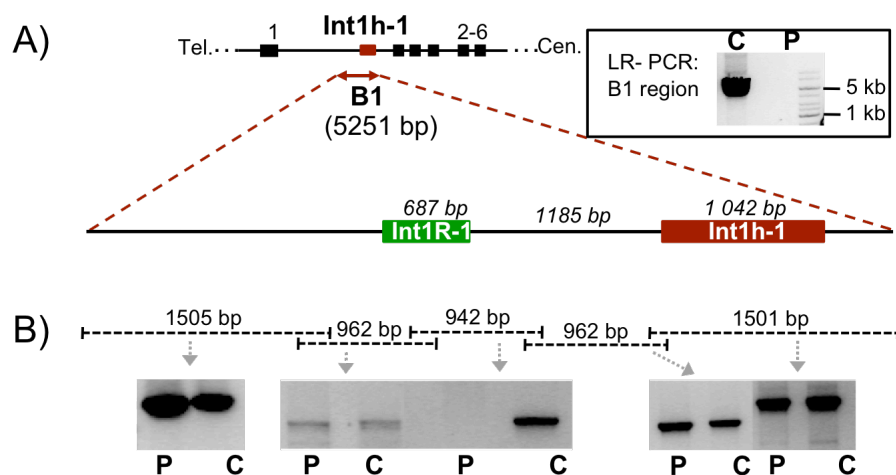
Figure 30: **Location of the breakpoint in *F8* intron 1**. A) The relative location of Int1h-1 and the Int1R-1 repeat in the non-amplifiable long-range PCR are shown; a schematic presentation of *F8* locus from exon 1 to exon 6 is shown at the upper part and the PCR amplification of the region B1 is shown in the box at the upper right. The 1kb ladder was used. B) Detailed localization of the breakpoint using overlapping short-range PCRs (C: control non-hemophilic DNA, P: Patient).

### 4.2.4.2 Sequence analysis near the breakpoint identified the presence of inverted repeats

Afterwards, a sequence homology search was performed on the sequence of the non-amplifiable region. Since both previously reported inversions in *F8* are caused by intrachromosomal recombination between highly homologous inverted repeats [23, 24], the hypothesis was investigated whether a recombination between homologous inverted sequences might also be involved in the cause of this breakpoint. A dot plot matrix analysis on the sequence of the breakpoint was performed (Figure 31 A). Analysis of the entire 5.25 kb intronic sequences flanking the breakpoint (Figure 31 B) revealed the presence of a 687 bp sequence which is highly homolog to a sequence at Xq28 in an inverted orientation, designated as Int1R-1 (chrX: 153,889,754 -153,890,442; hg18) and Int1R-2 (chrX: 153,433,817 -153,434,506; hg18), respectively (Figure 31 B). This result indicated a possible homologous recombination between the identified repeat in *F8* intron 1 and its highly homologous inverted repeat in intron 2 of inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase gamma (*IKBKG*).
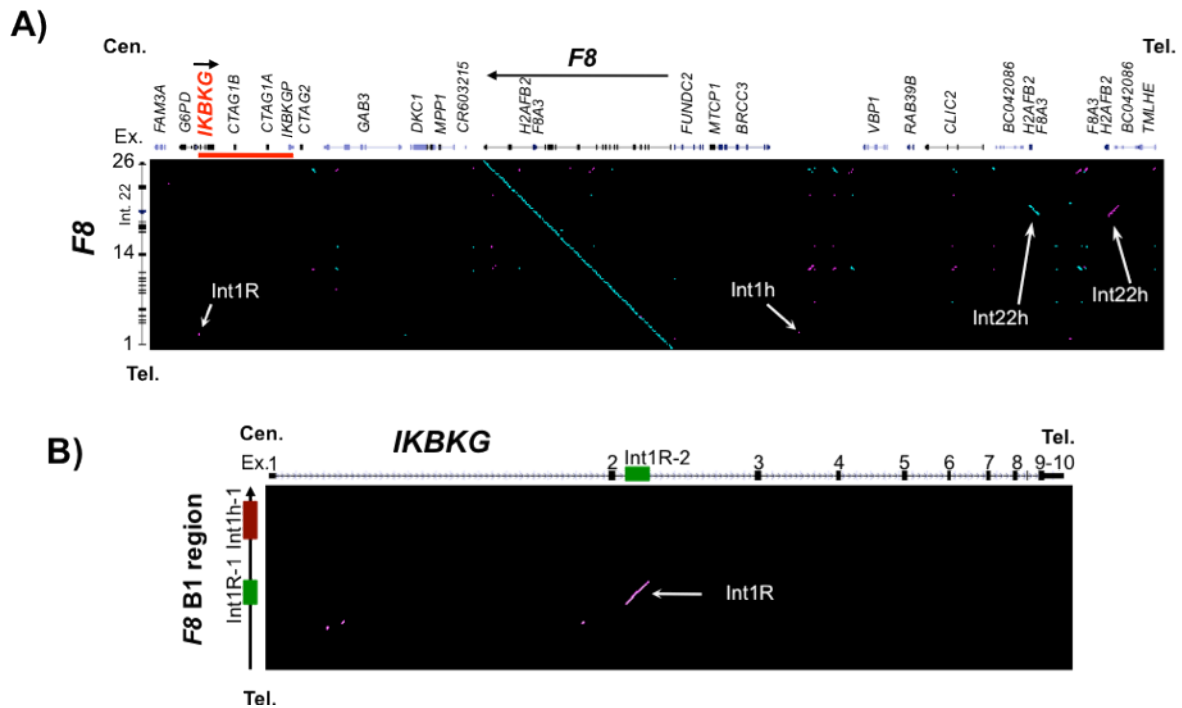
Figure 31: **Homology analysis to identify identical repeats within 1Mb (500 kb upstream and 500 kb downstream of the breakpoint).** A) Dot plot of 1 Mb region containing the B1 region at the center vs. *F8* locus. The identical repeats, Int22h and Int1h, known to be involved in inversions in *F8* and the newly identified Int1R, are labeled. The horizontal red bar at the top left of the figure represents the duplicated 94 kb region in the index patient. B) Dot plot of *IKBKG* vs. the *F8* B1 region; The Int1h-1 and Int1R repeats are labeled in red and green, respectively. Inverted repeats in dot plot are colored purple and direct repeats are blue. The word size and the tile size used for the dot plot are 50 and 2000, respectively. Geneious software was used for the analysis.

Subsequently, BLAT analysis was done on the sequence of the two repeats. Analysis revealed a 98.9% homology between the two repeats (Figure 32). This data strengths the possibility of an inversion resulted from a rearrangement between the two repeats that are localized on the Xq28 in an inverted manner.

Figure 32: **Alignment of the two repeats (Int1R-1 in *F8* to the Int1R-2 in *IKBKG*)**. The sequence highlighted in pink represents the consensus sequence. The aligned sequences correspond to *F8*: chrX: 153,889,755- 153,890,460 and *IKBKG*: chrX: 153,433,810- 153,434,496; hg18.

### 4.2.4.3  Southern blotting verified the presence of a rearrangement

Southern blot analysis was performed to verify the presence of the rearrangement within the intron 1 of *F8*. The proposed novel DNA rearrangement would result in a change in length of restriction fragments surrounding the breakpoint. To detect such changes, several restriction enzymes were used whose cleavage sites lie outside or within the region of homology. The probe was designed and amplified (Table 2) to bind to a sequence upstream of the non-amplifiable region, in a non-repetitive unique sequence upstream the Int1R-1 homologous region (Figure 33 B, C). Southern blotting verified the presence of a rearrangement in the patient and his mother after

digesting with different enzymes (*Bam*HI, *Eco*RI, *Hinc*II or *Nde*I), whose cutting sites lie outside the inverted repeat regions (Int1R-1). Altered band patterns were detected in the patient and his mother compared to the wild-type healthy male and female controls. The aberrant band sizes from the patient's samples were consistent with a rearrangement by which *F8* sequence is fused to a sequence in *IKBKG* with the breakpoints in the Int1R-1 repeat (Figure 34 A).
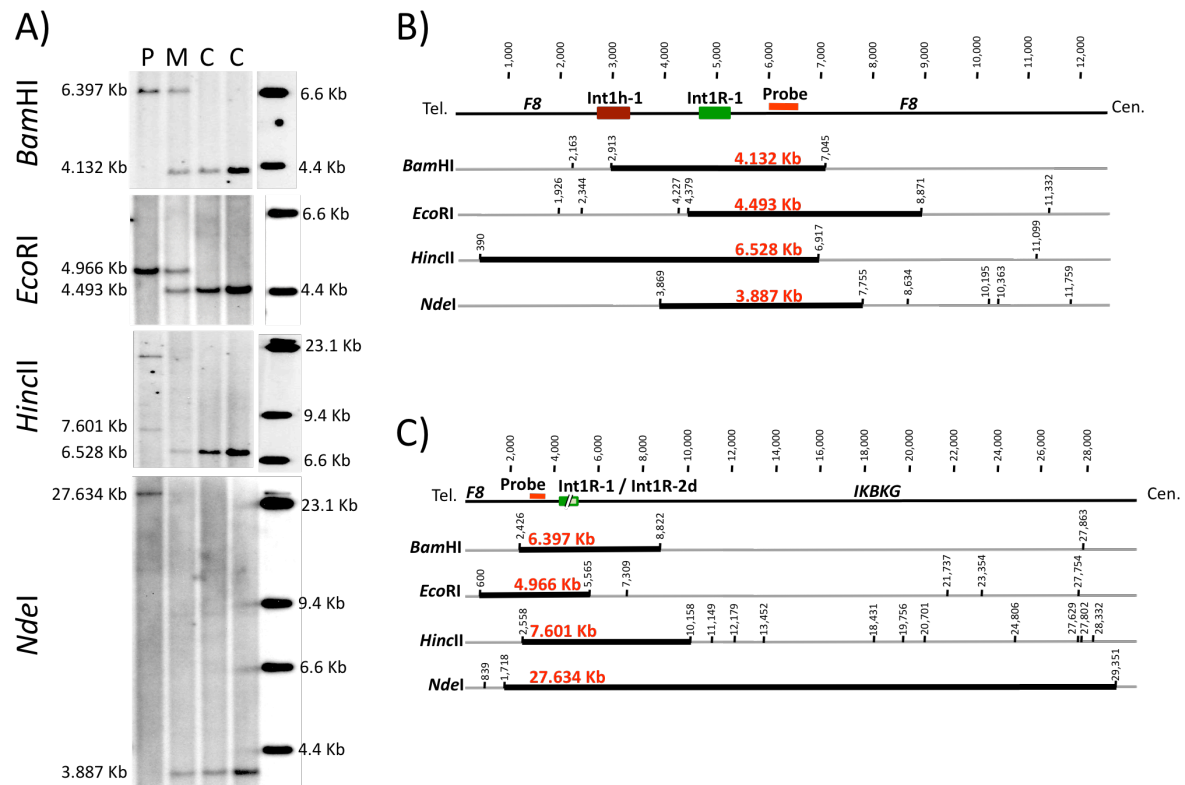


Figure 33: **Southern blot analysis.** A) Radiograms of 4 restriction enzymes used for the patient (P), mother (M) and two control (C) samples. The band in the M lane for *Hinc*II was weak due to the fact that this is a methylation sensitive enzyme and will give several bands when the cutting sites have a mosaic methylation pattern on a given allele; thus on the original radiogram several weak bands were seen with different sizes ranging from 7.6 up to 20,7 Kb. B) and C) Details of the restriction map around the Int1R repeat of the normal and mutant alleles respectively; orange horizontal bar indicates the location of the southern blot probe; Int1h-1 and Int1R-1 repeats are represented by red and green bars, respectively. Bold horizontal lines indicate the normal wild type (in B) and the mutant (in C) sizes of the restriction bands for each restriction enzyme. The horizontal scale at the top of panels B and C is in bp.

### 4.2.4.4  Inverse PCR identified the breakpoint junction *F8-IKBKG*

An inverse PCR approach was designed and applied in order to amplify the DNA fragment containing both breakpoint junctions associated with the rearrangement. Primers were designed to allow inverse PCR amplification of regions upstream and

downstream of the *F8*:Int1R-1 repeat (Figure 34). The inverse PCR amplification revealed aberrant upstream and downstream PCR products that were generated from the patient's DNA and were absent in healthy individuals. The DNA from his mother and sister were heterozygous for both PCR amplifications (Figure 34 B, C). Subsequently, the inserts were sequenced using a DNA walking strategy after being cloned into the pCR-XL-TOPO vector. Sequencing revealed that the sequence upstream from *F8* intron 1 was joined to intron 2 of *IKBKG* (Figure 34 B). The sequence downstream of *F8* intron 1 was joined to a sequence close to *IKBKG* (Figure 34 C) about 65 kb away from the region joined to the upstream breakpoint. However a sequence comprising several repetitive elements, with a high homology to a region in intron 2 of *FUNDC2* (chrX: 153,919,264-153,918,534; hg18), was inserted between the sequence downstream of *F8* intron 1 repeat (Int1R-1) and the duplicated sequences in *IKBKG* region (Figure 34 C).
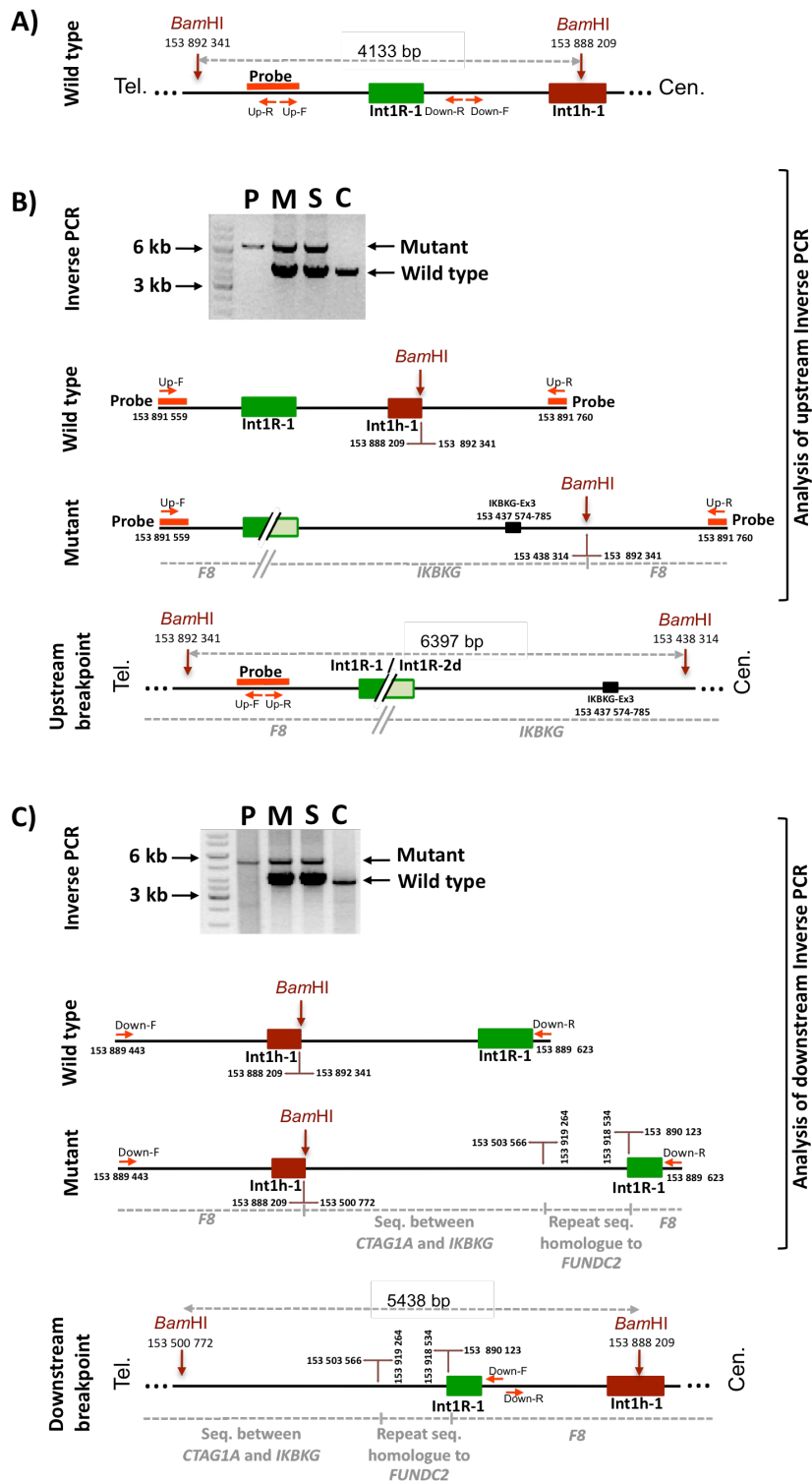
Figure 34: **Inverse PCR analysis.** A) Relative positions of the primers in the upstream and downstream inverse PCRs together with the surrounding *Bam*HI sites. Pictures of the upstream and downstream inverse PCRs are shown in B) and C), respectively; additionally, the map of both the wild-type and mutant sequenced inverse PCR products is shown; the maps of breakpoints are shown in the lower panels. All genomic positions are according to hg18 and based on the coordinates of the non-duplicated original sequences on X chromosome. The 1kb ladder is used (P: patient, M: mother, S: sister, C: control).

### 4.2.4.5 Multiplex PCR detects the rearrangement

A simple multiplex short PCR was established to provide a straightforward diagnostic test for this novel *F8-IKBKG* inversion in severe hemophilia A cases and/or carrier females (Figure 35 A). Primers that bind to the flanking regions of the rearrangement junctions would not be expected to amplify products from the wild-type allele. Also, primers flanking the Int1R-1 repeat in wild-type *F8* intron 1 will yield an amplified product for both, a homozygous wild type individual as well as for the patient's heterozygous mother and sister. For these primers, no amplification should take place in the patient. Based on this strategy, a multiplex PCR with three primers was established: two of them, 3F and 4R, amplify the wild-type *F8*:Int1R-1 fragment, while 2R and 4R together are required for amplifying the junction corresponding to the region on the recombined mutant allele (Figure 35 B).
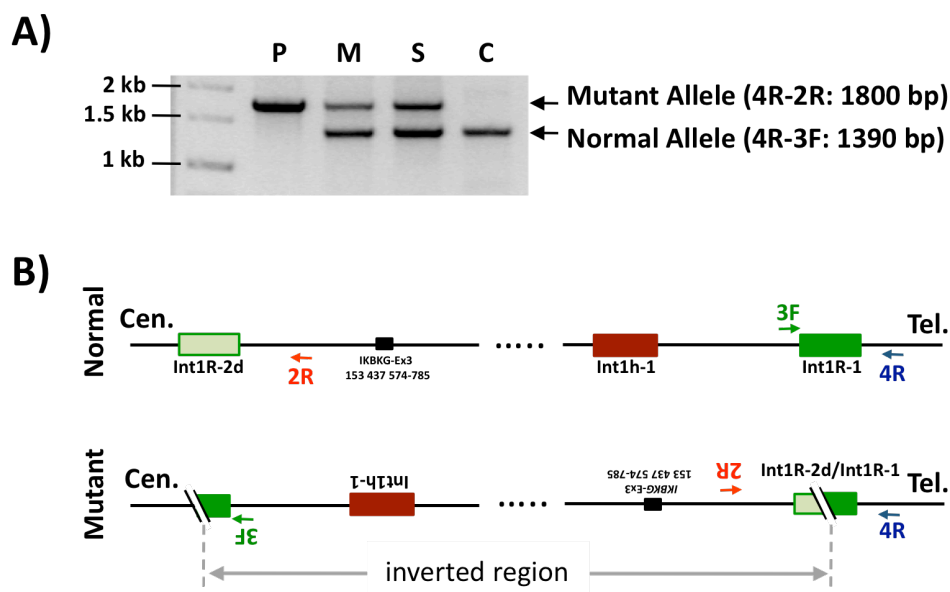


Figure 35: **Multiplex PCR results identifying the rearrangement.** A) PCR picture: the upper and lower bands correspond to the mutant and normal products, respectively (P: Patient, M: Mother, S: Sister, C: Normal control). B) Schematic diagram showing the relative positions of the primers 4R-3F and 4R-2R on both the wild type and mutant chromosomes, respectively. The 1kb ladder was used.

### 4.2.4.6 Detection of a 94 kb duplication on Xq28

The gene *IKBKG*, which is involved in the rearrangement reported here, encodes for the regulatory subunit of the inhibitor of KappaB kinase (IKK) complex required for the activation of the NF-κB pathway. This gene is mapped to a position 386 kb

proximal to *F8* locus. IKBKG is an essential protein and rearrangements in *IKBKG* that lead to a disrupted gene (equivalent to null mutations) are lethal [101, 102]. Therefore, the integrity of this locus in patient was confirmed by LR-PCRs. The genomic region involved in the rearrangement was amplified in the patient. PCR amplification was successful and a product across the Int1R-2 repeat in intron 2 of *IKBKG* was amplified. However, this result was in contrast to the hypothesized recombination event. Therefore, the possibility of a homologous recombination between inverted repeat copies of Int1R occurred between the *F8* copy and a third repeat outside the *IKBKG* was postulated.

However, the BLAT search revealed the existence of only two copies of the Int1R repeat. The next assumption was that of an independent duplication event to be responsible for the creation of a third Int1R repeat (designated here as Int1R-2d) in the patient's genome. To prove this, patient's X-chromosome was screened for duplications using a CGH-array. The results showed a large duplication of ~94.4 kb on the X chromosome where the duplicated region extends from Int1R-2 in the *IKBKG* gene to the *IKBKGP* (Figure 36). The duplication event provides an intact copy of the *IKBKG*. This data explains well why the patient does not suffer from any other health problems than hemophilia A.
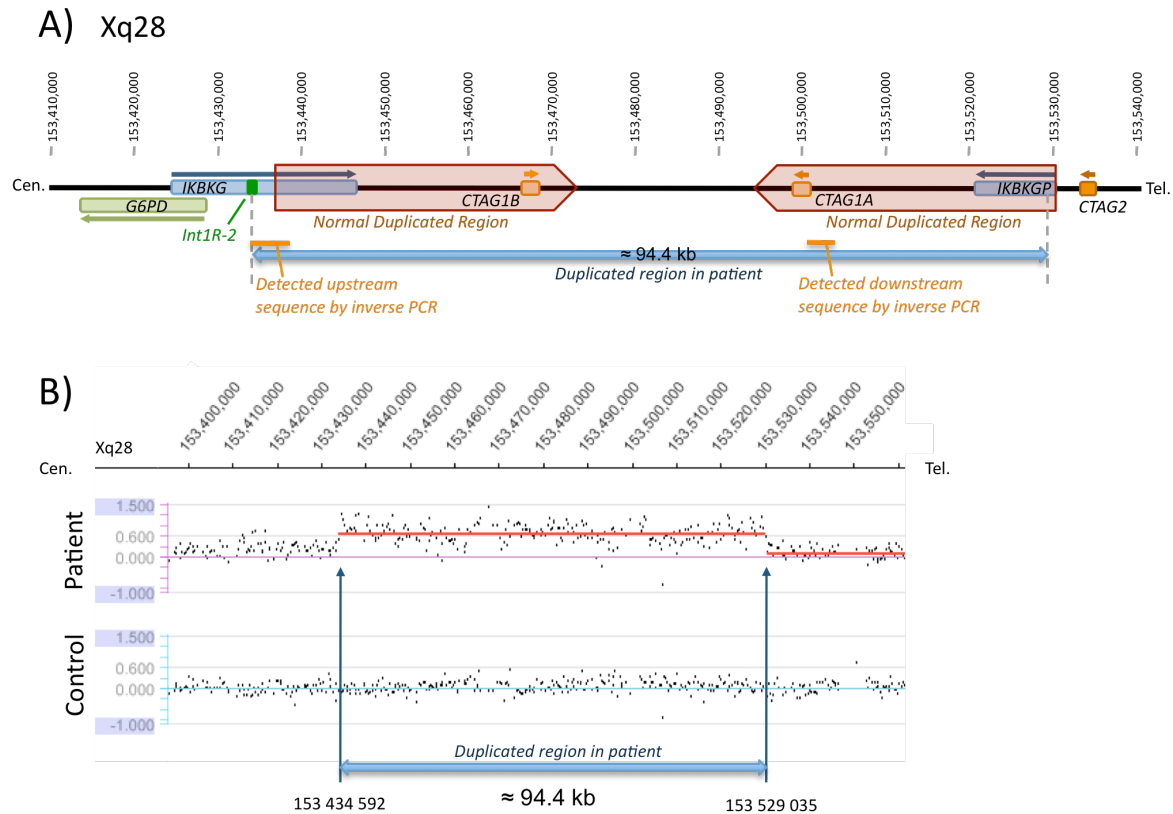
Figure 36: **Characterization of the duplication.** A) Map of the involved region; *G6PD*, *IKBKG*, *CTAG1B*, *CTAG1A* and *IKBKGP* are shown; the thin arrows above genes indicate direction of transcription; the thick pink/transparent horizontal arrows indicate the naturally occurring duplicated region; the extent of the duplication in the index patient is indicated by the blue line with arrows below the map; orange thin lines represent the detected sequence from the inverse PCRs. B) The CGH results of the patient (above) and a pool of male controls (below). All genomic positions are according to hg18.

Further analysis of the borders of the duplicated sequence revealed a rich occurrence of several repeats mainly of the endogenous retrovirus (ERV) family of repeats that could have facilitated the generation of the duplication. However, no clear motives could be found at the borders of the duplicated regions or the proposed insertion site to explain how the inversion happens (Figurer 37).
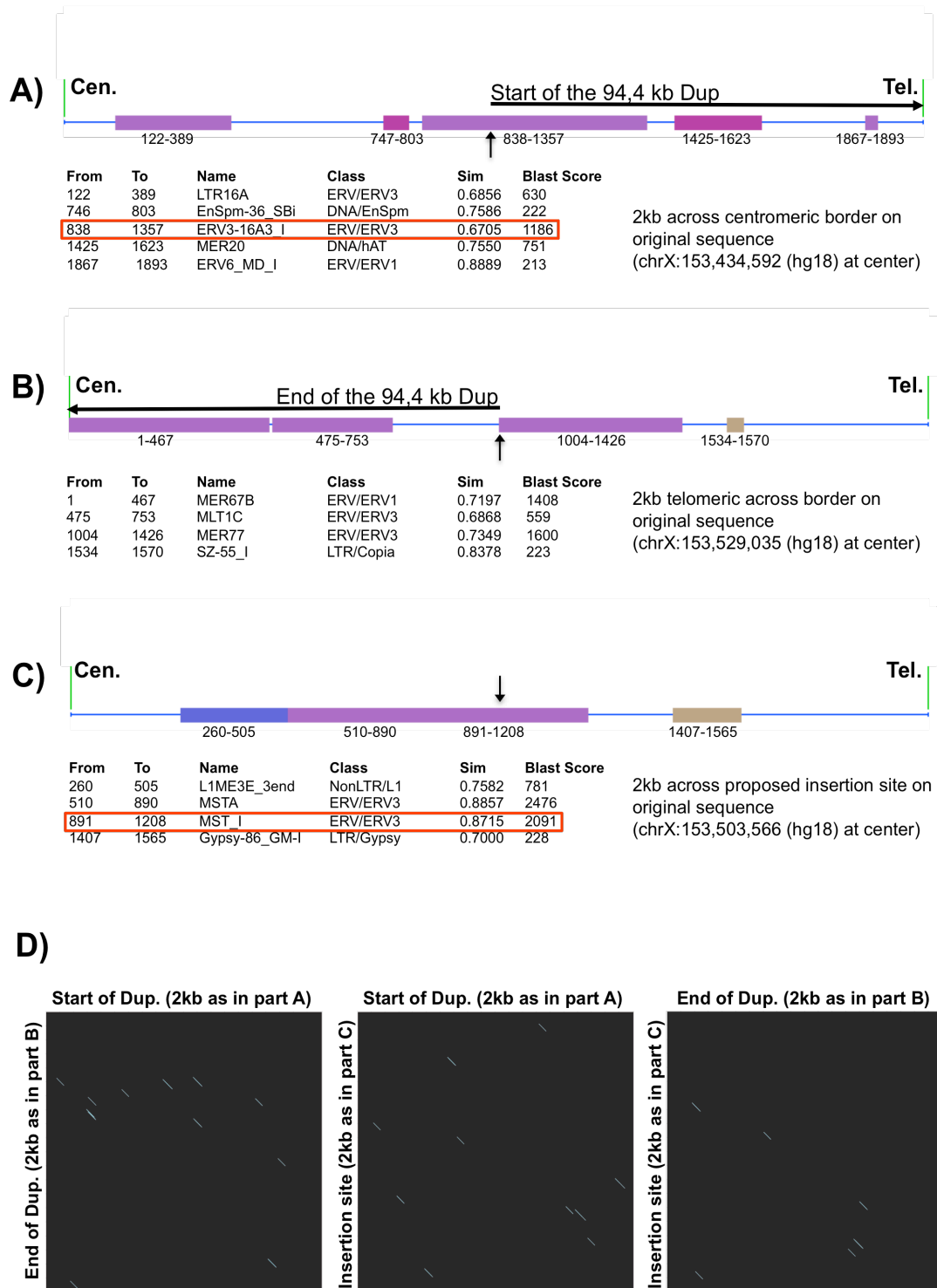
Figure 37: **Search for motives and similarities at duplication (Dup) and insertion borders.** Schematic presentation of the repeats in the centromeric end (A) and telemetric end (B) of the original sequence of the duplicated region and the proposed insertion site (C). Vertical arrows represent the start (A), the end (B) and the proposed insertion position. The position of each repeat in the 2 kb analyzed sequence is indicated together with the name and class of the repeat as well as similarity (Sim) and blast score. (D) Dot plot matrix analysis to identify similarity between sequences; three analysis are presented: on the left start of the duplication against the end of duplication, in the middle the start sequence of the duplication against the insertion site and on the right the end sequence of duplication against the insertion site.

