Accepted Manuscript online: 2023-11-27 Article published online: 2023-12-26

Complete *F9* Gene Deletion, Duplication, and Triplication Rearrangements: Implications for Factor IX Expression and Clinical Phenotypes

YuXin Ma[1,2,*] Yang Li[2,*] Jie Sun[3] Qian Liang[2] Runhui Wu[3] Qiulan Ding[2,4] Jing Dai[2,5]

[1] Department of Clinical Laboratory, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China
[2] Department of Laboratory Medicine, Shanghai Jiaotong University School of Medicine, Ruijin Hospital, Shanghai, China
[3] Haemophilia Comprehensive Care Center, Capital Medical University, Beijing Children's Hospital, Beijing, China
[4] Collaborative Innovation Center of Hematology, Shanghai Jiaotong University School of Medicine, Shanghai, China
[5] Department of Laboratory Medicine, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

Address for correspondence Qiulan Ding, MD, Department of Laboratory Medicine, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, No. 197 Ruijin Second Road, Shanghai, 200025, China (e-mail: qiulan_ding@126.com).

Jing Dai, MD, Department of Laboratory Medicine, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, No. 197 Ruijin Second Road, Shanghai, 200025, China (e-mail: dj40572@rjh.com.cn).

Thromb Haemost 2024;124:374–385.

Abstract

Background Factor IX (FIX) plays a critical role in blood coagulation. Complete deletion of *F9* results in severe hemophilia B, whereas the clinical implications of complete *F9* duplication and triplication remain understudied.

Objective To investigate the rearrangement mechanisms underlying complete *F9* deletion (cases 1 and 2), duplication (cases 3 and 4), and triplication (case 5), and to explore their association with FIX expression levels and clinical impacts.

Methods Plasma FIX levels were detected using antigen and activity assays. CNVplex technology, optical genome mapping, and long-distance polymerase chain reaction were employed to characterize the breakpoints of the chromosomal rearrangements.

Results Cases 1 and 2 exhibited FIX activities below 1%. Case 3 displayed FIX activities within the reference range. However, cases 4 and 5 showed a significant increase in FIX activities. Alu-mediated nonallelic homologous recombination was identified as the cause of *F9* deletion in case 1; FoSTeS/MMBIR (Fork Stalling and Template Switching/microhomology-mediated break-induced replication) contributed to both *F9* deletion and tandem duplication observed in cases 2 and 3; BIR/MMBIR (break-induced replication/microhomology-mediated break-induced replication) mediated by the same pair of low-copy repeats results in similar duplication–triplication/inversion–duplication (DUP–TRP/INV–DUP) rearrangements in cases 4 and 5, leading to complete *F9* duplication and triplication, respectively.

Conclusion Large deletions involving the *F9* gene exhibit no apparent pattern, and the extra-hematologic clinical phenotypes require careful analysis of other genes within the deletion. The impact of complete *F9* duplication and triplication on FIX expression

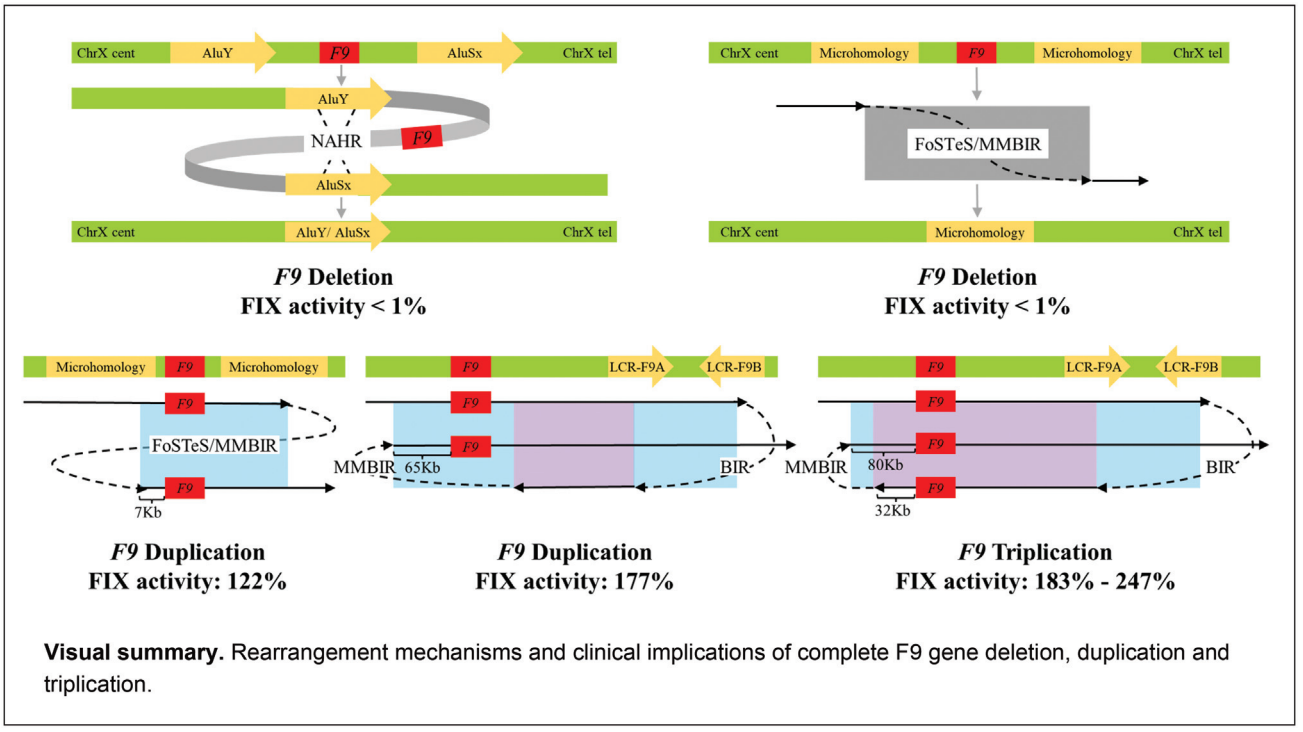
Keywords
- factor IX
- genomic structural variation
- hemophilia B
- low-copy repeats
- thrombosis

* *These authors contributed equally as co-first authors.*

received
September 8, 2023
accepted after revision
November 21, 2023
accepted manuscript online
November 27, 2023
article published online
December 26, 2023

DOI https://doi.org/10.1055/a-2217-9837.
ISSN 0340-6245.

Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany



Visual summary. Rearrangement mechanisms and clinical implications of complete F9 gene deletion, duplication and triplication.

might depend on the integrity of the *F9* upstream sequence and the specific rearrangement mechanisms. Notably, DUP–TRP/INV–DUP rearrangements significantly elevate FIX activity and are closely associated with thrombotic phenotypes.

Introduction

The *F9* gene, located on the X-chromosome at position q27.1, spans 34 kb and consists of eight exons and seven introns. This gene encodes coagulation factor IX (FIX), which plays a crucial role in activating downstream coagulation factors and forming stable fibrin clots.[1] Deficiency of FIX, known as hemophilia B (HB), is an X-linked recessive bleeding disorder associated with reduced plasma FIX activity.[2] Conversely, elevated FIX levels have been identified as an essential risk factor for venous thromboembolism (VTE).[3–5] The risk of VTE is significantly increased in individuals with FIX levels exceeding 150%, showing a more than 3-fold increase compared with those within the reference range of FIX levels.[4] Furthermore, specific variants such as FIX Padua (p.Arg338Leu) and FIX Shanghai (p.Arg338Gln) have been identified as a cause of juvenile thrombophilia due to significantly increased FIX activity and excessive clotting risk.[6,7]

The Factor IX Gene Variant Database (https://www.factorix.org/) serves as a comprehensive resource for *F9* variants, encompassing 1692 unique variants documented from 5,358 individual case reports.[8] While a significant portion of these variants are point mutations, advancements in understanding chromosomal structural rearrangements have revealed an increasing proportion of larger structural variants (SVs). Among these SVs, the database identified 43 types of gross deletions, with complete deletion of the *F9* gene observed in 92 patients across 36 articles.[8] In contrast, only four cases of gross duplications involving *F9* have been reported, all of which entail partial duplications of specific exons (exons 1–6, exons 2–6, exons 4–6, and exons 5–8).[9–11] Deletions of the *F9* gene often produce straightforward and severe consequences on FIX activity, leading to severe HB. However, the impact of these partial *F9* duplications on FIX activity varies, with reported levels ranging from below 1 to 8% of normal activity levels.[11] Notably, complete duplications of the *F9* gene have not been reported in this database, and the impact of whole-gene duplication on FIX expression remains unknown. Currently, most of the information regarding complete duplications of *F9* is incidentally reported during studies on large X chromosomal rearrangements.[12–14] Unfortunately, there is a lack of available clinical phenotype information associated explicitly with bleeding and thrombotic manifestations in individuals with complete *F9* duplications.

