European Journal of Human Genetics
https://doi.org/10.1038/s41431-018-0334-9

ARTICLE

Inverse PCR to perform long-distance haplotyping: main applications to improve preimplantation genetic diagnosis in hemophilia

Miguel Martín Abelleyro[1] · Vanina Daniela Marchione[1] · Micaela Palmitelli[1] · Claudia Pamela Radic[1] · Daniela Neme[2] · Irene Beatriz Larripa[1] · Enrique Medina-Acosta[3] · Carlos Daniel De Brasi[1,4] · Liliana Carmen Rossetti[1]

Received: 7 September 2018 / Revised: 28 November 2018 / Accepted: 4 December 2018

Abstract

Among other applications of long-distance haplotype phasing in clinical genetics, determination of linked DNA markers as surrogate for problematic structural variants (e.g., repeat-mediated rearrangements) is essential to perform diagnosis from low-quality DNA samples. We describe a next-of-kin-independent (physical) phasing approach based on inverse-PCR (iPCR) paired-end amplification (PI). This method enables typing the multialleles of the short tandem repeat (STR) *F8*Int21[CA]n at the *F8*-intron 21, as a surrogate DNA marker for the *F8*-intron 22 inversion (Inv22), the hemophilia A-causative hotspot, within the transmitted haplotype in informative carriers. We provide proof-of-concept by blindly validating the PI approach in 15 carrier mother/affected-son duos. Every *F8*Int21[CA]n STR allele determined in phase with the Inv22 allele in the female carriers from the informative duos was confirmed in the hemizygous proband ($P = 0.00003$). A second surrogate STR *locus* at the *F8*-IVS22 was obtained by the PI approach improving severe-HA preimplantation genetic diagnosis by augmenting heterozygosity in Inv22 carriers bypassing the requirement for family linkage analysis. The ability of the PI-assay to combine other marker pairs was demonstrated by haplotyping a SNV (*F8*:c.6118T > C) with a >28kb-distant *F8*-IVS22 STR. The PI approach has proven flexibility to target different marker pairs and has potential for multiplex characterization of iPCR products by massively parallel sequencing.

Introduction

Haplotype phase determination relies on the identification of alleles that segregate on the same chromosome. Haplotype phasing of physically distant linked alleles is crucial to ascertain which chromosome copy carries a known or unknown variants associated with disease by interrogating surrogate DNA *loci* within the transmitted haplotype. Haplotype phasing is used extensively in medical genetics, including association and disease risk estimation studies [1, 2], epigenetic regulation of gene expression [3], evolution and population genetic studies [4, 5], pharmacogenomics [6, 7], and to identify surrogate markers for complex rearrangements in prenatal diagnosis in which the reduced quality/quantity of DNA samples from such instances (e.g., preimplantation genetic diagnosis, PGD, and cell-free fetal DNA from maternal plasma) impedes its direct identification.

A method for high-throughput haplotyping by use of microfluidics-based linked-read sequencing technology has been recently described [8, 9]. Although this approach

These authors contributed equally: Miguel Martín Abelleyro, Vanina Daniela Marchione

These authors contributed equally: Carlos Daniel De Brasi, Liliana Carmen Rossetti

Supplementary information The online version of this article (https://doi.org/10.1038/s41431-018-0334-9) contains supplementary material, which is available to authorized users.

✉ Liliana Carmen Rossetti
rossetti@hematologia.anm.edu.ar

[1] Instituto de Medicina Experimental (IMEX), CONICET-Academia Nacional de Medicina, Buenos Aires, Argentina

[2] Fundación de la Hemofilia Alfredo Pavlovsky, Buenos Aires, Argentina

[3] Universidade Estadual do Norte Fluminense Darcy Ribeiro, Centro de Biociências e Biotecnologia, Laboratório de Biotecnologia, Núcleo de Diagnóstico e Investigação Molecular, Campos dos Goytacazes, Rio de Janeiro, Brazil

[4] Instituto de Investigaciones Hematológicas Mariano R Castex, Academia Nacional de Medicina, Buenos Aires, Argentina

Published online: 09 January 2019

resolves haplotype information for most genetic variants also including genomic rearrangements, it fails to identify, for example, duplicon-mediated inversions and applies state-of-the-art technology, which is not available for most genetic testing laboratories worldwide.