### 4.2.4.7 Proposed mechanism of the complex rearrangement

Based on the inverse PCR sequences of the downstream breakpoint, it appears that the 5´ end of the 94.4 kb duplicated region (before the inversion) is connected to a repeat sequence with a high homology to a repetitive sequence in intron 2 of *FUNDC2*, followed by a region between *CTAG1* and *IKBKG.* As the later regions are found normally duplicated [103]. This gives two possibilities for the position of the insertion: the first between *CTAG1B* and *IKBKG*, and the second between *CTAG1A* and *IKBKGP*. The first possibility will lead to an insertion generating Int1R-2d in same orientation as the Int1R-1 and thus incompatible with an inversion. Therefore, the second insertion position is more likely (Figure 38).
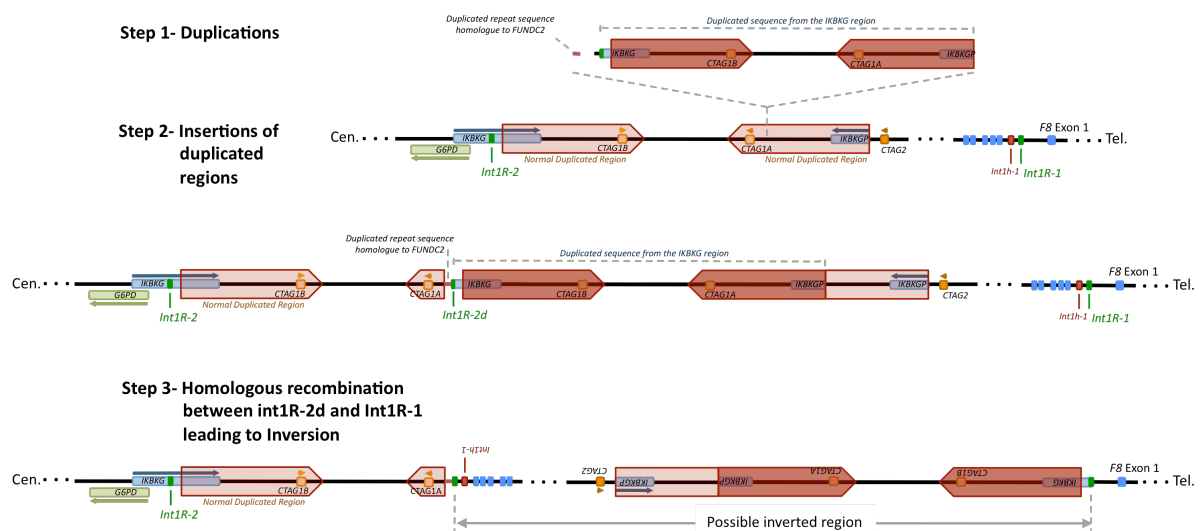


Figure 38: **Proposed mechanism of the rearrangement.** Three steps are schematically presented. The first is the duplication, the second is the insertion of the duplication, and the third is the intrachromosomal homologous recombination. Part of *F8* including exons 1-6 and the *IKBKG* region are shown.

The most clinically obvious effect of the rearrangement is the hemophilia A phenotype in the patient that removes the promoter and exon 1 of *F8*. This is sufficient to abolish FVIII synthesis and cause severe hemophilia A.

### 4.2.5 Identification of a novel duplication/ triplication in *F8* gene in patient HA#7

The index patient (HA#7) is a male hemophilia A patient with a severe phenotype (FVIII:C<1%). No inhibitors have been detected. The index patient was negative for intron 22 and 1 inversions. Complete sequencing of *F8* cDNA revealed no deleterious mutations. The vWF 2N mutations were excluded as well. FISH analysis revealed normal localization of *F8*.

#### 4.2.5.1 MLPA analysis revealed a novel triplication/duplication of *F8*

MLPA analysis for *F8* gene was done for the patient and an increased copy number of exons 1–25 (duplication of exons 1-4 and 23-25 and triplication of exons 5-22) were detected. The CNV pattern does not comprise exon 26 of *F8* (Figure 39).
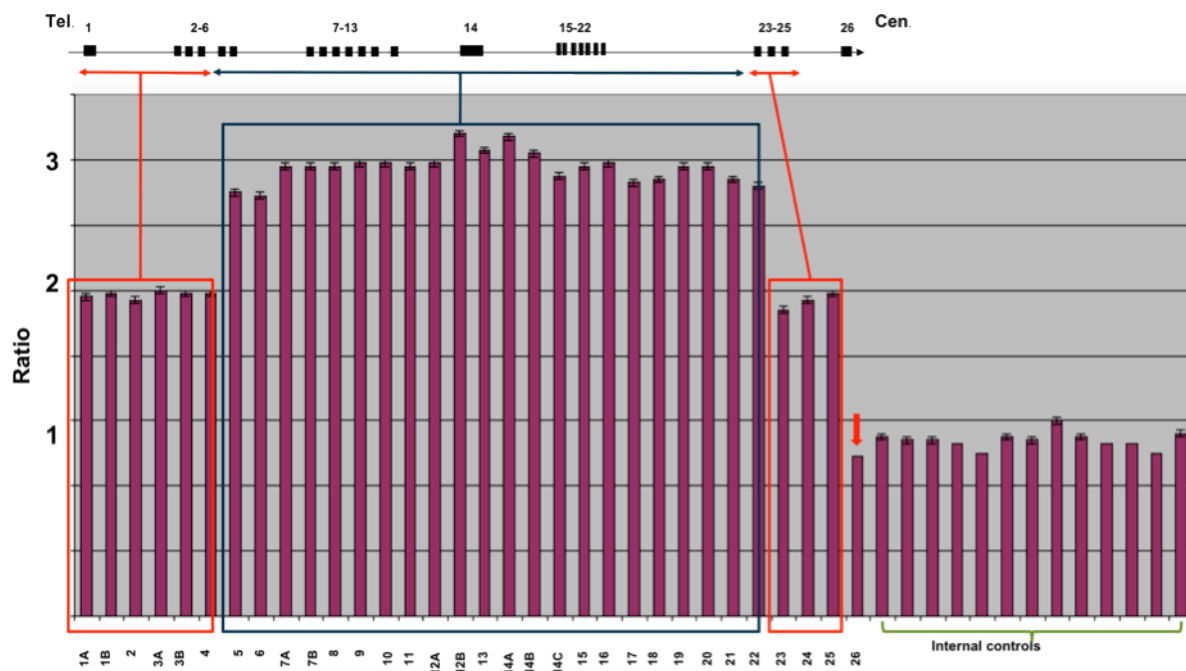


Figure 39: **MLPA results for patient HA#7.** Almost all exons on *F8* gene are involved in the duplication/ triplication mutation. The arrow indicates exon 26 where no increased probe signal was observed. The duplicated and triplicated regions are boxed in red and blue.

After performing MLPA on DNA material of his daughters, we could demonstrate that this imbalance is inherited in both obligate carriers.

### 4.2.5.2 The duplication/ triplication event comprises the complete intronic regions of *F8*

To verify the MLPA results and to investigate submicroscopic copy number changes, a CGH-array was applied. The results showed that the copy number change at Xq28 comprises only the genomic region of *F8* (Figure 40), and that the span of the duplication/ triplication is carried over the intronic regions of *F8*. These results are in agreement with the MLPA findings (Figure 39).
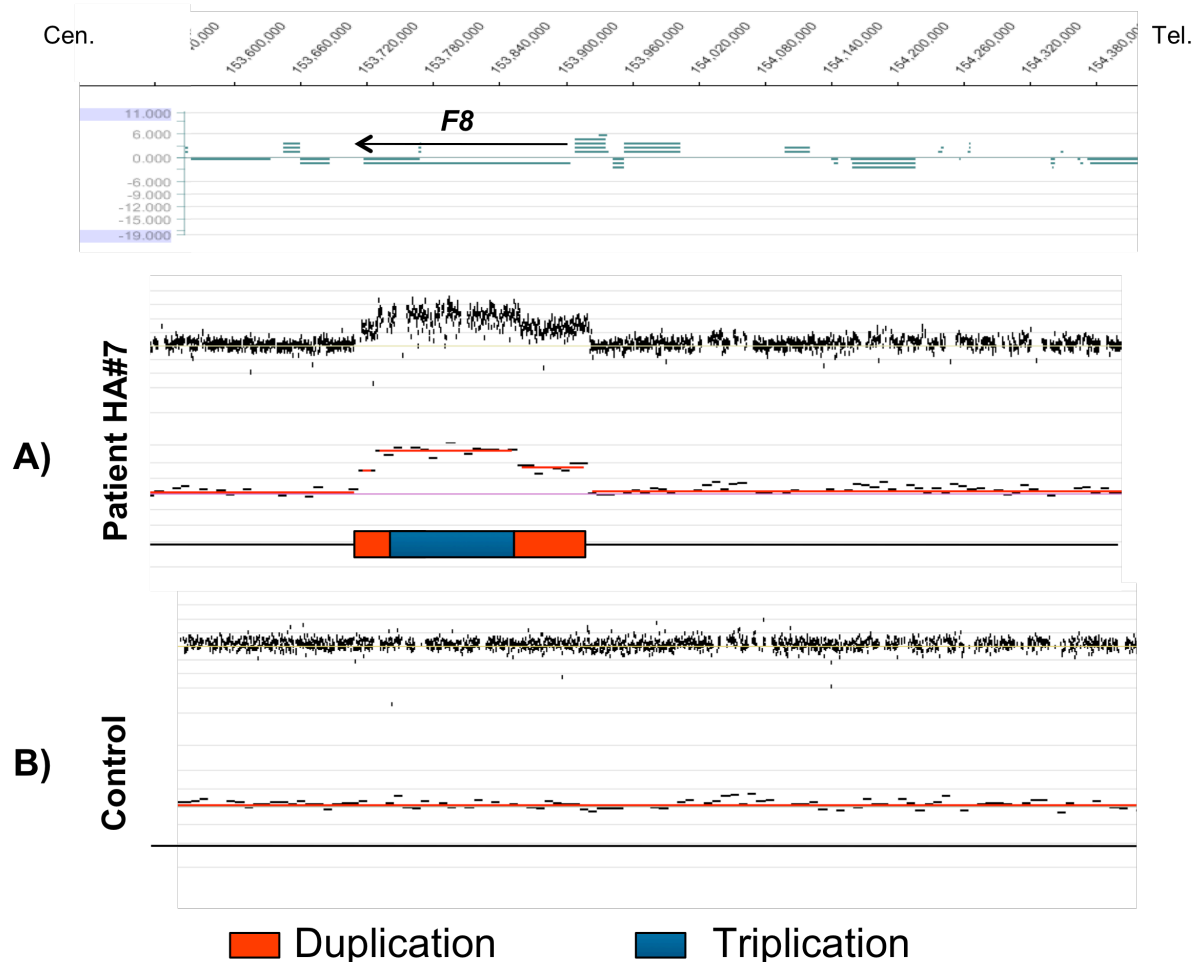


Figure 40: **Genomic region harboring alterations at Xq28.** The upper panel shows the position of *F8* on X chromosome (hg:18, NCBI build 36). The arrow indicates the direction of *F8* transcription. Results from CGH array show a large duplication and triplication of genomic region of *F8* in patient HA#7 (A) comparing to B) DNA from pool of male controls. The duplication/triplication event spans the complete genomic region of *F8* from exon 1-25 including the intronic regions.

### 4.2.5.3  Absence of FVIII protein despite high *F8* mRNA expression levels

When a deletion is found by MLPA, it is recommended to use an independent method to confirm the result. In case of duplication, the verification is more laborious. To analyze whether despite the duplication/triplication on the genomic level a mRNA transcript is generated; the transcription pattern was investigated and RT-PCR was performed as described above (3.3.1.2). After performing RT-PCR which is a semi-quantitative method, a much more intensive PCR band for the patient was observed compared to other hemophilia A patients and controls. However since RT-PCR a semi quantitative method is, a quantitative mRNA expression analysis was performed using a Taqman based assay (3.3.1.4). Quantitative mRNA analysis revealed an unexpected high *F8* mRNA expression level in comparison to healthy individuals (Figure 41).
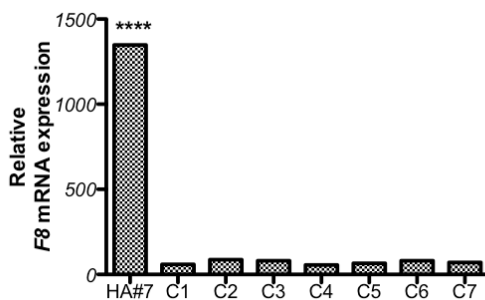


Figure 41: **Relative *F8* mRNA expression for patient HA#7.** The mRNA expression analysis reveals a significant higher mRNA expression of *F8* compared to healthy male individuals (C1-C7 are the controls, ttest ****= P<0.0001).

Despite the high mRNA expression levels, no protein is synthesized (FVIII:Ag<1%, Table 3). This could be a hint that no normal mRNA transcript is available, and the existing mRNA transcript is not translated into a functionally active FVIII protein.

### 4.2.5.4  Genomic gains originate from one chromosome homolog

FISH analysis showed no intra-chromosomal translocation of *F8* and all copies of the gene were located to Xq28. To test for potential inter-chromosomal exchanges during rearrangement product formation, marker haplotypes were evaluated from the genomic interval spanning the complex rearrangement. Notably, the patient lacked heterozygosity for all SNPs tested using this approach, including SNPs localized to both duplicated and triplicated genomic intervals. The lack of heterozygosity observed in these regions suggests that these complex genomic rearrangements originated from a single chromosome. In addition, linkage analysis showed that the haplotype associated with the disease in the index patient is shared between all affected members (Figure 42). The analysis revealed to date not described duplication/triplication pattern comprising the whole genomic region of *F8* locus leading to a severe HA phenotype where, despite high *F8* mRNA expression levels, FVIII protein is lacking.
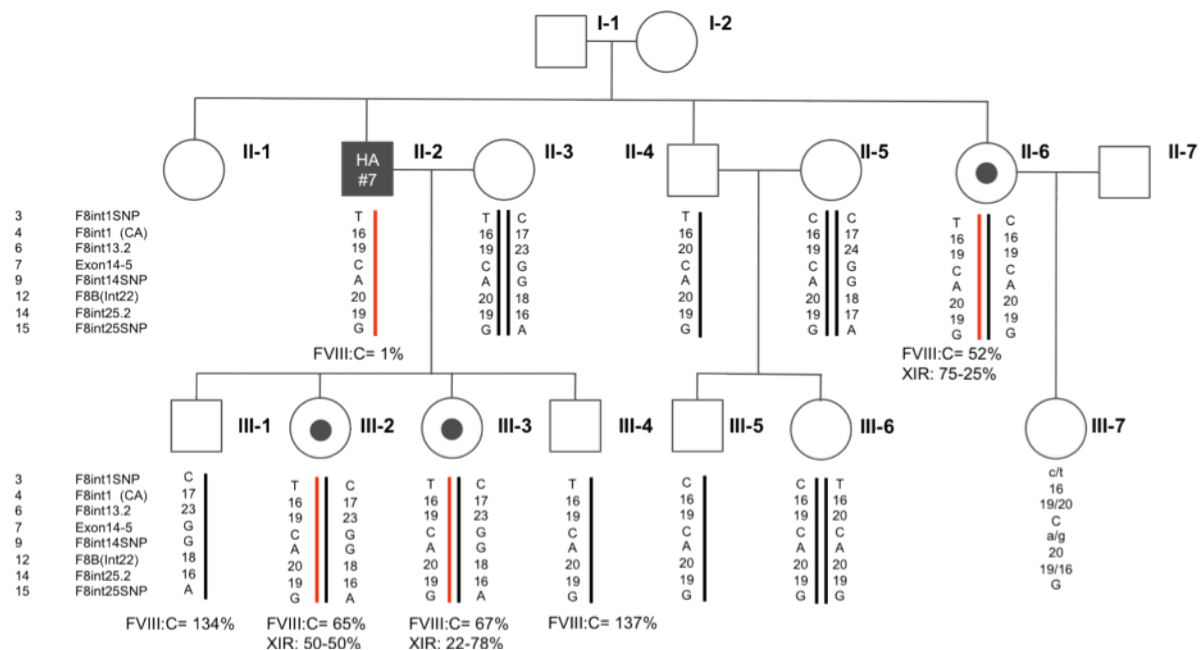


Figure 42: **Haplotype segregation analysis in family of patient HA#7.** FVIII:C levels are shown, where this information was available. The X-chromosome inactivation (XIR) ratio is shown for the potential carrier females. Where no haplotype is constructed no DNA material was available. For III-7 no haplotype can be constructed, since the allele information from her father is not available.

### 4.2.6 Identification of a novel breakpoint in exon 18 of *F8* gene in patient HA#14

The index patient (HA#14) is a male hemophilia A patient with a severe phenotype (FVIII:C<1%). No inhibitors have been detected. The index patient was negative for intron 22 and 1 inversions. Primers used for characterization of this breakpoint are listed in Table S12.

#### 4.2.6.1 Discriminative results for exon 18 in patient HA#14

The amplification of exon 18 of *F8* failed using the routinely used primers for diagnostic (Figure 43). Complete sequencing of all other *F8* exons revealed no mutations. This result indicates deletion of exon 18, which is consistent with the severe hemophilia A phenotype.
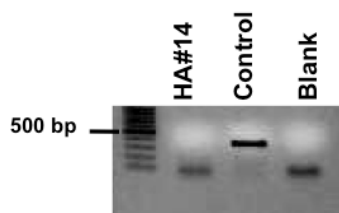


Figure 43: **PCR picture of amplification of exon 18.** The amplification failed in HA#14 using the primers located in intronic regions flanking exon 18 of *F8*. The 100 bp ladder was used.

To verify the deletion, MLPA analysis was carried out. Interestingly, MLPA analysis revealed contrary results to the PCR amplification. MLPA approach did not confirm the deletion of exon 18 and showed normal copy number for all exons of *F8.*

#### 4.2.6.2 Identification of a breakpoint in exon 18 of *F8* using LR-PCRs

The first assumption was that one of the primers used for the amplification of exon 18 fails to bind to this region due to a rare polymorphism. To investigate this, primers were designed to be amplified with the MLPA probe as forward and reverse primer (MLPA18F and MLPA18R). Both amplifications were using these primers (Figure 44). Sequencing of the PCR fragments revealed no changes in exon 18.
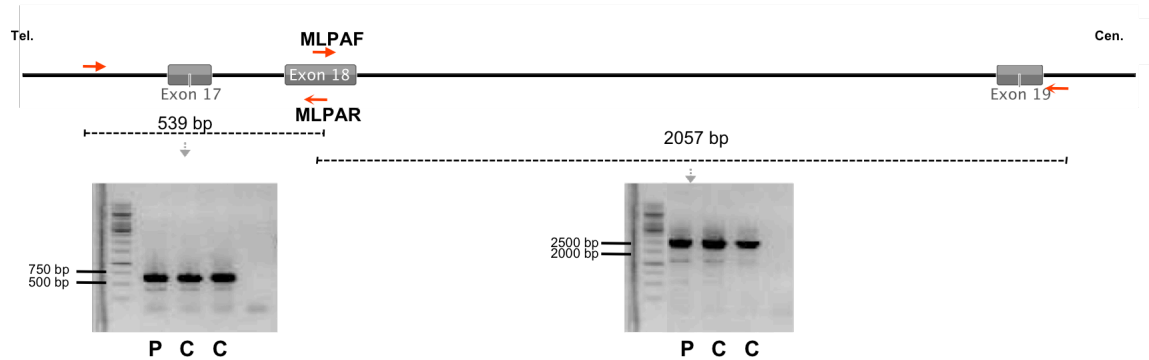
Figure 44: **Identification of a breakpoint in exon 18 of *F8*.** PCR amplification using MLPA probe as primers to amplify exon 18 for patient HA#14. All PCR fragments show normal amplicon sizes for both regions. The red arrows indicate the primers used for the amplification. The 1kb ladder was used (P=patient, C=Control).

The next step was to apply LR-PCR using primers of region N (chrX: 154,134,934–154,124,275; hg19). Inability to amplify region N, a 10.636 bp region located between intron 14 and intron 22 of *F8* (flanking exon 18) from the genomic DNA of the index patient, indicated the presence of a genomic breakpoint in this region (Figure 45).
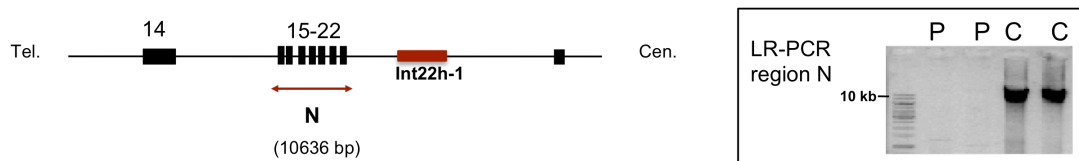


Figure 45: **LR-PCR across exon 18 using the primers in intron 14 and 22 of *F8*.** Schematic presentation of the amplified region (left) and LR-PCR picture of region N (left) is shown. LR-PCR reveals no amplification for patient and indicates the presence of a breakpoint in this region. The 1kb ladder was used (P=patient, C=control).

### 4.2.6.3  Inverse PCR identifies the junction of the breakpoint

To characterize the breakpoint junctions, an inverse PCR approach was designed to amplify the region upstream and downstream of exon 18 of *F8* (Figure 47 A). The genomic DNA of the index patient was digested using *Hind*III enzyme. The product was ligated and purified as described in section 3.3.3.3 and the ligated product was amplified. The PCR fragments, of both upstream and downstream PCRs, were sequenced.

The upstream inverse PCR revealed a smaller PCR fragment in size (1771 bp compared to the 2725 bp of the wild type sequence). Sequencing of the patients amplicon revealed a break within exon 18 of *F8* (chrX: 154,132,278; hg19). A short repetitive element of L1 family of repeats was inserted in the body of F8 gene followed by a junction to a sequence at Xq28 from chrX: 154,473,718 to 154,474,688; hg19 (Figure 47 B). The insertion junction is directly after the binding site of the MLPA probe (Figure 46).
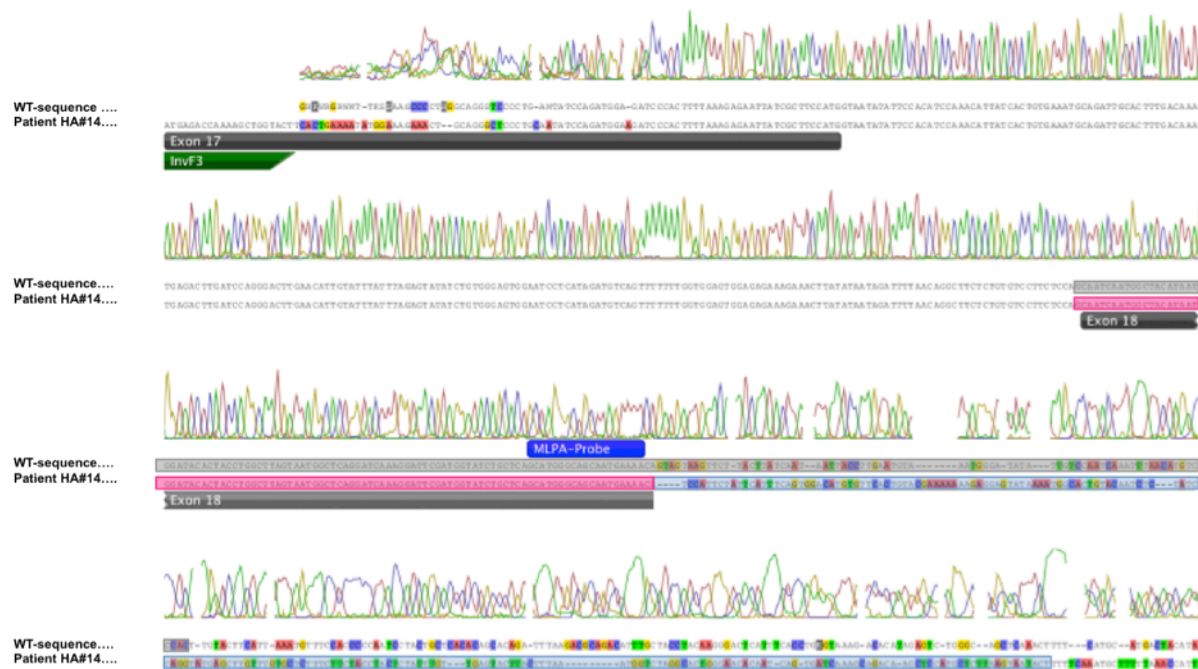


Figure 46: **Sequencing results of upstream inverse PCR product of patient HA#14.** The wild type sequence (WT-sequence) is aligned to the PCR product of patient. Exon 18 of *F8* is boxed grey and pink in wild type and patient's sequence, respectively. The breakpoint is located directly after the MLPA probe (dark blue box). The inserted regions in patients genomic DNA is boxed in transparent blue box. The position of InvF3 primer is shown in green in exon 17 of *F8*.

The downstream inverse PCR revealed a fragment of 3362 bp, which is bigger in size compared to the product obtained from healthy controls (2819 bp). Sequencing of the aberrant PCR product revealed an insertion of sequences located at Xq28 region (chrX: 154,487,749- 154,486,661) in intron 17 of *F8* (Figure 47 C).
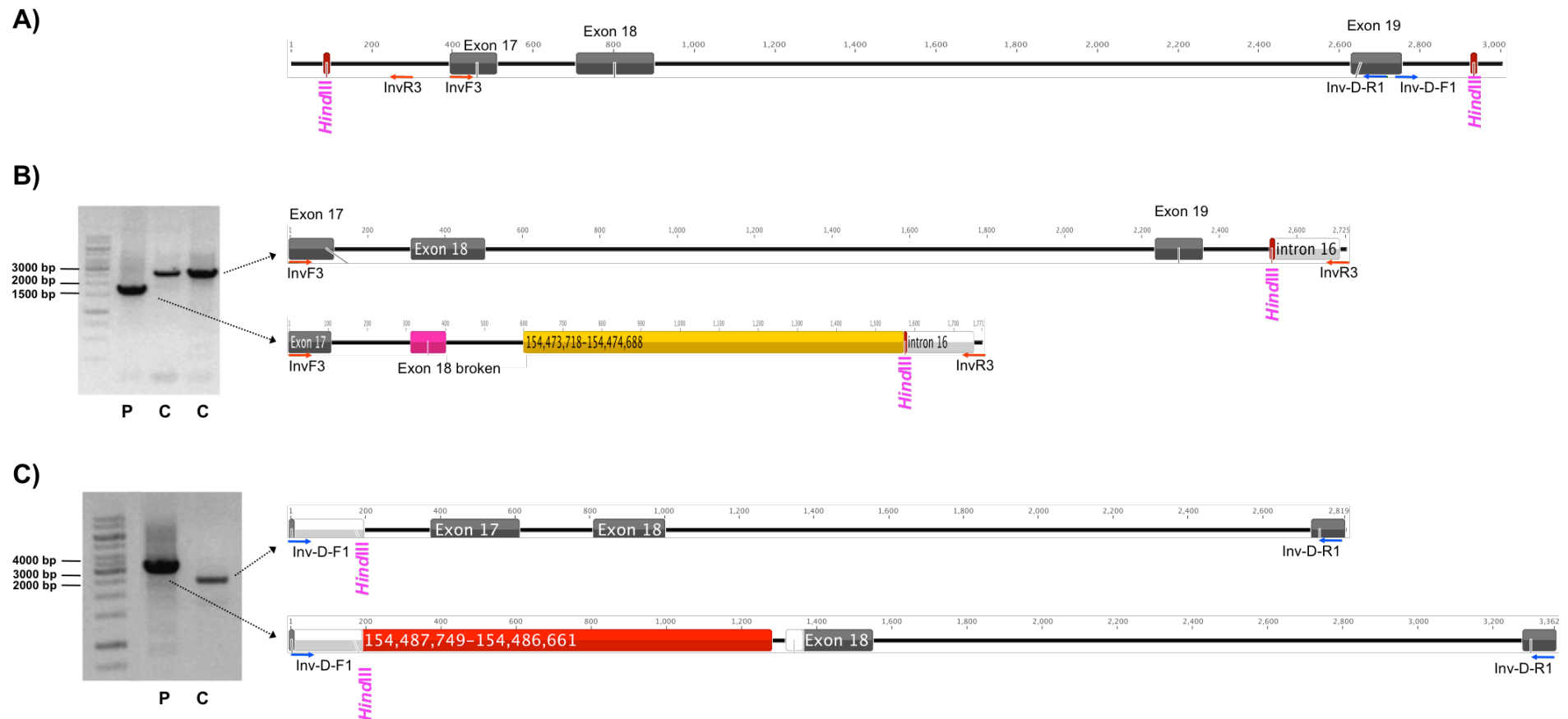
Figure 47: **Inverse PCR approach for patient HA#14.** A) Relative positions of the primers used for the upstream and downstream inverse PCRs together with the surrounding *Hind*III restriction sites (marked in red). B) Picture and map of the sequence of the upstream C) and downstream inverse PCRs are shown for both the wild type and patients inverse PCR product. All genomic positions are according to hg19. The exon 18 junction to the insertion is shown in partial exon 18 in pink and the inserted region is shown in yellow. The inserted region before intron 17 is shown in red for the downstream inverse PCR (P: Patient, C: Control). Red and purple arrows indicate the location of the primers used for the upstream and downstream PCRs, respectively. The 1kb ladder was used.

### 4.2.6.4  Multiplex PCR detects the rearrangement in HA#14

In order to provide a diagnostic test for detection of this breakpoint in other severe hemophilia A cases and/or carriers, a simple multiplex short PCR was established (Figure 48). The amplification was done using a forward primer in intron 17 of *F8* (AT-MPX-F8-F: chrX: 154,132,464- 154,132,445; hg19) with two reverse primers, one amplifying the wild type *F8* sequence located in intron 18 of *F8* (AT-MPX-F8-R: chrX: 154,132,035- 154,132,015; hg19) and one located in the inserted fragment in the genomic region of *F8* (AT-MPX-R2: chrX: 154,473,911- 154,473,931; hg19). The PCR products of wild type and mutant regions differ in size; 450 bp and 603 bp, respectively.
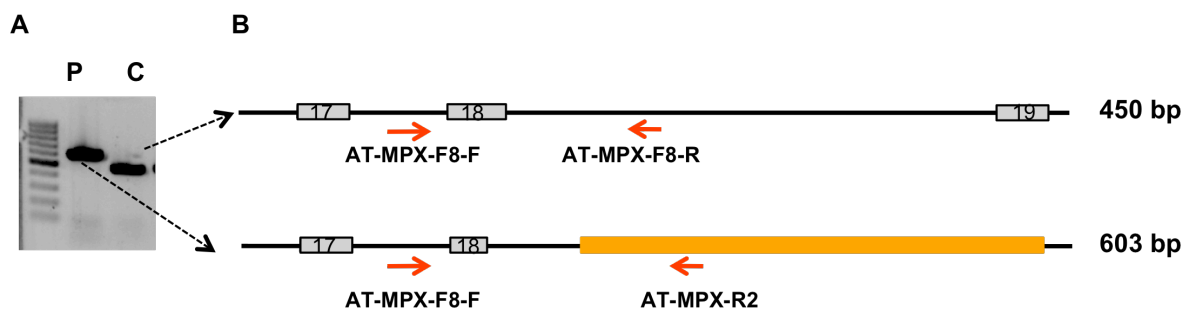


Figure 48: **Multiplex PCR for detection of the breakpoint in patient HA#14.** The PCR amplification was carried out using At-MPX-F8-F primer with both reverse primers (The arrows indicate the position of the primers used, the insertion is shown as an orange box. The 100 bp ladder is used (P: patient, C: control).