Understanding the specific breakpoints of *F9*-related SVs can provide valuable insight into the underlying rearrangement mechanisms and help to predict FIX expression and associated clinical phenotypes. Previous studies have suggested the involvement of DNA repair-based mechanisms, such as nonhomologous end joining and nonallelic homologous recombination (NAHR), and DNA replication-based mechanisms like Fork Stalling and Template Switching/

microhomology-mediated break-induced replication (FoSTeS/MMBIR) in *F9*-related SVs.[15–17] However, recurrent NAHR rearrangements mediated by low-copy repeats (LCRs), similar to the intron 22 inversion of *F8*, have not been identified in structural rearrangements involving *F9* to date.

The widespread adoption of next-generation sequencing (NGS) allows for the precise identification of some SV breakpoints by relying on the presence of chimeric reads spanning the junction. However, due to the limitations of short reads (150∼300 bp), NGS would generate ambiguous alignments in highly repetitive regions. Interestingly, these regions are more prone to forming SVs.[18] To overcome the limitations of NGS, long-read sequencing technologies such as PacBio and Oxford Nanopore Technologies have been commercialized and hold promise.[19] Nevertheless, the costs associated with achieving the necessary coverage levels for acceptable accuracy with long-read sequencing are exceedingly high, limiting their widespread use in breakpoint characterization. Recently, optical genome mapping (OGM) has recently emerged as a robust, high-throughput technology for detecting SVs.[20] OGM provides a more comprehensive and accurate analysis of SVs by directly analyzing long-range information from single DNA molecules. To date, OGM has been successfully applied in mapping SVs associated with hematologic malignancies, hemophilia A, and other genetic diseases.[21,22]

In this study, we combined OGM with long-distance polymerase chain reaction (LD-PCR) techniques to characterize the breakpoints of two severe HB patients with complete *F9* gene deletion and three VTE patients with complete *F9* duplication and triplication. Our findings reveal a shared duplication–-triplication/inversion–duplication (DUP–TRP/INV–DUP) rearrangements, resulting in *F9* duplication and triplication in two unrelated families. Additionally, we elucidate the underlying structural basis of these rearrangements. To our knowledge, this is the first report of complete *F9* duplication and triplication and their effects on FIX expression, which sheds crucial new light on *F9*-related SVs and their potential relationship with thrombotic phenotypes.

Methods

Subjects

Cases 1 and 2 belonged to the severe HB patients cohort admitted to the Beijing Children's Hospital from 2015 to 2022, who were previously reported to have a deletion spanning all eight exons of the *F9* gene.[23] Cases 3, 4, and 5 belonged to the VTE patients cohort registered in Shanghai Ruijin Hospital from 2017 to 2023. They were identified with additional copies of the complete *F9* gene through our Hemostasis and Thrombosis Gene Panel analysis. Informed consent was obtained from all patients and their families, and the study was approved by the Ethics Committee of Ruijin Hospital.

Plasma Analysis

Peripheral blood was collected from patients and their related family members using vacuum blood collection tubes containing 0.129 mol/L sodium citrate to obtain platelet-poor plasma. Coagulation factors activity levels, including FIX, were assessed using a one-stage clotting method on an automated ACL TOP analyzer (Instrumentation Laboratory, United States). FIX antigen levels were measured using an enzyme-linked immunosorbent assay following the manufacturer's instructions (Enzyme Research Laboratories, South Bend, Indiana, United States). Additional thrombophilia tests, including the activity levels of protein C (PC), protein S (PS), antithrombin III, and lupus anticoagulant, were performed as described previously.[24]

Hemostasis and Thrombosis Gene Panel Analysis

Five probands were subjected to the panel analysis, covering 33 genes associated with hemorrhage and thrombosis (►**Supplementary Table S1**, available in the online version). Copy number variants (CNVs) were evaluated using CNVplex technology, a high-throughput multiplex CNV analysis (Genesky Biotechnologies Inc., Shanghai, China).[25] Quantitative probes targeting the *F9* gene can be found in ►**Supplementary Table S2** (available in the online version). Additionally, targeted next-generation sequencing was employed to detect mutations in the promoter region, 5′ and 3′ untranslated regions, coding sequence, and exon–intron boundaries of these genes. The identified mutations were further validated using Sanger sequencing in the probands and their remaining family members.

Optical Genome Mapping Analysis

High-molecular weight (HMW) DNA was extracted from the blood samples of five probands using the Bionano Prep SP Blood and Cell Culture DNA Isolation Kit (Bionano Genomics, United States). The integrity and size of extracted HMW DNA were confirmed by pulsed-field gel electrophoresis. The DNA backbone was stained overnight with the Bionano Prep DLS Kit, where the DLE-1 enzyme specifically recognized the CTTAAG base sequence on genomic DNA and labeled them with the DL-Green fluorophore group. Then, Bionano Saphyr Instruments imaged genomic DNA molecules on a chip containing up to 120,000 nanochannels, each detecting DL-Green fluorescent signals within the blue DNA background. High-resolution charge-coupled device cameras generated real-time imaging data for all nanochannels, automatically converting them into molecular data and integrating them into a genome-wide optical map. De novo assembly analysis was performed using Bionano access V1.7 software, focusing on SVs within the X-chromosome *F9* gene and its nearby regions. Additionally, the Bionano software incorporates a database of SVs found in 180 phenotypically normal individuals, allowing for filtering SVs polymorphisms in the population.

Long-Distance Polymerase Chain Reaction and Breakpoint Sequencing

Genomic DNA was extracted using the QIAamp DNA blood purification kit (Qiagen, Venlo, The Netherlands). Based on the assembly results from Bionano, amplification primers were designed near the DL-Green fluorescent labels flanking the breakpoint. LD-PCR amplification was performed with the KOD One PCR Master Mix kit (TOYOBO, Shanghai, China) according to the manufacturer's instructions. The extension time was determined by the distance between the two fluorescent labels, with 5 seconds per kb for a range of 1 to 10 kb and 10 seconds per kb for

over 10 kb. After confirming the formation of LD-PCR amplified productions through agarose gel electrophoresis, smaller-spaced sequencing primers were designed between the amplification primers for Sanger sequencing. The sequences of the amplification and sequencing primers are provided in ►**Supplementary Tables S3** and **S4** (available in the online version).

X-chromosome Inactivation Pattern Detecting

The methylation of androgen receptor (*AR*) and retinitis pigmentosa 2 (*RP2*) genes were analyzed to assess X-chromosome inactivation (XCI) patterns, as described previously.[26] Genomic DNA of peripheral blood lymphocytes from female individuals with *F9* copy number variations was digested with the methylation-sensitive enzyme *Hha*I (New England Biolabs, United States). *Hha*I recognition sites are located near the CAG repeat region in *AR*, and the GAAA repeat region upstream of *RP2*. The digested DNA was amplified using specific fluorescent primers (►**Supplementary Table S5**, available in the online version) on the 2720 Thermal Cycler (Applied Biosystems, United States). Capillary electrophoresis was performed on the 3730xl genetic analyze (Applied Biosystems, United States) to analyze the amplified fragments. The peak fluorescence values for allele determination were calculated using the Gene Mapper 5.0 software (Applied Biosystems, United States). To eliminate the effect of allele length difference on the X-chromosome inactivation ratio (XCIR), the following formula was employed: $XCIR = (d1/u1)/(d1/u1 + d2/u2)$, where d and u represent the peak fluorescence value corresponding to the genomic DNA after and before digestion, respectively.[27] Each XCIR value was calculated as an average of two replicate measurements.

