Accepted Manuscript online: 2023-06-07 Article published online: 2023-07-10

Comprehensive Analysis of Hemophilia A (CAHEA): Towards Full Characterization of the *F8* Gene Variants by Long-Read Sequencing

Yingdi Liu[1,*] Dongzhi Li[2,*] Dongyi Yu[3,*] Qiaowei Liang[4,*] Guilan Chen[2] Fucheng Li[2] Lu Gao[3] Zhuo Li[1] Tiantian Xie[5] Le Wu[5] Aiping Mao[5] Lingqian Wu[1,4] Desheng Liang[1,4]

[1] Center for Medical Genetics and Hunan Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, Changsha, Hunan, China
[2] Prenatal Diagnostic Center, Guangzhou Women and Children's Medical Center, Guangzhou, Guangdong, China
[3] Center for Medical Genetics and Prenatal Diagnosis, Shandong Provincial Maternal and Child Health Care Hospital, Shandong Medicine and Health Key Laboratory of Birth Defect Prevention and Genetic Medicine, Key Laboratory of Birth Regulation and Control Technology of National Health Commission of China, Jinan, Shandong, China
[4] Department of Medical Genetics, Hunan Jiahui Genetics Hospital, Changsha, Hunan, China
[5] Berry Genomics Corporation, Beijing, China

Thromb Haemost 2023;123:1151–1164.

Address for correspondence Aiping Mao, PhD, Berry Genomics Corporation, Beijing 102200, China (e-mail: maoaiping@berrygenomics.com).

Lingqian Wu, MD, PhD, Center for Medical Genetics and Hunan Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, 110 Xiangya Road, Changsha 410078, Hunan, China (e-mail: wulingqian@sklmg.edu.cn).

Desheng Liang, MD, PhD, Center for Medical Genetics and Hunan Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, 110 Xiangya Road, Changsha 410078, Hunan, China (e-mail: liangdesheng@sklmg.edu.cn).

Abstract

Background Hemophilia A (HA) is the most frequently occurring X-linked bleeding disorder caused by heterogeneous variants in the F8 gene, one of the largest genes known. Conventional molecular analysis of F8 requires a combination of assays, usually including long-range polymerase chain reaction (LR-PCR) or inverse-PCR for inversions, Sanger sequencing or next-generation sequencing for single-nucleotide variants (SNVs) and indels, and multiplex ligation-dependent probe amplification for large deletions or duplications.

Materials and Methods This study aimed to develop a LR-PCR and long-read sequencing-based assay termed comprehensive analysis of hemophilia A (CAHEA) for full characterization of F8 variants. The performance of CAHEA was evaluated in 272 samples from 131 HA pedigrees with a wide spectrum of F8 variants by comparing to conventional molecular assays.

Results CAHEA identified F8 variants in all the 131 pedigrees, including 35 intron 22-related gene rearrangements, 3 intron 1 inversion (Inv1), 85 SNVs and indels, 1 large insertion, and 7 large deletions. The accuracy of CAHEA was also confirmed in another set of 14 HA pedigrees. Compared with the conventional methods combined altogether, CAHEA assay demonstrated 100% sensitivity and specificity for identifying various types of F8 variants and had the advantages of directly determining the break regions/points of large inversions, insertions, and deletions, which enabled analyzing the mechanisms of recombination at the junction sites and pathogenicity of the variants.

Conclusion CAHEA represents a comprehensive assay toward full characterization of F8 variants including intron 22 and intron 1 inversions, SNVs/indels, and large insertions and deletions, greatly improving the genetic screening and diagnosis for HA.

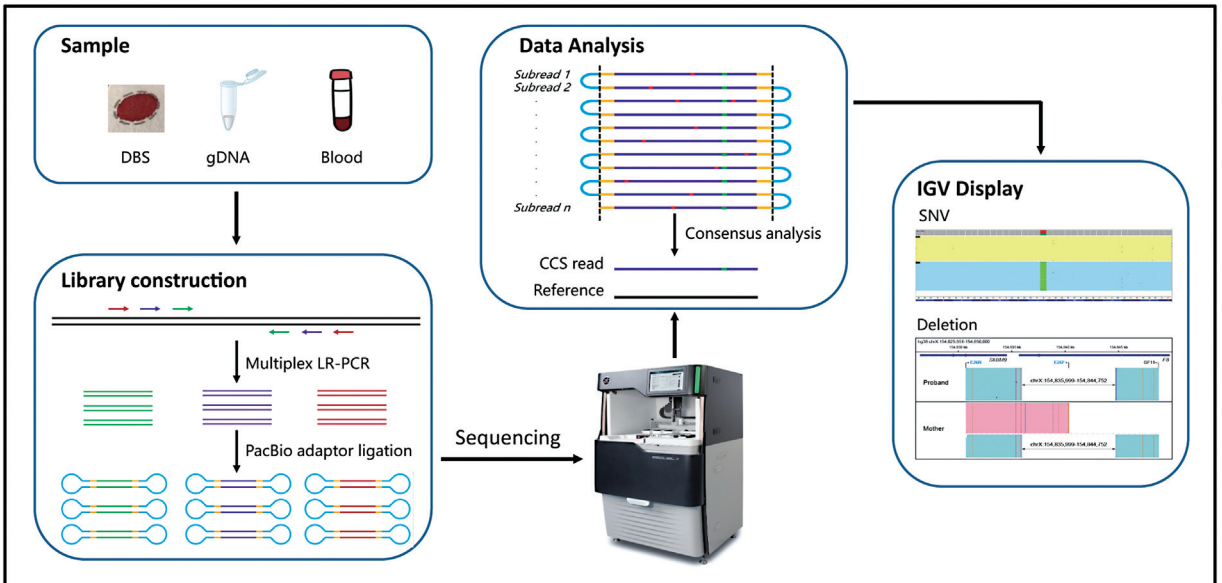
Keywords
- hemophilia A
- gene rearrangements
- long-range PCR
- long-read sequencing
- genetic testing

* *These authors contributed equally to this work.*

received
March 1, 2023
accepted after revision
May 15, 2023
accepted manuscript online
June 7, 2023
article published online
July 10, 2023

DOI https://doi.org/10.1055/a-2107-0702.
ISSN 0340-6245.

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



Visual Summary. CAHEA: comprehensive analysis of hemophilia A.

Introduction

Hemophilia A (HA; OMIM#306700) is an X-linked recessive disease caused by a deficiency in coagulation factor VIII (FVIII) and has a global incidence of approximately 1 in 5,000 male newborns.[1] According to baseline FVIII coagulant activity (FVIII:C), the phenotype of HA can be classified as severe (<1%), moderate (1–5%), and mild (>5–40%).[2] The FVIII gene (*F8*), one of the largest known human genes, maps to the distal end of the long arm of X-chromosome (Xq28) and spans 187 kb of genomic DNA. HA is caused by a wide spectrum of genetic abnormalities affecting *F8*, of which the most common defects are recurrent intron 22 (Int22) inversion (Inv22) and intron 1 (Int1) inversion (Inv1).[3,4] The homologous recombination between a 9.5-kb segment in intragenic Int22 (Int22h-1) and two extragenic homologous regions (Int22h-2 and Int22h-3) can lead to Inv22 Type I, Inv22 Type II, Int22 deletions (Del22) and duplications (Dup22).[5] There are also some rare Int22-related complex gene rearrangements (Int22-CGR), of which only a few cases have been delineated so far.[6–8] The Inv22 Type I is the most prevalent Int22-related gene rearrangement and responsible for 40 to 50% of the severe HA cases.[9] Inv1 caused by recurrent recombination between a 1-kb segment in intragenic Int1 (Int1h-1) and its extragenic repeat (Int1h-2) accounts for 2 to 5% of severe HA.[10] The nonrecurrent *F8* gene deletions and duplications are found in 3 to 7% of patients with severe HA.[11–13] In the remaining patients with HA, a large and heterogeneous spectrum of *F8* gene variants including over 3,000 single nucleotide variants (SNVs) and indels has been reported.[14–16]