We describe a cost-effective next-of-kin-independent practical approach for phasing alleles over long distances based on allele-specific inverse-PCR (iPCR) amplification. We termed the method "physical phasing by inverse-PCR" (PI). As a proof-of-concept, we used the hemophilia A (HA) (OMIM #306700) inversion hotspot, the *F8* intron 22 inversion, NC_000023.11:g.(154880816_154890330)_(155454664_155464178)inv (Inv22), a paradigmatic structural variant caused by non-allelic homologous recombination between inverted 10 kb duplicons (*int22h*) involved in half of the severe HA cases worldwide [10, 11]. We validated the determination of the gametic phase between the Inv22 and the *F8*Int21[CA]n (NC_000023.11:g.154896506AC[26]) STR *locus* [12] within the *F8*-IVS21 in female carriers. A second STR *locus* within the *F8*-IVS22 (NC_000023.11:g.154875736GT[19], *F8*Int22[GT]n) was phased with the Inv22 for improving the likelihoods for informative carriers. Finally, to show the adaptability of the PI approach to different practical scenarios, a single-nucleotide variant (SNV) in *F8*-exon 20, NM_000132.3:c.6118T > C, was phased with the 28.4 kb-distant *F8*Int22[GT]n STR.

Materials and methods

Subjects and samples

The PI assay that links the Inv22 and the wild-type (Wt) allele with the *F8*Int21[CA]n STR alleles was carried on genomic DNA (gDNA) of peripheral blood collected from 55 participating female donors (52 Inv22 carriers and 3 non-carriers) and 28 severe-HA male patients belonging to 49 Inv22-informative families (38 with Inv22 type 1 and 11 with Inv22 type 2), previously genotyped by inverse shifting-PCR (IS-PCR) version-2008 [13]. The PI approach validation was performed on gDNA samples from 15 informative duos (Inv22 carrier mother/affected child; 12 with Inv22 type 1 and 3 with Inv22 type 2) from the same family series.

Population parameters for the *F8*Int21[CA]n STR *locus* were determined in both the hemophilia family series and in 40 healthy controls (25 men and 15 women) from the same geographic regions in Argentina.

Genomic DNA was prepared from peripheral blood leukocytes either by phenol–chloroform [14] or salting-out extraction [15] and ethanol precipitation protocols. Purity, quality, and concentration of gDNA samples were estimated by UV-spectrophotometry (DeNovix DS-11 FX) and agarose gel electrophoresis.

A second application of PI associating the Inv22 allele and the Wt allele with the *F8*Int22[GT]n STR alleles was carried on gDNA from peripheral blood collected from one participating female *F8*-Inv22 type 1 carrier and her severe-HA affected son.

A third application of PI, phasing the variant allele c.6118T > C predicting the missense change p.Cys2040Arg, and the Wt allele with the *F8*Int22[GT]n STR alleles, was carried on gDNA from peripheral blood collected from one participating female carrier and her moderate HA affected brother.

Nomenclature

For the molecular analysis of variants and STR, we used the nucleotide/aminoacids RefSeq of the *F8* gene, mRNA and Protein, NG_011403.1, NM_000132.3, and NP_000123.1, respectively. We used the X-chromosome nucleotide sequence NC_000023.11, genomic build GRCh38.p12.

We reported the SNV NM_000132.3:c.6118T > C in a specific database (European Associations for Haemophilia and Allied Disorders (EAHAD), URL: www.factorviii-db.org/advance_search.php). Marker characteristics are described in HEMApSTR (version 2.2) (URL: www.uenf.br/Uenf/Pages/CBB/LBT/HEMApSTR.html) (Supplemental Table 1).

Long-distance haplotype phasing assay by inverse-PCR

Generation of *Bcl*I-rings: Circularization of *Bcl*I-digested gDNA molecules into DNA rings was performed as was described by Rossetti et al. [13, 16]. An updated circularization reaction entails (i) digestion of 1 µg of gDNA with 10 U of Fast Digest *Bcl*I restriction enzyme (Thermo Scientific, Argentina) in a volume of 30 µL by 30 min at 37 °C, followed by heat inactivation 20 min at 80 °C and ethanol precipitation; (ii) fragment-ends' self-ligation with 3 Weiss U/0.5-µg-gDNA of T4 DNA Ligase (Thermo Scientific, Argentina) in a volume of 200 µL (5 ng gDNA/µL) at 22 °C for 1 h, followed by ethanol precipitation and resuspension in 30 µL of distilled water (approximately 33 ng/µL).