### 4.2.6.5  Proposed arrangement of the genomic region of *F8*

Based on the results of the inverse PCR and due to the fact that no repetitive sequences were found in the border of the rearrangement, an inversion was excluded. The possible arrangement of the insertion in the body of *F8* gene is shown in Figure 49. Genomic region of *F8* is disrupted by a large insertion of about 13 kb. Interestingly, part of exon 18 (chrX: 154,132,363- 154,132,278; hg19) seems to be duplicated which is present in the upstream PCR product.
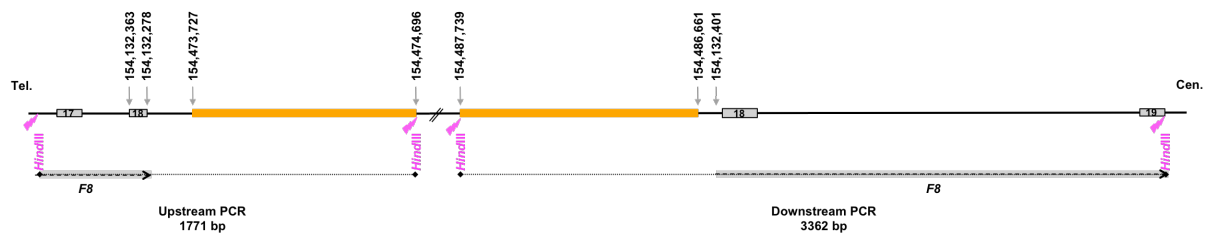
Figure 49: **Possible arrangement of the genomic region for patient HA#14.** The upstream and downstream PCR products with the *HindIII* cutting site are shown. The insertion is about 13 kb in size (orange boxes) and disrupts the genomic region of *F8*. The grey arrows indicate the transcription direction of F8. Exons of *F8* are boxed. All coordinates are according to hg 19.

This data clearly shows the presence of a large insertion in body of *F8*, which leads to disruption of *F8* genomic region and can explain the severe hemophilia A phenotype.

## 4.3 Summary of novel mutations causing hemophilia A presented in this work

In the context of this work, several causal mutations were identified in hemophilia A patients earlier classified as patients without mutation.

1. Identification and characterization of two breakpoints in genomic region of *F8*:
   - A novel rearrangement in intron 1 of *F8* (patient HA#1)
   - An insertion in intron 18 of *F8* (patient HA#14)
2. Identification of a novel duplication and triplication (patient HA#7)
3. Identification of deep intronic mutations
   - Characterization and prove of causality of deep intronic mutations (patient HA#8, HA#9, HA#10, HA#11, HA#13)
   - Identification of a novel mutation in 3´ UTR of *F8* mRNA (patient HA#3 and HA#4)

# 5 Discussion

## 5.1 Key considerations for diagnosis of hemophilia A

Hemophilia A together with VWD, account for the majority of congenital bleeding disorders. The main reason for referral of patients to molecular diagnosis is a family history of hemophilia A. However, in one third of cases, hemophilia A is a sporadic disease [104, 105]. Sporadic cases of severe hemophilia A are often diagnosed in about one year of age, after the first unexpected bleeding occurs. Quite contrary to severe phenotype, mild and moderate cases may not be diagnosed until adulthood.

Generally, the diagnosis of hemophilia A is based on FVIII:C levels. The severity of disease is critical for diagnosis and the choice of the genetic analysis strategy that needs to be applied. Two main issues should be considered in order not to misdiagnose hemophilia A. The first category is pre analytic considerations which are needed to be taken in account to determine the correct severity grade of the disease, e.g. correct storage of samples before FVIII:C measurements [106] and amount of citrate anticoagulant in blood collection tubes [106, 107]. These variables should be considered for mild and moderate phenotypes. The second category is misdiagnosis of hemophilia A in patients with VWD or combined FV/FVIII deficiency. For later, diagnosis of FV deficiency is essential for the correct treatment regiment. For all patients with a bleeding tendency and reduced FVIII:C levels, it is important to perform vWF assays to exclude the possibility of VWD, independent of the severity of the disease. The role of vWF as carrier and protector of FVIII protein is quite critical. A number of patients with a mild phenotype may actually suffer from vWF type 2N. Prior to correct diagnosis, all these patients would be treated with FVIII concentrates, and would show poor responses to the therapy that might have clinical consequences [108-110]. Therefore, mutation analysis in exons comprising the FVIII binding domains of *vWF* [111] and performance of vWF:FVIII binding assays [47, 112, 113] are necessary. This issue is a very important point in cases, where no mutation in *F8* is detected.

In case of combined FV/FVIII deficiency, the bleeding phenotype is due to disruption of cargo specific ER to Golgi transport proteins, LMAN1 and MCFD2. The combination of FV and FVIII deficiency does not seem to cause more bleeding. Patients usually show a mild phenotype. Nevertheless, since the causal mutation resulting in low amounts of FVIII and FV activity is either in *F8* or *F5* gene, it is important to assay FV levels in patients diagnosed as hemophilia A with a mild bleeding tendency as well [44, 45, 114, 115].

Only a few cases of hemophilia A are reported in females and the molecular mechanism in this cases are rarely investigated [21, 22]. On the other hand, the correct diagnosis of hemophilia A in carriers should not be neglected. It should be taken in consideration that not all carriers are symptomatic [116]. Only about 10% of female hemophilia A carriers show bleeding events [117]. Reduced FVIII:C levels in these cases may also be due to VWD type 2N or skewed X chromosome inactivation [18]. In these cases, vWF:FVIII binding assay and HUMARA test would help to achieve a correct diagnosis of hemophilia A carriership.

## 5.2   Hemophilia A diagnostic- beyond the routine tests

The diagnosis of hemophilia A is mainly done by blood tests on the basis of residual FVIII:C levels. The molecular cause of FVIII deficiency can be divided into 3 categories. The first category is the mutations in *F8* causing a structural change in FVIII protein, which are the most common reason of FVIII deficiency. The next category contains mutations in proteins involved in trafficking and transport of FVIII protein such as LMAN1/MCFD2 [44, 45] and vWF [47, 95]. These changes would also lead to reduction of FVIII:C levels. The data presented in this work, provide insight into the molecular basis of hemophilia A with special focus on the third group, consisting mutation negative patients. In a small group of clinically defined hemophilia A patients, a 100% mutation detection rate is not achieved [25, 48, 118]. For this group of patients, additional experimental protocols should be applied to find the causal mutation.

The lack or the faulty transcription of *F8* might result from a change in genomic regions of *F8* that are normally not investigated for mutations or it is the result of complex rearrangements that are difficult to identify. Current mutation screening flowchart applied in most genetic laboratories (Figure 4) includes only screening for inversions in intron 1 and 22 [24, 119] as well as sequencing all exons and their flanking intron boundaries.

To address the difficulties of finding the molecular mechanisms that can explain the clinically defined hemophilia A phenotype in such patients, a 4-step experimental protocol was designed. This protocol would serve as a guideline to detect novel abnormalities. The first step, after excluding all factors associated with reduced FVIII:C levels, is the exclusion of duplications. In step two, the complete *F8* locus is amplified by overlapping LR-PCRs. In the next step, all LR-PCR products would be sequenced using a NGS approach. Finally, mRNA analysis (RT-PCRs and qRT-PCR) may be applied to detect abnormal splicing across exon-exon boundaries, harboring deep intronic variations. Using this protocol several causative mutations were found in patients earlier classified as patients without detectable mutations.

The targeted capture NGS approach used in this study combines two techniques; a universal LR-PCR protocol and a massive pyrosequencing technology. This approach will help to identify sequence variants in regions that are overlooked using the conventional screening methods. Although both systematic false positive (because of incorrect genome mapping, systematic equipment sequencing errors) and false negative variation calls (because of low depth of coverage and difficulty in repetitive regions) exist, the clinical application of NGS as a diagnostic tool has become increasingly evident [120-122].

Most of NGS approaches are based on a pyrosequencing principle. Although pyrosequencing is widely used in various applications [123-125], it has several limitations. Beside its relative short read lengths, it has a limitation when reading a strain of more than five or six of the same nucleotides, known as homopolymer. These two limitations make NGS less attractive for diagnostic purposes. However, although a single read of NGS technology has higher individual-base calling error rates compared to Sanger sequencing, the false positive results are significantly

reduced by increasing the sequencing coverage [126]. Nevertheless, all variants found by NGS should be validated using Sanger sequencing. Moreover, the subsequent mRNA analysis performed in this study is a convenient complementary approach to the NGS strategy. Though, the limitations of NGS are the reasons why its applications in molecular diagnostics are still limited. However, the recent improvements in read lengths and coverage make this technology more attractive for diagnostic purposes.

The results obtained in this study highlight once more the complexity of mechanisms leading to hemophilia A. The presented data shows that CNVs, i.e. duplications and triplications and aberrant transcription as well as deep intronic mutations and rearrangements contribute to the mutation spectrum of hemophilia A patients classified as patients without mutation. These mechanisms can only be determined, when diagnosis tests beyond the routine applications are performed. Moreover, it should be considered that the large intronic regions of *F8* due to their high repetitive content [23] are prone to rearrangements. These abnormalities are easily missed using the conventional mutation screening methods.

Taken together, application of new mutation screening tools in the routine genetic diagnosis of hemophilia A, such as MLPA and CGH for detection of duplications and deletions as well as NGS technologies for re-sequencing of the whole genomic region of *F8* and mRNA analysis, is necessary to reduce the number of cases classified as mutation negative hemophilia A patients.

## 5.3   Disease associated intronic mutations

Quite a number of mutations are being missed, as some of these involve sequence variations in intronic regions that are not being assessed by current techniques. Even the more recent whole-exome sequencing methods do not cover the intronic regions. According to Human Gene Mutation Database (http://www.hgmd.org/), about 10% of reported mutations in disease-associated genes are splice site mutations. Most of these mutations affect consensus sites. However, the evidence of deeper intronic sequence variations affecting splicing process is now becoming more available.

Deep intronic sequence exonisation has been reported for several genes: *ATM* [127], *CFTR* [128], *RB1* [129], *NF1* [130], *CDKN2A* [131], *BRCA2* [132], *HRPT* [133], *USH2A* [134], *CF* [135] and *MLC1* [136]. All these reported intronic mutations are proved to be associated with diseases. Nevertheless, the importance of screening for splicing variations has been underlined in hemophilia A as well [42, 43, 75, 118]. Intronic mutations activate a cryptic donor or acceptor splice site and splicing between this novel splice site and a pre-existing splice site leads to the inclusion of a cryptic exon. These mutations would lead to splice defects that are more commonly identified in association with human diseases. For the time being, NGS and mRNA analysis are the most suitable approaches to identify and analyze such deep intronic mutations.

The results shown in this work highlight the importance of implementation of *F8* mRNA analysis in the clinical practice as well as the relevance of mRNA analysis in confirming whether changes found at the DNA level could lead to a disease. However, since FVIII is a liver-specific protein, tissue specific splicing should not be neglected [137]. The cDNA used in this study was synthesized from RNA obtained from blood. So far, there is no assurance that the effect of the variations found here will be the same in the liver.

Taken together, the study presented here highlights the necessity of looking beyond the coding, exon/intron boundary, and promoter regions, when attempting to find the molecular cause of a disease in the absence of pathogenic mutations. There is a need to sequence the complete *F8* to unveil the underlying mutation in such cases. This work clearly shows that targeted capture and NGS are likely to be very useful in a diagnostic setting to clarify the genetic basis of patients in which no mutation is detectable through conventional Sanger sequencing. Moreover, this data confirms that deep intronic mutations in *F8*, although rare, can cause abnormal splicing that leads to hemophilia A.

## 5.4 Heterogeneity of mutational events leading to hemophilia A

The majority of hemophilia A cases can be explained by specific defects in *F8*, but in a minority of cases (2%) no genetic *F8* variant can be identified [42, 138]. In such cases, after excluding large tandem duplications [41, 139] and the two previously reported intron 1 and intron 22 inversions [23, 140], a wide spectrum of mutation mechanisms have been reported. Among others, mutations in introns [118], insertions associated with deletions [141-143] and deletions arisen from both, unequal homologous recombination between Alu-derived sequences [144, 145] and non-homologous recombination [146], have been reported. Such complex rearrangements that evade standard PCR screening strategies need to be considered. Both inversions reported in severe hemophilia A cases, are arising from rearrangements because of homologous repeats [23, 24].

Tandem duplication of a genomic segment is one of the possible outcomes of unequal crossing over between paralogous sequences. Although genomic rearrangements normally do not result in gain or loss of DNA, there are some evidences of increase and/or decrease in copy number of genes [147, 148]. Some cases of even more complex mutational events have been reported involving different mechanisms. Ven der Water et al. reported a 38 bp insertion associated with a 20.7 kb deletion within *F8* involving LINE-1 retrotransposon-mediated DNA repair [141]. Later, a 6 bp insertion in junctions of a 316 bp deletion in *F8* was reported that resulted from a model of serial replication slippage (SRS) [149]. More recent reports demonstrated even a more complex model of break-induced SRS at Xq28 consisting of a 15.5 kb deletion/16 bp insertion together with a 28.1 kb deletion/263 kb insertion [143]. All these publications show that Xq28 is prone to such rearrangements.

The work here demonstrates techniques to identify and confirms the presence of novel rearrangement events as well as deep intronic variations. The severity of the disease seems to be well associated with the complexity of the causal mutation.

Deep intronic mutations found in this work and other studies [118, 150-152] are associated with mild to moderate hemophilia A phenotypes. On the other hand, complex rearrangements reported in this study and other studies lead to the severe phenotype of hemophilia A [23, 119, 143].

This study presents several novel rearrangements causing severe hemophilia A. Using the LR-PCR assay developed for amplification of whole genomic region of *F8,* two breakpoints were identified; one in intron 1 and the other one in intron 18 of *F8*. With the advances in LR-PCR techniques and the sequence information provided by the Human Genome Project, gross rearrangements responsible for a number of genetic diseases are becoming easier to be identified. Both features helped to precisely analyze these rearrangements without the need for laborious genomic cloning procedures. In addition to breakpoint characterization, a specific PCR test for family carrier detection and prenatal diagnosis of both rearrangements was developed, which can easily be implemented in routine genetic diagnosis of hemophilia A.

In accordance with the previously characterized inversions (comprising introns 1 and 22), the novel inversion identified in this work (patient HA#1), was also caused by a rearrangement between homologous repeats. This complex novel genomic rearrangement causes several changes in gene copy number in the Xq28 region. Interestingly, the *IKBKG* locus at Xq28 involved in the inversion is complex and contains an inverted segmental duplication of about 35.4 kb that includes the non-functional pseudogene, *IKBKGP,* extending from exons 3 to 10 and a complete duplication of the *CTAG1* gene. Mutations in *IKBKG* lead to Incontinentia Pigmenti (IP), a dominant neuroectodermal disorder. Null mutations in *IKBKG* are lethal in male fetuses [101]. Inverted orientations and the high homology of the segmental duplications at this locus make this region prone to rearrangements [153, 154]. In addition, the high frequency of micro/macro homologies and repetitive sequences at this locus predispose to pathological genomic rearrangements [103].

However, in contrast to the *IKBKG-F8* rearrangement described here, none of the previously characterized rearrangements are outside the *IKBKG* locus and the associated inverted segmental duplication. Interestingly, in about 27% of IP patients

no genetic defects in *IKBKG* could be identified [155]. Therefore, based on the *IKBKG-F8* rearrangement described here, it is questionable if a reciprocal phenomenon for which a homologous recombination with a duplicated *F8* intron 1 would also lead to breakage of *IKBKG*. Such a rearrangement does not lead to gain or loss of DNA or change in *IKBKG* DNA copy numbers and would have escaped the traditional screening methods in the *IKBKG*, which is based on short range PCR amplification of the coding sequence [155]. To detect all rearrangements in the *IKBKG* including inversions a long range PCR approach should be implemented. Also, a third possibility of recombination between Int1R-1 and Int1R-2 repeats exists. Although lethal in males, it would lead in females to two consequences: an IP phenotype and a hemophilia A carrier status with considerable reduction in FVIII:C levels.

The rearrangement event reported here is the result of two events. The first is the duplication of a large 94 kb region that extends from the functional *IKBKG* to *IKBKGP*. The duplicated region contains the Int1R repeat at one end (designated Int1R-2d) and appears to be inserted in the telomeric 35.4 kb segmental duplication. This rearrangement leads to the second event wherein an intra-chromosomal homologous recombination occurs between *F8*:Int1R-1 and Int1R-2d in the duplicated *IKBKG* gene. As a result, *F8* is split in intron 1 into two oppositely oriented fragments for which no functional mRNA is transcribed [75]. On the other hand, the essential *IKBKG* gene is still intact and transcribed, as the homologous recombination involves the duplicated copy and not the endogenous *IKBKG* or *IKBKGP*. This explains the absence of additional clinical symptoms to the hemophilia A phenotype for this patient.

For the second severe hemophilia A case characterized in this work (HA#14), first a simple deletion of exon 18 was thought to be the causal mutation. However, this case shows the importance of implementation of several complementary methods for verification of mutations. The discriminative results between the PCR and MLPA show the limitations of both applications. Some rare polymorphisms in PCR primers can lead to unsuccessful PCR amplifications. In such cases MLPA seems to be an efficient method to screen chromosomal abnormalities. Until now, single base changes in the probe binding sequence were the major reason of altered results of

MLPA analysis. The case studied in this work shows clearly that MLPA can also yield false negative results and a disruption of exons can be easily missed depending on the position of the probe. MLPA may be a fast and inexpensive method for detection of microdeletions and microduplications, but the method should be validated by additional methods. However, despite its high sensitivity and accuracy, for application of MLPA in a diagnostic setting, different probes for all specific exons should be used, to prevent false positive and false negative results.

The practical application of MLPA for diagnostic purposes of hemophilia A is demonstrated for patient HA#7. The novel duplication and triplication pattern was easily detected by MLPA analysis. Concerning the laborious karyotype analysis for patients and the limitations of FISH analysis in detection of microdeletions and micro duplications, the MLPA technique serves as a good and fast tool. However, when such large duplication and triplication pattern, as in patient HA#7, is observed, it is important to exclude extra-chromosomal translocation of the several *F8* copies.

Furthermore, to elucidate the extent of duplicated and triplicated segments, either a multicolor FISH analysis using several DNA probes or a CGH-array should be applied. The later is a less laborious method to identify genomic rearrangement complexity. However, this method provides information only about copy number and not about the orientation or genomic position of the CNV. Due to experimental challenges in determining the breakpoints in case of triplications, it difficult to infer the molecular mechanism. Therefore, the molecular mechanism for triplication formation in hemophilia A is poorly understood and investigated.

The high expression level of *F8* mRNA, in this patient HA#7, is contrary to the severe hemophilia A phenotype and can not explain the absence of FVIII protein (based on FVIII:Ag measurements). Only a few studies have analyzed the mRNA-protein expression correlation in human [156-158]. A recent study showed that gene expression at mRNA level is correlated with the protein level [159]. However, in case of this patient, despite the duplication/ triplication event, the transcription of *F8* mRNA is accelerated. However, the translation of this mRNA is failed.

The causal mutation for all severe hemophilia A cases studied in this work was detected. In all cases the defect was in *F8*. This data suggests that further characterization of other severe hemophilia A patients without detectable *F8* mutations, may reveal novel complex rearrangements involving either repetitive sequence hot spots [23, 160] or micro homology regions [41]. Complex recombination mechanisms might be the fundamental cause of a number of severe hemophilia A cases that, until now, have evaded specific genetic diagnosis. The special features of Xq28 make this region an unstable area of human genome, which is exposed to rearrangements [141, 143, 153, 154]. The presence of long, almost identical inverted repeats at Xq28 account for the high frequency at which these rearrangements occur.

## 5.5  Role of 5´ - and 3´- untranslated regions of mRNA in human diseases

The regulation of protein synthesis is critical for balancing the cellular metabolism. Proteins involved in the process of translation play an essential role in the regulation machinery of proteins. However, changes in the *cis*-regulatory elements and *trans*-acting factors of UTRs can disturb the protein synthesis, associated with various diseases. UTR sequences are highly diverse in sequence but they contain common regulatory motifs. The special features of UTRs, such as their GC rich regions and internal ribosome entry sites (IRESs), have a significant bearing on translation rate of mRNA. *Trans*-acting factors such as 5´-cap, multiple ORFs, multiple upstream AUG (uAUGs), IRESs and polyadenylisation signals, have an effect on the stability of mRNA, its accessibility to the ribosome and its general interaction with the other proteins involved in the translation machinery [161].

Many important genes encoding proteins that are involved in critical cellular processes are regulated at the level of translation, predominately through their UTRs [162]. Mutations in these regions can lead to a disease by different mechanisms. So far quite a number of mutations are reported in 5´ UTR of mRNAs. These mutations could affect the length and the secondary structure of 5´ UTR as in breast cancer [163]. A single G>C transition at position -3 downregulates the translation of *BRCA1*

mRNA [164]. Moreover, changes upstream ORFs are reported to be associated with Alzheimer [165] and bipolar affective disorder [166]. The X-linked Charcot-Marie-tooth disease [167], multiple myeloma [168, 169] and fragile X syndrome [170, 171] are affected by mutations in IRES elements of 5´ UTR. Additionally, changes in self-complementary stem-loop structures of 5´ UTR and RNA binding proteins are reported to be associated with hereditary hyperferritinemia/cataract syndrome [172].

The number of studies on disease-associated changes in 3´ UTR is much lesser in extent. Although, 3´ UTR plays an important role in translocation, localization and stability of mRNA, it is often not screened during genetic analysis [173]. MicroRNAs, 21-24 nt duplex RNAs, are associated with 3´ UTR diseases. They bring the RNA-induced silencing complex (RISC) to 3´ UTR of mRNAs. Mutations in 3´ UTR can lead to gain or loss of microRNA function that results in inhibition or promotion of translation [174]. Beside the role of microRNAs on regulation of mRNA translation [174, 175], three main mutation mechanisms are reported in this context. Diseases associated with mutations in 3´ UTR can be caused by mutations generating stop codons (e.g. Epidermolysis bullosa simplex [176], Aniridia [177]). Such mutations result in production of a protein that is larger in size. This can interfere with the function of the protein or its interaction with associated proteins. Mutations can affect the polyadenylation signal causing diseases such as hemoglobin H disease [178-180] and immune dysfunction polyendocrinopathy enteropathy X-linked [181]. The polyadenylation signal is a conserved sequence motif (AAUAAA), which is essential for transcriptional termination of mRNA. Furthermore, mutations affecting the secondary structure of mRNA have also been reported to be associated with various diseases. These mutations effect the interaction with associated proteins. Congenital heart disease is caused by mutations in mRNA of *GATA4*, a sequence motif that binds to GATA motif. These mutations alter the secondary structure of mRNA and affect the localization and translation of the mRNA [182]. A single transition in the 3´ UTR of TGF-β3 is reported to be associated with arrhythmogenic right ventricular cardiomayopathy [183].

Several groups have already reported mutations in 5´ UTR of *F8* mRNA [184, 185]. Figueiredo et al. analyzed the *cis*-acting elements and transcription factors that are involved in the promoter activity of *F8*. Until recently, no functional assays were

performed to characterize these mutations. Recently, Zimmermann et al. analyzed the causality of five *F8* promoter variants *in vitro*. All variants led to a decrease in promoter activity in accordance to the phenotype of patients [187].

So far, no pathogenic changes are reported in 3´ UTR of *F8* mRNA. The frequent non-pathogenic polymorphism rs1050705 (minor allele frequency G=0.3743), at position c.7055*+1223 of *F8* mRNA, is the only known variation in this region [69]. A unique G>T transition in 3´ UTR of *F8* was found in this study (patients HA#3 and HA#4).

Results obtained from the miRBase microRNA prediction software showed that no differences in binding of mature microRNAs to the wild type and mutant sequence. However, miRBase results showed the loss of binding site of three stem loop microRNAs (hsa-mir-1470, hsa-mir-493 and hsa-mir-607) and gain of one binding site (hsa-mir-4754) in presence of the mutation. However, to assume whether these microRNAs could affect the expression of *F8* mRNA, two issues should be considered. First, it is questionable if these microRNAs are expressed in the liver, which is the main site of *F8* expression and synthesis [50]. Second, a suitable cellular system from the patients is needed to perform functional assays to get conclusive results.

Moreover, the wild type mRNA has one more guanidine-cytosine pair than the mutant, and the length of the paired zone is one pair longer compared to mutant (Figure 27). The loss of the stronger-paired zones (G-C bound is stronger compared to the A-U bound) and the shorter paired zones would reduce the melting temperature of mRNA. This facilitates the break of the pairing in the mutant by ribosome and thereby processing the translation [188]. This could explain the reduction of FVIII:C levels due to presence of the alternative spliced mutant mRNA.

The mRNA analysis revealed alternative splicing in UTR of *F8* mRNA. The aberrant band showed a deletion of 159 bp of 3´ UTR downstream (+2 bp) of this variation. This mutation results in a change in length of *F8* mRNA. This would affect the structure and accessibility of *F8* mRNA to its interacting partners. Based on the revealed results, the impact of the length of 3´ UTR is more suggestive for the instability of *F8* mRNA than microRNA and structural changes. The presence of the

wild type *F8* mRNA in the patient would explain the mild to moderate phenotype observed in patient HA#3 and HA#4.

Taken together, mutations in both 5′ and 3′ UTR of mRNA can affect the regulation of protein synthesis. These mutations may result in a change in structure, length, and interaction of the mRNA. These changes may affect the accessibility of mRNA to microRNAs and its interacting proteins, as well as stability of mRNA.

## 5.6 Reliability and accuracy of splice site prediction softwares

When the causality of a DNA change is not clear, there is a need for prediction softwares to estimate the impact of these changes on splicing. In eukaryotes, when the mRNA is transcribed from DNA the introns are interspersed within the coding sequence. This immature mRNA is designated then as precursor mRNA (pre-mRNA). The mature mRNA is produced during a process known as RNA splicing. In most eukaryotes, the 5′ boundary or donor site and the 3′ boundary or acceptor site dinucleotides are conserved sequences. The donor site of introns is usually the dinucleotide GT (GU in pre-mRNA), while the acceptor site contains the dinucleotide AG. A pyrimidine-rich region precedes the AG at the acceptor site, a shorter consensus follows the GT at the donor site, and a very weak consensus sequence appears at the branch point ~30 nt upstream from the acceptor site (Figure 50). A complex of proteins and small nuclear RNAs, known collectively as the spliceosome, recognizes these consensus sequences. With the help of spliceosome, the introns from pre-mRNA are spliced out and the mature mRNA transcript is produced. In case of mutations within the consensus splice site, this site may still be used, although less efficient. In case of variations not affecting the consensus splice site, more splice sites compete, and alternative splicing generates mRNA variants that produce different proteins from the same pre-mRNA (reviewed in [189]).
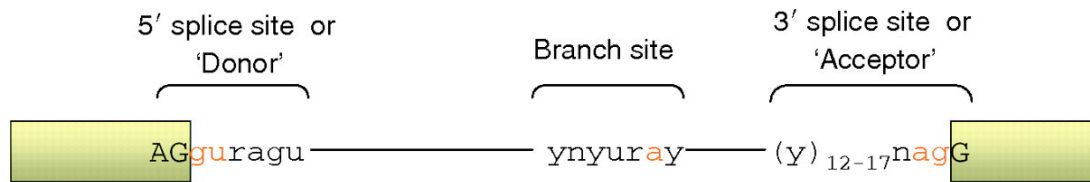
Figure 50: ***Cis*-acting sequences that control splicing.** Consensus sequences for the 5′ splice site (donor), branch site and 3′ splice site (acceptor). Exon sequences are boxed (modified from [190]).

Alternative splicing is a common event [191], which was described for the first time 1997 in the adenovirus [192] and in 1980 for *IgM* gene [193]. Splice site prediction programs are reliable tools to predict whether a mutation could disturb the correct pre-mRNA splicing and result in alternative splicing.

However, several algorithms should be applied to determine if a given variation could create a new dss and/or ass site. A number of computational methods have been developed to identify these splice sites. The most common resources are: NetGene2 [194], HSF (based on position weight matrices with position-dependent logic) [195] and FRUITFLY (based on neural networks) [196], which are online splice-site prediction programs and the program *Alamut*, which combines several splice site prediction tools (SpliceSiteFinder-like, MaxEntScan, NNSPLICE, HSF, GeneSplicer).

Zimmermann et al. analyzed the effect of several intronic variants in *F8* close to or within splice site consensus sequences. They showed that the majority of results obtained from splice prediction tools are in accordance with the experimental *F8* mRNA analysis done. However, a discrepancy was observed between *in silico* and *in vitro* analysis for some cases [197].

The data presented in this work also strongly suggests that the potential of intronic variations on splicing should be determined based on experimental mRNA analysis rather than relying on prediction softwares. The two mRNA approaches applied in this work investigate *F8* mRNA in two aspects. The qualitative nested RT-PCR approach reveals alternative splicing of *F8* mRNA. Moreover, the quantitative RT-PCR approach gives insight to the expression amount of normal *F8* mRNA transcript to the aberrant spliced mRNA. It should be taken into account that incorrect splicing

do not affect 100% of the transcripts. The amount of correctly spliced transcripts to aberrant transcripts could make a difference between phenotype of the disease. In conclusion, the causality of splice site variants needs to be validated by mRNA analysis (both, quantitatively and qualitatively).

## 5.7 Hemophilia A patients without mutation

Still, in some patients, it was not possible to detect a causal mutation explaining the hemophilia A phenotype. Here, some possibilities may be considered. The first is tissue mosaicism. A point mutation in *F8* could be present in the liver among other tissues but not in blood. Although mutations causing hemophilia A usually appear to have arisen in germ cells, a *de novo* mutation may also occur during early embryogenesis and thus may represent either germline and/or somatic mosaicism [198-201]. However, this could be excluded in familial cases as in some of our patients.

Another possibility is the rapid degradation of the FVIII protein in blood or its rapid uptake. In this respect, the clinical observations from the patients would argue against this possibility, as all patients receive similar amount of substitution regimens of either plasma-derived or rFVIII concentrates, compared to a hemophilia A patient with known mutation and the same severity.

The third possibility is a mutation in another gene encoding a novel interacting or modifying protein of FVIII. Other unknown mutations at other loci could affect the intracellular trafficking and/or secretion of FVIII molecule. It is possible that not only LMAN1 and MCFD2 are responsible for trafficking of FVIII protein. A third protein may be involved in VIII/vWF complex and could affect its protection from degradation. The last possibility is a still not identified regulatory region in *F8*. This is, however, in accordance with the observations in pedigrees. The disease seems to be caused by a mutation in *F8* gene itself or the defect is linked to *F8*.

To elucidate the molecular mechanism behind hemophilia A in the remaining patients, FVIII protein should be investigated at its major site of production, which is

the liver. However, taking a liver biopsy from these patients is almost impossible. Not only the ethical aspects are important but also the risk of hemorrhage in hemophilia A patients is high. Alternatively, a patient specific cellular system should be established to monitor the expression and secretion of FVIII in the remaining cases. The recent advances in generation of induced pluripotent stem cells (iPS), make them a promising tool to establish such patient specific cell lines in the future.