Due to the complexity and variety of *F8* gene variants, the genetic diagnosis of patients with HA often requires several different techniques. Multiplex long-range polymerase chain reaction (LR-PCR) assays have been developed to achieve the genetic diagnosis of Inv22,[17] or for discriminating among Inv22 Type I/II, Del22, and Dup22.[3] To overcome the difficulties of PCR amplification of Int22h duplicons, inverse-PCR (IS-PCR) approaches were designed to screen for Inv22 and other Int22-related gene rearrangements.[18,19] Similarly, specific multiplex LR-PCR and IS-PCR were also applied for genetic testing of Inv1.[10,19] Standard PCR followed by Sanger sequencing has been used conventionally to identify SNVs and indels of the essential *F8* regions including promoter, exons, splice junctions, and 3′-polyadenylation signal region.[20,21] In recent years, next-generation sequencing (NGS) has been employed for high-throughput genetic screening of SNVs and indels of *F8* including the deep intronic regions.[22–24] When PCR failure and NGS suggest the absence of one or more exons in male patients, multiplex ligation-dependent probe amplification (MLPA) allows confirmation of deletions. MLPA can also define the carrier status in female relatives for large gene deletions, being masked by the amplification of the normal allele using PCR, as well as duplications that are unidentifiable by standard PCR.[25] Genetic testing combining all the methods are expensive and labor-intensive. Alternatively, some laboratories choose to use a stepwise diagnosis strategy that tests the existence of Inv22 first. For patients negative for Inv22, different methods were then applied for Inv1, SNVs/indels, and large deletions/duplications.[20,22,26–28] While overall cost-effective, the whole diagnosis process can take a long time for patients negative for inversions. To overcome the limit of short-read length in detecting highly homologous regions, Johnsen et al developed a NGS-based approach that captured Inv22 and Inv1 from Ksp221 digested/ligated DNA, and other variants from untreated genomic DNA simultaneously.[29] Taking advantages of long sequencing reads that can discriminate between highly homologous regions, long-read sequencing (LRS)-based methods have been developed for

comprehensive genetic analysis of disorders like thalassemia,[30] congenital adrenal hyperplasia,[31] spinal muscular atrophy,[32] and fragile-X syndrome,[33] with high accuracy and cost-effectiveness. However, LRS has not been thoroughly explored for HA genetic testing.

Here, we developed a multiplex LR-PCR and LRS-based approach termed comprehensive analysis of hemophilia A (CAHEA) aiming at full characterization of *F8* variants. This approach was rigorously validated through retrospective analysis of 272 samples from 131 unrelated HA pedigrees with a wide spectrum of variants. In addition, the paralogous sequence variants (PSVs) among Int22 duplicons, junction regions of Inv22, and exact breakpoints of large deletions were analyzed.

Materials and Methods

Study Subjects

This study enrolled two sets of HA pedigrees. The first set included a total of 272 genomic DNA samples from 131 unrelated HA pedigrees archived in Hunan Jiahui Genetics Hospital, Guangzhou Women and Children's Medical Center, and Shandong Maternal and Child Care Hospital from March 2012 to July 2022, including 246 samples from peripheral blood and 26 from chorionic villus sampling (►**Supplementary Table S1,** available in the online version). The second set included a total of 22 genomic DNA samples from 14 unrelated HA pedigrees newly diagnosed in Hunan Jiahui Genetics Hospital from December 2022 to March 2023. The clinical diagnosis of all the probands was established by standard coagulation assays for FVIII:C and the genetic diagnosis was made by conventional methods. The first set of DNA samples was retrospectively sent to the genetic testing laboratory of Berry Genomic Corporation for blind CAHEA assay. The second set of DNA samples was subjected to conventional methods and CAHEA assay simultaneously. All procedures involved in this study complied with the ethical standards of the 1964 Helsinki Declaration and its later amendments. Study ethics approval was granted by the institutional review boards of Hunan Jiahui Genetics Hospital, Guangzhou Women and Children's Medical Center, and Shandong Maternal and Child Care Hospital. Informed written consent was obtained from all the subjects or their legal guardians.

Conventional Genetic Analysis for *F8* Variants

Screening for molecular defects in the *F8* gene of HA patients was performed with the following stepwise strategy. First, the presence of Inv22 was screened by IS-PCR or LR-PCR as previously described.[17,18] Second, for samples with negative Inv22, IS-PCR or LR-PCR was performed to analyze Inv1.[10,19] Third, for samples lacking either inversion, PCR followed by Sanger sequencing or panel-based NGS was performed to analyze the essential *F8* regions including promoter, exons, splice junctions, and 3′-polyadenylation signal region. Potential deletions were also identified in male patients when there were no PCR amplicons or NGS reads for one or more exons. Finally, MLPA (probemex P17 *F8*, MRC-Holland) was performed in patients with potential deletions and in patients without Inv22, Inv1, or pathogenic/likely pathogenic SNVs/indels to identify gross deletions and duplications of the *F8* gene.

Amplification of *F8* and PacBio sequencing

CAHEA utilized LR-PCR with four sets of primers in one reaction to comprehensively analyze over 3,000 *F8* gene variants including Int22h-related gene arrangements, Inv1, SNVs/indels, and deletions listed in public databases for *F8*[14–16] (►**Fig. 1A**, ►**Supplementary Table S2,** available in the online version). The first set included five primers (H1F, H1R, H2F, H3F, and H2/3R) to cover Int22h-1, Int22h-2, Int22h-3, as well as Int22h-related inversions, deletions, and duplication.[3] The second set included three primers (9cR, 9F, and int1h-2F) to cover Int1h-1 and Inv1.[10] The third set included 22 primers to cover the essential *F8* regions for SNVs/indels. The fourth set included 20 gap primers up to 50-kb upstream and 50-kb downstream of *F8* gene, which combined with the other three sets of primers could detect both known and novel large deletions. Multiplex LR-PCR was performed in 50-μL reactions containing 10 to 100 ng of genomic DNA, 1 × PCR buffer for KOD FX Neo, 0.4 mM of each dNTP, 1 μM of primer mixture, and 1 μL of KOD FX Neo

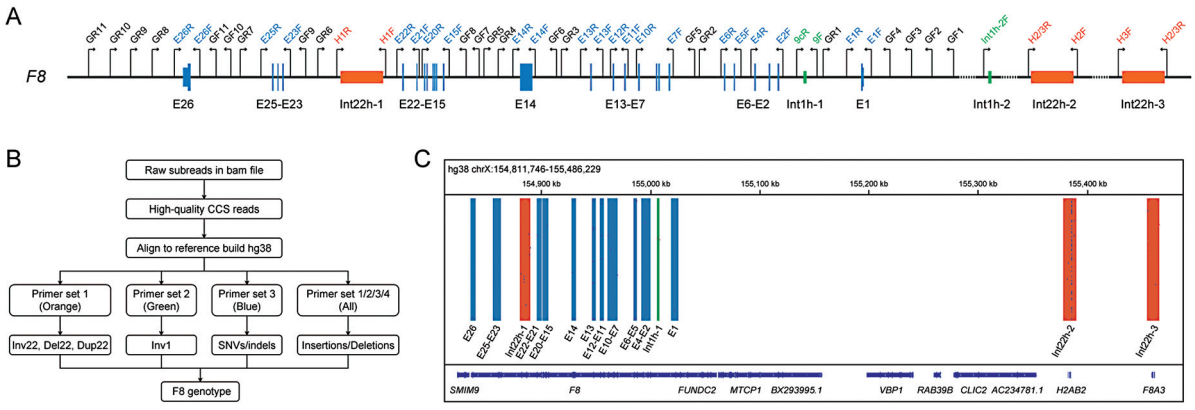


Fig. 1 Establishment and validation of the CAHEA assay. (**A**) Design of multiplex long-range primers for CAHEA. (**B**) Customized bioinformatics pipeline for *F8* variants analysis. (**C**) IGV plots showing the CCS reads of CAHEA for normal male sample JHA255 from pedigree 123. CAHEA, comprehensive analysis of hemophilia A; CCS, circular consensus sequencing; IGV, Integrative Genomics Viewer.