Inverse-PCR reactions were performed in separate reactions from 8 µL (roughly 260 ng) of *Bcl*I-circles for amplification of the *F8*-wild-type allele (primers 22B + ID) and the Inv22 allele (primers 22B + ED): yielding products of 2.5 and 2.6 kb, respectively, using 0.6 µM of each primer (Table 1), 1.5 U of Go Taq® DNA Polymerase (Promega) and standard PCR reagents in a total volume of 20 µL. Thermo-cycling conditions: 94 °C for 3 min (1 cycle), 94 °C

Table 1 Oligonucleotide primer sequences, physical map locations, and references

Name	Sequence 5′ to 3′	[a]GRCh38.p12 coordinates	Reference
ID	ACATACGGTTTAGTCACAAGT	154,877,137-17	[16]
ED	TCCAGTCACTTAGGCTCAG	155,374,502-485 155,466,227-09	[16]
22B	TGCCAACAGATCCACCTAC	154,896,235-17	This study
*F8*Int21F[b]	AGGTGTGGGCAAAGCATTAG	154,896,615-596	[12]
*F8*Int21R	AGCCTGCTTCTTGACAAACC	154,896,471-52	[12]
Int22PH	TAAAAAATCCCAGGTGTTC	154,876,039-21	This study
Int20PH	GATATAATCAGCCCAGGTTC	154,901,344-25	This study
Ex20[A]wt	ATTCTTAGGGGAGTCTGACA	154,901,439-20	This study
Ex20[G]var	ATTCTTAGGGGAGTCTGACG	154,901,439-20	This study
*F8*Int22F[c]	CCAGACATGTCAAGGTGTCAA	154,875,862-82	[17]
*F8*Int22R	GTACTGGGAATGCACAGCCTA	154,875,676-96	[17]

Underlined base T mismatch the reference sequence

[a]NCBI nucleotide sequence file NC_000023.11 genomic build GRCh38.p12 accessed 26/03/2018 [b]Primer used with and without FAM fluorochrome modification at the 5′ end [c]Primer used with FAM fluorochrome modification at the 5′ end

30 s; 52 °C 1 min and 72 °C 6 min (30 cycles) and 72 °C 6 min (1 cycle) on a thermal-cycler Life Express TC-96/G/H (Bioer). Inverse-PCR products were analyzed by ethidium bromide-stained 1% agarose gel electrophoresis.

*F8*Int21[CA]n STR *locus* genotyping was performed using 2 µL of 10^{-6} or 10^{-7} serial dilution of each Wt- and Inv22-allele-specific iPCR product as a substrate of a nested-PCR amplification with primers *F8*Int21F and *F8*Int21R (Table 1). Alternatively, agarose gel blocks containing each allele-specific iPCR band (i.e., Wt: 2529 bp and Inv22: 2601 bp) were excised and suspended separately in 300 µL of distilled water overnight. *F8*Int21[CA]n STR nested-PCR substrate analyses using ten-fold serial dilutions ranging from 10^{-1} to 10^{-3} for each allele-specific iPCR product showed that 2 µL of the 10^{-3} dilution of each 300 µL preparation of Wt- and Inv22-allele were required for optimal STR amplification.

Also, the *Pac*I restriction enzyme was used to generate *Pac*I-rings, substantially as described above, over *Pac*I digested long linear gDNA molecules of 28.5 kb. Inverse-PCR assays were performed separately from 8 µL of *Pac*I-rings for amplification of the *F8*-Wt allele not involved in Inv22 (primers Int20PH + Int22PH, Table 1) yielding a product of 3.1 kb in the presence of 0.6 µM of each primer, 1.75 mM $MgCl_2$, 0.75 U of Kapa Long-Range PCR System (BioSystems), and standard PCR reagents in a total volume of 20 µL. Thermo-cycling conditions were equal to those described for *Bcl*I-circles. Inv22 type 1 allele associates with an extremely long *Pac*I-ring of 69.9 kb, thus hindering their formation and impairing their iPCR analysis.

For the third PI application over the *Pac*I digested long linear gDNA molecules of 28.5 kb rings, iPCR assays were performed separately for amplification of the *F8*-wild-type allele (primers Ex20[A]wt + Int22PH) and the variant allele (primers Ex20[G]var + Int22PH) yielding products of 3.04 kb comprising either the c.6118T-wt or c.6118C-var allele (Table 1).