Bioinformatic Analysis

Breakpoint analysis was performed using a sequencing alignment-based approach with the hg38 reference genome. Sanger sequencing data containing the breakpoints and potential LCRs near the breakpoints were analyzed using the Human BLAT Search Tool (http://genome.ucsc.edu/cgi-bin/hgBlat). The Repeat Masker Tool (www.repeatmasker.org) was used to identify repetitive elements flanking the breakpoint junction. This tool can identify well-known repetitive elements associated with rearrangement mechanisms, such as long interspersed nuclear elements, short interspersed nuclear elements (SINEs), long terminal repeats, and DNA transposons. To assess clinical relevance of the identified variants, we consulted five databases: dbVAR (https://www.ncbi.nlm.nih.gov/dbvar/), ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), HGMD (http://www.hgmd.cf.ac.uk/ac/index.php), gnomAD (https://gnomad.broadinstitute.org/), and PubMed (http://www.ncbi.nlm.nih.gov/pubmed).

Results

Case Reports

This study involved the analysis of five unrelated families comprising a total of 12 individuals with *F9* copy number variations (►**Table 1**).

Table 1 Clinical characteristics and laboratory data of the family members

Participants	Sex	Clinical phenotypes	*F9* copy numbers	FIX: Ac (%)	FIX: Ag (%)	Additional variants in Hemostasis and Thrombosis Gene Panel	Corresponding protein activity (%)	XCIR (%)
Case 1								
Mother	F	Hemorrhage	1	22	16	ND	ND	58
Proband 1	M	Hemorrhage	0	<1	<1	NO	ND	ND
Case 2								
Mother	F	Hemorrhage	1	12	7	NO	ND	64
Sister	F	Asymptomatic	1	65	60	NO	ND	50
Proband 2	M	Hemorrhage	0	<1	<1	*VWF*: c.4499C > T, p.Ala1500Val	ND	ND
Case 3								
Mother	F	Asymptomatic	3	122	118	*PROS1*: c.200A > C, p.Glu67Ala	PS: Ac: 53	50
Proband 3	M	PE, CVST, MVT	2	*	*	*PROS1*: c.200A > C, p.Glu67Ala	*	ND
Case 4								
Mother	F	Asymptomatic	3	177	153	*F5*: c.1723C > T, p.Arg575Cys	FV: Ac: 101	54
Proband 4	M	CVST	2	*	*	*F5*: c.1723C > T, p.Arg575Cys	FV: Ac: 87	ND
Case 5								
Grandmother	F	Asymptomatic	4	236	217	*PROC*: c.1151A > G, p.Asn384Ser	PC: Ac: 84	52
Mother	F	Asymptomatic	4	182	178	*PROC*: c.1151A > G, p.Asn384Ser	PC: Ac: 83	53
Proband 5	M	DVT, MVT	3	247	238	*PROC*: c.1151A > G, p.Asn384Ser	PC: Ac: 63	ND

Abbreviations: FII, factor II; FV, factor V; FVII, factor VII; FIX, factor IX; FX, factor X; PS, protein S; PC, protein C; Ac, activity; Ag, antigen; F, female; M, male; PE, pulmonary embolism; CVST, cerebral venous sinus thrombosis; MVT, mesenteric vein thrombosis; DVT, deep vein thrombosis; NO, no carrier; ND, no detection; XCIR, X chromosome inactivation rate. Reference range: FII: Ac, FV: Ac, FVII: Ac, FIX: Ac, FX: Ac: 50%-150%; PS: Ac: 60%-130%; PC: Ac: 70%-140%.
*indicates that the test was conducted during the patient's administration of warfarin. For proband 3, FII: Ac, FVII: Ac, FIX: Ac, FX: Ac, and PS: Ac were 22%, 48%, 77%, 11%, and 24%, respectively. FIX: Ag was 80%. For proband 4, FII: Ac, FVII: Ac, FIX: Ac, FX: Ac were 15%, 35%, 69%, and 14%, respectively, and FIX: Ag was 78%.

Proband 1 and proband 2 presented with skin ecchymosis at around 6 months of age and were diagnosed with severe HB. During prothrombin complex concentrate infusion for prevention, both patients experienced allergic reactions in the form of rashes, which were effectively managed by suspending the infusion and employing appropriate desensitization therapy. No other abnormal clinical phenotypes were observed in either patient. Gene panel analysis revealed a complete deletion of the *F9* gene in both patients, whereas their mothers were found to have only one copy of the *F9* gene. Proband 1's mother reported a medical history including intractable epistaxis, menorrhagia, and gingival bleeding, and proband 2's mother reported excessive menstrual bleeding. Their FIX activity levels were only 22 and 12%, respectively. No additional mutations were found in the exonic regions and exon–intron boundaries of the *F9* gene, and no extremely skewed XCI was observed in either case. Furthermore, gene panel analysis detected a heterozygous *VWF* p.Ala1500Val mutation in proband 2, which might lead to a diagnosis of type 1 von Willebrand disease.[28]

Proband 3 experienced sudden chest pain at 15 and was subsequently diagnosed with pulmonary embolism based on computed tomography angiography. At 17, he developed a sudden headache and epilepsy. Magnetic resonance imaging revealed cerebral venous sinus thrombosis (CVST). At 26, he developed mesenteric vein thrombosis. Subsequently, he received long-term anticoagulant therapy with warfarin. Gene panel analysis identified a heterozygous *PROS1* p.Glu67Ala mutation in the patient, along with complete duplication of the *F9* gene. Both the *PROS1* variant and *F9* gene duplication were inherited from his asymptomatic mother. Under treatment with warfarin, the activities of vitamin K-dependent coagulation factors, including FIX and PS, were significantly suppressed in the patient. However, as demonstrated by his mother, the *PROS1* p.Glu67Ala mutation significantly reduced the PS activity, whereas the *F9* duplication did not result in a notable increase in FIX activity (►**Table 1**).

Proband 4 had a history of recurrent CVST involving the transverse sinus and right sigmoid sinuses. At 16, he experienced a brain hemorrhage in the left parietal lobe during one episode of CVST. Subsequently, the patient was prescribed warfarin treatment. Gene panel analysis revealed that the patient inherited both a duplication of the *F9* gene and a heterozygous *F5* p.Arg575Cys mutation from his asymptomatic mother. During warfarin therapy, the patient's FIX activity decreased to 62% of the normal level. However, his mother exhibited an elevated FIX activity of 177% due to the *F9* gene duplication. Interestingly, the *F5* variant present in proband 4 and his mother did not affect their factor V (FV) activity, which measured at 101 and 87%, respectively (►**Table 1**).

Proband 5 was diagnosed with mesenteric vein, splenic vein, and bilateral iliac vein thrombosis at age 15 and underwent long-term treatment with rivaroxaban for thromboprophylaxis. However, despite the treatment, the patient's symptoms did not significantly improve. Gene panel analysis revealed that the patient, his mother, and his grandmother all carried a triplication of the *F9* gene and a heterozygous *PROC* p.Asn384Ser mutation. After 1 week of drug discontinuation, protein activity assays were performed. While PC activity was within the normal range for his mother and grandmother, his PC activity was slightly below the lower limit of the normal range, measuring at 63%. Additionally, the three patients exhibited significantly elevated FIX activity due to the *F9* triplication. FIX activity levels ranged from 182 to 247% (►**Table 1**).

Unravelling Chromosomal Rearrangements with Optical Genome Mapping

To determine the specific size and orientation of the rearranged regions, we conducted an OGM analysis on five cases (►**Fig. 1**). Results revealed that the deletion region in case 1 reached 2.88 Mb, affecting 13 OMIM genes, including *F9*. In case 2, the deletion fragment was smaller, measuring 0.20 Mb and involving exons 2 to 26 of the *MCF2* gene and complete *F9* gene (►**Table 2**).

Case 3 exhibited a duplicated fragment with a length of 0.43 Mb, encompassing three complete OMIM genes: *F9*, *MCF2*, and *ATP11C*. Furthermore, OGM matching results only exhibited a single breakpoint, suggesting a tandem orientation for this duplication. Notably, cases 4 and 5 displayed similar rearrangement patterns, referred to as DUP–TRP/INV–DUP rearrangements (►**Fig. 1C**). In brief, a triplicated fragment was inserted in reverse orientation onto the same chromosome through two breakpoints. Interestingly, the downstream duplicated regions were completely identical, suggesting a shared rearrangement mechanism. We identified a pair of inverted LCRs (designated LCR-F9A and LCR-F9B, respectively) with approximately 16 kb length in this region. This pair of LCRs contained three highly homologous core sequences with over 98.5% identity: XIRT-31240, XIRT-31246, and XIRT-31253.[29] Within the 0.23 Mb region between LCR-F9A and LCR-F9B in both cases 4 and 5, duplications were observed, but this region did not contain any OMIM genes. The difference between the two cases was that the triplication region extended 0.27 Mb upstream of LCR-F9A in case 4 but 0.51 Mb in case 5, respectively. Additionally, there was another duplicated region with a length of 0.27 and 0.05 Mb on the other side of the triplicated regions, respectively (►**Fig. 1**). The *F9* gene was located 0.46 Mb upstream of LCR-F9A, explaining why case 5 exhibited *F9* triplication, whereas case 4 only had an *F9* duplication.