(TOYOBO). PCR cycling conditions for optimal fragment amplification were 94°C for 5 minutes (1 cycle); 98°C for 10 and 68°C for 12 minutes (30 cycles), and 68°C for 10 minutes (1 cycle). The single-molecule real-time dumbbell (SMRTbell) libraries were prepared and sequenced as previously described.[30] A one-step end-repair and ligation reaction was performed to ligate unique PacBio-barcoded adaptors to PCR products, which was followed by exonuclease digestion to remove unsuccessfully ligated DNA to get dumbbell-shaped pre-libraries. The uniquely barcoded pre-libraries were purified with AMpure PB beads (Pacific Biosciences), quantified with Qubit dsDNA HS assay (ThermoFisher scientific), and then pooled with equal mass. A SMRTbell library for sequencing was prepared by annealing sequencing primer and polymerase to the pooled pre-libraries using the Sequel II Binding Kit 2.0 and Internal Control Kit 1.0 (Pacific Biosciences) and sequenced with Sequel II Sequencing Kit 2.0 (Pacific Biosciences) for 30 hours under circular consensus sequencing (CCS) mode.

PacBio Data Analysis and Variant Calling

The raw subreads in the bam file were processed to get high-quality CCS reads, debarcoded to individual samples, and then aligned to reference build hg38 using the SMRT Link analysis software suite (Pacific Biosciences). The presence of Inv22 Type I, Inv22 Type II, Del22, Dup22, and Inv1 was determined as previously described.[3,5,10] SNVs and indels of the *F8* gene were called by FreeBayes1.3.4 with CCS reads $\geq$30 for each fragment (Biomatters, Inc., San Diego, California, United States). The large deletions were identified based on the length and primer sequences of the CCS reads. The exact breakpoints were determined by aligning the CCS reads to GRCH/hg38 (hg38) and then confirmed by Sanger sequencing. The reads for Int22 duplicons were also aligned to reference T2T-CHM13 (CHM13) for alignment comparison with hg38.

Display of CCS Reads Generated by PacBio Sequencing

For each PCR fragment, 30 CCS reads were randomly extracted and displayed in the Integrative Genomics Viewer (IGV). For CCS reads with inversions, the reads were clipped in the break regions manually and then aligned to reference genome hg38. For CCS reads with deletions, the blank region between colored regions represented the deletion. The vertical lines with different colors showed the presence of SNVs/indels, with blue line meaning C, green line meaning A, brown line meaning G, red line meaning T, and purple line meaning insertion. When $\geq$200 kb of DNA regions were displayed in IGV, the vertical lines of SNVs/indels could not be shown due to limited resolution.

Results

Establishment and Validation of CAHEA

The CAHEA assay containing four sets of primers in one reaction was designed to comprehensively analyze *F8* gene variants including Int22-related gene rearrangements, Inv1, SNVs/indels, and large insertions and deletions (►**Fig. 1A**). Multiplex LR-PCR was performed on genomic DNA samples and the amplified PCR products were subjected to PacBio SMRTbell library preparation, sequencing, and analyzing using a customized variant calling pipeline (►**Fig. 1B**). Successful amplification and sequencing were confirmed by displaying the CCS reads in the IGV plot (►**Fig. 1C**). CAHEA had full coverage of all the 26 exons and 15 out of 25 introns, while had 62 to 4,801 bp coverage in the exon–intron boundary regions for the other 10 introns (►**Supplementary Table S3,** available in the online version). Overall, CAHEA assay had 100% coverage for the exons and 36.0% coverage of the introns of *F8*.

To evaluate the performance of CAHEA assay in clinical samples, a total of 272 samples from 131 unrelated HA pedigrees diagnosed by conventional methods were enrolled, renamed, and tested by blind CAHEA assay. Following the diagnostic workflow, CAHEA identified disease-causing *F8* variants in all the 131 probands (►**Table 1**). These variants included 34 Inv22 (26.0%), 1 Int22-CGR (0.76%), 3 Inv1 (2.3%), 85 SNVs/indels (64.9%), 1 large insertion (0.76%), and 7 large deletions (5.3%). Out of 121 female family members enrolled, 85 were identified to be carriers by both conventional methods and CAHEA (►**Supplementary Table S1,** available in the online version). In total, *F8* variants were identified in 229 out of 272 samples (►**Supplementary Table S1,** available in the online version). Compared with the conventional methods combined altogether, CAHEA assay demonstrated 100% sensitivity and specificity, with the advantages of directly determining the types of inversions and the breakpoints of large deletions (►**Table 1**).

CAHEA Assay for Analysis of Int22-Related Gene Rearrangements

Out of 131 probands, CAHEA identified Int22-related gene rearrangements in 35 (26.3%), including 32 with Inv22 Type I, 2 with Inv22 Type II, and 1 with Inv22-CGR (►**Table 1**). To investigate the recombination regions of Int22 inversions, the PSVs among Int22h-1/2/3 need to be identified first. Since there were 32 discordant sites in Int22h duplicons between genome build hg38 and CHM13, including 1 in Int22h-1, 15 in Int22h-2, and 16 in Int22h-3 (►**Supplementary Fig. S1A,** available in the online version), we sought to find out which genome build was more suitable for our dataset. Overall, the LRS data from 272 samples supported hg38 at two sites (h1–1 and h3–9), CHM13 at four sites (h2–2, h2–10, h3–6, and h3–10), and both genome builds for the other 26 polymorphic sites located in Int22-hour2/h3 (►**Supplementary Fig. S1B,** available in the online version). Additionally, the corresponding positions in Int22-hour1 had the same sequences as one of the genome builds in these 26 polymorphic sites, suggesting that these were not signature variants among Int22 duplicons. After adjusting the reference build for h1–1, h3–9, h2–2, h2–10, h3–6, and h3–10, all the six sites were the same among Int22-hour1/2/3. Thus, only the sequences concordant between two genome builds were compared among Int22 duplicons, and six PSVs were found between Int22-hour1 and Int22-hour2/3 (►**Fig. 2A**). The recombination regions of Inv22 were

Table 1 Concordance of CAHEA and conventional methods for *F8* gene variants in 131 Chinese HA pedigrees

Type of variant	Conventional methods	CAHEA assay			Domain	Phenotype	No. of pedigree	Novel
	Nucleotide change	Nucleotide change	Protein change	Gene region				
Int22-related gene rearrangements ($n = 3$)								
	Inv22	Inv22 Type I	NA	Intron 22	NA	Moderate/severe	10/22	N
	Inv22	Inv22 Type II	NA	Intron 22	NA	Moderate/severe	1/1	N
	Int22-CGR	Int22-CGR	NA	Intron 22	NA	Severe	1	N
Inv1 ($n = 1$)	Inv1	Inv1	NA	Intron 1	NA	Moderate/severe	1/2	N
SNV/indel ($n = 75$)								
Nonsense	c.1202G > A	c.1202G > A	p.W401*	Exon 8	A2	Severe	1	N
	c.1726G > T	c.1726G > T	p.E576*	Exon 11	A2	Moderate	1	N
	c.1804C > T	c.1804C > T	p.R602*	Exon 12	A2	Severe	1	N
	c.1967G > A	c.1967G > A	p.W656*	Exon 13	A2	Severe	1	N
	c.2025T > A	c.2025T > A	p.Y675*	Exon 13	A2	Moderate	1	Y
	c.2440C > T	c.2440C > T	p.R814*	Exon 14	B	Moderate/severe	1/1	N
	c.3143G > A	c.3143G > A	p.W1048*	Exon 14	B	Moderate	1	N
	c.4132G > T	c.4132G > T	p.E1378*	Exon 14	B	Severe	1	Y
	c.4804C > T	c.4804C > T	p.Q1602*	Exon 14	B	Moderate	1	N
	c.5063C > A	c.5063C > A	p.S1688*	Exon 14	a3	Moderate	1	Y
	c.5080G > T	c.5080G > T	p.E1694*	Exon 14	a3	Moderate	1	N
	c.5143C > T	c.5143C > T	p.R1715*	Exon 14	A3	Severe	1	N
	c.5677C > T	c.5677C > T	p.Q1893*	Exon 17	A3	Severe	1	N
	c.5878C > T	c.5878C > T	p.R1960*	Exon 18	A3	Severe	1	N
	c.6496C > T	c.6496C > T	p.R2166*	Exon 23	C1	Moderate/severe	1/1	N
Frameshift	c.209_212delTTGT	c.209_212delTTGT	p.F70*	Exon 2	A1	Severe	2	N
	c.755_756delCA	c.755_756delCA	p.T252Sfs*32	Exon 6	A1	Severe	1	N
	c.1248delC	c.1248delC	p.L417*	Exon 8	A2	Moderate	1	Y
	c.1454delA	c.1454delA	p.K485Rfs*30	Exon 10	A2	Moderate	1	Y
	c.1487dupC	c.1487dupC	p.H497Sfs*5	Exon 10	A2	Severe	1	Y
	c.2534_2537delACAG	c.2534_2537delACAG	p.D845Vfs*8	Exon 14	B	Moderate	1	N
	c.3001delG	c.3001delG	p.A1001Lfs*3	Exon 14	B	Severe	1	N