*F8*Int22[GT]n STR genotyping with primers *F8*Int22F and *F8*Int22R was performed from agarose gel blocks containing each allele-specific iPCR band (i.e., *Pac*I Wt: 3,130 bp for Inv22 phasing; *Pac*I Wt or Var allele-specific amplification for *F8*:c.6118 T > C: 3,035 bp) excised and suspended separately in 300 µL of distilled water overnight. *F8*Int22[GT]n STR nested-PCR substrate analyses using ten-fold serial dilutions ranging from 10^{-1} to 10^{-3} for each allele-specific iPCR product showed that 2 µL of the 10^{-1} dilution of each 300 µL preparation of Wt for Inv22 phasing, and Wt and Var c.6118 T > C alleles were required for optimal STR amplification.

F8 IVS21 and IVS22 STR analysis

The informative *F8*Int21[CA]n STR *locus* [12] maps to the *F8* intron 21 within the *Bcl*I restriction fragment that also contains the *int22h*-1 copy. The *F8*Int21[CA]n STR *locus* characteristics in the Argentine population are shown in Table 2. The *F8*Int22[GT]n STR *locus* [17] maps to the *F8* intron 22 within the *Pac*I restriction fragment also containing the *int22h*-1 copy.

F8 IVS21 STR PCR amplification was performed as was described [12] with primers *F8*Int21F + *F8*Int21R and either directly on 10 ng of gDNA in informative Inv22 carriers or hemizygous Inv22 patients and normal controls, or by nested-PCR on an optimal dilution from the Inv22/Wt-allele-specific iPCR as indicated above.

Table 2 Population parameters for the *F8*Int21[CA]n STR locus in unrelated women from Argentina

Apparent allele size (bp) [CA]n[a]	Inv22 carriers Allele frequency (*n*)	Inv22 carriers and non-carriers Allele frequency (*n*)
161 $[CA]_{12}$	0.02 (2)	0.02 (2)
163 $[CA]_{13}$	0.34 (33)	0.33 (42)
165 $[CA]_{14}$	0.03 (3)	0.02 (3)
167 $[CA]_{15}$	0.61 (60)	0.63 (81)
Total	1.00 (98)	1.00 (128)
H_o^b	0.61	0.52
H_E^c	0.511	0.4934

[a]The apparent allele size (bp) varies in different genetic analyzers (ABI 310 versus ABI 3130xl) or 5′-modification with different fluorochromes. [CA] repeat units (*n*) were enumerated by Sanger sequencing (Supplemental Fig. 2). [b]H_o: observed heterozygosity [c]H_E: expected heterozygosity at Hardy–Weinberg equilibrium. Non-significant differences were observed between H_o and H_E rates (Pearson's chi-square test (6 DF) $p > 0.1$, Supplemental Table 2)

F8 IVS22 STR PCR amplification was performed as was described [17] with primers *F8*Int22F + *F8*Int22R (Table 1) using 100 ng of gDNA from informative Inv22 or c.6118T > C [+/−] carriers or hemizygous Inv22 or c.6118T > C [+] patients, or by nested PCR on an optimal dilution from each allele-specific iPCR product as indicated above.

STR alleles were separated by capillary electrophoresis on an ABI 3130xl Genetic Analyzer (Applied Biosystems), using GeneScan 500 LIZ size standard. Data were analyzed using GeneScan SoftGenetics GeneMarker AFLP/Genotyping software version 2.2.0 (SoftGenetics-LLC).

Assessment of the detection limit of the STR analysis

The minimal gDNA input required for the typing STR *locus* was assessed by testing, in triplicate, four ten-fold serial dilutions of *Eco*RI-digested gDNA to reduce sample viscosity and to ensure proper DNA sample dilution and homogeneity, from a control male (#1006) and a control female (#997) (range 2–0.02 ng) representing the gDNA content in 333, 33, 3, and 0.3 cells (considering that 1 ng of gDNA represents 167 diploid cells) as substrate of *F8*Int21[CA]n STR-based PCR assay.