## 5.8   Proposed flowchart for genetic analysis of mutation negative patients

Based on this work, the following strategy is proposed to elucidate the molecular mechanism in mutation negative hemophilia A patients. In the beginning of the analysis, the general algorithm for genetic analysis of patients with bleeding tendency and reduced FVIII:C levels should be applied for all patients. Hemophilia A should be reconfirmed in all mutation negative patients, on the basis of the flowchart presented in Figure 51.
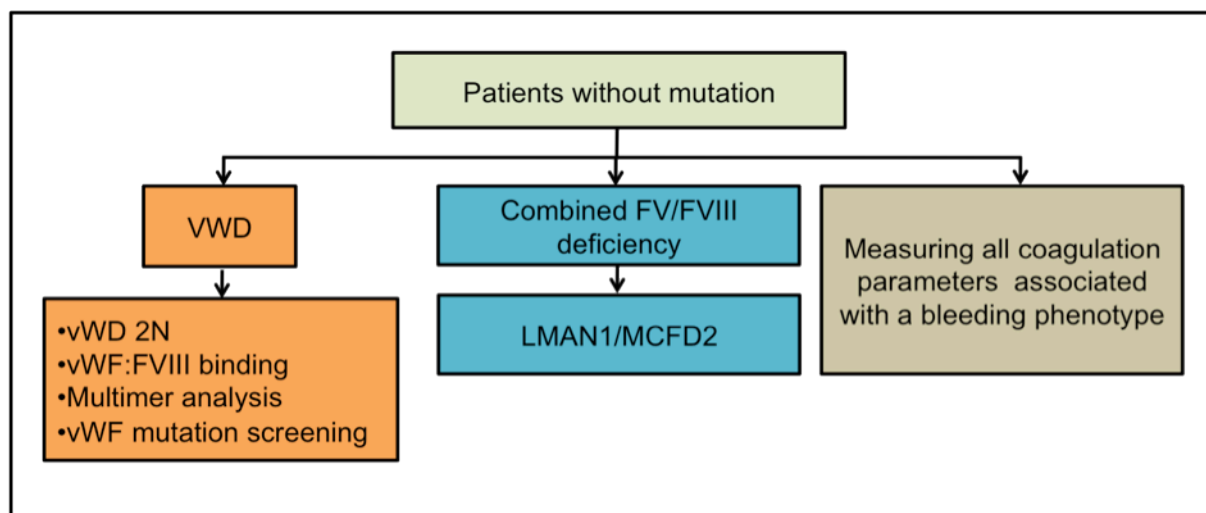


Figure 51: **Flowchart for reconfirmation of hemophilia A in patients without mutation.**

After performing this analysis, misdiagnosis of hemophilia A is excluded in all patients. Up to this point, hemophilia A is confirmed and the only parameter associated with the bleeding tendency would be lack or shortage of FVIII protein. The next step of analysis would include changes at non-coding region of *F8*, followed by mRNA analysis.

In the next step, extensive analysis on DNA level of *F8* should be performed to exclude CNVs. In case of abnormalities a CGH-array may be applied to verify MLPA results. If No CNVs are detected, FISH analysis is needed to exclude translocation of *F8.* The next step of analysis is the amplification of complete genomic region of *F8* using a LR-PCR approach. In cases, where a breakpoint in *F8* locus is detected, the junctions of the breakpoint can be characterized by an inverse PCR approach. Based on sequencing results of inverse PCR products, a multiplex Short-Range PCR can be designed. This could be used later for routine diagnostic purposes of mutation negative patients. In patients, where the integrity of *F8* locus is confirmed, NGS of the LR-PCR products could reveal intronic variants (Figure 52).

The analysis will be continued on mRNA level to evaluate the effect of these DNA changes on mRNA. Prior to mRNA analysis, all variants should be checked in SNP databases to determine the unique intronic variants, which should be analyzed using splice site prediction softwares. Irrespective to the *in silico* results, RT-PCRs of *F8* mRNA should be performed to detect abnormal RNA transcripts resulted from alternative splicing. To finalize the analysis, the amount of *F8* mRNA should be determined in exon-exon boundaries flanking the unique intronic variations. This is done by performing qRT-PCR analysis and calculation the ratio of wild type *F8* mRNA to the alternative spliced mRNA (Figure 52).
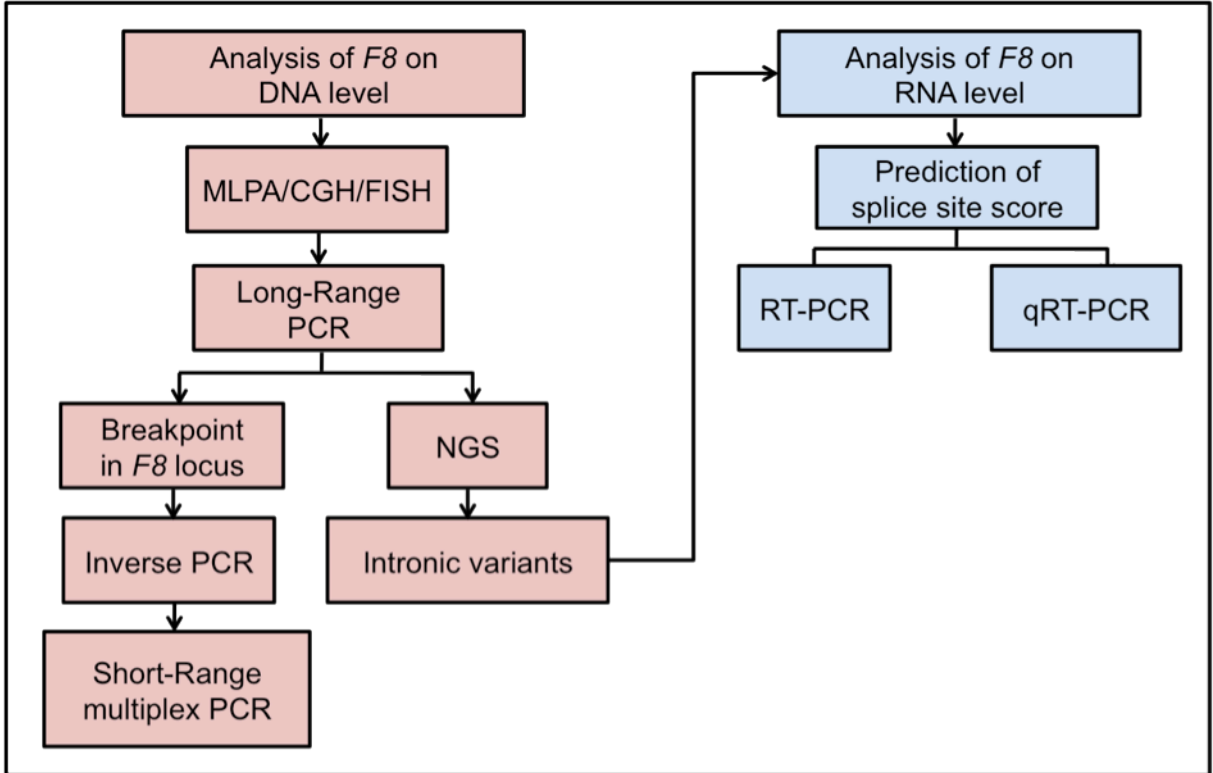
Figure 52: **Experimental protocol applied on *F8* for identification of novel mechanisms leading to hemophilia A.**

This proposed experimental protocol would narrow the number of hemophilia A patients so far classified as patients without mutation. Moreover, this practical approach could serve as a model for genetic analysis of other monogenic diseases, where no pathogenic mutation is found.

### References:

1 Mann KG. Biochemistry and physiology of blood coagulation. *Thromb Haemost*. 1999; **82**: 165-74.

2 Butenas S, Mann KG. Blood coagulation. *Biochemistry (Mosc)*. 2002; **67**: 3-12.

3 Monroe DM, Hoffman M. What does it take to make the perfect clot? *Arterioscler Thromb Vasc Biol*. 2006; **26**: 41-8.

4 Macfarlane RG. An Enzyme Cascade in the Blood Clotting Mechanism, and Its Function as a Biochemical Amplifier. *Nature*. 1964; **202**: 498-9.

5 Davie EW, Ratnoff OD. Waterfall Sequence for Intrinsic Blood Clotting. *Science*. 1964; **145**: 1310-2.

6 Mann KG, Brummel K, Butenas S. What is all that thrombin for? *J Thromb Haemost*. 2003; **1**: 1504-14.

7 Ichinose A. Physiopathology and regulation of factor XIII. *Thromb Haemost*. 2001; **86**: 57-65.

8 Furie B, Furie BC. Mechanisms of thrombus formation. *N Engl J Med*. 2008; **359**: 938-49.

9 Mannucci PM, Tuddenham EG. The hemophilias--from royal genes to gene therapy. *N Engl J Med*. 2001; **344**: 1773-9.

10 McGlynn LK, Mueller CR, Begbie M, Notley CR, Lillicrap D. Role of the liver-enriched transcription factor hepatocyte nuclear factor 1 in transcriptional regulation of the factor V111 gene. *Mol Cell Biol*. 1996; **16**: 1936-45.

11 Rosner F. Hemophilia in the Talmud and rabbinic writings. *Ann Intern Med*. 1969; **70**: 833-7.

12 Otto JC. An account of an hemorrhagic disposition existing in certain families. *Clin Orthop Relat Res*. 1996: 4-6.

13 McKusick VA. The Royal Hemophilia. *Sci Am*. 1965; **213**: 88-95.

14 Antonarakis SE. Molecular genetics of coagulation factor VIII gene and hemophilia A. *Thromb Haemost*. 1995; **74**: 322-8.

15 Antonarakis SE. Molecular genetics of coagulation factor VIII gene and haemophilia A. *Haemophilia*. 1998; **4 Suppl 2**: 1-11.

16 Bicocchi MP, Migeon BR, Pasino M, Lanza T, Bottini F, Boeri E, Molinari AC, Corsolini F, Morerio C, Acquila M. Familial nonrandom inactivation linked to the X inactivation centre in heterozygotes manifesting haemophilia A. *Eur J Hum Genet*. 2005; **13**: 635-40.

17 Renault NK, Dyack S, Dobson MJ, Costa T, Lam WL, Greer WL. Heritable skewed X-chromosome inactivation leads to haemophilia A expression in heterozygous females. *Eur J Hum Genet*. 2007; **15**: 628-37.

18 Knobe KE, Sjorin E, Soller MJ, Liljebjorn H, Ljung RC. Female haemophilia A caused by skewed X inactivation. *Haemophilia*. 2008; **14**: 846-8.

19 Windsor S, Lyng A, Taylor SA, Ewenstein BM, Neufeld EJ, Lillicrap D. Severe haemophilia A in a female resulting from two de novo factor VIII mutations. *Br J Haematol*. 1995; **90**: 906-9.

20 Cai XH, Wang XF, Dai J, Fang Y, Ding QL, Xie F, Wang HL. Female hemophilia A heterozygous for a de novo frameshift and a novel missense mutation of factor VIII. *J Thromb Haemost*. 2006; **4**: 1969-74.

21 Pavlova A, Forster T, Delev D, Schroder J, El-Maarri O, Muller-Reible C, Oldenburg J. Heterozygous large deletions of Factor 8 gene in females identified by multiplex PCR-LC. *Haemophilia*. 2008; **14**: 599-606.

22      Pavlova A, Brondke H, Musebeck J, Pollmann H, Srivastava A, Oldenburg J. Molecular mechanisms underlying hemophilia A phenotype in seven females. *J Thromb Haemost*. 2009; **7**: 976-82.

23      Bagnall RD, Waseem N, Green PM, Giannelli F. Recurrent inversion breaking intron 1 of the factor VIII gene is a frequent cause of severe hemophilia A. *Blood*. 2002; **99**: 168-74.

24      Lakich D, Kazazian HH, Jr., Antonarakis SE, Gitschier J. Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. *Nat Genet*. 1993; **5**: 236-41.

25      Oldenburg J, El-Maarri O. New insight into the molecular basis of hemophilia A. *Int J Hematol*. 2006; **83**: 96-102.

26      Jacquemin M, De Maeyer M, D'Oiron R, Lavend'Homme R, Peerlinck K, Saint-Remy JM. Molecular mechanisms of mild and moderate hemophilia A. *J Thromb Haemost*. 2003; **1**: 456-63.

27      Troisi CL, Hollinger FB, Hoots WK, Contant C, Gill J, Ragni M, Parmley R, Sexauer C, Gomperts E, Buchanan G, et al. A multicenter study of viral hepatitis in a United States hemophilic population. *Blood*. 1993; **81**: 412-8.

28      Eyster ME, Gail MH, Ballard JO, Al-Mondhiry H, Goedert JJ. Natural history of human immunodeficiency virus infections in hemophiliacs: effects of T-cell subsets, platelet counts, and age. *Ann Intern Med*. 1987; **107**: 1-6.

29      Gitschier J, Wood WI, Goralka TM, Wion KL, Chen EY, Eaton DH, Vehar GA, Capon DJ, Lawn RM. Characterization of the human factor VIII gene. *Nature*. 1984; **312**: 326-30.

30      Toole JJ, Knopf JL, Wozney JM, Sultzman LA, Buecker JL, Pittman DD, Kaufman RJ, Brown E, Shoemaker C, Orr EC, et al. Molecular cloning of a cDNA encoding human antihaemophilic factor. *Nature*. 1984; **312**: 342-7.

31      Chuah MK, Collen D, Vandendriessche T. Preclinical and clinical gene therapy for haemophilia. *Haemophilia*. 2004; **10 Suppl 4**: 119-25.

32      White GC, 2nd. Gene therapy in hemophilia: clinical trials update. *Thromb Haemost*. 2001; **86**: 172-7.

33      Levinson B, Kenwrick S, Lakich D, Hammonds Jr G, Gitschier J. A transcribed gene in an intron of the human factor VIII gene. *Genomics*. 1990; **7**: 1-11.

34      Levinson B, Kenwrick S, Gamel P, Fisher K, Gitschier J. Evidence for a third transcript from the human factor VIII gene. *Genomics*. 1992; **14**: 585-9.

35      Peters MF, Ross CA. Isolation of a 40-kDa Huntingtin-associated protein. *J Biol Chem*. 2001; **276**: 3188-94.

36      Naylor JA, Buck D, Green P, Williamson H, Bentley D, Giannelli F. Investigation of the factor VIII intron 22 repeated region (int22h) and the associated inversion junctions. *Hum Mol Genet*. 1995; **4**: 1217-24.

37      Liu J, Liu Q, Liang Y, Wang L, Nozary G, Xiao B, Zhu Z, Zhou Y, Liu L, Guan Y, Zhang J, Sommer SS. PCR assay for the inversion causing severe Hemophilia A and its application. *Chin Med J (Engl)*. 1999; **112**: 419-23.

38      Youssoufian H, Antonarakis SE, Bell W, Griffin AM, Kazazian HH, Jr. Nonsense and missense mutations in hemophilia A: estimate of the relative mutation rate at CG dinucleotides. *Am J Hum Genet*. 1988; **42**: 718-25.

39      El-Maarri O, Olek A, Balaban B, Montag M, van der Ven H, Urman B, Olek K, Caglayan SH, Walter J, Oldenburg J. Methylation levels at selected CpG sites in the factor VIII and FGFR3 genes, in mature female and male germ cells: implications for male-driven evolution. *Am J Hum Genet*. 1998; **63**: 1001-8.

40    Roth L, Marschalek R, Oldenburg J, Oyen F, Schneppenheim R. Characterisation of two novel large F8 deletions in patients with severe haemophilia A and factor VIII inhibitors. *Thromb Haemost*. 2011; **105**: 279-84.

41    Zimmermann MA, Oldenburg J, Muller CR, Rost S. Characterization of duplication breakpoints in the factor VIII gene. *J Thromb Haemost*. 2010; **8**: 2696-704.

42    El-Maarri O, Herbiniaux U, Graw J, Schroder J, Terzic A, Watzka M, Brackmann HH, Schramm W, Hanfland P, Schwaab R, Muller CR, Oldenburg J. Analysis of mRNA in hemophilia A patients with undetectable mutations reveals normal splicing in the factor VIII gene. *J Thromb Haemost*. 2005; **3**: 332-9.

43    Castaman G, Giacomelli SH, Mancuso ME, Sanna S, Santagostino E, Rodeghiero F. F8 mRNA studies in haemophilia A patients with different splice site mutations. *Haemophilia*. 2010; **16**: 786-90.

44    Nichols WC, Seligsohn U, Zivelin A, Terry VH, Hertel CE, Wheatley MA, Moussalli MJ, Hauri HP, Ciavarella N, Kaufman RJ, Ginsburg D. Mutations in the ER-Golgi intermediate compartment protein ERGIC-53 cause combined deficiency of coagulation factors V and VIII. *Cell*. 1998; **93**: 61-70.

45    Zhang B, Cunningham MA, Nichols WC, Bernat JA, Seligsohn U, Pipe SW, McVey JH, Schulte-Overberg U, de Bosch NB, Ruiz-Saez A, White GC, Tuddenham EG, Kaufman RJ, Ginsburg D. Bleeding due to disruption of a cargo-specific ER-to-Golgi transport complex. *Nat Genet*. 2003; **34**: 220-5.

46    Neels JG, Bovenschen N, van Zonneveld AJ, Lenting PJ. Interaction between factor VIII and LDL receptor-related protein. Modulation of coagulation? *Trends Cardiovasc Med*. 2000; **10**: 8-14.

47    Schneppenheim R, Budde U, Krey S, Drewke E, Bergmann F, Lechler E, Oldenburg J, Schwaab R. Results of a screening for von Willebrand disease type 2N in patients with suspected haemophilia A or von Willebrand disease type 1. *Thromb Haemost*. 1996; **76**: 598-602.

48    Uen C, Oldenburg J, Schroder J, Brackmann HJ, Schramm W, Schwaab R, Schneppenheim R, Graw J. [2% Haemophilia A patients without mutation in the FVIII gene]. *Hamostaseologie*. 2003; **23**: 1-5.

49    Wion KL, Kelly D, Summerfield JA, Tuddenham EG, Lawn RM. Distribution of factor VIII mRNA and antigen in human liver and other tissues. *Nature*. 1985; **317**: 726-9.

50    Hollestelle MJ, Thinnes T, Crain K, Stiko A, Kruijt JK, van Berkel TJ, Loskutoff DJ, van Mourik JA. Tissue distribution of factor VIII gene expression in vivo--a closer look. *Thromb Haemost*. 2001; **86**: 855-61.

51    Shahani T, Lavend'homme R, Luttun A, Saint-Remy JM, Peerlinck K, Jacquemin M. Activation of human endothelial cells from specific vascular beds induces the release of a FVIII storage pool. *Blood*. 2010; **115**: 4902-9.

52    Graw J, Brackmann HH, Oldenburg J, Schneppenheim R, Spannagl M, Schwaab R. Haemophilia A: from mutation analysis to new therapies. *Nat Rev Genet*. 2005; **6**: 488-501. nrg1617 [pii]
10.1038/nrg1617.

53    Kane WH, Davie EW. Cloning of a cDNA coding for human factor V, a blood coagulation factor homologous to factor VIII and ceruloplasmin. *Proc Natl Acad Sci U S A*. 1986; **83**: 6800-4.

54    Fuentes-Prior P, Fujikawa K, Pratt KP. New insights into binding interfaces of coagulation factors V and VIII and their homologues lessons from high resolution crystal structures. *Curr Protein Pept Sci*. 2002; **3**: 313-39.

55      Cripe LD, Moore KD, Kane WH. Structure of the gene for human coagulation factor V. *Biochemistry*. 1992; **31**: 3777-85.

56      Rehemtulla A, Kaufman RJ. Protein processing within the secretory pathway. *Curr Opin Biotechnol*. 1992; **3**: 560-5.

57      Fay PJ. Factor VIII structure and function. *Int J Hematol*. 2006; **83**: 103-8.

58      Curtis JE, Helgerson SL, Parker ET, Lollar P. Isolation and characterization of thrombin-activated human factor VIII. *J Biol Chem*. 1994; **269**: 6246-51.

59      Nogami K, Shima M, Nishiya K, Hosokawa K, Saenko EL, Sakurai Y, Shibata M, Suzuki H, Tanaka I, Yoshioka A. A novel mechanism of factor VIII protection by von Willebrand factor from activated protein C-catalyzed inactivation. *Blood*. 2002; **99**: 3993-8.

60      Bovenschen N, Mertens K, Hu L, Havekes LM, van Vlijmen BJ. LDL receptor cooperates with LDL receptor-related protein in regulating plasma levels of coagulation factor VIII in vivo. *Blood*. 2005; **106**: 906-12.

61      Green D. Immunosuppression of patients with acquired factor VIII inhibitors. *Semin Hematol*. 1994; **31**: 60-1.

62      Jeremic M, Weisert O, Gedde-Dahl TW. Factor VIII (AHG) levels in 1016 regular blood donors. The effects of age, sex, and ABO blood groups. *Scand J Clin Lab Invest*. 1976; **36**: 461-6.

63      McCallum CJ, Peake IR, Newcombe RG, Bloom AL. Factor VIII levels and blood group antigens. *Thromb Haemost*. 1983; **50**: 757.

64      Gill JC, Endres-Brooks J, Bauer PJ, Marks WJ, Jr., Montgomery RR. The effect of ABO blood group on the diagnosis of von Willebrand disease. *Blood*. 1987; **69**: 1691-5.

65      Conlan MG, Folsom AR, Finch A, Davis CE, Sorlie P, Marcucci G, Wu KK. Associations of factor VIII and von Willebrand factor with age, race, sex, and risk factors for atherosclerosis. The Atherosclerosis Risk in Communities (ARIC) Study. *Thromb Haemost*. 1993; **70**: 380-5.

66      Balleisen L, Bailey J, Epping PH, Schulte H, van de Loo J. Epidemiological study on factor VII, factor VIII and fibrinogen in an industrial population: I. Baseline data on the relation to age, gender, body-weight, smoking, alcohol, pill-using, and menopause. *Thromb Haemost*. 1985; **54**: 475-9.

67      Balleisen L, Assmann G, Bailey J, Epping PH, Schulte H, van de Loo J. Epidemiological study on factor VII, factor VIII and fibrinogen in an industrial population--II. Baseline data on the relation to blood pressure, blood glucose, uric acid, and lipid fractions. *Thromb Haemost*. 1985; **54**: 721-3.

68      Tracy RP, Bovill EG, Fried LP, Heiss G, Lee MH, Polak JF, Psaty BM, Savage PJ. The distribution of coagulation factors VII and VIII and fibrinogen in adults over 65 years. Results from the Cardiovascular Health Study. *Ann Epidemiol*. 1992; **2**: 509-19.

69      Viel KR, Machiah DK, Warren DM, Khachidze M, Buil A, Fernstrom K, Souto JC, Peralta JM, Smith T, Blangero J, Porter S, Warren ST, Fontcuberta J, Soria JM, Flanders WD, Almasy L, Howard TE. A sequence variation scan of the coagulation factor VIII (FVIII) structural gene and associations with plasma FVIII activity levels. *Blood*. 2007; **109**: 3713-24.

70      Lowe GD, Rumley A, Woodward M, Morrison CE, Philippou H, Lane DA, Tunstall-Pedoe H. Epidemiology of coagulation factors, inhibitors and activation markers: the Third Glasgow MONICA Survey. I. Illustrative reference ranges by age, sex and hormone use. *Br J Haematol*. 1997; **97**: 775-84.

71      Schlit AF, Grandjean P, Donnez J, Lavenne E. Large increase in plasmatic 11-dehydro-TxB2 levels due to oral contraceptives. *Contraception*. 1995; **51**: 53-8.

72      Geffken DF, Cushman M, Burke GL, Polak JF, Sakkinen PA, Tracy RP. Association between physical activity and markers of inflammation in a healthy elderly population. *Am J Epidemiol*. 2001; **153**: 242-50.

73      Moffat EH, Giddings JC, Bloom AL. The effect of desamino-D-arginine vasopressin (DDAVP) and naloxone infusions on factor VIII and possible endothelial cell (EC) related activities. *Br J Haematol*. 1984; **57**: 651-62.

74      Bloom AL, Peake IR, Furlong RA, Davies BL. High potency factor VIII concentrate: more effective than cryoprecipitate in a patient with von Willebrand's disease and inhibitor. *Thromb Res*. 1979; **16**: 847-52.

75      El-Maarri O, Singer H, Klein C, Watzka M, Herbiniaux U, Brackmann HH, Schroder J, Graw J, Muller CR, Schramm W, Schwaab R, Haaf T, Hanfland P, Oldenburg J. Lack of F8 mRNA: a novel mechanism leading to hemophilia A. *Blood*. 2006; **107**: 2759-65.

76      Castaman G, Lethagen S, Federici AB, Tosetto A, Goodeve A, Budde U, Batlle J, Meyer D, Mazurier C, Fressinaud E, Goudemand J, Eikenboom J, Schneppenheim R, Ingerslev J, Vorlova Z, Habart D, Holmberg L, Pasi J, Hill F, Peake I, Rodeghiero F. Response to desmopressin is influenced by the genotype and phenotype in type 1 von Willebrand disease (VWD): results from the European Study MCMDM-1VWD. *Blood*. 2008; **111**: 3531-9.

77      Kaufmann JE, Vischer UM. Cellular mechanisms of the hemostatic effects of desmopressin (DDAVP). *J Thromb Haemost*. 2003; **1**: 682-9.

78      Lethagen S, Frick K, Sterner G. Antidiuretic effect of desmopressin given in hemostatic dosages to healthy volunteers. *Am J Hematol*. 1998; **57**: 153-9.

79      Stoof SC, Sanders YV, Petrij F, Cnossen MH, de Maat MP, Leebeek FW, Kruip MJ. Response to desmopressin is strongly dependent on F8 gene mutation type in mild and moderate haemophilia A. *Thromb Haemost*. 2013; **109**.

80      Ling MM, Robinson BH. Approaches to DNA mutagenesis: an overview. *Anal Biochem*. 1997; **254**: 157-78.

81      Hopfner KP, Eichinger A, Engh RA, Laue F, Ankenbauer W, Huber R, Angerer B. Crystal structure of a thermostable type B DNA polymerase from Thermococcus gorgonarius. *Proc Natl Acad Sci U S A*. 1999; **96**: 3600-5.

82      Bernard P, Gabant P, Bahassi EM, Couturier M. Positive-selection vectors using the F plasmid ccdB killer gene. *Gene*. 1994; **148**: 71-4.

83      Rosenblum BB, Lee LG, Spurgeon SL, Khan SH, Menchen SM, Heiner CR, Chen SM. New dye-labeled terminators for improved DNA sequencing patterns. *Nucleic Acids Res*. 1997; **25**: 4500-4.

84      Tost J, Dunker J, Gut IG. Analysis and quantification of multiple methylation variable positions in CpG islands by Pyrosequencing. *Biotechniques*. 2003; **35**: 152-6.

85      Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*. 1988; **16**: 1215.

86      Machado FB, Medina-Acosta E. High-resolution combined linkage physical map of short tandem repeat loci on human chromosome band Xq28 for indirect haemophilia A carrier detection. *Haemophilia*. 2009; **15**: 297-308. 10.1111/j.1365-2516.2008.01866.x
HAE1866 [pii].

87      Oldenburg J, Schwaab R, Grimm T, Zerres K, Hakenberg P, Brackmann HH, Olek K. Direct and indirect estimation of the sex ratio of mutation frequencies in hemophilia A. *Am J Hum Genet*. 1993; **53**: 1229-38.

88      Tang MX, Redemann CT, Szoka FC, Jr. In vitro gene delivery by degraded polyamidoamine dendrimers. *Bioconjug Chem*. 1996; **7**: 703-14.

89      Turi DC, Peerschke EI. Sensitivity of three activated partial thromboplastin time reagents to coagulation factor deficiencies. *Am J Clin Pathol*. 1986; **85**: 43-9.

90      van Dieijen-Visser MP, van Wersch, J, Brombacher, P.J, Rosing, J, Hemker, H.C, van Dieijen, G. Use of Chromogenic Peptide Substrates in the Determination of Clotting Factors II, VII, IX and X in Normal Plasma and in Plasma of Patients Treated with Oral Anticoagulants. *Haemostasis*. 1982; **12**: 241-55.

91      Thomas KB, Sutor AH, Altinkaya N, Grohmann A, Zehenter A, Leititis JU. von Willebrand factor-collagen binding activity is increased in newborns and infants. *Acta Pædiatrica*. 1995; **84**: 697-700.

92      Budde U, Schneppenheim R, Plendl H, Dent J, Ruggeri ZM, Zimmerman TS. Luminographic detection of von Willebrand factor multimers in agarose gels and on nitrocellulose membranes. *Thromb Haemost*. 1990; **63**: 312-5.

93      Budde U, Schneppenheim R, Eikenboom J, Goodeve A, Will K, Drewke E, Castaman G, Rodeghiero F, Federici AB, Batlle J, Perez A, Meyer D, Mazurier C, Goudemand J, Ingerslev J, Habart D, Vorlova Z, Holmberg L, Lethagen S, Pasi J, Hill F, Peake I. Detailed von Willebrand factor multimer analysis in patients with von Willebrand disease in the European study, molecular and clinical markers for the diagnosis and management of type 1 von Willebrand disease (MCMDM-1VWD). *J Thromb Haemost*. 2008; **6**: 762-71.

94      Smith NL, Chen MH, Dehghan A, Strachan DP, Basu S, Soranzo N, Hayward C, Rudan I, Sabater-Lleal M, Bis JC, de Maat MP, Rumley A, Kong X, Yang Q, Williams FM, Vitart V, Campbell H, Malarstig A, Wiggins KL, Van Duijn CM, McArdle WL, Pankow JS, Johnson AD, Silveira A, McKnight B, Uitterlinden AG, Aleksic N, Meigs JB, Peters A, Koenig W, Cushman M, Kathiresan S, Rotter JI, Bovill EG, Hofman A, Boerwinkle E, Tofler GH, Peden JF, Psaty BM, Leebeek F, Folsom AR, Larson MG, Spector TD, Wright AF, Wilson JF, Hamsten A, Lumley T, Witteman JC, Tang W, O'Donnell CJ. Novel associations of multiple genetic loci with plasma levels of factor VII, factor VIII, and von Willebrand factor: The CHARGE (Cohorts for Heart and Aging Research in Genome Epidemiology) Consortium. *Circulation*. 2010; **121**: 1382-92.

95      Lollar P. The association of factor VIII with von Willebrand factor. *Mayo Clin Proc*. 1991; **66**: 524-34.

96      Vlot AJ, Koppelman SJ, Bouma BN, Sixma JJ. Factor VIII and von Willebrand factor. *Thromb Haemost*. 1998; **79**: 456-65.

97      Kaufman RJ, Dorner AJ, Fass DN. von Willebrand factor elevates plasma factor VIII without induction of factor VIII messenger RNA in the liver. *Blood*. 1999; **93**: 193-7.

98      Federici AB. The factor VIII/von Willebrand factor complex: basic and clinical issues. *Haematologica*. 2003; **88**: EREP02.

99      Eikenboom JC, Castaman G, Kamphuisen PW, Rosendaal FR, Bertina RM. The factor VIII/von Willebrand factor ratio discriminates between reduced synthesis and increased clearance of von Willebrand factor. *Thromb Haemost*. 2002; **87**: 252-7.