Elucidating Breakpoint Sequences through Long-Distance Polymerase Chain Reaction

To determine the nucleotide sequences at the breakpoints and the specific rearrangement mechanisms, we designed LD-PCR primers to amplify the rearranged regions containing the breakpoint junctions (►**Fig. 2**). In cases 1 and 2, the P1F/P1R and P2F/P2R primers were designed to target the upstream and downstream regions of the deletion, respectively. For case 3, the forward primer was designed downstream of the *F9* gene, whereas the reverse primer was designed upstream of the *F9* gene. In cases 4 and 5, where the breakpoint junctions were inverted, both LCRs9A and LCRs9B primers were created on the DNA positive strand, whereas both P4F/P4R and P5F/P5R primers were designed on the DNA negative strand. These primer design strategies would result in PCR products in

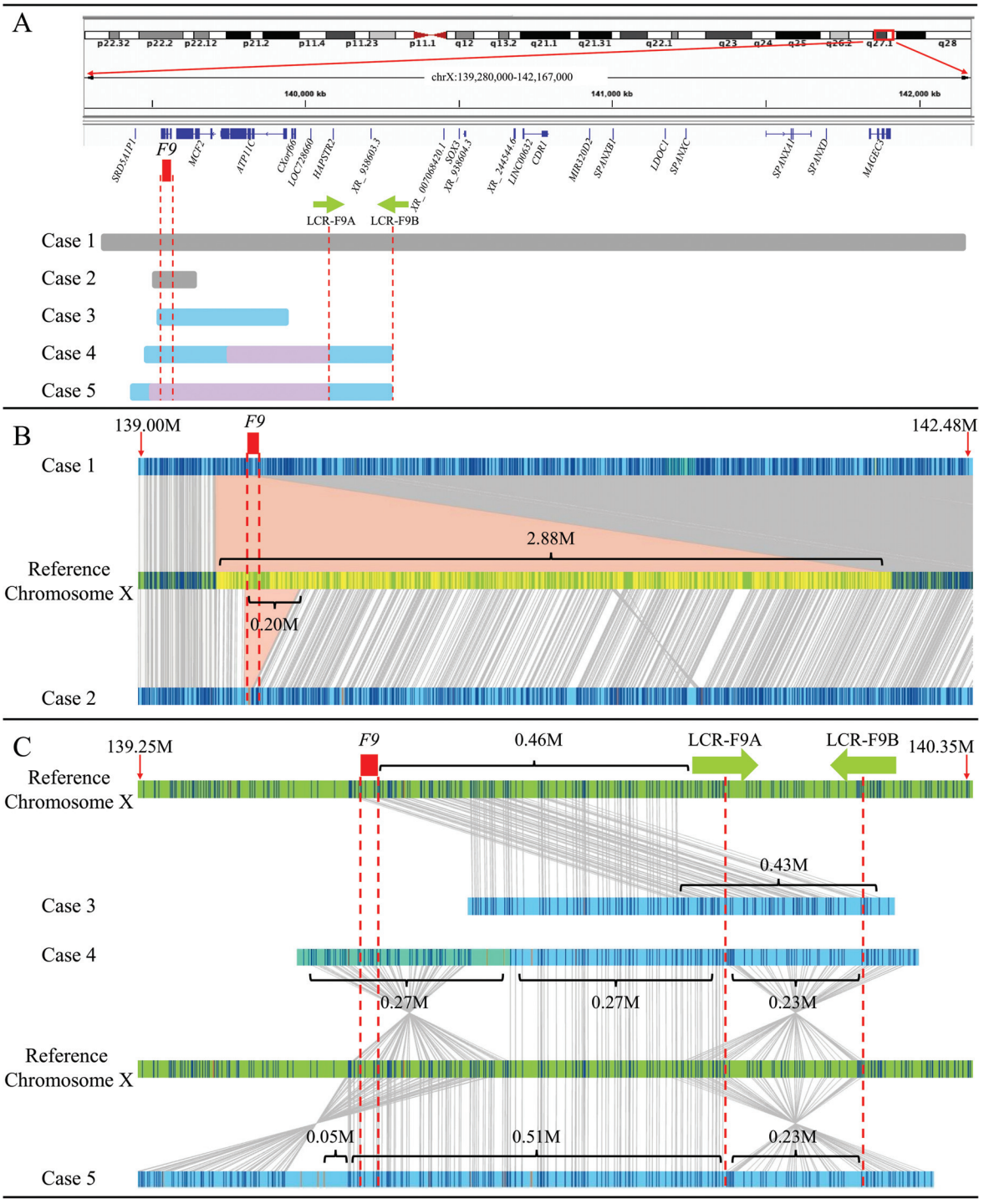


Fig. 1 The structural variations involving *F9* were identified by optical genome mapping. (**A**) Genomic region containing alterations involving *F9*. Deletion, duplication, and triplication regions are depicted as grey, blue, and purple bars, respectively. (**B**) Deletion rearrangement of *F9*. (**C**) Duplication and triplication rearrangements of *F9*. The blue bars represent rearranged X chromosomes in patients, while the green bars represent reference X chromosomes. The blue vertical lines on the X chromosome represent DL-Green fluorescent signals.

individuals carrying *F9*-related rearrangements but no product in the healthy controls.

The sequencing results across breakpoint junctions are shown in ►**Fig. 3** and ►**Supplementary Fig. S1** (available in the online version). In case 1, the breakpoint region exhibited approximately 86% identity with a homologous sequence spanning around 120 bp. Repeat Masker analysis revealed that both sequences constituting the breakpoint, ChrX:139282455–139282768 and ChrX:142160680–142160800, belong to SINEs, specifically AluY and AluSx subfamily, respectively. These two elements together formed a chimeric Alu element, mediating a 2.88Mb deletion encompassing *F9*. In contrast, no repetitive element was detected at the breakpoints of *F9* deletion in case 2 and *F9* duplication in case 3, with the two chromosome regions being joined solely by a 3 bp microhomology sequence. As predicted by OGM analysis, cases 4 and 5 shared the same

Table 2 Characterization of five *F9*-involved genomic rearrangements

	CNV abnormality	ChrX regions (hg38)	Sizes (Mb)	OMIM genes	Breakpoint elements	Recombination mechanism
Case 1	Deletion	139282588–142160682	2.88	*F9,MCF2,ATP11C, MIR505,SOX3, SPANXB1,LDOC1, SPANXC,SPANXA1, SPANXA2,SPANXD, MAGEC3,MAGEC1*	AluSx/AluY	NAHR
Case 2	Deletion	139500925–139700954	0.20	*F9,MCF2* (exon 2–26)	Microhomology	FoSTeS/MMBIR
Case 3	Duplication	139523429–139957837	0.43	*F9,MCF2,ATP11C*	Microhomology	FoSTeS/MMBIR
Case 4	Duplication	139465148–139731227	0.27	*F9,MCF2,ATP11C* (exon 30)	microhomology	MMBIR
	Triplication	139731227–140004210	0.27	*ATP11C* (exon 1–29)	Microhomology/LCR-F9A	BIR
	Duplication	140004210–140229413	0.23	No	LCR-F9A/LCR-F9B	BIR
Case 5	Duplication	139450301–139498754	0.05	No	Microhomology	MMBIR
	Triplication	139498754–140004210	0.51	*F9,MCF2,ATP11C*	Microhomology/LCR-F9A	BIR
	Duplication	140004210–140229413	0.23	No	LCR-F9A/LCR-F9B	BIR

Abbreviations: BIR, break-induced replication; ChrX, X-chromosome CNV, Copy number variants; FoSTeS Fork Stalling and Template Switching; LCR, low-copy repeat; MMBIR, microhomology-mediated break-induced replication. NAHR, nonallelic homologous recombination.