Results

Assay design: application to long-distance haplotype phasing between the F8-Inv22 and the F8Int21[CA]n STR locus

The juxtaposition of the cohesive ends of each *Bcl*I-restricted fragment achieved by self-circularization of the template DNA used in the inverse-PCR assay inspired the Physical Phasing by Inverse-PCR (PI) approach. Its first application permits the determination of the gametic phase of the F8 wild type or the Inv22 allele with alleles of the F8Int21[CA]n STR *locus* within the F8 intron 21 (Fig. 1). The iPCR-based method for Inv22 genotyping [16] takes advantage of amplifying short and target-specific DNA segments encompassing the *Bcl*I RFLP conventionally used in the Southern blot diagnostic analysis of HA [10]. We have modified the Inv22 iPCR system by changing the IU primer target site to the 22B site, thus covering the F8Int21[CA]n STR *locus*. As a result, the size of each allele-specific amplification product (i.e., Inv22 and Wt) increased 2.5 kb.

Figure 1 depicts the PI strategy for haplotyping the *F8*Int21[CA]n STR alleles with the Inv22 allele in informative severe-HA carriers. Briefly, the PI assay entails: the screening for and selection of informative subjects (i.e., Inv22 female carriers heterozygous for the *F8*Int21[CA]n STR *locus*) (Fig. 1a) by genotyping the relevant substrate *Bcl*I-rings (the 21.6 kb circle containing *int22h*-1 representing the Wt allele and the 20 kb circle with a chimerical *int22h*-1/-3(-2) representing the Inv22 allele) (Fig. 1b). Wt and Inv22 associated haplotypes are discriminated using allele-specific iPCR assays (Wt allele size: 2529 bp and Inv22 allele size: 2601 bp) (Fig. 1c). The gametic phase of the *F8*Int21[CA]n STR is then determined on each Wt or Inv22 allele-specific amplimer by nested-PCR analysis (Fig. 1d). In the representative example shown in Fig. 1, the *F8*Int21*163 bp allele is phased with the Inv22 allele while *F8*Int21*167 bp allele, with the Wt allele.

Validation and performance of Inv22-F8Int21[CA]n PI assay

We validated the *F8*-Inv22-*F8*Int21[CA]n phasing assay in 15 proband-mother informative duos (Supplemental Figure 1; Table 3). The *F8*Int21[CA]n alleles observed by gene tracking analysis in the hemizygous probands with the Inv22 allele perfectly matched the alleles reported by the PI assay in each corresponding carrier mother in a blind study (under the null hypothesis, the STR diagnosis at random yields $p = 3 \times 10^{-5}$) (Table 3).

We note that in PGD [18], the *F8*-Inv22 cannot be diagnosed in single or few cells given its structural complexity neither by Southern blot analysis [10], long-range PCR [19], nor by iPCR [16]. We thus included four *F8*Int21[CA]n STR-informative Inv22 carriers without affected relatives (i.e., obligate carriers or sisters of deceased probands) requesting genetic counseling to perform PGD and assessed them using the PI approach (Table 3). Each PI analysis was applied in two independent assays obtaining coincident results (e.g., family #2, family #17; Table 3) to ensure reproducibility and accuracy.