100     Souto JC, Almasy L, Borrell M, Blanco-Vaca F, Mateo J, Soria JM, Coll I, Felices R, Stone W, Fontcuberta J, Blangero J. Genetic susceptibility to thrombosis

and its relationship to physiological risk factors: the GAIT study. Genetic Analysis of Idiopathic Thrombophilia. *Am J Hum Genet*. 2000; **67**: 1452-9.

101    Lenz W. Letter: Half chromatid mutations may explain incontinentia pigmenti in males. *Am J Hum Genet*. 1975; **27**: 690-1.

102    Rudolph D, Yeh WC, Wakeham A, Rudolph B, Nallainathan D, Potter J, Elia AJ, Mak TW. Severe liver degeneration and lack of NF-kappaB activation in NEMO/IKKgamma-deficient mice. *Genes Dev*. 2000; **14**: 854-62.

103    Fusco F, Paciolla M, Napolitano F, Pescatore A, D'Addario I, Bal E, Lioi MB, Smahi A, Miano MG, Ursini MV. Genomic architecture at the Incontinentia Pigmenti locus favours de novo pathological alleles through different mechanisms. *Hum Mol Genet*. 2011.

104    Miller CH, Aledort LM. Discriminant analysis for probability of hemophilia A carriership. *Blood*. 1987; **69**: 704-5.

105    Kasper CK, Lin JC. Prevalence of sporadic and familial haemophilia. *Haemophilia*. 2007; **13**: 90-2.

106    Piedras J, Sanchez-Montero PE, Herrera FM, Cordova MS, Sanchez-Medal L. Effect of plasma freezing temperature, anticoagulant and time of storage on factor VIII:C activity in cryoprecipitate. *Arch Med Res*. 1993; **24**: 23-6.

107    Marlar RA, Potts RM, Marlar AA. Effect on routine and special coagulation testing values of citrate anticoagulant adjustment in patients with high hematocrit values. *Am J Clin Pathol*. 2006; **126**: 400-5.

108    Gupta M, Lillicrap D, Stain AM, Friedman KD, Carcao MD. Therapeutic consequences for misdiagnosis of type 2N von Willebrand disease. *Pediatr Blood Cancer*. 2011; **57**: 1081-3.

109    Gaucher C, Mercier B, Jorieux S, Oufkir D, Mazurier C. Identification of two point mutations in the von Willebrand factor gene of three families with the 'Normandy' variant of von Willebrand disease. *Br J Haematol*. 1991; **78**: 506-14.

110    Mazurier C, Gaucher C, Jorieux S, Goudemand M. Biological effect of desmopressin in eight patients with type 2N ('Normandy') von Willebrand disease. Collaborative Group. *Br J Haematol*. 1994; **88**: 849-54.

111    Keeney S, Bowen D, Cumming A, Enayat S, Goodeve A, Hill M. The molecular analysis of von Willebrand disease: a guideline from the UK Haemophilia Centre Doctors' Organisation Haemophilia Genetics Laboratory Network. *Haemophilia*. 2008; **14**: 1099-111.

112    Caron C, Dautzenberg MD, Delahousse B, Droulle C, Pouzol P, Dubanchet A, Rothschild C. A blinded in vitro study with Refacto mock plasma samples: similar FVIII results between the chromogenic assay and a one-stage assay when using a higher cephalin dilution. *Haemophilia*. 2002; **8**: 639-43.

113    Zhukov O, Popov J, Ramos R, Vause C, Ruden S, Sferruzza A, Dlott J, Sahud M. Measurement of von Willebrand factor-FVIII binding activity in patients with suspected von Willebrand disease type 2N: application of an ELISA-based assay in a reference laboratory. *Haemophilia*. 2009; **15**: 788-96.

114    Zhang B, McGee B, Yamaoka JS, Guglielmone H, Downes KA, Minoldo S, Jarchum G, Peyvandi F, de Bosch NB, Ruiz-Saez A, Chatelain B, Olpinski M, Bockenstedt P, Sperl W, Kaufman RJ, Nichols WC, Tuddenham EG, Ginsburg D. Combined deficiency of factor V and factor VIII is due to mutations in either LMAN1 or MCFD2. *Blood*. 2006; **107**: 1903-7.

115    Spreafico M, Peyvandi F. Combined FV and FVIII deficiency. *Haemophilia*. 2008; **14**: 1201-8.

116    Plug I, Mauser-Bunschoten EP, Brocker-Vriends AH, van Amstel HK, van der Bom JG, van Diemen-Homan JE, Willemse J, Rosendaal FR. Bleeding in carriers of hemophilia. *Blood*. 2006; **108**: 52-6.

117    Miesbach W, Alesci S, Geisen C, Oldenburg J. Association between phenotype and genotype in carriers of haemophilia A. *Haemophilia*. 2011; **17**: 246-51.

118    Castaman G, Giacomelli SH, Mancuso ME, D'Andrea G, Santacroce R, Sanna S, Santagostino E, Mannucci PM, Goodeve A, Rodeghiero F. Deep intronic variations may cause mild hemophilia A. *J Thromb Haemost*. 2011; **9**: 1541-8.

119    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.

120    Voelkerding KV, Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. *Clin Chem*. 2009; **55**: 641-58.

121    Voelkerding KV, Dames S, Durtschi JD. Next generation sequencing for clinical diagnostics-principles and application to targeted resequencing for hypertrophic cardiomyopathy: a paper from the 2009 William Beaumont Hospital Symposium on Molecular Pathology. *J Mol Diagn*. 2010; **12**: 539-51.

122    Gargis AS, Kalman L, Berry MW, Bick DP, Dimmock DP, Hambuch T, Lu F, Lyon E, Voelkerding KV, Zehnbauer BA, Agarwala R, Bennett SF, Chen B, Chin EL, Compton JG, Das S, Farkas DH, Ferber MJ, Funke BH, Furtado MR, Ganova-Raeva LM, Geigenmuller U, Gunselman SJ, Hegde MR, Johnson PL, Kasarskis A, Kulkarni S, Lenk T, Liu CS, Manion M, Manolio TA, Mardis ER, Merker JD, Rajeevan MS, Reese MG, Rehm HL, Simen BB, Yeakley JM, Zook JM, Lubin IM. Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nat Biotechnol*. 2012; **30**: 1033-6.

123    Tost J, Gut IG. DNA methylation analysis by pyrosequencing. *Nat Protoc*. 2007; **2**: 2265-75.

124    Ahmadian A, Gharizadeh B, Gustafsson AC, Sterky F, Nyren P, Uhlen M, Lundeberg J. Single-nucleotide polymorphism analysis by pyrosequencing. *Anal Biochem*. 2000; **280**: 103-10.

125    Doostzadeh J, Shokralla S, Absalan F, Jalili R, Mohandessi S, Langston JW, Davis RW, Ronaghi M, Gharizadeh B. High throughput automated allele frequency estimation by pyrosequencing. *PLoS One*. 2008; **3**: e2693.

126    Koboldt DC, Ding L, Mardis ER, Wilson RK. Challenges of sequencing human genomes. *Brief Bioinform*. 2010; **11**: 484-98.

127    Pagani F, Buratti E, Stuani C, Bendix R, Dork T, Baralle FE. A new type of mutation causes a splicing defect in ATM. *Nat Genet*. 2002; **30**: 426-9.

128    Chillon M, Dork T, Casals T, Gimenez J, Fonknechten N, Will K, Ramos D, Nunes V, Estivill X. A novel donor splice site in intron 11 of the CFTR gene, created by mutation 1811+1.6kbA-->G, produces a new exon: high frequency in Spanish cystic fibrosis chromosomes and association with severe phenotype. *Am J Hum Genet*. 1995; **56**: 623-9.

129    Dehainault C, Michaux D, Pages-Berhouet S, Caux-Moncoutier V, Doz F, Desjardins L, Couturier J, Parent P, Stoppa-Lyonnet D, Gauthier-Villars M, Houdayer C. A deep intronic mutation in the RB1 gene leads to intronic sequence exonisation. *Eur J Hum Genet*. 2007; **15**: 473-7.

130    Raponi M, Upadhyaya M, Baralle D. Functional splicing assay shows a pathogenic intronic mutation in neurofibromatosis type 1 (NF1) due to intronic sequence exonization. *Hum Mutat*. 2006; **27**: 294-5.

131    Harland M, Mistry S, Bishop DT, Bishop JA. A deep intronic mutation in CDKN2A is associated with disease in a subset of melanoma pedigrees. *Hum Mol Genet*. 2001; **10**: 2679-86.

132    Anczukow O, Buisson M, Leone M, Coutanson C, Lasset C, Calender A, Sinilnikova OM, Mazoyer S. BRCA2 deep intronic mutation causing activation of a cryptic exon: opening toward a new preventive therapeutic strategy. *Clin Cancer Res*. 2012; **18**: 4903-9.

133    Corrigan A, Arenas M, Escuredo E, Fairbanks L, Marinaki A. HPRT deficiency: identification of twenty-four novel variants including an unusual deep intronic mutation. *Nucleosides Nucleotides Nucleic Acids*. 2011; **30**: 1260-5.

134    Vache C, Besnard T, le Berre P, Garcia-Garcia G, Baux D, Larrieu L, Abadie C, Blanchet C, Bolz HJ, Millan J, Hamel C, Malcolm S, Claustres M, Roux AF. Usher syndrome type 2 caused by activation of an USH2A pseudoexon: implications for diagnosis and therapy. *Hum Mutat*. 2012; **33**: 104-8.

135    Costa C, Pruliere-Escabasse V, de Becdelievre A, Gameiro C, Golmard L, Guittard C, Bassinet L, Bienvenu T, Georges MD, Epaud R, Bieth E, Giurgea I, Aissat A, Hinzpeter A, Costes B, Fanen P, Goossens M, Claustres M, Coste A, Girodon E. A recurrent deep-intronic splicing CF mutation emphasizes the importance of mRNA studies in clinical practice. *J Cyst Fibros*. 2011; **10**: 479-82.

136    Mancini C, Vaula G, Scalzitti L, Cavalieri S, Bertini E, Aiello C, Lucchini C, Gatti RA, Brussino A, Brusco A. Megalencephalic leukoencephalopathy with subcortical cysts type 1 (MLC1) due to a homozygous deep intronic splicing mutation (c.895-226T>G) abrogated in vitro using an antisense morpholino oligonucleotide. *Neurogenetics*. 2012.

137    Xu Q, Modrek B, Lee C. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res*. 2002; **30**: 3754-66.

138    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.

139    Rost S, Loffler S, Pavlova A, Muller CR, Oldenburg J. Detection of large duplications within the factor VIII gene by MLPA. *J Thromb Haemost*. 2008; **6**: 1996-9.

140    Lakich D, Kazazian HH, Jr., Antonarakis SE, Gitschier J. Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. *Nat Genet*, 1993/11/01 edn, 1993, 236-41.

141    Van de Water N, Williams R, Ockelford P, Browett P. A 20.7 kb deletion within the factor VIII gene associated with LINE-1 element insertion. *Thromb Haemost*. 1998; **79**: 938-42.

142    Chen JM, Chuzhanova N, Stenson PD, Ferec C, Cooper DN. Complex gene rearrangements caused by serial replication slippage. *Hum Mutat*. 2005; **26**: 125-34.

143    Sheen CR, Jewell UR, Morris CM, Brennan SO, Ferec C, George PM, Smith MP, Chen JM. Double complex mutations involving F8 and FUNDC2 caused by distinct break-induced replication. *Hum Mutat*. 2007; **28**: 1198-206.

144    Rossetti LC, Goodeve A, Larripa IB, De Brasi CD. Homeologous recombination between AluSx-sequences as a cause of hemophilia. *Hum Mutat*. 2004; **24**: 440.

145    Vidal F, Farssac E, Tusell J, Puig L, Gallardo D. First molecular characterization of an unequal homologous alu-mediated recombination event responsible for hemophilia. *Thromb Haemost*. 2002; **88**: 12-6.

146    Woods-Samuels P, Kazazian HH, Jr., Antonarakis SE. Nonhomologous recombination in the human genome: deletions in the human factor VIII gene. *Genomics*. 1991; **10**: 94-101.

147    Beckmann JS, Estivill X, Antonarakis SE. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nat Rev Genet*. 2007; **8**: 639-46. nrg2149 [pii]
10.1038/nrg2149.

148    Nelson DL, Aradhya S, Woffendin H, Bonnen P, Heiss NS, Yamagata T, Esposito T, Bardaro T, Poustka A, D'Urso M, Kenwrick S. Physical and genetic characterization reveals a pseudogene, an evolutionary junction, and unstable loci in distal Xq28. *Genomics*. 2002; **79**: 31-40. Doi 10.1006/Geno.2001.6680.

149    Tavassoli K, Eigel A, Horst J. A deletion/insertion leading to the generation of a direct repeat as a result of slipped mispairing and intragenic recombination in the factor VIII gene. *Hum Genet*. 1999; **104**: 435-7.

150    Santacroce R, Santoro R, Sessa F, Iannaccaro P, Sarno M, Longo V, Gallone A, Vecchione G, Muleo G, Margaglione M. Screening of mutations of hemophilia A in 40 Italian patients: a novel G-to-A mutation in intron 10 of the F8 gene as a putative cause of mild hemophilia A in southern Italy. *Blood Coagul Fibrinolysis*. 2008; **19**: 197-202.

151    Gau JP, Hsu HC, Chau WK, Ho CH. A novel splicing acceptor mutation of the factor VIII gene producing skipping of exon 25. *Ann Hematol*. 2003; **82**: 175-7.
10.1007/s00277-002-0592-y.

152    David D, Tavares A, Lavinha J. Characterization of a splicing mutation in the factor VIII gene at the RNA level. *Hum Genet*. 1995; **95**: 109-11.

153    Aradhya S, Bardaro T, Galgoczy P, Yamagata T, Esposito T, Patlan H, Ciccodicola A, Munnich A, Kenwrick S, Platzer M, D'Urso M, Nelson DL. Multiple pathogenic and benign genomic rearrangements occur at a 35 kb duplication involving the NEMO and LAGE2 genes. *Hum Mol Genet*. 2001; **10**: 2557-67.

154    Fusco F, D'Urso M, Miano MG, Ursini MV. The LCR at the IKBKG locus is prone to recombine. *Am J Hum Genet*. 2010; **86**: 650-2; author reply 2-3.

155    Fusco F, Pescatore A, Bal E, Ghoul A, Paciolla M, Lioi MB, D'Urso M, Rabia SH, Bodemer C, Bonnefont JP, Munnich A, Miano MG, Smahi A, Ursini MV. Alterations of the IKBKG locus and diseases: an update and a report of 13 novel mutations. *Hum Mutat*. 2008; **29**: 595-604.

156    Anderson L, Seilhamer J. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis*. 1997; **18**: 533-7.

157    Chen G, Gharib TG, Huang CC, Taylor JM, Misek DE, Kardia SL, Giordano TJ, Iannettoni MD, Orringer MB, Hanash SM, Beer DG. Discordant protein and mRNA expression in lung adenocarcinomas. *Mol Cell Proteomics*. 2002; **1**: 304-13.

158    Lichtinghagen R, Musholt PB, Lein M, Romer A, Rudolph B, Kristiansen G, Hauptmann S, Schnorr D, Loening SA, Jung K. Different mRNA and protein expression of matrix metalloproteinases 2 and 9 and tissue inhibitor of metalloproteinases 1 in benign and malignant prostate tissue. *Eur Urol*. 2002; **42**: 398-406.

159    Guo Y, Xiao P, Lei S, Deng F, Xiao GG, Liu Y, Chen X, Li L, Wu S, Chen Y, Jiang H, Tan L, Xie J, Zhu X, Liang S, Deng H. How is mRNA expression predictive for protein expression? A correlation study on human circulating monocytes. *Acta Biochim Biophys Sin (Shanghai)*. 2008; **40**: 426-36.

160    Pezeshkpoor B, Rost S, Oldenburg J, El-Maarri O. Identification of a third rearrangement at Xq28 that causes severe hemophilia A as a result of homologous recombination between inverted repeats. *J Thromb Haemost*. 2012; **10**: 1600-8.

161    Chatterjee S, Pal JK. Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. *Biol Cell*. 2009; **101**: 251-62. 10.1042/BC20080104
BC20080104 [pii].

162    Pickering BM, Willis AE. The implications of structured 5' untranslated regions on translation and disease. *Semin Cell Dev Biol*. 2005; **16**: 39-47. S1084-9521(04)00110-7 [pii]
10.1016/j.semcdb.2004.11.006.

163    Arrick BA, Lee AL, Grendell RL, Derynck R. Inhibition of translation of transforming growth factor-beta 3 mRNA by its 5' untranslated region. *Mol Cell Biol*. 1991; **11**: 4306-13.

164    Signori E, Bagni C, Papa S, Primerano B, Rinaldi M, Amaldi F, Fazio VM. A somatic mutation in the 5'UTR of BRCA1 gene in sporadic breast cancer causes down-modulation of translation efficiency. *Oncogene*. 2001; **20**: 4596-600. 10.1038/sj.onc.1204620.

165    Zhou W, Song W. Leaky scanning and reinitiation regulate BACE1 gene expression. *Mol Cell Biol*. 2006; **26**: 3353-64. 26/9/3353 [pii]
10.1128/MCB.26.9.3353-3364.2006.

166    Niesler B, Flohr T, Nothen MM, Fischer C, Rietschel M, Franzek E, Albus M, Propping P, Rappold GA. Association between the 5' UTR variant C178T of the serotonin receptor gene HTR3A and bipolar affective disorder. *Pharmacogenetics*. 2001; **11**: 471-5.

167    Hudder A, Werner R. Analysis of a Charcot-Marie-Tooth disease mutation reveals an essential internal ribosome entry site element in the connexin-32 gene. *J Biol Chem*. 2000; **275**: 34586-91. 10.1074/jbc.M005199200
M005199200 [pii].

168    Chappell SA, LeQuesne JP, Paulin FE, deSchoolmeester ML, Stoneley M, Soutar RL, Ralston SH, Helfrich MH, Willis AE. A mutation in the c-myc-IRES leads to enhanced internal ribosome entry in multiple myeloma: a novel mechanism of oncogene de-regulation. *Oncogene*. 2000; **19**: 4437-40. 10.1038/sj.onc.1203791.

169    Chappell SA, Edelman GM, Mauro VP. A 9-nt segment of a cellular mRNA can function as an internal ribosome entry site (IRES) and when present in linked multiple copies greatly enhances IRES activity. *Proc Natl Acad Sci U S A*. 2000; **97**: 1536-41. 97/4/1536 [pii].

170    Chiang PW, Carpenter LE, Hagerman PJ. The 5'-untranslated region of the FMR1 message facilitates translation by internal ribosome entry. *J Biol Chem*. 2001; **276**: 37916-21. 10.1074/jbc.M105584200
M105584200 [pii].

171    Tassone F, Hagerman RJ, Taylor AK, Hagerman PJ. A majority of fragile X males with methylated, full mutation alleles have significant levels of FMR1 messenger RNA. *J Med Genet*. 2001; **38**: 453-6.

172    Cazzola M. Role of ferritin and ferroportin genes in unexplained hyperferritinaemia. *Best Pract Res Clin Haematol*. 2005; **18**: 251-63. S1521-6926(04)00091-X [pii]
10.1016/j.beha.2004.08.025.

173    Conne B, Stutz A, Vassalli JD. The 3' untranslated region of messenger RNA: A molecular 'hotspot' for pathology? *Nat Med*. 2000; **6**: 637-41. 10.1038/76211.

174    Soifer HS, Rossi JJ, Saetrom P. MicroRNAs in disease and potential therapeutic applications. *Mol Ther*. 2007; **15**: 2070-9. 6300311 [pii] 10.1038/sj.mt.6300311.

175    Roush SF, Slack FJ. Micromanagement: a role for microRNAs in mRNA stability. *ACS Chem Biol*. 2006; **1**: 132-4. 10.1021/cb600138j.

176    Gu LH, Ichiki Y, Sato M, Kitajima Y. A novel nonsense mutation at E106 of the 2B rod domain of keratin 14 causes dominant epidermolysis bullosa simplex. *J Dermatol*. 2002; **29**: 136-45.

177    Chao LY, Huff V, Strong LC, Saunders GF. Mutation in the PAX6 gene in twenty patients with aniridia. *Hum Mutat*. 2000; **15**: 332-9. 10.1002/(SICI)1098-1004(200004)15:4<332::AID-HUMU5>3.0.CO;2-1 [pii] 10.1002/(SICI)1098-1004(200004)15:4<332::AID-HUMU5>3.0.CO;2-1.

178    Prior JF, Lim E, Lingam N, Raven JL, Finlayson J. A moderately severe alpha-thalassemia condition resulting from a combination of the alpha2 polyadenylation signal (AATAAA-->AATA- -) mutation and a 3.7 Kb alpha gene deletion in an Australian family. *Hemoglobin*. 2007; **31**: 173-7. 778207962 [pii] 10.1080/03630260701288997.

179    Higgs DR, Goodbourn SE, Lamb J, Clegg JB, Weatherall DJ, Proudfoot NJ. Alpha-thalassaemia caused by a polyadenylation signal mutation. *Nature*. 1983; **306**: 398-400.

180    Losekoot M, Fodde R, Harteveld CL, van Heeren H, Giordano PC, Went LN, Bernini LF. Homozygous beta+ thalassaemia owing to a mutation in the cleavage-polyadenylation sequence of the human beta globin gene. *J Med Genet*. 1991; **28**: 252-5.

181    Bennett CL, Brunkow ME, Ramsdell F, O'Briant KC, Zhu Q, Fuleihan RL, Shigeoka AO, Ochs HD, Chance PF. A rare polyadenylation signal mutation of the FOXP3 gene (AAUAAA-->AAUGAA) leads to the IPEX syndrome. *Immunogenetics*. 2001; **53**: 435-9. 10.1007/s002510100358.

182    Reamon-Buettner SM, Cho SH, Borlak J. Mutations in the 3'-untranslated region of GATA4 as molecular hotspots for congenital heart disease (CHD). *BMC Med Genet*. 2007; **8**: 38. 1471-2350-8-38 [pii] 10.1186/1471-2350-8-38.

183    Beffagna G, Occhi G, Nava A, Vitiello L, Ditadi A, Basso C, Bauce B, Carraro G, Thiene G, Towbin JA, Danieli GA, Rampazzo A. Regulatory mutations in transforming growth factor-beta3 gene cause arrhythmogenic right ventricular cardiomyopathy type 1. *Cardiovasc Res*. 2005; **65**: 366-73. S0008-6363(04)00440-7 [pii] 10.1016/j.cardiores.2004.10.005.

184    Riccardi F, Rivolta GF, Franchini M, Pattacini C, Neri TM, Tagliaferri A. Characterization of a novel mutation in the F8 promoter region associated with mild hemophilia A and resistance to DDAVP therapy. *J Thromb Haemost*. 2009; **7**: 1234-5. 10.1111/j.1538-7836.2009.03468.x JTH3468 [pii].

185    Dai L, Cutler JA, Savidge GF, Mitchell MJ. Characterization of a causative mutation of hemophilia A identified in the promoter region of the factor VIII gene (F8). *J Thromb Haemost*. 2008; **6**: 193-5. JTH2806 [pii] 10.1111/j.1538-7836.2007.02806.x.

186    Figueiredo MS, Brownlee GG. cis-acting elements and transcription factors involved in the promoter activity of the human factor VIII gene. *J Biol Chem*. 1995; **270**: 11828-38.

187    Zimmermann MA, Meier D, Oldenburg J, Muller CR, Rost S. Identification and characterization of mutations in the promoter region of the factor VIII gene. *J Thromb Haemost*. 2012; **10**: 314-7. 10.1111/j.1538-7836.2011.04574.x.

188    Kozak M. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc Natl Acad Sci U S A*. 1986; **83**: 2850-4.

189    Black DL. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem*. 2003; **72**: 291-336. 10.1146/annurev.biochem.72.121801.161720 121801.161720 [pii].

190    Srebrow A, Kornblihtt AR. The connection between splicing and cancer. *J Cell Sci*. 2006; **119**: 2635-41. 119/13/2635 [pii] 10.1242/jcs.03053.

191    Patel AA, Steitz JA. Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol*. 2003; **4**: 960-70. 10.1038/nrm1259 nrm1259 [pii].

192    Berget SM, Moore C, Sharp PA. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A*. 1977; **74**: 3171-5.

193    Early P, Rogers J, Davis M, Calame K, Bond M, Wall R, Hood L. Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA processing pathways. *Cell*. 1980; **20**: 313-9. 0092-8674(80)90617-0 [pii].

194    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. 6s0185 [pii].

195    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. gkp215 [pii] 10.1093/nar/gkp215.

196    Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol*. 1997; **4**: 311-23.

197    Zimmermann MA, Gehrig A, Oldenburg J, Muller CR, Rost S. Analysis of F8 mRNA in haemophilia A patients with silent mutations or presumptive splice site mutations. *Haemophilia*. 2013; **19**: 310-7. 10.1111/hae.12039.

198    Bakker E, Van Broeckhoven C, Bonten EJ, van de Vooren MJ, Veenema H, Van Hul W, Van Ommen GJ, Vandenberghe A, Pearson PL. Germline mosaicism and Duchenne muscular dystrophy mutations. *Nature*. 1987; **329**: 554-6.

199    Higuchi M, Kochhan L, Olek K. A somatic mosaic for haemophilia A detected at the DNA level. *Mol Biol Med*. 1988; **5**: 23-7.

200    Maddalena A, Sosnoski DM, Berry GT, Nussbaum RL. Mosaicism for an intragenic deletion in a boy with mild ornithine transcarbamylase deficiency. *N Engl J Med*. 1988; **319**: 999-1003.

201    Oldenburg J, Rost S, El-Maarri O, Leuer M, Olek K, Muller CR, Schwaab R. De novo factor VIII gene intron 22 inversion in a female carrier presents as a somatic mosaicism. *Blood*. 2000; **96**: 2905-6.

202    Abdrakhmanova A, Zwicker K, Kerscher S, Zickermann V, Brandt U. Tight binding of NADPH to the 39-kDa subunit of complex I is not required for catalytic activity but stabilizes the multiprotein complex. *Biochim Biophys Acta*. 2006; **1757**: 1676-82. S0005-2728(06)00277-5 [pii] 10.1016/j.bbabio.2006.09.003.

**Supplementary tables:**

Table S 1: List of polymorphisms found in *vWF* in hemophilia A patients without detectable mutation.

nd= no data, Ex= Exon, Int= Intron, W=T/A, R=G/A, M=C/A, Y=C/T, K=G/T, S=C/G

| Nr. | Exon/ Intron | Nuclotide change | rs number | aa change | HA#2 | HA#3 | HA#4 | HA#5 | HA#6 | HA#7 | HA#8 | HA#9 | HA#10 | HA#11 | HA#12 | HA#13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Ex 8 | T954A | 1800387 | N318K | T | T | T | T | T | T | W | T | T | T | T | T |
| 2 | Ex 11 | A1173T | 1800375 | T391 | W | W | W | A | A | A | A | A | A | A | A | A |
| 3 | Ex 12 | G1411A | 1800377 | V471I | R | R | R | G | G | R | R | G | G | G | G | G |
| 4 | Ex 13 | G1451A | 1800378 | R484H | G | G | G | G | G | R | G | R | R | R | G | R |
| 5 | Ex 14 | T1548C | 7312411 | P538 | T | T | T | T | T | C | T | C | T | C | T | T |
| 6 | Ex 14 | G1626A | 35365059 | A542 | G | G | G | G | G | G | G | G | G | R | G | G |
| 7 | Int 17 | c.2282-42 | 21693 | C>A | M | C | C | A | A | C | M | C | A | M | A | A |
| 8 | Ex 18 | A2365G | 1063856 | T789A | R | R | R | A | A | G | A | G | A | R | A | A |
| 9 | Ex 18 | T2385C | 1064008 | Y795 | Y | Y | Y | T | T | C | T | C | T | Y | T | T |
| 10 | Int18 | c.2442+38 | 77372340 | G>T | T | G | G | G | G | K | G | G | K | G | G | G |
| 11 | Int 19 | c.2547- 146 | 12810426 | C>T | Y | C | Y | T | T | C | Y | C | T | Y | T | T |
| 12 | Ex 20 | A2555G | 216321 | R852Q | G | R | R | G | G | G | R | G | G | G | G | G |
| 13 | Ex 21 | G2771A | 33978901 | R924Q | G | G | G | R | R | G | G | G | G | G | G | G |
| 14 | Ex 22 | G2880A | 1800380 | R960R | R | R | G | G | G | G | G | A | G | R | G | G |
| 15 | Ex 28 | A4141G | 216311 | A1381T | G | A | A | G | G | A | G | A | A | G | G | *nd* |
| 16 | Ex 28 | G4414C | 1800383 | D1472H | G | G | G | G | G | G | G | G | S | G | S | G |
| 17 | Ex 28 | T4641C | 216310 | T1547 | Y | T | C | C | C | T | Y | T | Y | Y | C | C |
| 18 | Ex 28 | A4665C | 1800384 | A1555 | A | A | A | A | M | A | A | A | A | A | A | M |
| 19 | Ex 28 | G4693T | 1800385 | V1565L | G | G | G | G | K | G | G | G | G | G | G | K |
| 20 | Ex 39 | A6846G | 10849371 | T2282 | A | A | A | A | A | R | A | A | A | A | A | A |
| 21 | Ex 42 | C7239T | 216867 | T2413 | T | T | Y | Y | T | *nd* | Y | C | Y | *nd* | *nd* | C |
| 22 | Ex 45 | T7682A | 35335161 | F2561 | T | T | T | T | T | W | T | W | T | W | T | W |
| 23 | Int 50 | c.8155+50 | 2270151 | C>T | C | C | C | C | C | Y | C | Y | Y | Y | C | T |

Table S 2: Summary of results obtained after NGS. The number of total reads and the mapped reads are shown.

The average coverage was between 287-554. The total rate of nucleotide mismatches for all patients was 2%. In total 3.126 variants were identified, from which 1510 and 1.616 were SNPs and Indels, respectively (Indel= Insertions and deletions).

| Samples | Nr. of reads | Nr. of mapped reads | Gaps (bp) | Total bases (bp) | Average Coverage | Nucleotide mismatches % | Total variants | SNPs | Indels |
|---|---|---|---|---|---|---|---|---|---|
| HA#2 | 2.036.106 | 1.912.611 | 3.302 | 71.263.710 | 346 | 0,16 | 302 | 141 | 161 |
| HA#3 | 3.391.020 | 3.066.420 | 3.787 | 118.685.700 | 554 | 0,19 | 366 | 140 | 226 |
| HA#6 | 1.903.566 | 1.744.525 | 5.215 | 66.624.810 | 315 | 0,16 | 323 | 176 | 147 |
| HA#8 | 4.083.402 | 3.829.590 | 4.501 | 142.919.070 | 692 | 0,28 | 541 | 328 | 213 |
| HA#9 | 2.681.898 | 2.490.669 | 2.473 | 93.866.430 | 450 | 0,15 | 285 | 133 | 152 |
| HA#10 | 2.621.230 | 2.299.893 | 7.580 | 91.743.050 | 416 | 0,22 | 407 | 167 | 240 |
| HA#11 | 1.897.122 | 1.753.722 | 5.149 | 66.399.270 | 317 | 0,14 | 274 | 116 | 158 |
| HA#12 | 2.294.748 | 1.849.058 | 10.329 | 80.316.180 | 334 | 0,21 | 394 | 209 | 185 |
| HA#13 | 1.735.616 | 1.534.895 | 5.148 | 60.746.560 | 278 | 0,12 | 234 | 100 | 134 |
| Total | 22.644.708 | 20.481.383 | 47.484 | 792.564.780 | 3.702 | 2 | 3.126 | 1.510 | 1.616 |

Table S 3: Distribution of known SNPS associated with FVIII:C in healthy controls.