(*Continued*)

Table 1 (*Continued*)

	c.3441delA	c.3441delA	p.Q1147Hfs*2	Exon 14	B	Moderate	1	Y
	c.3528dupT	c.3528dupT	p.T1177Yfs*21	Exon 14	B	Severe	1	Y
	c.3583delC	c.3583delC	p.T1196Lfs*22	Exon 14	B	Severe	1	Y
	c.3637delA	c.3637delA	p.I1213Ffs*5	Exon 14	B	Moderate	2	N
	c.3637dupA	c.3637dupA	p.I1213Nfs*28	Exon 14	B	Moderate	1	N
	c.3851_3852delCA	c.3851_3852delCA	p.T1284Sfs*35	Exon 14	B	Severe	1	N
	c.3870dupA	c.3870dupA	p.G1291Rfs*29	Exon 14	B	Severe	1	N
	c.4052_4053ins GAAAAAAGGT	c.4052_4053ins GAAAAAAGGT	p.I1351Mfs*7	Exon 14	B	Moderate	1	Y
	c.4121_4124delTAGA	c.4121_4124delTAGA	p.I1374Tfs*49	Exon 14	B	Moderate	1	N
	c.4379dupA	c.4379dupA	p.N1460Kfs*2	Exon 14	B	Moderate	1	N
	c.4825dupA	c.4825dupA	p.T1609Nfs*4	Exon 14	B	Moderate/severe	1/1	N
	c.5084_5087delATTT	c.5084_5087delATTT	p.D1695Vfs*35	Exon 14	a3	Moderate	1	N
	c.6267delG	c.6267delG	p.W2089*	Exon 21	C1	Severe	1	Y
Missense	c.262A > G	c.262A > G	p.M88V	Exon 2	A1	Moderate	1	N
	c.266_267delinsTG	c.266_267delinsTG	p.G89V	Exon 3	A1	Moderate	1	Y
	c.361G > A	c.361G > A	p.G121S	Exon 3	A1	Mild	1	N
	c.389G > A	c.389G > A	p.G130E	Exon 4	A1	Moderate	1	Y
	c.398A > C	c.398A > C	p.Y133S	Exon 4	A1	Moderate	1	N
	c.536C > G	c.536C > G	p.S179C	Exon 4	A1	Severe	1	N
	c.568G > A	c.568G > A	p.G190S	Exon 4	A1	Severe	1	Y
	c.569G > A	c.569G > A	p.G190D	Exon 4	A1	Severe	1	N
	c.974T > G	c.974T > G	p.F325C	Exon 7	A1	Moderate	1	Y
	c.976C > G	c.976C > G	p.L326V	Exon 7	A1	Mild	1	Y
	c.1238A > G	c.1238A > G	p.D413G	Exon 8	A2	Moderate	1	N
	c.1316G > A	c.1316G > A	p.G439D	Exon 9	A2	Severe	1	N
	c.1420G > A	c.1420G > A	p.G474R	Exon 9	A2	Moderate	1	N
	c.1505T > G	c.1505T > G	p.V502G	Exon 10	A2	Moderate	1	N
	c.1615G > A	c.1615G > A	p.G539R	Exon 11	A2	Mild/moderate	1/1	N
	c.1991A > G	c.1991A > G	p.Q664R	Exon 13	A2	Severe	1	N
	c.2080G > A	c.2080G > A	p.G694R	Exon 13	A2	Moderate	1	N
	c.5183A > G	c.5183A > G	p.Y1728C	Exon 14	A3	Severe	1	N

Table 1 (*Continued*)

	c.5246T > C	c.5246T > C	p.F1749S	Exon 15	A3	Moderate	1	N
	c.5308G > A	c.5308G > A	p.E1770K	Exon 15	A3	Severe	1	N
	c.5336G > A	c.5336G > A	p.G1779E	Exon 15	A3	Moderate	1	N
	c.5381T > C	c.5381T > C	p.F1794S	Exon 16	A3	Severe	1	N
	c.5399G > A	c.5399G > A	p.R1800H	Exon 16	A3	Moderate	1	N
	c.5405A > C	c.5405A > C	p.Y1802S	Exon 16	A3	Moderate/severe	1/1	N
	c.5417C > T	c.5417C > T	p.S1806F	Exon 16	A3	Mild/moderate	1/2	Y
	c.5426T > G	c.5426T > G	p.I1809S	Exon 16	A3	Mild	1	Y
	c.5572T > C	c.5572T > C	p.S1858P	Exon 16	A3	Severe	1	Y
	c.5843T > C	c.5843T > C	p.L1948P	Exon 18	A3	Moderate	1	N
	c.6103G > A	c.6103G > A	p.V2035M	Exon 19	A3	Moderate	1	N
	c.6134G > T	c.6134G > T	p.G2045V	Exon 20	C1	Moderate	1	N
	c.6200C > T	c.6200C > T	p.P2067L	Exon 21	C1	Moderate	1	N
	c.6520C > G	c.6520C > G	p.H2174D	Exon 23	C1	Moderate	1	N
	c.6544C > T	c.6544C > T	p.R2182C	Exon 23	C1	Moderate	1	N
	c.6977G > T	c.6977G > T	p.R2326L	Exon 26	C2	Moderate	1	N
Splicing	c.144–26A > C	c.144–26A > C	NA	Intron 1	NA	Severe	1	Y
	c.1538–1G > A	c.1538–1G > A	NA	Intron 10	NA	Severe	1	N
	c.5815 + 1G > A	c.5815 + 1G > A	NA	Intron 17	NA	Severe	1	N
	c.6273 + 1G > A	c.6273 + 1G > A	NA	Intron 21	NA	Moderate/severe	1/1	N
	c.6574 + 1G > C	c.6574 + 1G > C	NA	Intron 23	NA	Moderate	1	N
Inframe deletion	c.2015_2017delTCT	c.2015_2017delTCT	p.F672del	Exon 13	A2	Severe	1	N
Large insertion (***n*** = 1)	c.4342_4343ins300bp	c.4342_4343ins300bp	p.Q1448Lfs*29	Exon 14	B	Severe	1	Y
Large deletion (***n*** = 7)	Exon26del	c.6900 + 5519_*1806?del (chrX:154825172–154854913)	NA	Intron 25- *SMIM9*	C2	Severe	1	Y
	Exon15–26del	c.5020–7853_*1806?del (chrX:154832502–154914426)	NA	Intron 14- *SMIM9*	B	Severe	1	Y
	Exon26del	c.6901–7000_*1598del (chrX:154835999–154844752)	NA	Intron 25- Exon 26	C2	Severe	1	Y

(*Continued*)

Table 1 (*Continued*)

	Exon23–25del	c.6550_6900 + 3106del (chrX:154857328–154863111)	NA	Exon 23- Intron 25	C1	Moderate	1	Y
	Exon20–22del	c.6115 + 95_6429 + 10973del (chrX:154885104–154901956)	NA	Intron 19- Intron 22	C1	Severe	1	Y
	Exon14del	c.2114–1981_4503delinsTG (chrX:154929287–154933657)	NA	Intron 13- Exon 14	NA	Severe	1	Y
	Exon6del	c.671–718_c.787 + 5619 (chrX:154979068–154985521)	NA	Intron 5- Intron 6	NA	Severe	1	Y
Total ($n = 87$)							131	

Abbreviations: CGR, complex gene rearrangement; N, no; NA, not applicable; Y, yes.

identified (►**Fig. 2B**) and selected Type I and Type II inversions were displayed by IGV plots (►**Fig. 2C, D**). This study also enrolled a male proband with rare Int22-CGR, which had fragments of Int22-hour1, h2, h3, and h1–3 by CAHEA assay (►**Fig. 2E**), correlated to fragments PQ, AB, AB, and AQ in conventional LR-PCR assay (►**Fig. 2F**), respectively. This rare Int22-CGR was similar to a previously reported case,[34] but the exact mechanism of rearrangement was not elucidated.