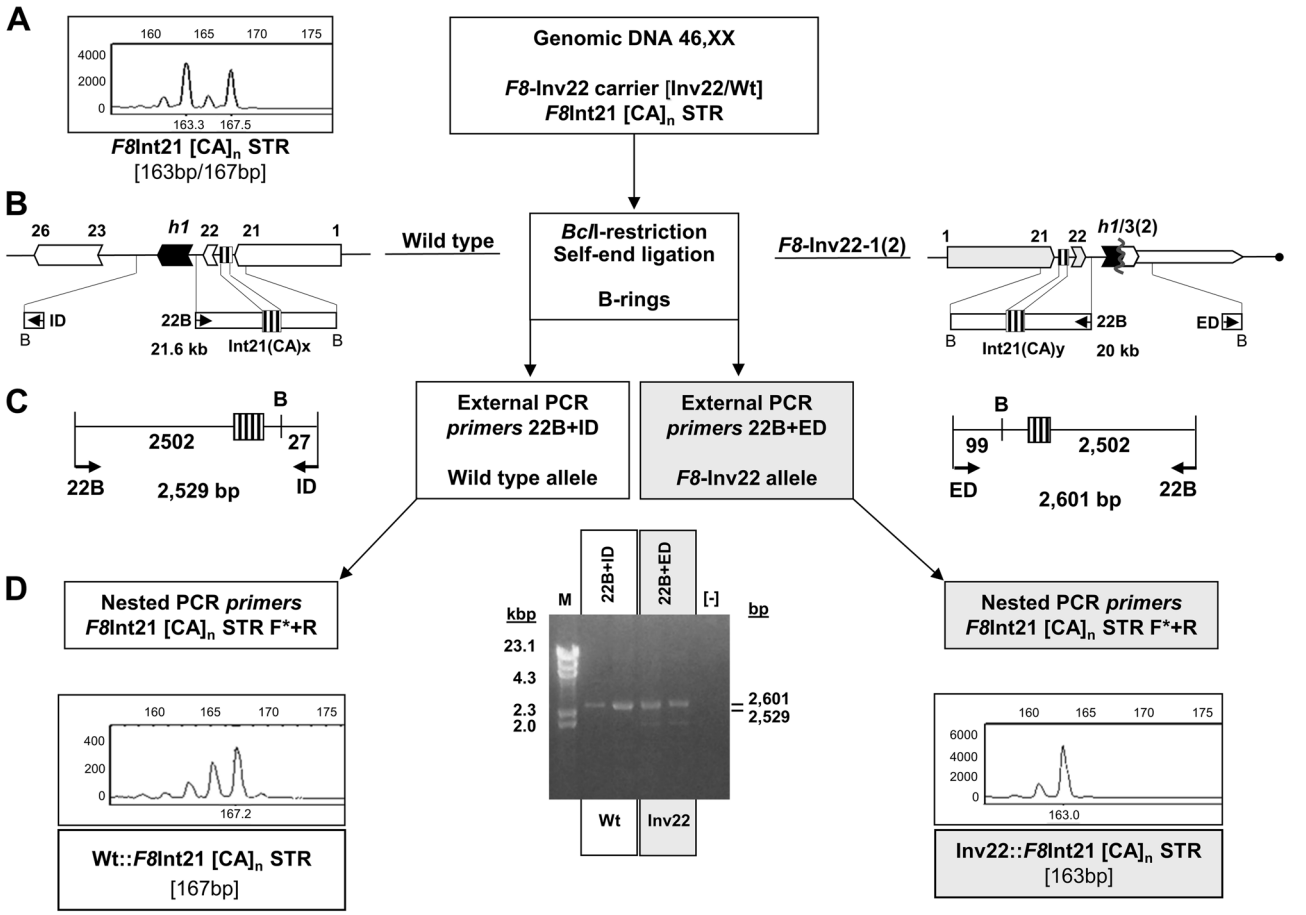


Fig. 1 PI assay for haplotyping the Inv22 and the *F8*Int21[CA]n STR. **a** Substrate: STR-informative Inv22-carrier DNA sample. **b–d** Schematic workflow of the approach showing the alelle-specific inverse-PCR primers, *Bcl*I-sites (B), *F8*Int21[CA]n STR (striped-box), and the wild type (white) and Inv22 (gray) *F8*-configuration. Wt- and Inv22-specific nested-PCR assays for *F8*Int21[CA]n are performed separately using the 22B + ID 2519bp- and 22B + ED 2601bp-inverse-PCR products, respectively. Result: Inv22::*F8*Int21[163 bp]

F8 intron 21 [CA]n STR characteristics

The usefulness of this particular application of PI relates to the capacity of a *F8*Int21[CA]n allele to act as a surrogate marker of the Inv22 in cases in which the direct interrogation of the latter is not possible in practice due to the quality/quantity of the obtained gDNA sample (e.g., PGD and NIPD, non-invasive prenatal diagnosis). Therefore, the probability of *F8*Int21[CA]n to be informative in our population of Inv22 carriers is an essential factor to measure the actual value of this specific PI application.

Because the Inv22 frequently originates de novo in each informative family, we observed the same distribution rates of heterozygous for the *F8*Int21[CA]n alleles within the Inv22 carriers versus non-carriers (Supplemental Table 2). Thirty out of 49 Inv22 carriers (61%) were found to be heterozygous at the *F8*Int21[CA]n STR *locus* (Table 2).