MAF= Minor allele frequency, 1C-96C=healthy controls, R=A/G, Y=C/T, K=G/T, nd= no data available

| SNP Nr. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Gene | *F8* | *F8* | *F8* | *STAB2* | *SCARA5* | *STBXP5* | *AB0* | *vWF* |
| rs number | rs7058826 | rs1800291 | rs1800292 | rs12229292 | rs9644133 | rs9390459 | rs687289 | rs1063856 |
| Chromosome | Xq28 | Xq28 | Xq28 | 12q23 | 8p21 | 6q24 | 9q34 | 12p13 |
| Location | NM_000132.3 c.1010-27 | NM_000132.3 c.3780 | NM_000132.3 c.3864 | NM_017564.9 c.7248+582 | NM_173833.5 c.241+9448 | NM_139244.4 c.2337 | NM_020469.2 c.99-329 | NM_000552.3 c.2365 |
| MAF | A=0.0928 | C=0.2459 | G=0.1049 | T=0.2729 | T=0.2418 | G=0.4826 | A=0.3668 | G=0.3429 |
| (hg:19) | 154194989 | 154158285 | 154158201 | 104153633 | 27814483 | 147680359 | 136137106 | 6153534 |
| 1C | A | G | A | *nd* | C | R | R | R |
| 2C | G | C | A | G | Y | A | G | R |
| 3C | G | C | A | G | C | G | A | R |
| 4C | G | C | A | G | C | R | R | G |
| 5C | G | G | A | G | Y | A | G | R |
| 6C | A | G | A | G | Y | G | R | A |
| 7C | G | C | C | G | C | R | R | A |
| 8C | G | C | C | *nd* | C | A | R | A |
| 9C | G | C | A | *nd* | C | R | R | R |
| 10C | G | C | A | K | C | R | R | A |
| 11C | G | C | C | G | C | A | R | A |
| 12C | G | C | A | G | Y | R | R | A |
| 13C | A | G | A | K | C | R | G | A |
| 14C | G | C | A | G | Y | R | G | A |
| 15C | G | C | A | K | Y | R | A | *nd* |
| 16C | G | C | C | *nd* | C | R | G | R |
| 17C | G | C | A | T | T | R | *nd* | A |
| 18C | A | G | A | K | Y | R | R | G |
| 19C | G | C | A | *nd* | C | R | A | R |
| 20C | G | C | A | *nd* | Y | R | R | R |
| 21C | G | C | A | K | C | A | A | R |
| 22C | G | C | A | K | C | A | R | A |
| 24C | G | C | A | G | C | R | G | A |
| 25C | G | C | A | *nd* | *nd* | *nd* | *nd* | *nd* |
| 26C | G | C | A | *nd* | *nd* | *nd* | *nd* | *nd* |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 27C | G | C | A | K | C | A | R | A |
| 28C | G | C | A | K | Y | R | R | A |
| 29C | A | G | A | *nd* | *nd* | *nd* | *nd* | *nd* |
| 30C | A | G | A | K | C | R | R | R |
| 31C | G | C | A | *nd* | C | R | R | *nd* |
| 32C | G | C | C | K | *nd* | R | G | A |
| 33C | G | C | C | K | *nd* | R | R | A |
| 34C | G | C | A | G | *nd* | R | G | G |
| 35C | G | C | A | K | C | R | A | R |
| 36C | A | G | A | K | C | G | G | A |
| 37C | G | C | A | K | Y | G | G | A |
| 38C | G | C | A | G | Y | G | G | G |
| 39C | G | C | A | K | *nd* | G | R | R |
| 40C | G | C | A | K | T | R | R | R |
| 41C | G | C | A | K | C | *nd* | R | R |
| 42C | G | C | A | T | C | *nd* | A | A |
| 43C | G | C | A | K | Y | R | R | A |
| 44C | G | C | A | G | Y | *nd* | G | A |
| 45C | G | C | A | *nd* | C | A | G | A |
| 46C | G | C | A | K | T | A | R | G |
| 47C | G | C | A | G | C | G | R | A |
| 48C | G | C | A | G | Y | G | A | R |
| 49C | G | C | A | K | Y | G | G | R |
| 50C | G | C | A | G | C | G | A | A |
| 51C | A | G | A | G | C | R | G | R |
| 52C | A | G | A | *nd* | Y | G | R | R |
| 53C | A | G | A | G | C | G | R | A |
| 54C | G | C | A | G | Y | G | R | A |
| 55C | G | C | A | *nd* | Y | A | G | A |
| 56C | A | G | A | T | Y | A | A | A |
| 57C | G | C | C | K | Y | A | G | R |
| 58C | G | G | A | G | C | R | R | R |
| 59C | A | G | A | T | C | A | R | G |
| 60C | G | G | A | K | C | G | R | A |
| 61C | G | C | C | T | Y | G | R | A |
| 62C | G | C | A | G | Y | G | G | R |
| 63C | G | C | A | G | C | G | R | A |
| 64C | G | C | C | G | Y | R | R | *nd* |
| **SNP Nr.** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |

| 65C | G | C | C | K | C | A | A | A |
|---|---|---|---|---|---|---|---|---|
| 66C | A | G | A | K | Y | G | R | R |
| 67C | G | C | A | G | Y | R | R | *nd* |
| 68C | G | C | A | G | C | R | R | R |
| 69C | A | G | A | G | C | A | R | R |
| 70C | G | C | A | *nd* | C | G | R | A |
| 71C | G | C | A | G | C | *nd* | A | A |
| 72C | G | C | A | K | Y | A | R | A |
| 73C | A | G | A | G | C | R | R | R |
| 74C | G | C | A | G | C | R | A | A |
| 75C | G | C | A | G | C | R | *nd* | R |
| 76C | A | G | A | G | C | A | R | R |
| 77C | A | G | A | G | C | R | G | *nd* |
| 78C | A | G | A | K | C | A | R | R |
| 79C | G | C | A | G | T | G | G | A |
| 80C | G | C | A | K | Y | G | G | A |
| 81C | G | C | A | K | C | G | G | R |
| 82C | G | C | A | K | C | R | R | A |
| 83C | G | C | A | T | C | A | R | R |
| 84C | G | C | A | G | Y | R | R | A |
| 85C | G | G | A | K | C | R | G | A |
| 86C | G | C | A | K | Y | G | G | A |
| 87C | G | C | A | K | C | R | G | R |
| 88C | G | C | A | K | C | G | G | A |
| 89C | G | C | A | *nd* | Y | R | R | A |
| 90C | G | C | C | G | C | R | G | *nd* |
| 91C | G | C | A | *nd* | Y | G | G | R |
| 92C | G | C | A | G | C | A | A | A |
| 93C | G | C | A | *nd* | C | R | R | A |
| 94C | G | G | A | K | C | R | R | R |
| 95C | A | G | A | G | C | G | A | G |
| 96C | G | C | A | *nd* | C | A | R | R |

Table S 4: Association of studied SNPs of Table S 3 with FVIII:C, FVIII:Ag, *F8* mRNA expression, vWF activity and vWF:Ag levels in healthy individuals.

P values are calculated using the Krustal-Wallis test for the autosomal loci and Mann Whitney test for the X-lined loci, exp.=expression.

| SNP Nr. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Gene | *F8* | *F8* | *F8* | *STAB2* | *SCARA5* | *STBXP5* | *AB0* | *vWF* |
| rs number | rs7058826 | rs1800291 | rs1800292 | rs12229292 | rs9644133 | rs9390459 | rs687289 | rs1063856 |
| Chromosome | Xq28 | Xq28 | Xq28 | 12q23 | 8p21 | 6q24 | 9q34 | 12p13 |
| Location | NM_000132.3 c.1010-27 | NM_000132.3 c.3780 | NM_000132.3 c.3864 | NM_017564.9 c.7248+582 | NM_173833.5 c.241+9448 | NM_139244.4 c.2337 | NM_020469.2 c.99-329 | NM_000552.3 c.2365 |
| MAF | A=0.0928 | C=0.2459 | G=0.1049 | T=0.2729 | T=0.2418 | G=0.4826 | A=0.3668 | G=0.3429 |
| (hg:19) | 154194989 | 154158285 | 154158201 | 104153633 | 27814483 | 147680359 | 136137106 | 6153534 |
| *F8* mRNA exp. | 0,5954 | 0,4079 | 0,6289 | - | - | - | - | - |
| FVIII:C | 0,293 | 0,703 | 0,0627 | 0,6593 | 0,4005 | 0,5736 | **0,0004** | 0,4498 |
| FVIII:Ag | 0,3996 | 0,8572 | 0,5765 | **0,0592** | 0,2352 | 0,4553 | **0,0022** | 0,4724 |
| vWF activity | 0,6754 | 0,7063 | 0,058 | 0,7571 | 0,7987 | 0,4292 | **< 0,0001** | 0,3582 |
| vWF:Ag | 0,9148 | 0,5458 | 0,1097 | 0,948 | 0,9116 | 0,4624 | **0,0002** | 0,2441 |

Table S 5: Association of known SNPs associated with FVIII:C levels in patients without mutation.

MAF= Minor allele frequency, 1C-96C=healthy controls, R=A/G, Y=C/T, K=G/T

| SNP Nr. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---------|---|---|---|---|---|---|---|---|
| Gene | *F8* | *F8* | *F8* | *STAB2* | *SCARA5* | *STBXP5* | *AB0* | *vWF* |
| rs number | rs7058826 | rs1800291 | rs1800292 | rs12229292 | rs9644133 | rs9390459 | rs687289 | rs1063856 |
| Chromosome | Xq28 | Xq28 | Xq28 | 12q23 | 8p21 | 6q24 | 9q34 | 12p13 |
| Location | NM_000132.3 c.1010-27 | NM_000132.3 c.3780 | NM_000132.3 c.3864 | NM_017564.9 c.7248+582 | NM_173833.5 c.241+9448 | NM_139244.4 c.2337 | NM_020469.2 c.99-329 | NM_000552.3 c.2365 |
| MAF | A=0.0928 | C=0.2459 | G=0.1049 | T=0.2729 | T=0.2418 | G=0.4826 | A=0.3668 | G=0.3429 |
| (hg:19) | 154194989 | 154158285 | 154158201 | 104153633 | 27814483 | 147680359 | 136137106 | 6153534 |
| HA#2 | G | C | A | *nd* | C | R | G | A |
| HA#5 | G | C | A | G | C | *nd* | *nd* | A |
| HA#6 | G | C | A | G | Y | R | R | A |
| HA#8 | G | C | A | K | Y | R | R | A |
| HA#9 | G | C | A | G | T | *nd* | R | G |
| HA#10 | G | C | A | G | C | R | G | R |
| HA#11 | G | C | A | K | C | R | G | R |
| HA#12 | G | C | A | G | C | G | A | A |
| HA#13 | G | C | A | K | Y | R | G | A |

Table S 6: Analyzing the distribution of SNPs found in patients after NGS in healthy controls

| SNP | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Intron | 5'UTR | 1 | 1 | I1 | 4 | 4 | 4 | I6 | 6 | 6 | 6 | 6 | I7 | 10 | 10 | 12 | 13 | 13 | 13 | 14 | 14 | I14 | 14 | 22 | 22 | 22 | 22 | 22 | 22 | 22 | 25 | 25 | 25 | 25 | 25 | 3'UTR |
| rs Nr. | rs6649625 | rs55807428 | rs14170586 | rs141897310 | rs5945269 | rs73641115 | rs144231135 | rs5945128 | rs28857481 | rs3861554 | rs201280173 | rs59835535 | rs149954517 | rs76063559 | rs190639729 | rs150203712 | rs66643622 | rs34552198 | rs78362479 | rs113224419 | rs66643714 | rs28845018 | rs147346816 | rs9305 | rs113683431 | rs111525585 | rs2292089 | rs28810103 | rs78327897 | rs28835013 | rs112922881 | rs285360050 | rs66443711 | rs3076842 | rs1509787 | rs1050705 |
| NG_11403.1 g. | 7716 | 15614 | 17323 | 25961 | 36681 | 37752 | 37942 | 45117 | 45260 | 46841 | 49012:49014 | 55023 | 60195 | 68425 | 69579 | 74668 | 82469 | 88822:88823 | 90648 | 103221 | 106551 | 113955 | 117255 | 140194 | 142941 | 141745 | 142472 | 148356 | 161095 | 161935 | 168273 | 174645 | 177957 | 186709Ddel TT | 187365 | 191799 |
| MAF | T=0,247 | C=0,084 | A=0,210 | C=0,021 | C=0,1802 | C=0,1519 | T=0,0084 | C=0,223 | T=0,3924 | G=0,4696 | -=0,024 | A=0,0476 | A=0,0072 | G=0,0874 | C=0,0018 | G=0,0205 | C=0,4063 | rare | G=0,0886 | G=0,0995 | A=0,4907 | A=0,4756 | A=0,0241 | rare | rare | rare | rare | A=0,4834 | A=0,088 | C=0,4943 | A=0,0928 | T=0,495 | A=0,4623 | -=0,0187 | T=0,1248 | G=0,3743 |
| Chromosomal location (hg:19) | 154248283 | 154240385 | 154238676 | 154230038 | 154219318 | 154218247 | 154218057 | 154210882 | 154210739 | 154209158 | 154211329 | 154200976 | 154195804 | 154187574 | 154186420 | 154181331 | 154173530 | 154167176:77 | 154165351 | 154152778 | 154149448 | 154142044 | 154138744 | 154115804 | 154114300 | 154114508 | 154113527 | 154107643 | 154094904 | 154094064 | 154087726 | 154081354 | 154078042 | 154069291 | 154068634 | 154064200 |
| 1C | C | A | A | A | T | T | G | T | T | G | - | A | G | nd | A | T | nd | TA | A | A | G | G | G | C | G | C | A | G | G | T | G | C | A | nd | T | G |
| 2C | C | A | G | A | T | T | nd | T | T | A | - | A | G | A | A | nd | A | - | A | A | A | A | G | nd | G | T | A | A | G | C | G | T | G | T | C | A |
| 3C | C | A | G | A | T | T | G | T | A | A | - | A | G | nd | A | T | C | - | A | A | A | A | G | nd | G | C | A | A | G | C | G | T | G | T | C | A |
| 4C | C | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | C | - | A | A | A | A | G | nd | G | T | A | A | G | C | G | T | G | T | C | A |
| 5C | C | A | G | A | C | T | G | T | T | G | - | A | G | A | A | T | C | TA | A | A | G | G | G | C | G | C | A | G | G | T | G | C | A | T | C | G |

| SNP | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 6C | C | A | A | A | T | T | G | T | T | G | - | A | G | A | A | T | C | TA | A | A | nd | G | G | C | G | C | A | G | G | T | G | C | A | nd | T | nd |
| 7C | C | A | A | A | T | T | G | C | T | G | - | A | G | A | A | T | C | - | A | A | G | G | G | nd | G | T | A | G | G | T | G | C | A | nd | C | G |
| 8C | C | A | A | A | C | C | G | C | nd | G | - | A | G | A | A | T | C | - | A | A | G | G | nd | nd | G | T | A | G | nd | T | G | C | A | T | C | G |
| 9C | C | nd | G | A | T | T | G | C | A | A | - | A | G | A | A | T | A | - | A | A | G | G | G | C | C | C | A | G | G | T | G | C | G | T | C | nd |
| 10C | C | C | G | A | C | C | G | C | T | G | - | A | G | G | A | T | A | TA | G | G | G | G | G | nd | G | C | A | G | G | T | A | C | A | T | C | A |
| 11C | nd | A | A | A | T | T | G | T | T | G | - | A | G | A | A | T | C | - | A | nd | G | G | G | C | G | C | A | G | G | T | G | C | A | T | C | nd |
| 12C | C | A | G | A | T | T | G | C | A | A | - | A | G | A | A | T | A | - | A | A | G | G | G | C | C | C | A | G | G | T | G | C | G | T | C | A |
| 13C | C | A | A | A | T | T | G | T | nd | G | - | A | G | A | A | T | C | TA | A | A | G | G | G | C | G | C | A | G | G | T | G | C | A | T | T | G |
| 14C | T | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | A | - | A | nd | A | A | G | A | G | C | G | A | G | C | G | T | G | T | C | A |
| 15C | T | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | A | - | A | A | A | A | nd | A | G | C | A | A | G | C | G | T | G | - | C | A |
| 16C | C | A | A | A | C | C | G | C | T | G | - | A | G | A | A | T | C | - | A | A | G | G | G | C | nd | C | A | G | G | T | G | C | A | T | C | G |
| 17C | T | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | A | - | A | A | A | A | G | A | G | C | A | A | G | C | G | T | G | T | C | A |
| 18C | C | A | A | A | T | T | G | T | T | G | - | A | G | A | A | T | C | TA | A | A | G | G | G | C | G | C | A | G | G | T | G | C | A | T | T | G |
| 19C | T | A | G | A | T | T | T | T | A | A | - | A | G | A | A | T | A | - | A | A | A | A | G | A | G | C | A | A | G | C | G | T | G | T | C | A |
| 20C | T | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | A | - | A | A | A | A | G | A | G | T | A | A | G | C | G | T | G | T | C | A |
| 21C | T | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | A | - | A | A | A | A | G | A | G | C | A | A | G | C | G | T | G | T | C | A |
| 22C | C | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | C | - | A | A | A | A | nd | nd | G | T | A | A | G | C | G | T | G | T | C | A |
| 24C | T | A | nd | A | nd | T | T | T | A | A | - | A | G | A | A | T | A | - | A | A | A | A | nd | C | G | C | A | A | G | C | G | T | G | T | C | A |
| 25C | C | C | G | A | C | C | G | T | T | G | - | A | G | nd | A | G | A | TA | nd | nd | G | G | G | nd | G | nd | G | G | A | nd | G | C | A | T | C | nd |
| 26C | nd | A | A | A | T | T | G | T | T | G | - | A | G | nd | A | nd | C | - | A | nd | G | G | G | nd | G | nd | A | G | G | T | G | C | A | T | C | G |
| 27C | T | A | G | A | T | T | G | T | nd | A | - | A | G | A | A | nd | A | - | A | A | A | A | G | A | G | C | A | A | G | C | G | T | G | T | C | A |
| 28C | T | A | G | A | T | T | G | T | A | A | - | A | G | nd | A | nd | A | - | A | A | A | A | G | A | G | T | A | A | G | C | G | T | G | T | C | A |

| SNP | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 29C | T | *nd* | *nd* | A | T | T | G | T | T | G | - | A | G | *nd* | **nd** | *nd* | C | TA | A | *nd* | G | G | G | *nd* | G | *nd* | A | G | *nd* | T | G | C | A | *nd* | T | G |
| 30C | T | A | A | A | T | T | G | T | T | G | - | A | G | A | A | T | C | TA | A | A | G | G | G | C | G | C | A | G | G | *nd* | G | C | A | T | T | G |
| 31C | C | *nd* | *nd* | A | C | C | G | T | T | G | - | A | G | A | A | T | A | TA | G | A | G | G | G | A | G | C | A | G | *nd* | T | A | *nd* | A | T | C | A |
| 32C | C | A | A | A | T | T | G | T | T | G | - | A | G | A | A | *nd* | C | - | A | A | G | G | G | C | G | C | A | G | G | T | G | C | A | T | C | G |
| 33C | C | A | A | A | T | T | G | T | T | G | - | A | G | A | A | T | C | - | *nd* | A | G | G | *nd* | A | G | *nd* | *nd* | G | G | T | G | C | A | T | C | G |
| 34C | C | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | C | - | A | A | A | A | G | A | G | C | A | A | G | C | G | T | G | T | C | A |
| 35C | C | C | G | G | C | C | G | C | T | G | - | A | G | G | A | G | A | TA | G | G | G | G | *nd* | C | G | T | A | G | A | T | A | C | A | T | C | A |
| 36C | C | C | G | G | C | C | G | C | T | G | - | A | G | A | A | T | C | TA | A | A | G | G | G | C | G | C | A | G | G | T | G | C | A | T | T | G |
| 37C | T | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | A | - | A | A | A | A | G | A | G | T | A | A | G | C | G | T | G | T | C | A |
| 38C | C | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | C | - | A | A | A | A | G | *nd* | G | T | A | A | G | C | G | T | G | T | C | A |
| 39C | C | A | G | A | T | T | G | C | A | A | - | A | G | A | A | T | A | - | A | A | G | G | G | C | C | C | A | G | G | T | G | C | G | *nd* | C | A |
| 40C | C | C | G | A | C | *nd* | G | C | A | G | - | A | G | G | A | T | A | TA | G | G | G | G | G | A | G | C | A | G | A | T | A | C | A | T | C | A |
| 41C | T | A | G | A | T | T | G | T | A | A | - | A | G | A | A | *nd* | A | - | A | A | A | A | G | A | G | C | A | A | G | C | G | T | G | T | C | A |
| 42C | T | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | A | - | A | A | A | A | G | A | G | C | A | A | G | C | G | T | G | T | C | A |
| 43C | T | A | G | A | T | T | G | T | A | A | - | A | G | A | A | *nd* | A | - | A | A | A | A | G | A | G | *nd* | A | A | G | C | G | T | G | *nd* | C | A |
| 44C | C | A | G | A | T | T | G | T | A | A | - | A | G | A | A | *nd* | A | - | A | A | A | A | G | A | G | T | G | A | G | C | G | T | G | T | C | A |
| 45C | C | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | A | - | A | A | A | G | G | A | G | *nd* | *nd* | A | G | C | G | T | G | T | C | A |
| 46C | C | A | G | A | T | T | G | T | T | A | - | A | G | A | A | T | A | - | A | A | G | G | G | C | C | C | A | G | G | T | G | C | G | T | C | A |
| 47C | *nd* | C | G | G | C | C | G | C | *nd* | G | - | A | G | A | A | *nd* | A | TA | G | G | G | G | A | C | C | T | A | G | A | T | A | C | A | T | C | A |
| 48C | C | A | G | A | T | T | G | T | *nd* | A | - | A | G | A | A | T | A | - | A | A | A | A | G | A | G | C | A | A | G | C | G | T | G | T | C | A |
| 49C | C | A | G | A | T | T | G | T | T | A | - | A | G | A | A | T | A | - | A | A | A | A | G | A | G | *nd* | A | A | G | C | A | T | G | T | C | *nd* |
| 50C | C | C | G | A | C | C | G | C | T | G | - | A | G | G | A | T | A | TA | G | G | G | G | G | A | G | C | A | G | A | *nd* | G | C | A | T | C | *nd* |
| 51C | T | A | A | A | T | T | G | T | T | G | - | A | G | A | A | T | C | TA | A | A | G | G | G | C | G | C | G | G | G | T | G | C | A | T | T | G |

| SNP | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 |
|-----|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 52C | C | A | A | A | T | T | G | T | T | G | - | A | G | A | *nd* | T | C | TA | A | A | G | G | G | C | G | C | G | G | G | T | G | C | A | T | T | G |
| 53C | C | A | A | A | T | T | G | T | T | G | - | A | G | A | A | T | C | TA | A | A | G | G | G | C | G | C | A | G | G | T | G | C | A | T | T | G |
| 54C | T | A | A | A | T | T | G | T | A | G | - | A | G | G | *nd* | G | A | TA | G | G | G | G | A | A | G | T | A | G | A | T | A | C | A | T | C | A |
| 55C | C | C | G | G | C | C | G | C | T | G | - | A | G | G | A | G | A | TA | G | G | G | G | A | C | G | T | A | G | A | T | A | C | A | T | C | A |
| 56C | C | A | A | A | T | T | G | T | *nd* | G | - | A | G | A | A | T | C | TA | A | A | G | G | G | C | G | C | G | G | G | T | G | C | A | T | *nd* | G |
| 57C | C | A | A | A | T | T | G | T | *nd* | G | - | A | G | A | A | T | C | - | A | A | G | G | G | C | G | C | G | G | G | T | G | C | A | T | C | G |
| 58C | C | A | G | A | T | T | G | T | A | G | - | A | G | A | A | T | C | - | A | A | G | G | G | A | G | T | A | G | G | T | G | C | A | T | C | G |
| 59C | C | A | A | A | T | T | G | T | T | G | - | A | G | A | A | T | C | TA | A | A | G | G | G | C | G | C | A | G | G | T | G | C | A | T | T | *nd* |
| 60C | C | A | G | A | T | T | G | T | A | G | - | A | G | A | A | T | C | - | A | A | G | G | G | A | G | C | A | G | G | T | G | C | A | T | C | G |
| 61C | C | A | G | A | T | T | G | T | T | G | - | A | G | A | A | T | C | - | A | A | G | G | G | A | G | C | G | G | G | T | G | C | A | T | C | *nd* |
| 62C | T | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | A | - | A | A | A | A | G | A | G | C | G | A | G | C | G | T | G | T | C | A |
| 63C | T | A | G | A | T | T | G | T | *nd* | R | - | A | G | A | A | T | *nd* | - | A | A | A | G | G | A | G | T | G | *nd* | G | *nd* | G | *nd* | *nd* | T | C | *nd* |
| 64C | C | A | G | A | C | C | G | C | *nd* | G | - | A | G | A | A | T | C | - | A | A | G | G | G | A | G | T | G | G | G | T | G | C | A | T | C | *nd* |
| 65C | C | A | A | A | C | C | G | T | T | G | - | A | G | A | A | T | C | - | A | A | G | G | G | C | G | C | A | G | G | *nd* | G | C | A | T | C | *nd* |
| 66C | C | A | A | A | T | T | G | T | *nd* | G | - | A | G | A | A | *nd* | C | TA | A | A | G | G | G | C | G | C | A | G | G | T | G | C | A | T | T | *nd* |
| 67C | T | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | A | - | A | A | A | A | G | A | G | *nd* | A | A | G | C | G | T | G | T | C | A |
| 68C | T | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | A | - | A | A | A | A | G | A | G | C | A | A | G | C | G | T | G | T | C | A |
| 69C | C | A | A | A | T | T | G | T | T | G | - | A | G | A | A | *nd* | C | TA | A | A | G | G | G | C | G | C | A | G | G | T | G | C | A | T | T | *nd* |
| 70C | T | A | G | A | T | T | G | T | A | A | delTAA | A | G | A | A | T | A | - | A | A | A | A | G | A | G | C | G | A | G | C | G | T | G | T | C | A |
| 55C | C | C | G | G | C | C | G | C | T | G | - | A | G | G | A | G | A | TA | G | G | G | G | A | C | G | T | A | G | A | T | A | C | A | T | C | A |
| 56C | C | A | A | A | T | T | G | T | *nd* | G | - | A | G | A | A | T | C | TA | A | A | G | G | G | C | G | C | G | G | G | T | G | C | A | T | *nd* | G |
| 57C | C | A | A | A | T | T | G | T | *nd* | G | - | A | G | A | A | T | C | - | A | A | G | G | G | C | G | C | G | G | G | T | G | C | A | T | C | G |
| 58C | C | A | G | A | T | T | G | T | A | G | - | A | G | A | A | T | C | - | A | A | G | G | G | A | G | T | A | G | G | T | G | C | A | T | C | G |

| SNP | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 59C | C | A | A | A | T | T | G | T | T | G | - | A | G | A | A | T | C | TA | A | A | G | G | G | C | G | C | A | G | G | T | G | C | A | T | T | nd |
| 60C | C | A | G | A | T | T | G | T | A | G | - | A | G | A | A | T | C | - | A | A | G | G | G | A | G | C | A | G | G | T | G | C | A | T | C | G |
| 61C | C | A | G | A | T | T | G | T | T | G | - | A | G | A | A | T | C | - | A | A | G | G | G | A | G | C | G | G | G | T | G | C | A | T | C | nd |
| 62C | T | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | A | - | A | A | A | A | G | A | G | C | G | A | G | C | G | T | G | T | C | A |
| 63C | T | A | G | A | T | T | G | T | nd | R | - | A | G | A | A | T | nd | - | A | A | A | G | G | A | G | T | G | nd | G | nd | G | nd | nd | T | C | nd |
| 64C | C | A | G | A | C | C | G | C | nd | G | - | A | G | A | A | T | C | - | A | A | G | G | G | A | G | T | G | G | G | T | G | C | A | T | C | nd |
| 65C | C | A | A | A | C | C | G | T | T | G | - | A | G | A | A | T | C | - | A | A | G | G | G | C | G | C | A | G | G | nd | G | C | A | T | C | nd |
| 66C | C | A | A | A | T | T | G | T | nd | G | - | A | G | A | A | nd | C | TA | A | A | G | G | G | C | G | C | A | G | G | T | G | C | A | T | T | nd |
| 67C | T | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | A | - | A | A | A | A | G | A | G | nd | A | A | G | C | G | T | G | T | C | A |
| 68C | T | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | A | - | A | A | A | A | G | A | G | C | A | A | G | C | G | T | G | T | C | A |
| 69C | C | A | A | A | T | T | G | T | T | G | - | A | G | A | A | nd | C | TA | A | A | G | G | G | C | G | C | A | G | G | T | G | C | A | T | T | nd |
| 70C | T | A | G | A | T | T | G | T | A | A | delTAA | A | G | A | A | T | A | - | A | A | A | A | G | A | G | C | G | A | G | C | G | T | G | T | C | A |
| 71C | T | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | A | - | A | A | A | A | G | A | G | nd | A | A | G | nd | G | T | G | T | C | A |
| 72C | T | A | G | A | nd | T | G | T | A | A | - | A | G | A | A | T | A | - | A | A | A | A | G | A | G | C | A | A | G | C | G | T | G | T | C | A |
| 73C | C | A | A | A | T | T | G | T | nd | G | - | A | G | A | A | nd | C | TA | A | A | G | G | G | C | G | C | A | G | G | T | G | C | A | T | T | nd |
| 74C | nd | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | A | - | A | A | A | A | G | A | C | C | A | A | G | C | G | T | G | T | C | A |
| 75C | nd | A | G | A | T | T | G | T | T | G | - | A | G | A | A | T | C | TA | A | A | G | G | G | C | G | C | A | G | G | T | G | C | A | T | C | nd |
| 76C | nd | A | A | A | T | T | G | T | T | G | - | A | G | A | A | nd | C | TA | A | A | G | G | G | C | G | nd | A | G | G | T | G | C | A | T | T | nd |
| 77C | nd | A | A | A | T | T | G | T | T | G | - | A | G | A | A | T | C | TA | A | A | G | G | G | C | G | C | A | G | G | T | G | C | A | T | T | nd |
| 78C | nd | A | A | A | T | T | G | T | T | G | - | A | G | A | A | T | C | TA | A | A | G | G | G | C | G | C | A | G | G | T | G | C | nd | T | T | nd |
| 79C | nd | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | A | - | A | A | G | G | G | A | G | C | A | G | G | T | G | C | A | T | C | nd |
| 80C | C | C | G | A | C | C | G | C | A | G | - | A | G | G | nd | T | A | TA | G | G | G | G | G | A | G | C | A | G | A | T | G | C | nd | T | C | A |
| 81C | T | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | A | - | A | A | A | G | G | A | G | nd | A | A | G | C | G | T | G | T | C | A |

| SNP | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 82C | nd | C | G | A | C | C | G | C | A | G | - | A | G | G | A | T | A | TA | G | G | G | G | G | A | G | nd | A | G | A | T | A | C | A | T | C | A |
| 83C | nd | A | G | A | T | T | G | T | T | A | - | A | G | A | A | nd | A | - | A | A | A | A | G | A | G | C | A | A | G | C | G | T | G | T | C | nd |
| 84C | T | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | A | - | A | A | A | A | G | A | G | nd | A | A | G | C | G | T | G | T | C | nd |
| 85C | T | A | G | A | T | T | G | T | nd | nd | - | A | G | A | A | T | C | - | A | A | G | G | G | A | G | T | A | G | G | T | G | C | A | T | C | G |
| 86C | T | A | G | A | T | T | G | T | T | A | - | A | G | A | A | T | A | - | A | A | A | A | G | A | G | C | A | A | G | C | G | T | G | T | C | A |
| 87C | T | C | G | A | C | C | G | C | T | G | - | A | G | G | A | T | A | TA | G | G | G | G | G | A | G | C | A | G | A | T | A | C | A | T | C | nd |
| 88C | nd | C | G | G | C | C | G | C | nd | G | - | A | G | G | A | G | A | TA | G | G | G | G | A | C | G | T | A | G | A | T | A | C | A | T | C | nd |
| 89C | C |  | G | A | T | T | G | T | T | A | - | A | G | A | A | nd | C | - | A | A | A | A | G | A | G | nd | A | A | G | C | G | T | G | T | C | A |
| 90C | T | A | A | A | C | C | G | C | T | G | - | A | G | A | A | T | C | - | A | A | G | G | G | A | G | nd | nd | G | G | T | G | C | A | T | C | nd |
| 91C | nd | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | A | - | A | A | A | A | G | A | G | nd | nd | A | G | C | G | T | G | T | C | A |
| 92C | nd | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | A | - | A | A | A | A | G | A | G | C | nd | A | G | C | G | T | G | T | C | A |
| 93C | T | A | G | A | T | T | G | T | A | A | - | A | G | A | A | T | A | - | A | A | A | A | nd | A | G | C | nd | A | G | C | G | T | G | T | C | A |
| 94C | nd | A | G | A | T | T | G | T | A | G | - | A | G | A | A | T | C | - | A | A | G | G | G | A | G | C | nd | G | G | T | G | C | A | T | C | nd |
| 95C | nd | A | A | A | T | T | G | T | nd | G | - | A | G | A | A | T | C | TA | A | A | G | G | G | C | G | C | A | G | G | T | G | C | A | T | T | nd |
| 96C | T | A | G | A | T | nd | G | T | A | nd | - | A | G | A | A | T | A | - | A | A | nd | A | G | A | G | T | A | A | G | C | G | T | nd | T | C | A |

Table S 7: Association of studied SNPs of Table S 6 with FVIII:C, FVIII:Ag, *F8* mRNA expression in healthy individuals.