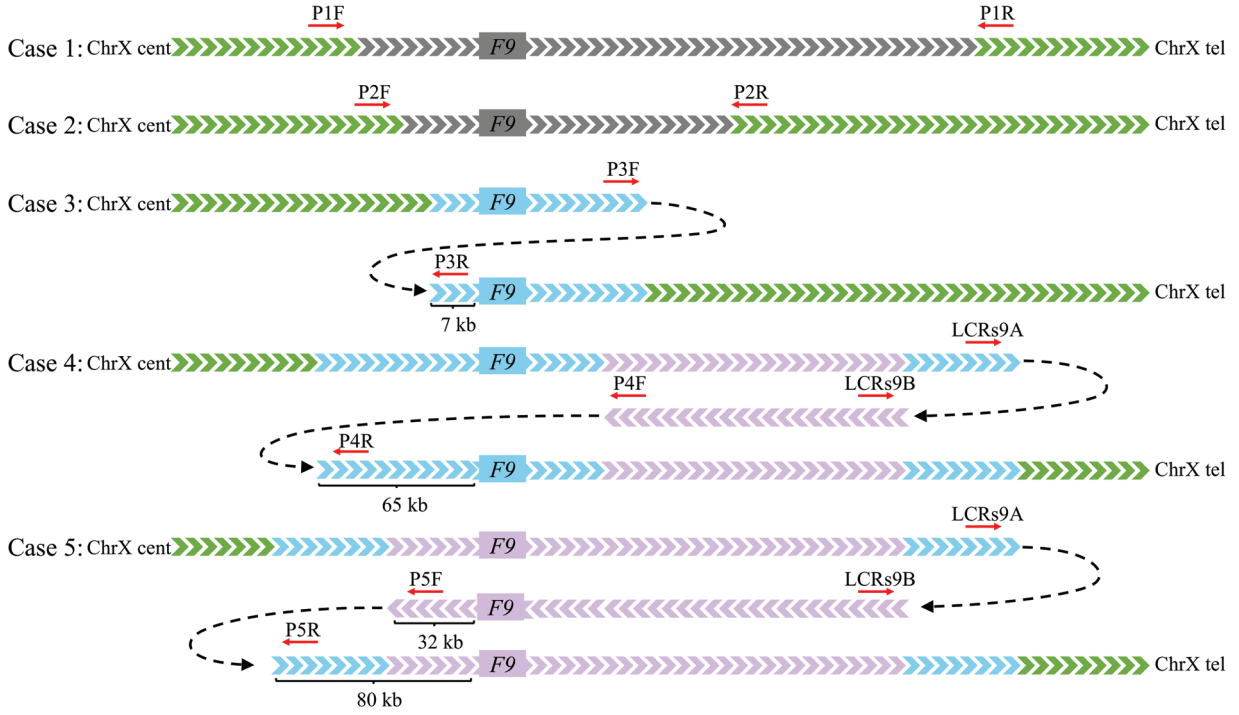


Fig. 2 Schematic representation of long-range primer design strategy. ChrX cent: X chromosome centromere; ChrX tel: X chromosome telomere. The green, grey, blue, and purple arrows represent normal copy, deletion, duplication, and triplication regions. The direction of these arrows indicates the direction of DNA synthesis. The black dashed lines represent the breakpoint junctions of chromosome rearrangement. Primers used for breakpoint identification are depicted as red arrows, with their direction indicating the amplification direction. Primer names are labeled above the arrows. Black brackets denote the region upstream of the additional *F9* copy gained after chromosome rearrangement, and the region's size is indicated below the brackets.

breakpoint, resulting from a rearrangement mediated by the hybrid LCRs element formed by LCR-F9A and LCR-F9B. Due to the length (16 kb) and high similarity (95%) between LCR-F9A and LCR-F9B, Sanger sequencing could not determine the exact rearrangement breakpoints. Therefore, the rearrangement was confirmed by a specific 17 kb PCR product unique to the LCR-F9A/LCR-F9B recombination. Different positions were observed for the other breakpoints

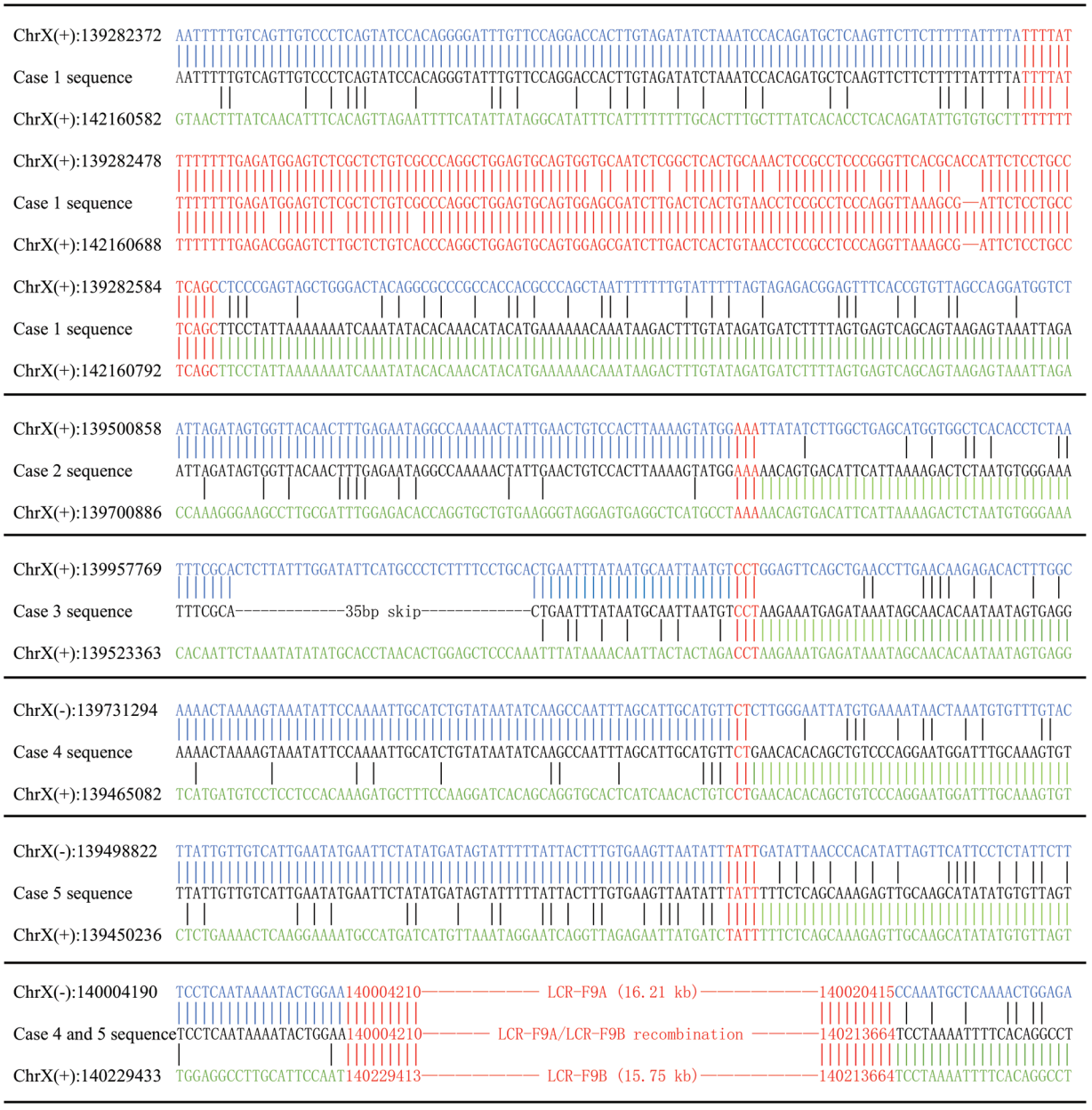


Fig. 3 Alignment of patients' rearranged sequences with the reference sequences. The top and bottom lines are the reference sequences flanking the breakpoint, while the middle line is the rearranged sequence of the patients. Homologous sequences at the breakpoints are highlighted in red. Vertical lines between bases indicate identical bases, while horizontal lines between bases indicate the absence or omission of corresponding bases.

involved in the DUP–TRP/INV–DUP rearrangement in cases 4 and 5. Again, no repetitive element was identified, and only 2 and 4-bp microhomology sequence was found at the breakpoint junction.