CAHEA Assay for SNVs/Indels

A total of 75 different SNVs/indels were identified by CAHEA in 85 HA pedigrees, including 15 nonsense, 20 frameshift, 34 missense, and 5 splicing-site variants, as well as one inframe deletion (►**Table 1**). Twenty-three (30.3%) variants were located in Exon 14, and the others spread all over the *F8* gene sequence. Of all the 75 variants, 20 (26.3%) had not been previously described including 3 nonsense, 8 frameshift, 8 missense, and 1 splicing-site variant. All the 19 novel exonic variants were predicted to be probably damaging by PolyPhen and intolerant by SIFT. One novel intronic variant c.144–26C > A of splicing site was identified in a patient with severe HA. Eight recurrent SNVs/indels including three in Exon 14 were identified in 17 HA pedigrees.

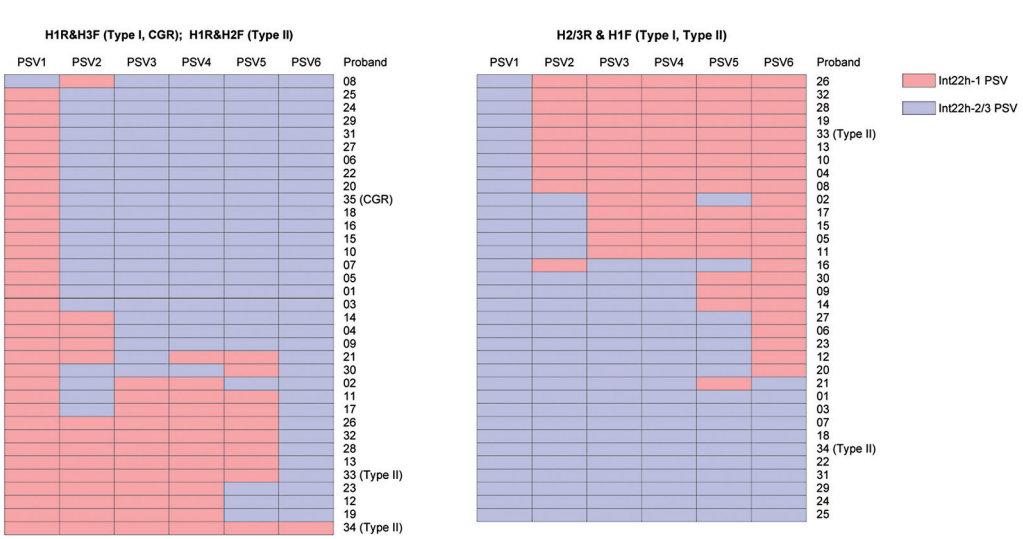
CAHEA Assay for Large Insertions and Deletions

Our studies enrolled one pedigree with a 300-bp insertion and seven pedigrees with gross large deletions (►**Table 1**, ►**Supplementary Table S1,** available in the online version). CAHEA directly identified an insertion sequence of 300-bp in Exon 14 (►**Fig. 3A**), which was confirmed by Sanger sequencing (►**Fig. 3B**). The insertion was a retrotransposition of an Alu element in antisense orientation, with the polyT tail and 22-bp target-site duplication. The insertion led to translational frameshift and severe HA in proband 124. Gross large deletions were identified by MLPA in seven pedigrees, of which all the 14 breakpoints were directly characterized at single-nucleotide resolution by CAHEA (►**Fig. 4A**). Two centromeric breakpoints were located in *SMIM9* and the other 10 breakpoints in *F8*, encompassing from intron 5 to exon 26. Overall, the deletions ranged from 4.4 to 81.9 kb with an average size of 22.0 kb. CAHEA identified a 5.8-kb deletion through primer par E23F and E25R, which amplified a 7.5-kb fragment in normal control, but only a 1.7-kb fragment in proband 128 (►**Fig. 4B**). CAHEA further determined the exact deletion region was chrX:154,835,999–154,844,752, which was confirmed by Sanger sequencing (►**Fig. 4C**). CAHEA identified the 8.8-kb deletion encompassing intron 25 to exon 26 in proband 127 through two levels of evidence. First, the 9.5-kb fragment amplified by primer pair E26F and E26R in normal control was absent, suggesting the existence of a deletion in this region. Second, the primer pair GF11 and E26R was 18.0 kb apart and could not amplify any fragments in normal control but amplified an 8.5-kb fragment in proband 127, suggesting a 9.5-kb deletion in this region (►**Fig. 4D**). Sanger sequencing confirmed the deletion region chrX:154,835,999–154,844,752 identified by CAHEA assay (►**Fig. 4E**).

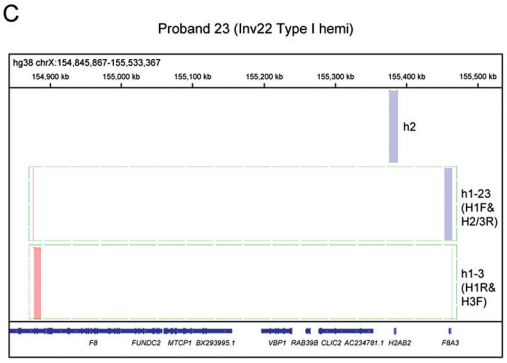
A

PSV	1	2	3	4	5	6
Int22-h1 position (hg38 chrX:)	154,880,923	154,885,252	154,885,988	154,886,026	154,886,234	154,889,872
Int22-h1 reference	A	C	T	G	T	T
Int22-h2 position (hg38 chrX:)	155,376,614	155,380,989	155,381,728	155,381,766	155,381,974	155,385,604
Int22-h2 reference	G	T	C	C	C	C
Int22-h3 position (hg38 chrX:)	155,464,100	155,459,725	155,458,987	155,458,949	155,458,741	155,455,106
Int22-h3 reference (reverse complement)	G	T	C	C	C	C

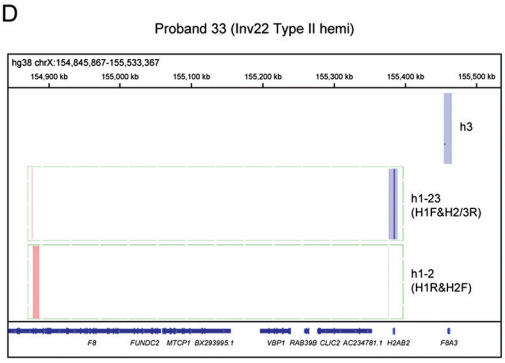
B



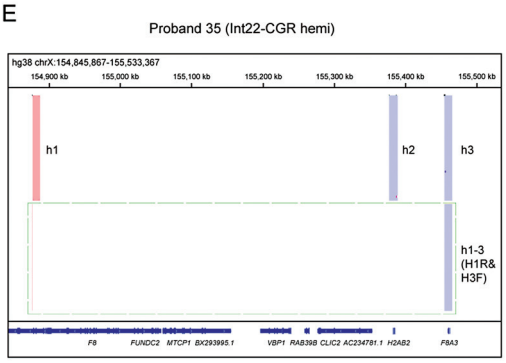
C



D



E



F

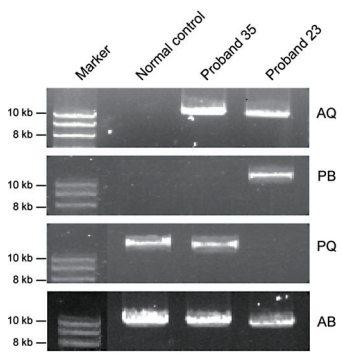


Fig. 2 Analysis of recombination regions of Inv22. (**A**) The positions and sequences of six PSVs identified among three Int22 duplicons. (**B**) Heatmaps showing the Int22 PSVs of CCS reads from CAHEA in 35 probands with Int22-related gene rearrangements. Left: CCS reads with primer pair of H1R&H3F for Inv22 Type I, Int22-CGR, or H1R&H2F for Inv22 Type II; right: CCS reads with primer pair H1F&H2/3R for Inv22 Type I and II. (**C**) IGV plots displaying the CCS reads of Inv22 Type I in proband 23. (**D**) IGV plots displaying the CCS reads of Inv22 Type II in proband 33. (**E**) IGV plots displaying the CCS reads of Int22-CGR in proband 35. (**F**) Conventional LR-PCR followed by agarose gel electrophoresis for proband 23 and 35. CAHEA, comprehensive analysis of hemophilia A; CCS, circular consensus sequencing; IGV, Integrative Genomics Viewer; LR-PCR, long-range polymerase chain reaction; PSV, paralogous sequence variant.