*P values are shown calculated according to Mann Whitney test for X-lined loci

| SNP | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *F8* mRNA exp. | 0.3845 | 0.6012 | 0.3181 | 0.8872 | 0.8449 | 0.8383 | - | 0.2331 | 0.0326 | 0.4062 | - | - | - | 0.1296 | - | 0.6142 | 0.0071 | 0.9349 | 0.12 | 0.0617 | 0.717 | 0.8187 | 0.1131 | 0.1859 | 0.0856 | 0.3401 | 0.4614 | 0.5099 | 0.2236 | 0.4751 | 0.2372 | 0.6568 | 0.3191 | - | 0.5037 | 0.1624 |
| FVIII:C | 0.4359 | 0.773 | 0.6651 | 0.8349 | 0.2773 | 0.5541 | - | 0.3378 | 0.1045 | 0.7632 | - | - | - | 0.6353 | - | 0.1646 | 0.3364 | 0.049 | 0.2115 | 0.3491 | 0.8824 | 0.8367 | 0.1355 | 0.8056 | 0.3109 | 0.6039 | 0.6571 | 0.9367 | 0.0611 | 0.924 | 0.7846 | 0.8609 | 0.8516 | - | 0.3038 | 0.0535 |
| FVIII:Ag | 0.3375 | 0.3008 | 0.2044 | 0.4631 | 0.1669 | 0.3524 | - | 0.3448 | 0.1487 | 0.7871 | - | - | - | 0.8966 | - | 0.4683 | 0.519 | 0.2503 | 0.586 | 0.744 | 0.8731 | 0.76 | 0.3944 | 0.3619 | 0.3071 | 0.6237 | 0.3245 | 0.8886 | 0.3283 | 0.8747 | 0.9629 | 0.9039 | 0.9239 | - | 0.5164 | 0.4036 |

Table S 8: Distribution of SNPs found after NGS in patients without mutation.

SNP numbering is according to Table S 6, del= deletion, nd=not data avaliable

| SNP | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HA#2 | C | C | G | G | C | C | G | C | T | G | - | A | A | G | A | G | nd | TA | G | G | G | G | A | C | C | T | A | G | A | T | A | C | A | - | C | A |
| HA#5 | T | A | G | A | T | T | G | T | A | A | - | G | G | A | A | T | A | - | A | A | A | A | G | A | G | C | G | A | G | C | G | T | G | T | C | A |
| HA#6 | T | A | G | A | T | T | G | T | A | A | - | G | G | A | A | T | A | - | A | A | A | A | G | A | G | C | G | A | G | C | G | T | G | T | C | A |
| HA#8 | T | A | G | A | T | T | G | T | A | A | delTAA | G | G | A | C | T | A | - | A | A | A | A | G | A | G | C | G | A | G | C | G | T | G | - | C | A |
| HA#9 | C | C | A | G | T | T | G | C | T | G | - | G | A | G | A | G | A | TA | G | G | G | G | A | C | C | T | A | G | A | T | A | C | A | - | C | A |
| HA#10 | T | A | G | A | T | T | T | T | nd | A | - | G | G | A | A | T | A | - | A | A | A | A | G | A | G | C | G | nd | G | C | G | T | G | - | C | A |
| HA#11 | T | A | G | A | nd | T | T | T | A | A | - | G | G | A | A | T | A | - | A | A | A | A | G | A | G | C | G | A | G | C | G | T | G | - | C | A |
| HA#12 | T | A | G | A | T | T | G | T | nd | A | delTAA | A | G | A | A | T | nd | - | A | A | A | A | G | A | G | C | G | A | G | C | G | T | nd | - | C | A |
| HA#13 | T | A | G | A | T | T | T | T | nd | A | - | G | G | A | A | T | A | - | A | A | A | A | G | A | G | C | G | A | G | C | G | T | G | - | C | A |

Table S 9: Characteristics of healthy individual controls included in this study.

| Sample | Age | FVIII:C | vWF activity | FVIII:Ag | vWF:Ag | *F8* mRNA exp. | ABO | Rhesus |
|---|---|---|---|---|---|---|---|---|
| 1C | 36 | 105 | 97.5 | 103 | 109 | 0.09 | A1 | Pos |
| 2C | 31 | 91 | 45.3 | 90 | 55 | 0.06 | O | Pos |
| 3C | 28 | 102 | 94.5 | 89 | 102 | 0.09 | A | pos |
| 4C | 27 | 159 | 151.8 | 140 | 198 | 0.08 | B | Pos |
| : 5C | 37 | 57 | 35.9 | 55 | 40 | 0.06 | O | pos |
| 6C | 57 | 129 | 119 | 116 | 133 | 0.07 | A1 | Pos |
| 7C | 44 | 126 | 66.2 | 95 | 94 | 0.10 | B | Pos |
| 8C | 56 | 89 | 80.6 | 85 | 95 | 0.10 | A2 | Pos |
| 9C | 62 | 85 | 43.7 | 83 | 97 | 0.16 | A2 | Pos |
| 10C | 40 | 91 | 75.5 | 81 | 87 | 0.26 | A1 | Pos |
| 11C | 28 | 111 | 79 | 93 | 85 | 0.07 | A | Pos |
| 12C | 35 | 139 | 106.3 | 121 | 133 | 0.06 | A1 | Pos |
| 13C | 32 | 67 | 44.4 | 59 | 83 | 0.21 | O | Pos |
| 14C | 54 | 113 | 90.9 | 96 | 152 | 0.05 | O | Pos |
| 15C | 22 | 114 | 95.4 | 96 | 235 | 0.12 | A | Pos |
| 16C | 50 | 129 | 88.3 | 96 | 94 | 0.05 | O | Neg |
| 17C | 40 | 119 | 98.8 | 89 | 103 | 0.11 | AB | Pos |
| 18C | 40 | 159 | 121.7 | 112 | 139 | 0.16 | B | Pos |
| 19C | 49 | 118 | 82.7 | 90 | 79 | 0.24 | A1B | Pos |
| 20C | 56 | 164 | 75.9 | 141 | 86 | 0.24 | A1 | Pos |
| 21C | 48 | 193 | 185 | 163 | 197 | 0.19 | B | Pos |
| 22C | 57 | 139 | 114.2 | 85 | 139 | 0.07 | A | Pos |
| 24C | 51 | 118 | 76.1 | 97 | 85 | 0.18 | O | pos |
| 25C | 50 | 168 | 144.5 | 119 | 168 | 0.05 | A1 | Pos |
| 26C | 39 | 98 | 78.1 | 78 | 84 | 0.06 | B | Pos |

| Sample | Age | FVIII:C | vWF activity | FVIII:Ag | vWF:Ag | *F8* mRNA exp. | ABO | Rhesus |
|--------|-----|---------|--------------|----------|--------|----------------|-----|--------|
| 27C | 50 | 107 | 85.5 | 76 | 84 | 0.08 | A2 | Neg |
| 28C | 29 | 118 | 90.9 | 88 | 106 | 0.09 | B | Neg |
| 29C | 42 | 123 | 82 | 67 | 81 | 0.13 | O | Pos |
| 30C | 33 | 117 | 70.9 | 81 | 72 | 0.11 | A2 | Pos |
| 31C | 41 | 147 | 112.6 | 97 | 118 | 0.07 | A1 | Pos |
| 32C | 28 | 106 | 76.9 | 81 | 84 | 0.12 | O | Pos |
| 33C | 48 | 114 | 78.7 | 86 | 88 | 0.05 | A1 | Pos |
| 34C | 48 | 128 | 93.1 | 100 | 108 | 0.04 | O | Pos |
| 35C | 27 | 119 | 92.3 | 89 | 103 | 0.04 | A1 | Pos |
| 36C | 48 | 85 | 54.6 | 50 | 55 | 0.10 | O | pos |
| 37C | 54 | 93 | 58.6 | 70 | 70 | 0.49 | O | Pos |
| 38C | 36 | 109 | 77.1 | 69 | 75 | 0.07 | O | Pos |
| 39C | 59 | 149 | 98.6 | 91 | 99 | 0.47 | A | Pos |
| 40C | 30 | 117 | 111.4 | 83 | 114 | 0.31 | A1 | Pos |
| 41C | 29 | 155 | 118.8 | 135 | 122 | 0.15 | A1 | Pos |
| 42C | 57 | 129 | 103.2 | 85 | 113 | 0.13 | A1B | Pos |
| 43C | 48 | 169 | 136.9 | 125 | 147 | 0.75 | B | Pos |
| 44C | 46 | 100 | 63.4 | 149 | 70 | 0.32 | O | Neg |
| 45C | 46 | 110 | 86.2 | 89 | 95 | 0.26 | O | Pos |
| 46C | 34 | 109 | 88.2 | 74 | 103 | 0.33 | A1 | Pos |
| 47C | 33 | 163 | 147.7 | 132 | 138 | 0.29 | B | Pos |
| 48C | 38 | 147 | 114.2 | 119 | 119 | 0.52 | A1 | Pos |
| 49C | 41 | 96 | 63.9 | 80 | 70 | 0.16 | O | neg |
| 50C | 55 | 144 | 97.7 | 102 | 111 | 0.28 | A1B | Pos |
| 51C | 46 | 111 | 63 | 90 | 72 | 0.43 | O | pos |
| 52C | 44 | 144 | 120.1 | 183 | 139 | 0.19 | A1 | Pos |
| 53C | 57 | 161 | 123 | 173 | 141 | 0.16 | A1 | Pos |

| Sample | Age | FVIII:C | vWF activity | FVIII:Ag | vWF:Ag | *F8* mRNA exp. | ABO | Rhesus |
|--------|-----|---------|--------------|----------|--------|----------------|-----|--------|
| **54C** | 50 | 245 | 275.7 | 143 | 250 | 0.41 | B | Neg |
| **55C** | 28 | 95 | 47.3 | 64 | 54 | 0.26 | O | Pos |
| **56C** | 30 | 125 | 75.2 | 115 | 88 | 0.23 | A1B | Pos |
| **57C** | 39 | 138 | 88.1 | 120 | 107 | 0.32 | O | Pos |
| **58C** | 41 | 118 | 84.6 | 96 | 94 | 0.05 | A1 | Pos |
| **59C** | 27 | 131 | 94.7 | 96 | 108 | 0.29 | A2 | Pos |
| **60C** | 51 | 101 | 68.8 | 85 | 75 | 0.43 | O | Pos |
| **61C** | 36 | 123 | 93.7 | 106 | 97 | 0.15 | A2 | Pos |
| **62C** | 31 | 90 | 45.6 | 75 | 57 | 0.16 | O | Pos |
| **63C** | 28 | 121 | 88.3 | 80 | 94 | 0.05 | A1 | Pos |
| **64C** | 29 | 96 | 62.9 | 74 | 85 | 0.32 | A2 | Neg |
| **65C** | 42 | 102 | 109.9 | 83 | 107 | 0.12 | A1B | Pos |
| **66C** | 44 | 126 | 124.7 | 79 | 133 | 0.12 | A1 | Pos |
| **67C** | 26 | 144 | 119 | 118 | 123 | 0.22 | A1 | Pos |
| **68C** | 37 | 108 | 65 | 116 | 70 | 0.17 | A2 | Pos |
| **69C** | 41 | 142 | 111 | 108 | 126 | 0.10 | A1 | Pos |
| **70C** | 53 | 134 | 112 | 65 | 136 | 0.08 | B | Neg |
| **71C** | 29 | 139 | 103.3 | 88 | 118 | 0.03 | B | Pos |
| **72C** | 35 | 121 | 86.5 | 82 | 102 | 0.13 | A1 | Pos |
| **73C** | 28 | 159 | 145.3 | 106 | 144 | 0.04 | A1 | Pos |
| **74C** | 32 | 191 | 159.1 | 119 | 103 | 0.24 | A1B | Pos |
| **75C** | 48 | 161 | 146.8 | 107 | 87 | 0.07 | A1 | Pos |
| **76C** | 58 | 189 | 134.6 | 112 | 73 | 0.07 | A1B | Neg |
| **77C** | 40 | 106 | 53.9 | 73 | 81 | 0.10 | O | Pos |
| **78C** | 55 | 146 | 59.8 | 86 | 87 | 0.05 | A2 | Pos |
| **79C** | 26 | 100 | 59 | 72 | 65 | 0.10 | O | Pos |

| Sample | Age | FVIII:C | vWF activity | FVIII:Ag | vWF:Ag | *F8* mRNA exp. | ABO | Rhesus |
|--------|-----|---------|--------------|----------|--------|----------------|-----|--------|
| 80C | 27 | 146 | 96 | 89 | 102 | 0.23 | O | Pos |
| 81C | 53 | 124 | 103.1 | 81 | 108 | 0.26 | O | Neg |
| 82C | 35 | 161 | 131.8 | 112 | 130 | 0.18 | A1 | Pos |
| 83C | 45 | 128 | 96.1 | 74 | 97 | 0.05 | A2 | Pos |
| 84C | 47 | 133 | 117.2 | 84 | 148 | 0.38 | A2 | Pos |
| 85C | 28 | 111 | 70.7 | 74 | 89 | 0.06 | O | Neg |
| 86C | 39 | 123 | 92.1 | 75 | 105 | 0.26 | O | Pos |
| 87C | 50 | 96 | 82.3 | 80 | 84 | 0.13 | O | Pos |
| 88C | 48 | 158 | 88 | 95 | 158 | 0.10 | O | Pos |
| 89C | 38 | 88 | 59.3 | 70 | 65 | 0.11 | A2 | Pos |
| 90C | 31 | 97 | 57.7 | 55 | 61 | 0.23 | O | Pos |
| 91C | 52 | 143 | 131.1 | 89 | 138 | 0.26 | O | Pos |
| 92C | 26 | 124 | 115.9 | 81 | 113 | 0.23 | A1B | Neg |
| 93C | 28 | 134 | 111.7 | 98 | 135 | 0.11 | A1 | Pos |
| 94C | 47 | 156 | 113.1 | 95 | 118 | 0.24 | B | Neg |
| 95C | 56 | 154 | 125.4 | 112 | 126 | 0.16 | A1 | Neg |
| 96C | 44 | 114 | 82.3 | 78 | 95 | 0.08 | A1 | Pos |

Table S 10: List of primers used for LR-PCR amplification of genomic region of *F8.*

| Name | Sequence 5´-->3´ | T$_m$ | X:Chr (hg18) | length (bp) |
|---|---|---|---|---|
| 5´-F | TAC ATC ACA GAG CTC AGC GTG | 64 | 153908320 | 5323 |
| 5´-R | CTC ACT CTG GTT ATA CTG GTC | | 153904797 | |
| A-F | ATG CCA GGT GGT TAT GAG TGC | 60 | 153905062 | 12796 |
| A-R | ACC TGG GCC TCT ACT TTC AAG | | 153892266 | |
| B1-F | AGT TCG TTC TAC AGC CTA TAT AGC | 60 | 153892311 | 5250 |
| B1-R | GGA ACC ATT CCA GAA CTT ATG AAG | | 153887061 | |
| B2-F | GTG TGC TAA GAT TGT GTG GCA G | 60 | 153887666 | 5450 |
| B2-R | ACT TAG ATG TGA AGC AGT TGT TCG | | 153882216 | |
| C-F | ATC CTG GTA ATC TGA GGC ATG G | 60 | 153882470 | 2861 |
| C-R | TCA GGT ATT AAC CAA CAG TAT CAC TC | | 153879609 | |
| D-F | ATG CAG TGT ATG GTT ATT GAT TTT ATG ATT CTT A | 60 | 153881259 | 14052 |
| D-R | CTG ATG ATA GAG AGG AAG ATA CAG | | 153867207 | |
| E-F | AAG TAA CTG ATG AAG AAC TGT CTC | 60 | 153868842 | 5993 |
| E-R | CAG CAT GCA CTA CTT AAT CAC AAG | | 153862849 | |
| F-F | ATA AGG TCA AAC AAA ATG TGG TTG AG | 55 | 153864554 | 12198 |
| F-R | AAT CTG TCC CTC CAC TAT AAG G | | 153852356 | |
| G-F | AGC AAC CCA AGA GAA TAA TGA AG | 60 | 153852504 | 9911 |
| G-R | AGG GTA GAT GTT ATA TGG TCT GC | | 153842593 | |
| H-F | AGT AAG AGA CTT GAG CAT CAC AGA TTG | 60 | 153842844 | 7541 |
| H-R | GAT TAC ATC AAT TTT TCT TTA TTC ACC ACC | | 153835303 | |
| I-F | ATG AAG AGA ATA AGG CTA CAG ACT G | 60 | 153835798 | 5655 |
| I-R | AGT CTA TGA AGG GAA TTG TTT CTG | | 153830143 | |
| J-F | ACA CAG TAC CTA GTA ACA GAG TAG | 60 | 153830215 | 11202 |
| J-R | AGA CAT CAG TAC AGA CAT TAT AAG G | | 153819013 | |
| K-F | ACT GTT GTC TTT GCC ATG AAT CTG | 60 | 153819771 | 10614 |
| K-R | AGC AAT AAG GGC TTA CTT GAG C | | 153809157 | |
| L-F | AGT TGA GAC CTA AAT GAT GAT AAC AG | 60 | 153809322 | 8830 |
| L-R | ATG ACC AGT AGT GTT GAA CTT CC | | 153800492 | |
| M1-F | ATG AAG AAC AAT GCC ACC AAG C | 60 | 153801065 | 4504 |
| M1-R | AGG TGG CAG TGG GAA TAA GTG | | 153796561 | |
| M2-F | ATC TAG GAA CAA ATT TAA CCA AAG AGG | 60 | 153797003 | 9136 |
| M2-R | AAT TCC ACT GTC CTT AAC TCA CC | | 153787867 | |
| N-F | CAG GCA CCT AGG AAA ATG AGG ATG | 60 | 153788105 | 10636 |
| N-R | GTG TTT GTC CAA TAT CTG AAA TCT GCC | | 153777469 | |
| O-F | AGG TGT AGC AGT TCA TCA CTG A | 60 | 153778148 | 9043 |
| O-R | CTG CGT TTT GTT ATT CCT ACT GC | | 153769105 | |
| P1-F | AGT ACA AGA GTA GCT TTA TCA TGG ATT TAG GGC | 60 | 153769493 | 4040 |
| P1-R | ACT CTC ATG CAC TAC TGT TGG GAA TGC GA | | 153765453 | |
| P2-F | GAG GCC GTG TGT GTC TGG CTG CTT TCA | 60 | 153765806 | 3050 |
| P2-R | AGC CAC ATC AAC CAC TAC TGC | | 153762756 | |
| Q-F | AAG TGG CAT TCC TGA TCT TAG G | 57 | 153763253 | 6132 |
| Q-R | AGA GGA CAC AGT CAG CTG TG | | 153757121 | |

| R-F | ATG TTA TGT AAT TTT GTC CAG AAC CCA G | 60 | 153757348 | 13293 |
| R-R | ATG TGT CCA TGT GCA TTT AAT GTT TAG | | 153744055 | |
| S1-F | ACA ATA GCA AAG GCA TGG AGT C | 64 | 153744273 | 4567 |
| S1-R | CCA TGG TGG CAG GGA TAG AG | | 153739706 | |
| S2-F | GGT GAC CCA AAC CTT TAT TCC TG | 58 | 153740100 | 6129 |
| S2-R | ATC AAT TGA CCA TAG ATG TAT AGG C | | 153733971 | |
| T1-F | ACA GAC ATA GAT CAG CAA ACA TGC | 60 | 153734957 | 4204 |
| T1-R | ACA TTC TTA ATC TTC AGC AAT ACT TTC | | 153730753 | |
| T2-F | ACC TAA ATG AAA GCA AGC AAA GAG AG | 60 | 153731207 | 8804 |
| T2-R | CTC AAG AAC ATT AAT ACC ATC TGC | | 153722403 | |
| U-F | ATG ACT AAA ACC ACC CAT AAC CC | 60 | 153722807 | 5644 |
| U-R | AGA GTG TCC ATC TTG CTA TTC AG | | 153717163 | |
| 3´-F | GGC AAA TG G AAA ACA GGA GA | 64 | 153717702 | 4750 |
| 3´-R | GAA CTT CCA CAC AGC AGC AA | | 153712952 | |

Table S 11: Primers used for characterization of the rearrangement (Patient HA#1)

List of primers used for mapping PCRs in intron 1 of *F8.*

| Name | Sequence 5´-->3´ | T$_m$ | X:Chr (hg18) | Length bp |
|------|------------------|-------|--------------|-----------|
| B1-F1 | AGT TCG TTC TAC AGC CTA TAT AGC | 59 | 153892311 | 1505 |
| B1-R1 | CAT ATC CTT GGT TCC ACT ACG T | | 153890807 | |
| B1-F2 | TGT TGG CCC GGC TGG TGT | 59 | 153891024 | 959 |
| B1-R2 | GAA CTC CCC AGT CCC AGC | | 153890063 | |
| B1-F3 | GTC TTG GGA AGG AAG GTG C | 59 | 153890164 | 938 |
| B1-R3 | GCG GAC AGT TTC ATG CCG TA | | 153889223 | |
| B1-F4 | CCA CCG AGA AGC TGG TTT G | 59 | 153889353 | 983 |
| B1-R4 | CCT TGA AAC TGC AGA CCT CTC T | | 153888392 | |
| B1-F5 | GAA ATG CAG GGC TAC GTG G | 59 | 153888454 | 1500 |
| B1-R5 | ATG ATG AGT ACC AGT TGC AGA | | 153886954 | |

List of primers used for inverse PCRs.

| Name | Sequence 5´-->3´ | T$_m$ | X:Chr hg 18 |
|------|------------------|-------|-------------|
| Up-F | CTG GGG AGC TGG ATA TTG TC | 60 | 153891540 |
| Up-R | ATT TGG CAA CCA TTG TCC TT | | 153891779 |
| Down-F | CCT GAA GGG AGG AAA GGC TA | 60 | 153889443 |
| Down-R | TGG TAT TGT CCA TCA CAG AGA C | | 153889353 |

List of primers used for multiplex PCR

| Name | Sequence 5´-->3´ | T$_m$ | X:Chr hg18 |
|------|------------------|-------|------------|
| 3F | TGG GGA ATC ATT GTA CAG CA | 62 | 153890934 |
| 4R | TGG GTC TGG CCT TTA ATC AG | | 153889544 |
| 2R | GGA CTC TGC TCC ACA CCC TA | | 153435095 |

Table S 12: Primers used for characterization of the breakpoint (Patient HA#14)

List of primers used for identification of breakpoint. MLPA specific adapters are underlined. *F8* specific sequences are in capital.

| Name | Sequence 5´-->3 | $T_m$ | Length (bp) |
|---|---|---|---|
| MLPA18F | gtagcgcgacggccagtAGCATGGGCAGCAATGAAAA | 58 | 2057 |
| 19R | cagggcgcagcgatgacGGAAGAAAGCTGTAAAGAAGTAGGC | | |
| 17F | gtagcgcgacggccagtCAGGTTGGACTGGCATAAAAA | 58 | 539 |
| MLPA18R | cagggcgcagcgatgacTTTTCATTGCTGCCCATGCT | | |

List of primers used for inverse PCRs.

| Name | Sequence 5´-->3 | $T_m$ | Length (bp) |
|---|---|---|---|
| InvR3 | CTA GGG AGG GAA GAC ATC AAT C | 55 | 2725 |
| InvF3 | ATG AGA CCA AAA GCT GGT ACT | | |
| Inv-D-F1 | CAA TAG TGA GTA GCA ATG TG | 55 | 2819 |
| Inv-D-R1 | CAT GTA GAT GCT CGC CAA TA | | |

Table S 13: Primers used for haplotype analysis.

| Marker | Polymorphism | rs number | Genotype [c] | Position | Name | Sequence 5' to 3' | Position (hg 19) | $T_m$ | Length |
|---|---|---|---|---|---|---|---|---|---|
| 1 | ATCT [a] | rs71969354 | Indel | 154854179 | HEMA154507.3 | GGG ATT CAT CTG GCA TTT ATT T | 154853916 | 57 | 356-364 |
| | | | | | HEMA154507.3 | GGG ATT TTG TAG TGT TGC TAT GG | 154854271 | | |
| 2 | TTTTA [a] | rs34754890 | Indel | 154739902 | TMLHEInt5 | CCA AAA CTA ATG GTT AGG CAA T | 154739770 | 57 | 276-286 |
| | | | | | TMLHEInt5 | GAG GGA TAG CAT CGG GAG AT | 154740055 | | |
| 3 | C/T | rs6649625 | CC: 0.345; CT: 0.195; TT: 0.460 | 154248283 | F8Int1SNP-F | ATC TAC TAT TAT CTT GCT ATG CCA | 154248447 | 57 | 258 |
| | | | | | F8Int1SNP-R | ATA GCA TCA CAC TTA TAA AAT ACA C | 154248190 | | |
| 4 | CA repeat [b] | rs72130835 | Indel | 154230888 | F8Int1-F | CTG CCC TTG GAC ATA AGC AT | 154230814 | 59 | 218-220 |
| | | | | | F8Int1-R | Fam-CCA TAT GAT CCA GCA ACT CG | 154231033 | | |
| 5 | G/A | rs7058826 | GG: 0.867;  GA: 0.080;  AA: 0.053 | 154194989 | F8-Ex 8-F | CCC ATA TAG TAG CCT GCA GAA ACA TGA | 154195086 | 57 | 454 |
| | | | | | F8-Ex 8-R | TGG CTT CAG GAT TTG TTG GT | 154194633 | | |
| 6 | CA (STR) [a] | rs10626536 | Indel | 154164258 | F8Int13.2-F | GCC TAG AGA ATG CCA AAG TAA CA | 154164237 | 61 | 150 |
| | | | | | F8Int13.2-R | Fam-TTG TTG TAA TTC CCA TTT GCA T | 154164386 | | |
| 7-8 | C/G | rs1800291 | GG: 0.797; GC: 0.102; CC: 0.102 | 154158285 | F8-Ex14-5-F | AAA GAG ATG GTT TTT CCA AGC A | 154158497 | 57 | 520 |
| | C/A | rs1800292 | AA:  0.833; AC: 0.083; CC: 0.083 | 154158201 | F8-Ex14-5-R | CTC GGG GTC AAA TGT TTC AT | 154157978 | | |
| 9 | G/A | rs6643714 | AA: 0.522; GA: 0.221; GG: 0.257 | 154149448 | F8Int14SNP-F | CCT GAC TTT TTA ATG ATC GCC A | 154149540 | 59 | 209 |
| | | | | | F8Int14SNP-R | AGA TCT AGC TTT TGT CTT TCT GAG | 154149332 | | |
| 10 | T/A | rs4898352 | AA: 0.688; AT: 0.152; TT: 0.161 | 154132090 | F8Int18SNP-F | ATG TGT TCA CTG TAC GA | 154131971 | 50 | 367 |
| | | | | | F8Int18SNP-R | AAT ATC TTG GGA TGG AC | 154131605 | | |
| 11 | T/C | rs4074307 | AA: 0.696; AG: 0.152; GG: 0.152 | 154130223 | F8Int19SNP-F | GTG AGT AGC AAT GTG GGC AG | 154130306 | 59 | 229 |
| | | | | | F8Int19SNP-R | ATC TTC GAG CTT TAC CAA GTT G | 154130078 | | |
| 12 | CA repeat [a] | rs5818972 | Indel | 154104012 | LR-F8B-F | ATG TTA TGT AAT TTT GTC CAG | 154104162 | 57 | 217 |
| | | | | | LR-F8B-R | Fam-AGA GGA CAC AGT CAG CTG TG | 154103946 | | |
| 13 | GA repeat [b] | REN90957 | Indel | 154087437 | F8Int25.1-F | CTG CCT TTG GAC TGG AAC AT | 154087369 | 57 | 145 |
| | | | | | F8Int25.1-R | TGT GTT AGT CAG GGA TAT CCA AA | 154087513 | | |
| 14 | ./CA [a] | REN90841 | Indel | 154082866 | F8Int25.2-F | CTC TAG GCT GCC AAC TCA CC | 154082796 | 59 | 144 |
| | | | | | F8Int25.2-R | Fam-TCA TGG TTC CCA AAG AAA CA | 154082939 | | |
| 15 | G/A | rs 6643711 | AA: 0.25;  GA: 0.222;  GG: 0.528 | 154078042 | F8Int25SNP-F | GAT TAA TGT TGT AGT TTA TAG GAT AG | 154078123 | 57 | 342 |
| | | | | | F8Int25SNP-R | GCC ATC ACA CCC AGC TAC | 154078464 | | |

Table S 14: Primers used for analyis of SNPs found after NGS.