Discussion

Breakpoint characterization provides crucial information for understanding rearrangement mechanisms. In case 1, we identified a chimeric Alu element at the breakpoint, resulting from NAHR between an AluSx and an AluY element. The 2.88Mb chromosomal segment between these two Alu elements was deleted during recombination (►**Fig. 4A**). The rearrangements observed in cases 2 and 3 exhibit more random breakpoints flanked by only 3-bp microhomology sequences. Replication-based mechanisms such as FoSTeS/MMBIR can explain this phenomenon (►**Fig. 4B**): during DNA replication, when replication fork stalling or breakage occurs, the lagging strand disengages from the original template switches to another template via microhomology and restarts DNA synthesis.[30] When invading a new template downstream of the replication fork, a deletion similar to case 2 is generated; conversely, switching to an upstream template results in the observed tandem duplication in case 3.[31] Furthermore, these dissociations and invasions can occur multiple times.[32] As shown in case 3, the first template

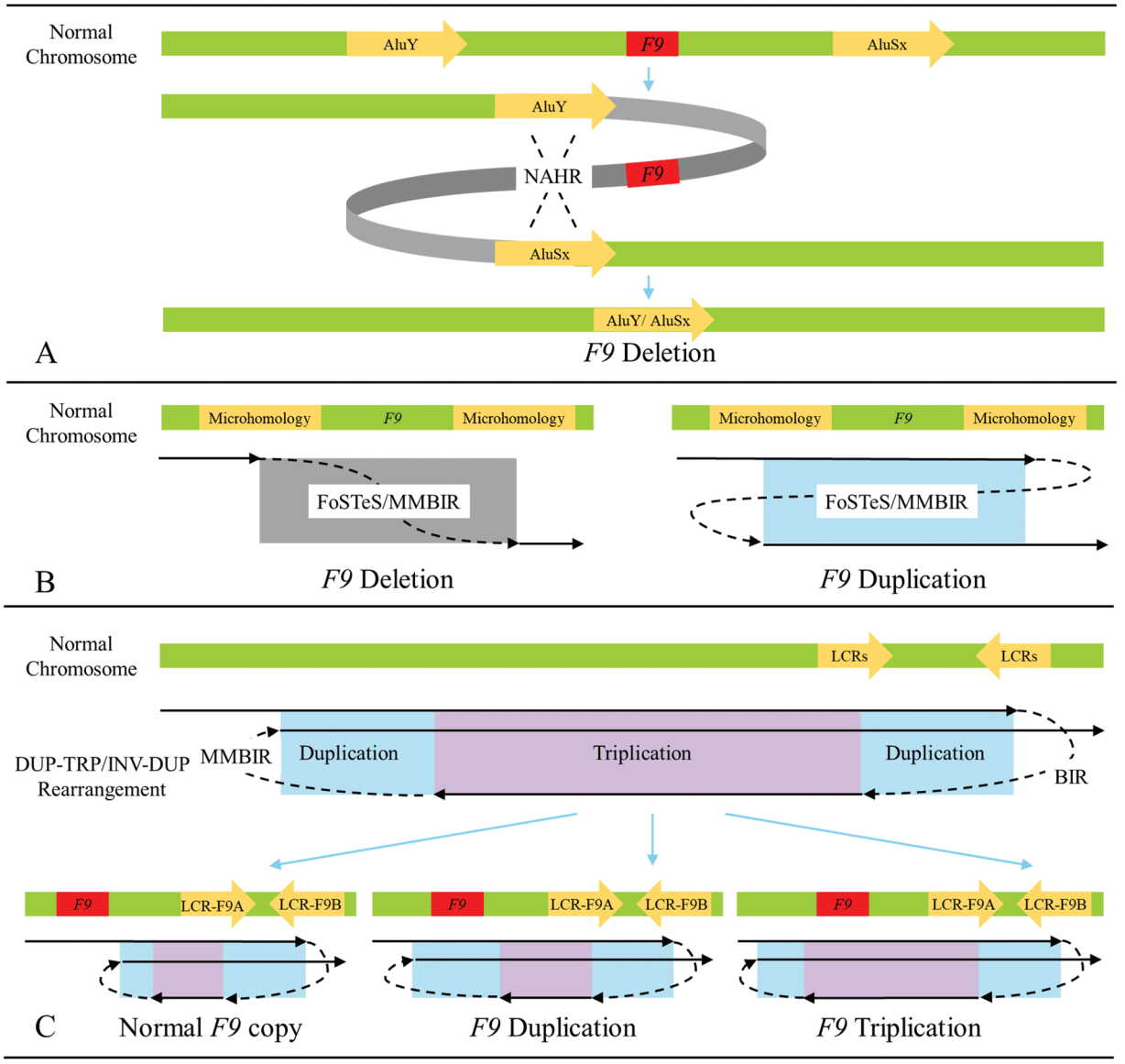


Fig. 4 Schematic of the complete *F9* deletion, duplication, and triplication genomic rearrangements. Vital elements at the breakpoint are shown in gold. Normal, deletion, duplication, and triplication regions are represented by green, grey, blue, and purple boxes. The black arrows indicate the direction of DNA synthesis. (**A**) Non-allelic homologous recombination (NAHR) events occur between similar, same-oriented AluY and AluSx pairs, resulting in complete *F9* gene deletion. (**B**) DNA replication-associated template switching processes, including fork stalling and template switching (FoSTeS) and microhomology-mediated break-induced replication (MMBIR), lead to complete *F9* gene deletion and duplication. (**C**) Break-induced replication (BIR) within low-copy repeats (LCRs), specifically LCR-F9A/LCR-F9B, and random MMBIR incidents, give rise to DUP-TRP/INV-DUP (duplication-triplication/inversion-duplication), characterized by inverted triplication located between two duplications. The length of the inverted triplication and the varying locations of MMBIR incidents contribute to the generation of normal *F9* copy, *F9* duplication, and *F9* triplication.

switch leads to a 35-bp skipping at the breakpoint, whereas the second template switch results in a tandem duplication, including the complete *F9* gene (►**Fig. 3**).

Remarkably, we have identified similar DUP–TRP/INV–DUP rearrangements in cases 4 and 5, which are mediated by a pair of inverted LCRs. Similar DUP–TRP/INV–DUP rearrangements facilitated by inverted LCRs have been observed in other genes, such as *MECP2*, *DMD*, and *PLP1*.[33–35] The proposed mechanism involves two template switches during the DNA replication process (►**Fig. 4C**): the first template switch occurs between inverted LCRs via break-induced replication, reversing the replication direction. After a distance of reverse replication, the second template switch occurs through MMBIR, restoring the original direction of replication for subsequent DNA synthesis.[35,36] In other words, one rearrangement breakpoint is fixed within inverted LCRs, whereas the second breakpoints are relatively random. Therefore, the impact of LCR-F9A/LCR-F9B-induced DUP–TRP/INV–DUP rearrangements on the *F9* copy number depends on the length of backward-directed replication and the position of the second template switch. This variability can result in no change, duplication, or triplication of the *F9* gene (►**Fig. 4C**).

Understanding the mechanisms underlying SVs facilitates clinical interpretation and further risk prediction. The complete

F9 deletions in probands 1 and 2 explain their reduced FIX activity and diagnosis of severe HB. According to Radic et al patients develop extra-hematologic phenotypes such as partial hypopituitarism, intellectual disability, and developmental delay when the deletion region encompasses the *SOX3* gene.[15] However, proband 1 had a chromosomal deletion affecting multiple genes, including *SOX3*, but did not exhibit any unique medical history records beyond HB. One possible explanation could be attributed to compensatory mechanisms within the genome that can mitigate the functional loss of the *SOX3* gene. Furthermore, some phenotypes may only manifest in adulthood,[37] whereas proband 1 is only 5 years old. Another noteworthy fact in cases 1 and 2 is that their mothers, both had low FIX activity despite the absence of skewed XCI. Interestingly, we found four additional female HB patients (FIX: Ac: 10–20%) with only *F9* gene deletions in different literature sources, one of whom had undergone XCI analysis in peripheral blood leukocytes, also showing no evidence of skewed XCI.[38,39] One possible explanation could be that the XCI patterns observed in a peripheral blood sample do not necessarily align with the pattern observed in the liver.[40] Another more intriguing hypothesis is that the expression of the *F9* gene on the normal chromosome is repressed by the other chromosome with large deletion.