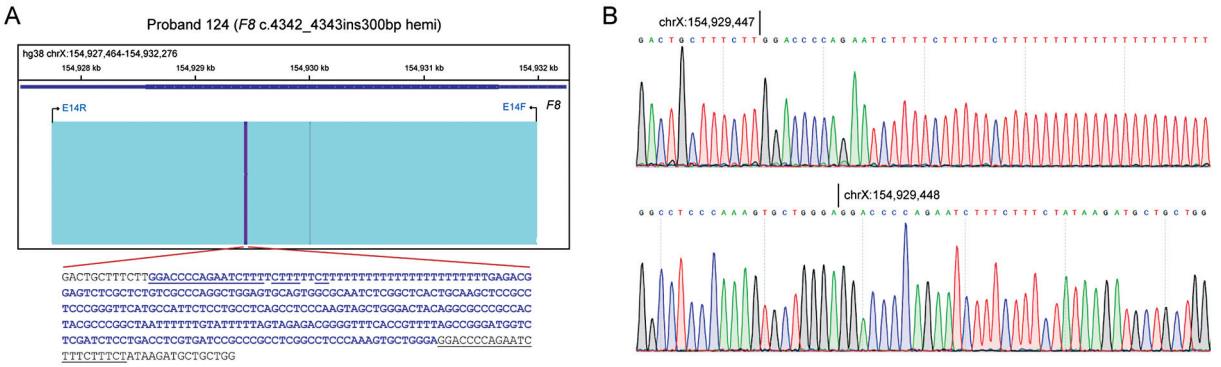


Fig. 3 Identification of a large insertion by CAHEA. (**A**) IGV plots showing the 300-bp insertion in proband 124. The 300-bp-inserted sequences were shown in blue, and the sequences with microhomology were underlined. (**B**) Confirmation of the 300-bp insertion by Sanger sequencing. CAHEA, comprehensive analysis of hemophilia A; IGV, Integrative Genomics Viewer.

The breakpoint junctions of the seven large deletions were then characterized to speculate the possible mechanisms for large deletions (►**Fig. 4F**). The 29.7-kb deletion in proband 125 had 64-bp homology between the two breakpoints due to a pair of Alu elements. The deletions in proband 127, 128, and 131 displayed 2-bp, 4-bp, and 3-bp of matching sequence at junctions, respectively. These suggested that a microhomology-mediated mechanism such as alternative end-joining and replication-based mechanisms might be involved in the formation of these four deletions. In contrast, the other three deletions had either simple blunt ends, only 1-bp matching sequence, or 2-bp insertion at junctions, which suggested the involvement of nonhomologous end-joining mechanisms.

Validation of CAHEA Assay for Prospective *F8* Genetic Testing

To further validate the clinical utility, 22 samples from 14 pedigrees with HA history were subjected to *F8* genetic testing by conventional methods and CAHEA assay simultaneously. CAHEA identified nine Inv22 Type I and nine SNVs/indels, which were all concordant with conventional methods (►**Table 2**). One novel variant c.6116–2delA and one deep intronic variant c.143 + 1567A > G of splicing-site were identified in two pedigrees, respectively.

Discussion

Accurate and complete genetic characterization of *F8* variants requires a combination of different conventional methods like LR-PCR, IS-PCR, Sanger sequencing, NGS, and MLPA. In this pilot study, the performance of LRS-based CAHEA approach was first evaluated with 272 retrospective clinical samples from 131 unrelated HA pedigrees for *F8* genetic analysis. Then this novel approach was prospectively validated with 22 clinical samples from 14 unrelated HA pedigrees. The accuracy of CAHEA was confirmed by showing 100% concordance with control methods in terms of the variant types of *F8*. CAHEA offered some additional advantages compared with conventional combined methods. First, CAHEA allowed for full characterization of the *F8* gene variants in one assay. Second, with the ability to distinguish among Int22 duplicons and identify signature PSVs, CAHEA enabled better understanding of the junction regions of Int22-related gene rearrangements. Third, CAHEA directly identified the breakpoints of large gene deletions at single-nucleotide resolution, which helped speculating the mechanisms involved in the deletions.

Multiplex LR-PCR with two to six fragments combined with LRS has achieved comprehensive genetic analysis for disorders like thalassemia.[30–33] Still, full characterization of HA using similar technique was particularly challenging because of the huge gene length and complicated variants of *F8*. Multiplex LR-PCRs for 15 fragments with the sizes between 3.0 and 12.0 kb were needed to cover the essential gene regions, and more gap primers were added to ensure identification of large deletions. The key points were to screen for primers that could work well and compatibly, and to test the dosage of each primer in the reaction to allow for somewhat balanced amplification of each fragment. Here the maximum difference of read depths among different fragments was less than 50-fold for DNA samples with good quality, and over 95% of samples including both peripheral blood and chorionic villus sampling passed quality control for read number at first experiment. With 48 barcoded samples pooled in one PacBio Sequel II flowcell, the reagent and sequencing cost is estimated to be approximately $80 per sample, with a turnaround time of 8 days. For conventional genetic testing, the costs for Inv22/1 by LR-PCR or IS-PCR, SNVs/indels by Sanger or NGS panel, and duplication/deletion by MLPA are approximately $30, $200 to $400, and $100, respectively. The total cost for conventional genetic testing could add up as high as $330 to $530.

Due to high homology between Int22h-1 and Int22h-2/3, the junction regions of Int22-related inversions have not been investigated. Here by extensive comparison between genome build hg38 and CHM13, between LRS data and reference, among three Int22 duplicons, six PSVs between Int22h-1 and Int22h-2/3 were identified and the junction regions in each Inv22 were determined.

Nonetheless, this study is not without any limitations. First, unlike MLPA that utilizes probe quantification to