*Some SNPs were amplified using biotin labeled primers and were subsequently applied on a Pyrosequencer. The rest of the SNPs were amlified ans sequenced on an *ABI* capillary sequencer.

| SNP Nr. | Ref Base->SNP | F8 Position | SNP coordinates | NG_0014031:g. | X-Chr / F8 Position | Primer name | Sequence 5' to 3' | Position on | Length (bp) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | C>T | 5´UTR | rs6649625 | 7716 | 154210739 | 154210739-F | GAC ACT TTA TTC TTC CAC ATA AAA CTG | 154248343 | 219 |
| | | | | | | 154210739-R | Bio-GCA CTC ATA AAT GGA AAT AGG AGA | 154248125 | |
| | | | | | | 154210739-SQ | TTA AAA AGA TCC AAG C | | |
| 2 | A>C | Intron 1 | rs55807428 | 15614 | 154240385 | 154240385-F | TAG GCA AAA CCA CAG TGA CG | 154240212 | 591 |
| | | | | | | 154240385-R | ATC AGC CTG GGC AAC ATA AC | 154240802 | |
| 3 | G>A | Intron 1 | rs1470586 | 17323 | 154238676 | 154238676-F | TCT GGA CAC ATC CTG ACA GC | 154238754 | 572 |
| | | | | | | 154238676-R | TGG AGA AAC AAG ATT GTG AGA CA | 154238183 | |
| 4 | A>C | intron 1 | rs141897310 | 25961 | 154230038 | 154230038-F | ACT GGA GGA GGT GAG AGT TAT TGT | 154230267 | 252 |
| | | | | | | 154230038-R | Bio-TGG GCA ACA CAG CAG TAC CT | 154230016 | |
| | | | | | | 154230038-SQ | ACC ACC ACG CCT GGC | | |
| 5 | T>C | Intron 4 | rs5945269 | 36681 | 154219318 | 154219318-F | GGA GGG GAA CAT TAT TCT TCA TT | 154219366 | 153 |
| | | | | | | 154219318-R | Bio-CCA GTT CTG TGA GCT CCC TTT | 154219214 | |
| | | | | | | 154219318-SQ | AGA TCT AAG ACA T | | |
| 6 | T>C | Intron 4 | rs73641115 | 37752 | 154218247 | 154218247-F | TGA TGC AAA CAT GAA TAT GAA GAA | 154218280 | 148 |
| | | | | | | 154218247-R | Bio-TGG TTC CAA GCA TTT CAG GT | 154218133 | |
| | | | | | | 154218247-SQ | CAA ATC TGT AAT CAC AAC | | |
| 7 | G>T | Intron 4 | rs144231135 | 37942 | 154218057 | 154218057-F | AAC GTT TCA GAT TTC AGA TAC TGG | 154218088 | 146 |
| | | | | | | 154218057-R | Bio-TGG TGT CTT TTA TAA AGC ATA TGG A | 154217943 | |
| | | | | | | 154218057-SQ | CAT ATA TAA TGA GAT AG | | |
| 8 | T>C | Intron 6 | rs 5945128 | 45117 | 154210882 | 154210882-F | CAC AAT GAA ACC CCG TCT CT | 154210536 | 556 |
| | | | | | | 154210882-R | GGG TCC CAA GTA TCC TTC AA | 154211091 | |
| 9 | A>T | Intron 6 | rs28857481 | 45260 | 154210739 | 154210739-F | CGT TTG TTT ACA TTT GTC CCA AC | 154210919 | 223 |
| | | | | | | 154210739-R | Bio-CTG GGC GAC AGC AAG ACT | 154210696 | |
| | | | | | | 154210739-SQ | GTG TAT ACA TAT ATA TAT A | | |
| 10 | A>G | Intron 6 | rs3861554 | 46841 | 154209158 | 154209158-F | GAG TGC CCT CTG TGG ATC TC | 154209346 | 201 |
| | | | | | | 154209158-R | Bio-CAG CCT GAG CAA CAT AGC AA | 154209146 | |
| | | | | | | 154209158-SQ | CCT GGC TGG GCT CAC | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 11 | TTA-del | Intron 6 | rs201280173 | 49012:49014 | 154211329 | 154211329-F | TAC CCA AGC AAC ACC ACA GA | 154207051 | 312 |
| | | | | | | 154211329-R | Bio-CCC TCA AGC ACT TAT CCT TTG | 154206740 | |
| | | | | | | 154211329-SQ | AAA TTG GAA AAA GTG A | | |
| 12 | C>A | Intron 6 | rs59835535 | 55023 | 154200976 | 154200976-F | ACC TGT TGG CCA TTT GTA GG | 154200571 | 576 |
| | | | | | | 154200976-R | TGT GGT GGC ACA CCT GTA GT | 154201146 | |
| 13 | G>A | Intron 7 | rs148954517 | 60195 | 154195804 | 154195804-F | TTT TTC GAT GCA GAC TGA GTT | 154195829 | 109 |
| | | | | | | 154195804-R | Bio-TTA TAC CAC AGG GCA ATT TTG A | 154195721 | |
| | | | | | | 154195804-SQ | AAG GCT ACA TGA ACC CC | | |
| 14 | A>G | Intron 10 | rs76063559 | 68425 | 154187574 | 154187574-F | Bio-GAT GCA AAA TTC CAG CTA GAT AGG | 154187904 | 362 |
| | | | | | | 154187574-R | ATG GTA GTG TGC GTC TGT GGT C | 154187543 | |
| | | | | | | 154187574-SQ | GCT GAA GTG GGA GGA | | |
| 15 | G>T | Intron 10 | rs190639729 | 69579 | 154186420 | 154186420-F | GGG TGA GGG AGG GTA GGT AA | 154186252 | 307 |
| | | | | | | 154186420-R | CTG AGG GGA TAG CAG AAC CA | 154186558 | |
| 16 | T>G | Intron 12 | rs150203712 | 74668 | 154181331 | 154181331-F | TCC TAC AGG GAA ATG CGA AA | 154181343 | 127 |
| | | | | | | 154181331-R | Bio-CAG AAT TTC AAA AGG CAT ATA TTG G | 154181217 | |
| | | | | | | 154181331-SQ | CGA AAT AGA ATA ACA A | | |
| 17 | C>A | Inron 13 | rs6643622 | 82469 | 154173530 | 154173530-F | TTT GCA TCA ATG TTC ATC AGA GA | 154173599 | 85 |
| | | | | | | 154173530-R | Bio-CAA GGG GGA AAA ATA CTT CCA A | 154173515 | |
| | | | | | | 154173530-SQ | TTG GTT TCA GGG TAA TAC T | | |
| 18 | TA-del | Inron 13 | rs34552198 | 88822:88823 | 154167176:77 | 154167176-F | Bio-CAT ATA TTT TAA ACC CCA TAA GAC ATT | 154167205 | 136 |
| | | | | | | 154167176-R | GCC AAC ACA CAC TGA AGG AA | 154167070 | |
| | | | | | | 154167176-SQ | AAA GAA ATG AAA GG | | |
| 19 | A>G | Inron 13 | rs78362479 | 90648 | 154165351 | 154165351-F | GTT TGT CGC CTT TGT TAT AGG TTC | 154165631 | 283 |
| | | | | | | 154165351-R | Bio-TTG TGA TAT GGG GTT AAA GCA GT | 154165349 | |
| | | | | | | 154165351-SQ | AGC ATA AAT TTC CCT T | | |
| 20 | A>G | Intron 14 | rs113224419 | 103221 | 154152778 | 154152778-F | ATG GCT GAA AGG GGT CAA TGT A | 154153130 | 358 |
| | | | | | | 154152778-R | Bio-TGG CAT TTT GCT CCT GTC CTA | 154152773 | |
| | | | | | | 154152778-SQ | AAT TCG AAG TTC CAC AA | | |
| 21 | G>A | Intron 14 | rs6643714 | 106551 | 154149448 | 154149448-F | TTT TGA GAA GTG TCT GTT CAT ATC C | 154149517 | 158 |
| | | | | | | 154149448-R | Bio-TTT ATC ACC CCT TCA TTT CTG A | 154149360 | |
| | | | | | | 154149448-SQ | TTT CTC AGA AGA AAT ATT GT | | |
| 22 | T>C | Intron 14 | 28845018 | 113955 | 154142044 | 154142044-F | CTC TGG ATA GGC AGG AAT TTT TAC | 154141549 | 725 |
| | | | | | | 154142044-R | CGT GGA GAA CCT CTG CTA GG | 154142273 | |

| | | | | | | 154138744-F | CGG CCT TCC GCA GTG TAT | 154138911 | |
|---|---|---|---|---|---|---|---|---|---|
| 23 | G>A | Intron 14 | rs147346816 | 117255 | 154138744 | 154138744-R | Bio-CAC AGG AGG GGG AAC ATC A | 154138660 | 252 |
| | | | | | | 154138744-SQ | GCT GCA CCC ACT AAC TC | | |
| | | | | | | 154115804-F | AAA TAA TTC CTA TGG CGG CAG T | 154115929 | |
| 24 | C>A | Intron 22 | rs9305 | 140194 | 154115804 | 154115804-R | Bio-AAG GAC ACG GAA GCC ATC AA | 154115662 | 268 |
| | | | | | | 154115804-SQ | ACT GAT GCA AAA AGT CC | | |
| | | | | | | 154114253F | Bio-GCT GCA ATG CAA GTG GTC TA | 154114349 | |
| 25 | C>T | Intron 22 | rs11152585 | 141745 | 154114253 | 154114253-R | ATG GCA CAG AGG AGC AGT CT | 154113988 | 362 |
| | | | | | | 154114253-SQ | GGG GAC AAC AGT | | |
| | | | | | | 154113527-F | Bio-TTC CGC TCT CCG CTG TTC T | 154113550 | |
| 26 | A>G | Intron 22 | rs2292089 | 142472 | 154113527 | 154113527-R | GTC TAC CTC GCT GCG GTT ATT G | 154113505 | 46 |
| | | | | | | 154113527-SQ | CCT GAC GGC CAA GGT | | |
| 27 | G>C | Intron 22 | rs113683431 | 142941 | 154114301 | 154114301-F | GCC TGC CCG GGA AAG TCC TC | 154114061 | 451 |
| | | | | | | 154114301-R | TGA TGA GGT GCA AAG AGC GG | 154114511 | |
| | | | | | | 154107643-F | Bio-TGA AAG TGA AAG GCA GAA TGG | 154107761 | |
| 28 | G>A | Intron 22 | rs28810103 | 148356 | 154107643 | 154107643-R | GTT ATC AAT GCT TTG TGG TAC TGG | 154107554 | 208 |
| | | | | | | 154107643-SQ | ATG GGT ATG TAC CTC AAA | | |
| | | | | | | 154094904-F | Bio-CCT GGA GGT GAT AAG TTG GG | 154095041 | |
| 29 | G>A | Intron 22 | rs78327897 | 161095 | 154094904 | 154094904-R | CCC AAC AGT AGA ATT ATC AAG C | 154094718 | 324 |
| | | | | | | 154094904-SQ | ACT GGG TTA CCA G | | |
| | | | | | | 154094064-F | GAC TTT CTG TTG CCA TTC TGT TAA | 154094099 | |
| 30 | T>C | Intron 22 | rs28835013 | 161935 | 154094064 | 154094064-R | Bio-TTC AAT TTG ATC CCT GTG GTA A | 154093960 | 140 |
| | | | | | | 154094064-SQ | TTG TAG TTT TCT TGT TCC TT | | |
| 31 | G>A | Intron 25 | rs112922881 | 168273 | 154087726 | 154087726-F | CAC AGT GAA ACC CCA TCT CT | 154087781 | 176 |
| | | | | | | 154087726-R | CTC ACT CTG TTG CCC AGG TT | 154087606 | |
| | | | | | | 154081354-F | GAG GTG CAA GTC TTG TCC ACT AAA | 154081404 | |
| 32 | C>T | Intron 25 | rs28536050 | 174645 | 154081354 | 154081354-R | Bio-ACC CAG AAT GCT TGC TTT GAA | 154081199 | 206 |
| | | | | | | 154081354-SQ | AAG AAC TGT CTG GTC TGA | | |
| | | | | | | 154078042-F | Bio-ACA ATG TAA GAC TCC CAA AAT GGT | 154078077 | |
| 33 | G>A | Intron 25 | rs6643711 | 177957 | 154078042 | 154078042-R | TCT CTT TGC TTG CTT TCA TTT AGG | 154078037 | 41 |
| | | | | | | 154078042-SQ | GCT TGC TTT CAT TTA GGT | | |

| 34 | TT_del | Intron 25 | rs3076842 | 186708;09 | 154069291 | 154069291-F | CCT CAT GTC TGC CTG CAT AA | 154069300 | 121 |
| | | | | | | 154069291-R | Bio-GGC CAT GCC TAA AGT TGA TG | 154069180 | |
| | | | | | | 154069291-SQ | CTG GCA TAA ATC ACC T | | |
| 35 | C>T | Intron 25 | rs1509787 | 187365 | 154068634 | 154068634-F | GCC AGA AGA TTA ATG GGA TCA | 154068658 | 158 |
| | | | | | | 154068634-R | Bio-TCC AAT GGT ATT AGC AAC CCT TA | 154068501 | |
| | | | | | | 154068634-SQ | CAC AGA AAC TGG | | |
| 36 | A>G | 3′ UTR | rs1050705 | 191799 | 154064200 | 154064200-F | GGT GAT ATG GTT TTA TTT CCT GTT A | 154064211 | 136 |
| | | | | | | 154064200-R | Bio-TCA AAG GCA TTT GTT TGT ATG TG | 154064076 | |
| | | | | | | 154064200-SQ | GTT ATG TTT AAC TT | | |

Table S 15: Primers used for amplification of unique variants found after NGS.

SQ= Sequencing primer

| SNP Nr. | Ref Base->SNP | F8 Position | SNP | NG_0014031:g. | X-Chr / F8 Position | Name | Sequence 5' to 3' | Position on X Chr hg19 | Length (bp) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | C>G | Intron 13 | new | 88055 | 154167944 | 154167944 | ATG TCC TTT CTG TTT GTT AG | 154168261 | 463 |
| | | | | | | 154167944 | GAA CTG GGT GGA GCC CCC CA | 154167799 | |
| 2 | G_ins | Intron 13 | new | 94667 | 154161332 | 154161332-F | BIO-GGT TTT TCC CTC AGC ACT TTA A | 154164433 | 99 |
| | | | | | | 154161332-R | GAG CTT GAA GCC ATC CAG ATT T | 154164335 | |
| | | | | | | 154161332-SQ | TTG AAT GCA AAT GGG | | |
| 3 | C>T | Intron 16 | new | 123107 | 154132892 | 154132892-F | GTAACTTTCAGAAATCAGGCCTCT | 154133275 | 558 |
| | | | | | | 154132892-R | CAG AGC AAA TTC CTG TAC TGT CAC | 154132718 | |
| 4 | C>T | Intron 18 | new | 124347 | 154131652 | 154131652-F | AGG GCA CCA GTA GTC ATC CA | 154131865 | 344 |
| | | | | | | 154131652-R | TGC AGT GGC ACT TTC ATA GC | 154131522 | |
| 5 | G>A | Intron 18 | new | 124759 | 154131240 | 154131240-F | GCC CTG TAA CTT TTC TGC TCA | 154131377 | 450 |
| | | | | | | 154131240-R | CCA CCC AGA GTA AAG GTG GA | 154130928 | |
| 6 | A>C | Intron 21 | new | 130722 | 154125277 | 154125277-F | BIO-TCC CAT TAT TCA CTC ATT CAT TCC | 154125402 | 216 |
| | | | | | | 154125277-R | CAG TGA ACC CAT TTG AGT CAC CT | 154125187 | |
| | | | | | | 154125277-SQ | GGG GTT AAA GAG TTA ATA CA | | |
| 7 | G>T | Intron 22 | new | 138108 | 154117891 | 154117891-F | GAATACCTCACAATGGGGAG | 154118130 | 400 |
| | | | | | | 154117891-R | TTG GGC TTC AGG AGA AGC CTC CTT | 154117731 | |
| 8 | A>G | Intron 22 | new | 160166 | 154095833 | 154095833-F | TCC TGG ATT CAA GCG ATT CT | 154095809 | 302 |
| | | | | | | 154095833-R | AAG AAA AGC CCA GGA CCA AT | 154095508 | |
| 9 | G>T | 3´UTR | new | 190183 | 154065816 | 154065816-F | Bio-CTG TGC TTT GCA GTG ACC AT | 154066071 | 342 |
| | | | | | | 154065816-R | CCA CCA AAG AAA TGC AGG AC | 154065730 | |
| | | | | | | 154065816-SQ | CAG GGA GGG ACA CTG | | |

List of Primers used for the qRT-PCR experiments

Table S 16: Primers used for mRNA expression analysis for association studies with coagulation parameters.

BHQ1= 3'-terminaler BlackHol

| Gene | Name | Sequence 5' to 3' | Location |
|------|------|-------------------|----------|
| F8c | Probe-F8-Taq | Fam- TGG CTA CAT AAT GGA TAC ACT ACC TGG CT -BHQ1 | 153785522 |
| | F-F8Ex17-18-Taq | AGA GAA TTA TCG CTT CCA TGC A | 153785765 |
| | R-F8Ex17-18-Taq | CAG ATA CCA TCG AAT CCT TTG ATC | 153785484 |
| PBGD | Probe-PBGD-Taq | Joe- CAA CGG CGG AAG AAA ACA GCC -BHQ1 | 118460979 |
| | F-PBGD-Taq | GGT AAC GGC AAT GCG GC | 118460960 |
| | R-PBGD-Taq | CCC ACG CGA ATC ACT CTC AT | 118464198 |

Table S 17: Primers used for investigating the effect of intronic variations on *F8* mRNA expression.

BHQ1= 3'-terminaler BlackHol

| Analyzed region | *F8* region | Name | Sequence 5' to 3' | Location |
|---|---|---|---|---|
| SNP1 SNP2 | Intron 13 | qRT-ex13-F-N | CTA TGA AGA CAC ACT CAC CCT | 154176020 |
| | | qRT-ex14-R-N | AAC TAG AAA CCT TCA GTA AG | 154159896 |
| | | Probe-ex13-14-F | FAM-ACC CAG GTC TAT GGA TTC TGG GGT G-BHQ1 | 154159934 |
| SNP3 | Intron 16 | qRT-ex16-F | AGA TGA GTT TGA CTG CAA AG | 154133115 |
| | | qRT-ex17-R | CCT GTA CTG TCA CTT GTC TCC | 154132726 |
| | | Probe-ex16-17-F | FAM-TGA TGT TGA CCT GGA AAA AGA TGT G-BHQ1 | 154133074 |
| SNP4 SNP5 | Intron 18 | qRT-ex17-F | AGA GAA TTA TCG CTT CCA TGC A | 154132676 |
| | | qRT-ex18-R | CAG ATA CCA TCG AAT CCT TTG ATC | 154132313 |
| | | Probe-ex18-F | FAM-TGG CTA CAT AAT GGA TAC ACT ACC TGG C-BHQ1 | 154132328 |
| SNP6 | Intron 21 | qRT-ex20-21-F | CAG GAC AAT ATG GAC AGT GGG | 154129646 |
| | | qRT-ex22-R | GGG TCT TGA TGC CGT GAA TA | 154124465 |
| | | Probe-ex21-22-F | FAM-TCT TGG ATC AAG GTG GAT CTG TTG G-BHQ1 | 154128206 |
| SNP7 SNP8 | Intron 22 | qRT-ex22-F | GTG GCA GAC TTA TCG AGG AAA | 154124371 |
| | | qRT-ex23-R | CGA ATG CTA TAA TGA GTT GGG TG | 154091421 |
| | | Probe-ex22-23-F | FAM-GAA CCT TAA TGG TCT TCT TTG GCA ATG-BHQ1 | 154091358 |
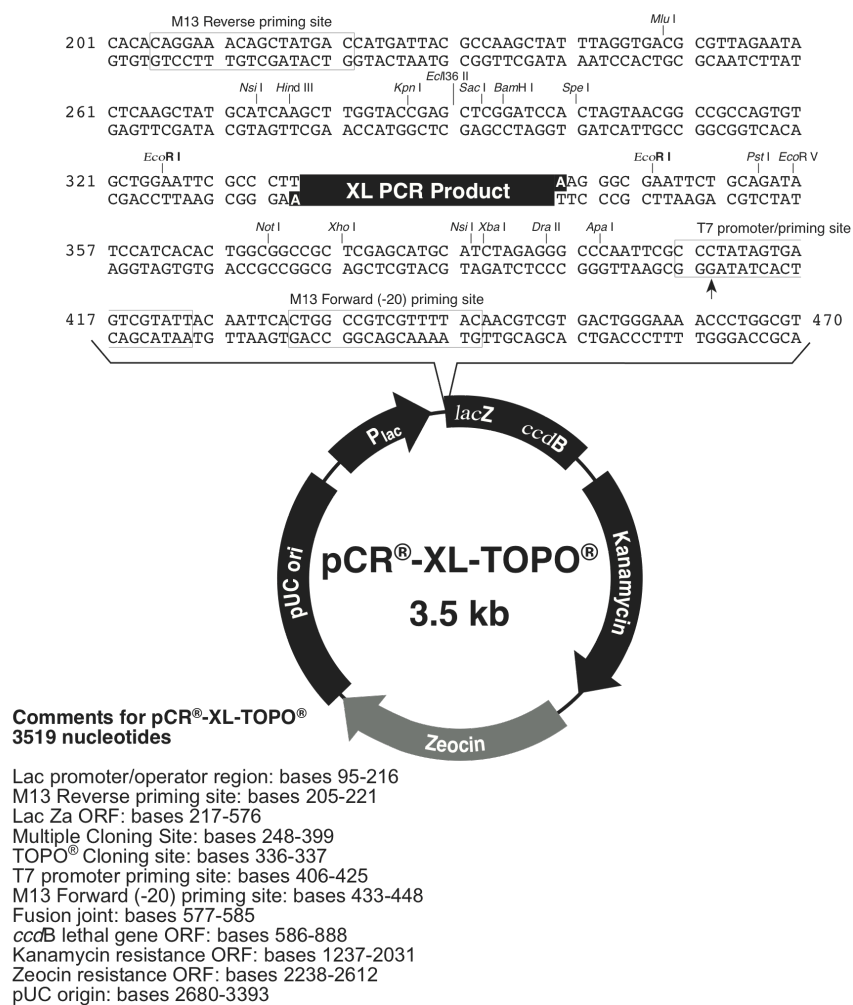
Figure S 1: Map of *pCR*-XL-TOPO® plasmid.

(Reference: https://products.invitrogen.com/ivgn/product/K470010)

**List of publications:**

- *Hämostaseologie, Volume: 30, Issue: 4A/2010, Supplement: 1, S158-S161*

  *Long-range PCR screening for large rearrangements in the FVIII gene in patients without detectable mutations in the coding sequence*
  **Pezeshkpoor B**., Nüsgen N., Dermer H., Vidovic N., Niemann B., Pavlova A., Oldenburg J., El-Maarri O.

- *Hämostaseologie, Volume: 30, Issue: 4A/2010, Supplement: 1, S162-S163*

  *F8a*, *F8b* and *F8c* expression and its association with 3 SNPs within the *F8* gene in healthy controls and HA patients with no mutation.
  O. El-Maarri; **B. Pezeshkpoor**; N. Nüsgen; J. Müller; J. Oldenburg

- *Journal of Thrombosis and Haemostasis*, 10: 1600–1608,*2012*

  Identification of a third rearrangement at Xq28 that causes severe hemophilia A as a result of homologous recombination between inverted repeats.
  **Pezeshkpoor B**., Rost, S., Oldenburg, J. and El-Maarri, O*.*

**Prices:**

- Winner of poster price at 56. Jahrestagung der Gesellschaft für Thrombose und Hämostaseologie-Forschung (GTH), st.gallen 2012.

**Behnaz Pezeshkpoor**, Simone Rost, Johannes Oldenburg, Osman El-Maarri.
***Titel:*** Genomic rearrangement between two homologous repeats at Xq28 disrupts F8 gene in a severe Haemophila A patient.

- Winner of poster price at Hämophilie Symposium, Hamburg 2012.

**Pezeshkpoor B**., Zimmer N., Marquardt N., Muller J., Nanda I., Haaf T., Budde U., Oldenburg J. and El-Maarri O.
***Titel:*** *Deep intronic 'mutations' cause mild hemophilia A: Application of Next Generation Sequencing in Patients without Detectable Mutation in F8 cDNA.*

**Oral presentations:**

- 40. Hämophilie Symposium, Hamburg 2009

*Titel: Long-range PCR screening for large rearrangements in patients without mutations in the F8 cDNA*

- Annual retreat of NRW international graduate school (Biotech Pharma), 2010

*Titel: Investigation of Mechanisms causing F8 Deficiency in Haemophilia A patients without Mutation in the F8 cDNA*

- Annual retreat of NRW international graduate school (Biotech Pharma), 2011

*Titel: Investigation of Mechanisms causing F8 Deficiency in Haemophilia A patients without Mutation in the F8 cDNA (update of PHD work)*

- 44. Jahreskongress der deutschen Gesellschaft für Transfusionsmedizin und Immunhämatologie (DGTI), Hannover 2011

*Titel: Factor VIII locus itself likely explains all hemophilia A cases without mutation in the F8 cDNA*


**Posters:**

- 40. Hämophilie Symposium, Hamburg 2009.

**Pezeshkpoor B.**, Nüsgen N., Muller J., El-Maarri O., Oldenburg J.

*Titel: F8a. F8b and F8c Expression levels and its association with three major SNPs (D1241E. S1269S and int7 G to A) within the F8 gene in healthy controls and in Hemophilia A patients with no mutations in the F8 cDNA.*

- 54. Jahrestagung der Gesellschaft für Thrombose und Hämostaseologie-Forschung GTH-NVTH, Nürnberg 2010.

**Pezeshkpoor B**, Nüsgen N, Dermer H, Vidovic N, Niemann B, Pavlova A, Oldenburg J and El-Maarri O.

*Titel: Long-range PCR screening for large rearrangements in patients without detectable mutations in the F8 coding sequence.*

- WFH (World Federation of Hemophilia) congress. Buenos Aires 2010.

**Behnaz Pezeshkpoor**, Nicole Nüsgen, Natacha Vidovic, Barbara Niemann, Anni Pavlova, Johannes Oldenburg and Osman El-Maarri.

*Titel: Long-range PCR screening for large rearrangements in the F8 gene in patients without detectable mutations in the coding sequence.*

- 41. Hämophilie Symposium, Hamburg 2010.

**Pezeshkpoor B.,** Vidovic N., Nusgen N., Niemann B., Nanda I., Haaf T., Budde U., Oldenburg J. and El-Maarri O.

*Titel: Investigation of putative determinants of F8 deficiency in Haemophilia A patients without detectable mutation in F8 cDNA.*

- 55. Jahrestagung der Gesellschaft für Thrombose und Hämostaseologie-Forschung Wiesbaden 2011.

**Pezeshkpoor B.,** Vidovic N., Nusgen N., Niemann B., Nanda I., Haaf T., Budde U., Oldenburg J. and El-Maarri O.

***Titel:*** *Investigation of putative determinants of F8 deficiency in Haemophilia A patients without detectable mutation in F8 cDNA.*

- XXIII Congress of the International Society on Thrombosis and Haemostasis. Kyoto. 2011

**Pezeshkpoor B.**, Vidovic N., Nüsgen N., Nanda I., Haaf T., Budde U., Oldenburg J. and El-Maarri O.
***Titel:*** *F8 locus itself seems to explain all cases of hemophilia A patients without mutations in the F8 cDNA.*

- 42. Hämophilie Symposium, Hamburg 2011.

**Behnaz Pezeshkpoor**, Simone Rost, Johannes Oldenburg, Osman El-Maarri.
***Titel:*** *Genomic rearrangement between two homologous repeats at Xq28 disrupts F8 gene in a severe Haemophila A patient.*

- 56. Jahrestagung der Gesellschaft für Thrombose und Hämostaseologie-Forschung St.gallen, GTH 2012.

**Behnaz Pezeshkpoor**, Simone Rost, Johannes Oldenburg, Osman El-Maarri.
***Titel:*** *Genomic rearrangement between two homologous repeats at Xq28 disrupts F8 gene in a severe Haemophila A patient.*

- Perspectives in cell and gene-based medicines (CGM), Frankfurt. 2012

**Behnaz Pezeshkpoor**, Simone Rost, Natasha Vidovic, Nicole Zimmer, Johannes Oldenburg and Osman El-Maarri.
***Titel:*** *F8 Locus incorporates Causal Mutations Leading to F8 Deficiency in Patients without Detectable Mutation in the F8 cDNA.*

- WFH (World Federation of Hemophilia) congress, Paris. 2012.

**Behnaz Pezeshkpoor**, Simone Rost, Johannes Oldenburg and Osman El-Maarri.
***Titel:*** *Identification of a third rearrangement at Xq28 that causes severe hemophilia A due to homologous recombination between inverted repeats.*

- WFH (World Federation of Hemophilia) congress, Paris. 2012.

**Behnaz Pezeshkpoor**, Nicole Zimmer, Natasha Vidovic, Johannes Oldenburg and Osman El-Maarri.
***Titel:*** *Absence of active FVIII protein despite high F8 mRNA expression levels in a severe Hemophilia A patient.*

- Hämophilie Symposium, Hamburg 2012.

**Pezeshkpoor B**., Zimmer N., Marquardt N., Muller J., Nanda I., Haaf T., Budde U., Oldenburg J. and El-Maarri O.
***Titel:*** *Deep intronic 'mutations' cause mild hemophilia A: Application of Next Generation Sequencing in Patients without Detectable Mutation in F8 cDNA.*

- 57. Jahrestagung der Gesellschaft für Thrombose und Hämostaseologie-Forschung St.gallen, München 2012.

**Pezeshkpoor B**., Zimmer N., Marquardt N., Muller J., Nanda I., Haaf T., Budde U., Oldenburg J. and El-Maarri O.
***Titel:*** *Deep intronic 'mutations' cause mild hemophilia A*

# Figure legend

**List of tables:**

**List of supplementary tables and figures:**