Unlike deletions directly affecting protein expression and function, complete gene duplications are expected to enhance protein expression through a dosage effect. The intuitive hypothesis would suggest that duplication and triplication of *F9* would lead to a doubling or tripling of FIX expression levels. However, our findings indicated that the *F9* tandem duplication in case 3 did not significantly increase FIX activity; only the additional *F9* copies, generated by the DUP–TRP/INV–DUP rearrangement, showed a notable increase in FIX activity in cases 4 and 5. Interestingly, compared with *F9* duplication in case 4, the triplication of *F9* in case 5 did not exhibit a markedly higher FIX activity. These observations suggest that FIX expression might be influenced by regulatory mechanisms beyond the *F9* gene.

Further investigation of the genomic breakpoints revealed potential structural differences. *F9* triplication in case 5 includes the upstream 32- and 80-kb regions of the *F9* gene, and *F9* duplication in case 4 encompasses the upstream 65-kb region. However, the duplication region in case 3 only extends up to 7-kb upstream (►**Fig. 2**). In addition, we reported an *F9* partial duplication in a mild HB patient before.[11] Despite retaining an intact *F9* gene, the patient exhibited FIX levels of only 8%. Breakpoint analysis showed this duplication contained only a 13-kb upstream sequence extension of the intact *F9* gene. Therefore, we speculate the presence of long-range gene regulatory elements like enhancers between the 32- and 65-kb region upstream of the *F9* gene. Despite being far from the target gene, these elements may engage in looping interactions with the *F9* gene promoter.[41,42] In other words, the additional *F9* copies can achieve sufficient expression only when the duplicated segments retain these long-range regulatory interactions. Johnsen et al conducted sequencing of the *F9* coding regions, splice sites, and 300-bp upstream and 1417-bp downstream untranslated sequences in 1,616 HB patients, revealing that 31 patients still could not receive a definitive genetic diagnosis.[43] Although deep intronic variation remains to be examined, we suggest detecting long-range gene regulatory elements upstream of the *F9* gene may also provide some diagnostic value.

Elevated FVIII, FIX, and FXI levels have been identified as a common risk factor for VTE.[3,4,44,45] Previous studies have shown that additional copies of the *F8* gene are associated with increased FVIII levels and an augmented VTE risk.[44] Similarly, duplication of the *F11* gene leads to elevated plasma levels of FXI and an increased susceptibility to VTE.[45] The ISTH Gold Variants project reported a patient with complete *F9* duplication who also experienced recurrent thrombosis, yet details regarding FIX levels or the specific duplication rearrangement were not provided.[46] Recently, Turkut et al reported a 554-kb X-chromosome duplication encompassing the *F9* gene, contributing to the elevated FIX activity level and subsequent cerebral venous thrombosis development.[47] Our study provides further evidence that the duplication or triplication of the *F9* gene may result in the upregulation of FIX expression, thereby contributing to the manifestation of thrombotic phenotypes. Furthermore, like HB, extra copies of the *F9* gene tend to have a greater impact on males than on female carriers. Theoretically, if the expression of the *F9* gene on a single X-chromosome is doubled, the plasma FIX levels in females would only increase to 150% due to the presence of a normal X-chromosome and random inactivation providing dosage compensation,[48] rather than the theoretical 200% in males. This may explain why clinical manifestations are predominantly observed in affected males within families exhibiting elevated FIX activity.[6,7,47]

Of note, the pathogenesis of thrombosis is complex, involving multiple genetic and environmental elements.[3] In our study, tandem *F9* duplication in case 3 did not significantly elevate FIX levels. Instead, it is possible that the patient's coexisting *PROS1* (p.Glu67Ala) mutation, a hotspot mutation in the Chinese population leading to type 1 PS deficiency,[24] may be the actual cause of the patient's thrombophilia phenotype. Additionally, the thrombotic phenotypes observed in probands 4 and 5 could be attributed to the combined effects of increased *F9* copy number and coexisting heterozygous mutations in *F5* (p.Arg575Cys) and *PROC* (p.Asn384Ser). *F5* (p.Arg575Cys) has not been previously reported. Although the associated FV activity in case 4 was within the normal range, *F5* (p.Arg575Cys) was predicted to be "disease causing" by online bioinformatics tools (SIFT, PolyPhen-2 and Mutation taster, ►**Supplementary Table S6**, available in the online version). It is worth noting that although *F5* (p.Arg575Cys) does not belong to any known activated protein C (APC) resistance-associated site,[49] the introduced cysteine residue may potentially form additional disulfide bonds and influence APC binding and hydrolysis through conformational changes, thereby increasing the risk for VTE. Similarly, *PROC* (p.Asn384Ser) only slightly reduced the PC activity in proband 5, but *PROC* (p.Asn384Lys) has been reported in thrombosis

patients with PC levels at the lower limit of the normal range.[50] Therefore, these mutations may have contributed to thrombus formation alongside elevated FIX activity.

In conclusion, our study comprehensively characterized SVs breakpoints in five unrelated families, exhibiting deletions, duplications, and triplications encompassing the *F9* gene. Using OGM technology, we have provided the first report on DUP–TRP/INV–DUP rearrangements that result in *F9* duplication and triplication. Additionally, we analyzed the impact of increased *F9* copy number on FIX expression and clinical phenotypes. These findings have important implications for the clinical management and risk assessment of individuals with *F9* duplications and triplications, particularly those with a history of thrombosis.

What is known about this topic?

- Complete *F9* gene deletion can lead to severe HB, whereas complete *F9* gene duplication/triplication rearrangements remain understudied.
- It has been established that elevated FIX level is an essential risk factor for VTE and an increased rate of thrombotic recurrence.
- No recurrent chromosomal structural rearrangements affecting the *F9* gene have been identified yet.

What does this paper add?

- We inaugural explore the complete *F9* gene duplication and triplication rearrangement mechanisms employing OGM.
- In two cases, we discovered complex DUP–TRP/INV–DUP rearrangements mediated by the same pair of inverted LCRs, resulting in *F9* duplication and triplication.
- *F9* duplication and triplication, resulting from DUP–TRP/INV–DUP rearrangements, can substantially elevate FIX levels and potentially augment the risk of VTE.

Authors' Contribution

Y.M. and Y.L. designed and performed the experiments and contributed to the writing of the manuscript. J.S., Q.L., and R.W., collected and organized the clinical cases. Q.D. and J.D. analyzed the data, reviewed and revised the manuscript, and provided supervision for the entire project.

Funding

This study was supported by the National Natural Science Foundation of China (grant numbers 81970126, 81970127, 82170128, and 82270128).

Conflict of Interest

None declared.