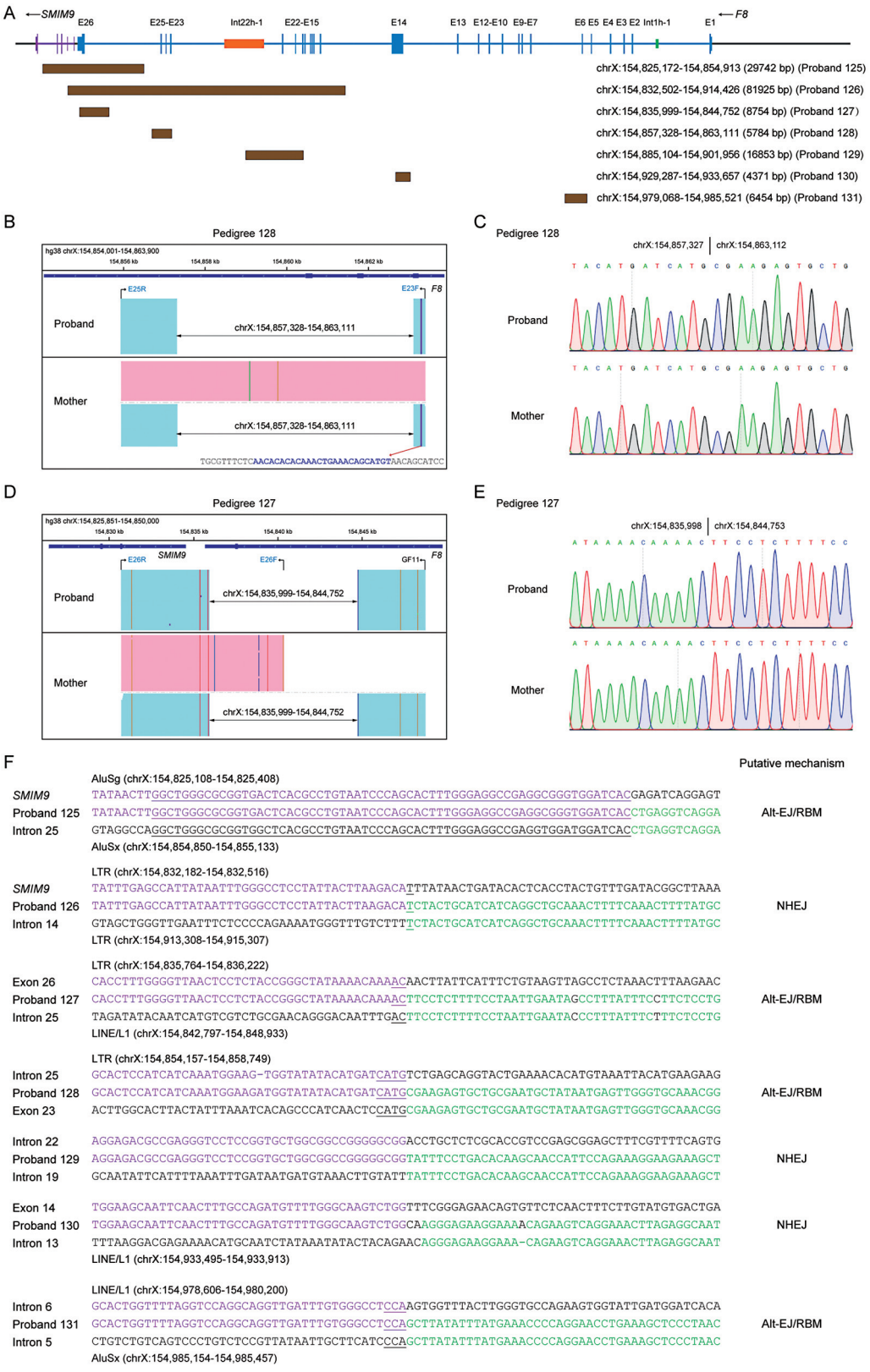


Fig. 4 Identification of large deletions by CAHEA. (**A**) Mapping of deletion breakpoints according to the reference genome build hg38. (**B**) IGV plots showing the 5.8-kb deletion in the proband and mother of pedigree 128 by CAHEA. (**C**) Confirmation of the 5.8-kb deletion in the proband and mother of pedigree 128 by Sanger sequencing. (**D**) IGV plots showing the 8.8-kb deletion in the proband and mother of pedigree 127 by CAHEA. (**E**) Confirmation of the 8.8-kb deletion in the proband and mother of pedigree 127 by Sanger sequencing. (**F**) Sequence of breakpoint junctions of seven HA probands with large *F8* deletions. Reference sequences encompassing the breakpoints were shown above and below the sequences of probands. The sequences with microhomology in the breakpoint junctions were underlined. Repetitive elements encompassing the breakpoint junctions were also labeled. CAHEA, comprehensive analysis of hemophilia A; IGV, Integrative Genomics Viewer.

Table 2 Validation of CAHEA with a new set of HA pedigrees

Pedigree	Sample	Type	Gender	Phenotype	Conventional methods, nucleotide change	CAHEA assay,nucleotide change	Protein change	Gene region
B01	BHA001	Carrier	F	NA	Inv22 het	Inv22 Type I het	NA	Intron 22
B02	BHA002	Carrier	F	NA	Inv22 het	Inv22 Type I het	NA	Intron 22
	BHA003	Fetus of BHA002	NA	NA	ND	ND	NA	NA
B03	BHA004	Carrier	F	NA	Inv22 het	Inv22 Type I het	NA	Intron 22
B04	BHA005	Carrier	F	NA	c.2080G > A het	c.2080G > A het	p.G694R	Exon 13
	BHA006	Fetus of BHA005	NA	NA	ND	ND	NA	NA
	BHA007	Carrier (sister of BHA005)	F	NA	c.2080G > A het	c.2080G > A het	p.G694R	Exon 13
	BHA008	Fetus of BHA007	F	NA	c.2080G > A het	c.2080G > A het	p.G694R	Exon 13
B05	BHA009	Carrier	F	NA	Inv22 het	Inv22 Type I het	NA	Intron 22
	BHA010	Fetus of BHA009	F	NA	ND	ND	NA	NA
B06	BHA011	Proband	M	Severe	Inv22 hemi	Inv22 Type I hemi	NA	Intron 22
B07	BHA012	Carrier	F	NA	c.5392G > A het	c.5392G > A het	p.A1798T	Exon 16
B08	BHA013	Carrier	F	NA	Inv22 het	Inv22 Type I het	NA	Intron 22
B09	BHA014	Carrier (mother)	F	NA	Inv22 het	Inv22 Type I het	NA	Intron 22
	BHA015	Patient (proband)	M	Severe	Inv22 hemi	Inv22 Type I hemi	NA	Intron 22
B10	BHA016	Carrier	F	NA	c.361G > A het	c.361G > A het	p.G121S	Exon 3
B11	BHA017	Carrier (mother)	F	NA	c.6116–2delA* het	c.6116–2delA het	NA	Intron 19
	BHA018	Patient (proband)	M	Severe	c.6116–2delA hemi	c.6116–2delA hemi	NA	Intron 19
B12	BHA019	Carrier	F	NA	Inv22 het	Inv22 Type I het	NA	Intron 22
B13	BHA020	Carrier	F	NA	c.143 + 1567A > G het	c.143 + 1567A > G het	NA	Intron 1
	BHA021	Fetus of BHA020	M	NA	ND	ND	NA	NA
B14	BHA022	Fetus	M	NA	c.3583delC hemi	c.3583delC hemi	p.T1196Lfs*22	Exon14

Abbreviations: F, female; M, male; NA, not applicable; ND, not detected.
*Novel variant.

determine the existence of large deletions, CAHEA assay relies on the combination of primers to directly identify the breakpoints of deletions, which is more straightforward but could miss very large deletions not covered by the primers. With the addition of more gap primers, CAHEA has the potential to cover more large deletions. Second, this study had no cases with large duplication of *F8* gene locus, which could also cause severe HA. There are no reports of recurrent *F8* duplication, and it is difficult to directly amplify large nonrecurrent duplications by LR-PCR and identify the breakpoints by CAHEA assay. However, long sequencing reads make it possible for haplotype analysis. One or more *F8* fragments with two haplotypes in males or three haplotypes in females would suggest the existence of large duplications. Similar approach has been applied to identify duplication of α-globin gene locus for thalassemia genetic testing.[35] Third, though CAHEA is cost-effective compared with conventional methods combined, the PacBio sequencing platform is very expensive. Developing a less-expensive benchtop platform would make LRS-based assays more approachable clinically. Therefore, prospective cohort studies would be needed in the future to further explore the advantages and clinical utility of CAHEA.

Conclusion

Multiplex LR-PCR and LRS-based CAHEA represents a comprehensive approach for identifying Int22 and Int1-related gene rearrangements, SNVs/indels, and large insertions/deletions in the *F8* gene. With further improvement on performance, decrease of the sequencing cost, and more clinical studies to prove its clinical feasibility, CAHEA is envisaged to have great clinical application for carrier screening and genetic diagnosis of HA.

What is known about this topic?

- Hemophilia A is caused by a large and heterogeneous spectrum of *F8* gene variants.
- *F8* genetic testing is complicated and often requires combination of different methods.

What does this paper add?

- A single long-read sequencing approach termed CAHEA to detect the majority of *F8* gene variants.
- The approach can annotate the long-range PCR to detect intron 22 and intron 1 inversions and analyze polymorphisms in the repeat regions.