References

1 Smith SA, Travers RJ, Morrissey JH. How it all starts: initiation of the clotting cascade. Crit Rev Biochem Mol Biol 2015;50(04): 326–336
2 Berntorp E, Fischer K, Hart DP, et al. Haemophilia. Nat Rev Dis Primers 2021;7(01):45
3 Khachidze M, Buil A, Viel KR, et al. Genetic determinants of normal variation in coagulation factor (F) IX levels: genome-wide scan and examination of the FIX structural gene. J Thromb Haemost 2006;4(07):1537–1545
4 van Hylckama Vlieg A, van der Linden IK, Bertina RM, Rosendaal FR. High levels of factor IX increase the risk of venous thrombosis. Blood 2000;95(12):3678–3682
5 Teixeira Mello TB, Machado TF, Montavão SA, Ozello MC, Annichino-Bizzacchi JM. Assessing the coagulation factor levels, inherited thrombophilia, and ABO blood group on the risk for venous thrombosis among Brazilians. Clin Appl Thromb Hemost 2009;15 (04):408–414
6 Simioni P, Tormene D, Tognin G, et al. X-linked thrombophilia with a mutant factor IX (factor IX Padua). N Engl J Med 2009;361(17): 1671–1675
7 Wu W, Xiao L, Wu X, et al. Factor IX alteration p.Arg338Gln (FIX Shanghai) potentiates FIX clotting activity and causes thrombosis. Haematologica 2021;106(01):264–268
8 Xu Z, Spencer HJ, Harris VA, Perkins SJ. An updated interactive database for 1692 genetic variants in coagulation factor IX provides detailed insights into hemophilia B. J Thromb Haemost 2023;21(05):1164–1176
9 Chan V, Au P, Lau P, Chan TK. A novel haemophilia B defect due to partial duplication of the factor IX gene. Br J Haematol 1994;86 (03):601–609
10 Wheeler RB, Cutler JA, Alamelu J, Mitchell MJ. The first report of a multi-exon duplication in the F9 gene causative of severe haemophilia B. Haemophilia 2015;21(05):e433–e435
11 Xie X, Chen C, Liang Q, et al. Characterization of two large duplications of F9 associated with mild and severe haemophilia B, respectively. Haemophilia 2019;25(03):475–483
12 Hol FA, Schepens MT, van Beersum SE, et al. Identification and characterization of an Xq26-q27 duplication in a family with spina bifida and panhypopituitarism suggests the involvement of two distinct genes. Genomics 2000;69(02):174–181
13 Du C, Wang F, Li Z, et al. Xq26.3-q27.1 duplication including SOX3 gene in a Chinese boy with hypopituitarism: case report and two years treatment follow up. BMC Med Genomics 2022;15(01):19
14 Kaminsky EB, Kaul V, Paschall J, et al. An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. Genet Med 2011;13(09):777–784
15 Radic CP, Abelleyro MM, Ziegler B, et al. Haemophilia B, severe childhood obesity and other extra-haematological features associated with similar 4Mb-deletions on Xq27: clinical findings, molecular insights and literature update. Haemophilia 2023;29 (03):844–854
16 Wu X, Lu Y, Ding Q, et al. Characterisation of large F9 deletions in seven unrelated patients with severe haemophilia B. Thromb Haemost 2014;112(03):459–465
17 Nakamura Y, Ando Y, Takagi Y, et al. Distinct X chromosomal rearrangements in four haemophilia B patients with entire F9 deletion. Haemophilia 2016;22(03):433–439
18 Onishi-Seebacher M, Korbel JO. Challenges in studying genomic structural variant formation mechanisms: the short-read dilemma and beyond. BioEssays 2011;33(11):840–850
19 Mantere T, Kersten S, Hoischen A. Long-read sequencing emerging in medical genetics. Front Genet 2019;10:426
20 Yuan Y, Chung CY, Chan TF. Advances in optical mapping for genomic research. Comput Struct Biotechnol J 2020;18:2051–2062

21 Smith AC, Neveling K, Kanagal-Shamanna R. Optical genome mapping for structural variation analysis in hematologic malignancies. Am J Hematol 2022;97(07):975–982
22 Fahiminiya S, Oikonomopoulos S, Rivard GE, et al. Deciphering a novel complex inversion affecting F8 in a family with severe haemophilia A by optical genome mapping. Haemophilia 2023;29 (03):921–924
23 Liu G, Sun J, Li Z, Chen Z, Wu W, Wu R. F9 mutations causing deletions beyond the serine protease domain confer higher risk for inhibitor development in hemophilia B. Blood 2023;141(06): 677–680
24 Li L, Wu X, Wu W, Ding Q, Cai X, Wang X. Clinical Manifestation and Mutation Spectrum of 53 Unrelated Pedigrees with Protein S Deficiency in China. Thromb Haemost 2019;119(03):449–460
25 Cao P, Yang A, Wang R, et al. Germline duplication of SNORA18L5 increases risk for HBV-related hepatocellular carcinoma by altering localization of ribosomal proteins and decreasing levels of p53. Gastroenterology 2018;155(02):542–556
26 Miranda-Furtado CL, Luchiari HR, Chielli Pedroso DC, et al. Skewed X-chromosome inactivation and shorter telomeres associate with idiopathic premature ovarian insufficiency. Fertil Steril 2018;110(03):476–485.e1
27 Dardik R, Avishai E, Lalezari S, et al. Molecular mechanisms of skewed X-chromosome inactivation in female hemophilia patients-lessons from wide genome analyses. Int J Mol Sci 2021;22(16):9074
28 Veyradier A, Boisseau P, Fressinaud E, et al; French Reference Center for von Willebrand disease. A laboratory phenotype/genotype correlation of 1167 French patients from 670 families with von Willebrand disease: a new epidemiologic picture. Medicine (Baltimore) 2016;95(11):e3038
29 Dittwald P, Gambin T, Gonzaga-Jauregui C, et al. Inverted low-copy repeats and genome instability–a genome-wide analysis. Hum Mutat 2013;34(01):210–220
30 Zhang F, Khajavi M, Connolly AM, Towne CF, Batish SD, Lupski JR. The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. Nat Genet 2009;41(07):849–853
31 Burssed B, Zamariolli M, Bellucco FT, Melaragno MI. Mechanisms of structural chromosomal rearrangement formation. Mol Cytogenet 2022;15(01):23
32 Lee JA, Carvalho CM, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell 2007;131(07):1235–1247
33 Carvalho CM, Ramocki MB, Pehlivan D, et al. Inverted genomic segments and complex triplication rearrangements are mediated by inverted repeats in the human genome. Nat Genet 2011;43 (11):1074–1081
34 Ishmukhametova A, Chen JM, Bernard R, et al. Dissecting the structure and mechanism of a complex duplication-triplication rearrangement in the DMD gene. Hum Mutat 2013;34(08):1080–1084
35 Bahrambeigi V, Song X, Sperle K, et al. Distinct patterns of complex rearrangements and a mutational signature of microhomeology are frequently observed in PLP1 copy number gain structural variants. Genome Med 2019;11(01):80
36 Li Y, Ding B, Mao Y, Zhang H, Wang X, Ding Q. Tandem and inverted duplications in haemophilia A: breakpoint characterisation provides insight into possible rearrangement mechanisms. Haemophilia 2023;29(04):1121–1134
37 Anson DS, Blake DJ, Winship PR, Birnbaum D, Brownlee GG. Nullisomic deletion of the mcf.2 transforming gene in two haemophilia B patients. EMBO J 1988;7(09):2795–2799
38 Jourdy Y, Chatron N, Carage ML, et al. Study of six patients with complete F9 deletion characterized by cytogenetic microarray: role of the SOX3 gene in intellectual disability. J Thromb Haemost 2016;14(10):1988–1993
39 Stoof SCM, Kersseboom R, de Vries FAT, Kruip MJHA, Kievit AJA, Leebeek FWG. Hemophilia B in a female with intellectual disability caused by a deletion of Xq26.3q28 encompassing the F9. Mol Genet Genomic Med 2018;6(06):1220–1224
40 Mason JA, Robertson JD. Extreme skewing of X-inactivation: rethinking severe haemophilia in women and girls. Haemophilia 2019;25(04):e286–e287
41 Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. Nature 2012;489(7414): 109–113
42 Yao L, Berman BP, Farnham PJ. Demystifying the secret mission of enhancers: linking distal regulatory elements to target genes. Crit Rev Biochem Mol Biol 2015;50(06):550–573
43 Johnsen JM, Fletcher SN, Dove A, et al. Results of genetic analysis of 11 341 participants enrolled in the My Life, Our Future hemophilia genotyping initiative in the United States. J Thromb Haemost 2022;20(09):2022–2034
44 Shen W, Gu Y, Zhu R, Zhang L, Zhang J, Ying C. Copy number variations of the F8 gene are associated with venous thromboembolism. Blood Cells Mol Dis 2013;50(04):259–262
45 Van Laer C, Peerlinck K, Jacquemin M, et al. F11 gene duplication causes elevated FXI plasma levels and is a risk for venous thrombosis. Thromb Haemost 2022;122(06):1058–1060
46 Megy K, Downes K, Morel-Kopp MC, et al. GoldVariants, a resource for sharing rare genetic variants detected in bleeding, thrombotic, and platelet disorders: communication from the ISTH SSC Subcommittee on Genomics in Thrombosis and Hemostasis. J Thromb Haemost 2021;19(10):2612–2617
47 Turkut Tan T, Pariltay E, Avci Durmusaliogu E, et al. A unique case of thrombophilia: the role of F9 gene duplication and increased factor IX activity in cerebral venous thrombosis. J Thromb Haemost 2023;21(10):2913–2916
48 Fang H, Deng X, Disteche CM. X-factors in human disease: impact of gene content and dosage regulation. Hum Mol Genet 2021;30 (R2):R285–R295
49 Kujovich JL. Factor V Leiden thrombophilia. Genet Med 2011;13 (01):1–16
50 Rovida E, Merati G, D'Ursi P, et al. Identification and computationally-based structural interpretation of naturally occurring variants of human protein C. Hum Mutat 2007;28(04):345–355