Note

Dr. Qiaowei Liang's present address is Department of Human Genetics, Graduate School of Medicine, Yokohama City University, Yokohama, Japan.

Funding

This work was supported by the National Key Research and Development Program of China (2022YFC2703700, 2022YFC2703400, 2021YFC1005302) and the National Natural Science Foundation of China (82271752).

Conflict of Interest

A.M., T.X., and L.W. are employees of Berry Genomics Corporation. The authors declare that they have no conflict of interest.

Acknowledgment

We thank all the subjects who participated in this study.

References

1 Bolton-Maggs PH, Pasi KJ. Haemophilias A and B. Lancet 2003;361 (9371):1801–1809
2 White GC II, Rosendaal F, Aledort LM, Lusher JM, Rothschild C, Ingerslev JFactor VIII and Factor IX Subcommittee. Definitions in hemophilia. Recommendation of the scientific subcommittee on factor VIII and factor IX of the scientific and standardization committee of the International Society on Thrombosis and Haemostasis. Thromb Haemost 2001;85(03):560
3 Bagnall RD, Giannelli F, Green PM. Int22h-related inversions causing hemophilia A: a novel insight into their origin and a new more discriminant PCR test for their detection. J Thromb Haemost 2006;4(03):591–598
4 Bagnall RD, Giannelli F, Green PM. Polymorphism and hemophilia A causing inversions in distal Xq28: a complex picture. J Thromb Haemost 2005;3(11):2598–2599
5 Rossetti LC, Radic CP, Abelleyro MM, Larripa IB, De Brasi CD. Eighteen years of molecular genotyping the hemophilia inversion hotspot: from southern blot to inverse shifting-PCR. Int J Mol Sci 2011;12(10):7271–7285
6 Mühle C, Zenker M, Chuzhanova N, Schneider H. Recurrent inversion with concomitant deletion and insertion events in the coagulation factor VIII gene suggests a new mechanism for X-chromosomal rearrangements causing hemophilia A. Hum Mutat 2007;28(10):1045
7 Fujita J, Miyawaki Y, Suzuki A, et al. A possible mechanism for Inv22-related F8 large deletions in severe hemophilia A patients with high responding factor VIII inhibitors. J Thromb Haemost 2012;10(10):2099–2107
8 Chen C, Xie X, Wu X, et al. Complex recombination with deletion in the F8 and duplication in the TMLHE mediated by int22h copies during early embryogenesis. Thromb Haemost 2017;117(08): 1478–1485
9 Antonarakis SE, Rossiter JP, Young M, et al. Factor VIII gene inversions in severe hemophilia A: results of an international consortium study. Blood 1995;86(06):2206–2212
10 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(01):168–174
11 Abelleyro MM, Radic CP, Marchione VD, et al. Molecular insights into the mechanism of nonrecurrent F8 structural variants: full breakpoint characterization and bioinformatics of DNA elements implicated in the upmost severe phenotype in hemophilia A. Hum Mutat 2020;41(04):825–836
12 Lannoy N, Hermans C. Principles of genetic variations and molecular diseases: applications in hemophilia A. Crit Rev Oncol Hematol 2016;104:1–8
13 Jourdy Y, Chatron N, Fretigny M, Dericquebourg A, Sanlaville D, Vinciguerra C. Comprehensive analysis of F8 large deletions: characterization of full breakpoint junctions and description of

a possible DNA breakage hotspot in intron 6. J Thromb Haemost 2022;20(10):2293–2305
14 Payne AB, Miller CH, Kelly FM, Michael Soucie J, Craig Hooper W. The CDC Hemophilia A Mutation Project (CHAMP) mutation list: a new online resource. Hum Mutat 2013;34(02):E2382–E2391
15 McVey JH, Rallapalli PM, Kemball-Cook G, et al. The European Association for Haemophilia and Allied Disorders (EAHAD) coagulation factor variant databases: important resources for haemostasis clinicians and researchers. Haemophilia 2020;26(02):306–313
16 Kemball-Cook G, Tuddenham EG, Wacey AI. The factor VIII structure and mutation resource site: HAMSTeRS version 4. Nucleic Acids Res 1998;26(01):216–219
17 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(04):1458–1459
18 Rossetti LC, Radic CP, Larripa IB, De Brasi CD. Genotyping the hemophilia inversion hotspot by use of inverse PCR. Clin Chem 2005;51(07):1154–1158
19 Rossetti LC, Radic CP, Larripa IB, De Brasi CD. Developing a new generation of tests for genotyping hemophilia-causative rearrangements involving int22h and int1h hotspots in the factor VIII gene. J Thromb Haemost 2008;6(05):830–836
20 Santacroce R, Acquila M, Belvini D, et al; AICE-Genetics Study Group. Identification of 217 unreported mutations in the F8 gene in a group of 1,410 unselected Italian patients with hemophilia A. J Hum Genet 2008;53(03):275–284
21 Citron M, Godmilow L, Ganguly T, Ganguly A. High throughput mutation screening of the factor VIII gene (F8C) in hemophilia A: 37 novel mutations and genotype-phenotype correlation. Hum Mutat 2002;20(04):267–274
22 Chen J, Li Q, Lin S, et al. The spectrum of FVIII gene variants detected by next generation sequencing in 236 Chinese non-inversion hemophilia A pedigrees. Thromb Res 2021;202:8–13
23 Inaba H, Shinozawa K, Amano K, Fukutake K. Identification of deep intronic individual variants in patients with hemophilia A by next-generation sequencing of the whole factor VIII gene. Res Pract Thromb Haemost 2017;1(02):264–274
24 Pezeshkpoor B, Zimmer N, Marquardt N, et al. Deep intronic 'mutations' cause hemophilia A: application of next generation sequencing in patients without detectable mutation in F8 cDNA. J Thromb Haemost 2013;11(09):1679–1687
25 Rost S, Löffler S, Pavlova A, Müller CR, Oldenburg J. Detection of large duplications within the factor VIII gene by MLPA. J Thromb Haemost 2008;6(11):1996–1999
26 Lu Y, Xin Y, Dai J, et al. Spectrum and origin of mutations in sporadic cases of haemophilia A in China. Haemophilia 2018;24(02):291–298
27 Repessé Y, Slaoui M, Ferrandiz D, et al. Factor VIII (FVIII) gene mutations in 120 patients with hemophilia A: detection of 26 novel mutations and correlation with FVIII inhibitor development. J Thromb Haemost 2007;5(07):1469–1476
28 Guo Z, Yang L, Qin X, Liu X, Zhang Y. Spectrum of molecular defects in 216 Chinese families with hemophilia A: identification of noninversion mutation hot spots and 42 novel mutations. Clin Appl Thromb Hemost 2018;24(01):70–78
29 Johnsen JM, Fletcher SN, Huston H, et al. Novel approach to genetic analysis and results in 3000 hemophilia patients enrolled in the My Life, Our Future initiative. Blood Adv 2017;1(13):824–834
30 Liang Q, Gu W, Chen P, et al. A more universal approach to Comprehensive Analysis of Thalassemia Alleles (CATSA). J Mol Diagn 2021;23(09):1195–1204
31 Liu Y, Chen M, Liu J, et al. Comprehensive analysis of congenital adrenal hyperplasia using long-read sequencing. Clin Chem 2022;68(07):927–939
32 Li S, Han X, Xu Y, et al. Comprehensive analysis of spinal muscular atrophy: SMN1 copy number, intragenic mutation, and 2 + 0 carrier analysis by third-generation sequencing. J Mol Diagn 2022;24(09):1009–1020
33 Liang Q, Liu Y, Liu Y, et al. Comprehensive analysis of fragile X syndrome: full characterization of the FMR1 locus by long-read sequencing. Clin Chem 2022;68(12):1529–1540
34 Zimmermann MA, Oldenburg J, Müller CR, Rost S. Unusual genomic rearrangements in introns 1 and 22 of the F8 gene. Hamostaseologie 2011;31(Suppl 1):S69–S73
35 Jiang F, Mao AP, Liu YY, et al. Detection of rare thalassemia mutations using long-read single-molecule real-time sequencing. Gene 2022;825:146438