Assessing the minimal DNA input for F8Int21[CA]n STR genotyping

To confirm previous results about the usefulness of STR analysis in PGD, we performed an estimation of the lower limit of DNA input associated with the minimal number of cells necessary to determine the *F8*Int21[CA]n STR genotype in individuals from families affected with the Inv22. Serial dilutions of control male and female gDNA samples approximately representing 333, 33, 3, and 0.3 cell equivalents were *F8*Int21[CA]n STR genotyped (Supplemental Fig. 2). The control female gDNA sample showed clear positive allele signals associated up to an equivalent of three cells, while the same cell equivalent was barely visible in the control male sample (Supplemental Fig. 2). Although our laboratory are not focused on PGD studies, the results above confirmed that *F8*Int21[CA]n STR has potential to be

Table 3 PI assay validation and performance in phasing the Inv22 allele to *F8* Int21[CA]n STR multiallelic-informative Inv22 carriers

Family ID	Direct detection index mother STR genotype	PI analysis WT allele	PI analysis Inv22 allele	Direct detection proband STR hemizygous genotype
Validation				
#1-844	161/163	161	163	163
#2-209	163/165	163	165	165
#3-538	163/165	163	165	165
#4-636	163/167	163	167	167
#5-1081	163/167	167	163	163
#6-1053	163/167	163	167	167
#7-1049	163/167	163	167	167
#8-862	163/167	167	163	163
#9-655	163/167	167	163	163
#10-625	163/167	163	167	167
#11-631	163/167	167	163	163
#12-263	163/167	163	167	167
#13-67	163/167	167	163	163
#14-41	163/167	167	163	163
#15-3	163/167	167	163	163
Performance				
#16-1128	161/163	163	161	NA
#17-997	163/167	163	167	NA
#18-845	163/167	163	167	NA
#19-716	163/167	167	163	NA

investigated in few cells by fluorescent-labeled capillary electrophoresis as a surrogate marker of the Inv22 allele suitable for PGD in HA.

Phasing the Inv22 with the F8Int22[GT]n STR: an additional surrogate marker for PGD

A second application of the PI approach was developed to get an additional surrogate STR-marker for the Inv22, *F8*Int22[GT]n STR. This development was aimed to augment the potential heterozygosity in carriers improving the usefulness of the PI approach for indirect PGD of the Inv22.

In this case, the PI approach targeted the circularized *Pac*I-restricted fragment encompassing the *F8*Int22[GT]n STR linked to the non-rearranged structure of *F8* intron 22 (Wt allele) (Supplemental Figure 3). As a result, the Wt allele-specific iPCR amplification product spans 3.1 kb which can be obtained from the 28.4 kb *Pac*I-ring. Supplemental Figure 3 depicts the PI strategy for haplotyping the *F8*Int22[GT]n STR allele with the Wt (NoInv22) allele in informative carriers.

Proving the flexibility of the PI approach: phasing the F8:c.6118T > C with the F8Int22[GT]n STR

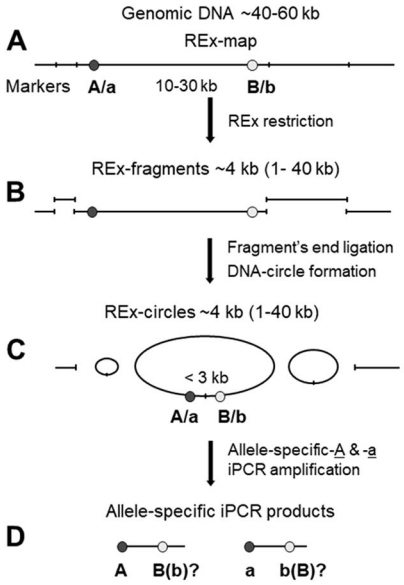
To demonstrate the ability of the PI approach to involve an allele-specific iPCR amplification based on a single-nucleotide difference, we developed an assay linking an exonic SNV associated with moderate HA, *F8*:c.6118 T > C to the *F8*Int22[GT]n STR alleles, using the same *F8*-specific *Pac*I-ring. In this case, the PI assay was applied to resolve the phase of an informative carrier, heterozygous for the *F8*:c.6118T > C variant and the *F8*Int22[CA]n STR *locus* by genotyping the relevant *Pac*I-ring separately by allele-specific iPCR amplification (3035 bp-iPCR product obtained from the 28.4 kb-*Pac*I-ring containing the thymine (T) representing the Wt and those with the cytosine (C) corresponding to the variant allele). Then, the gametic phase of the *F8*Int22[CA]n STR alleles was determined on each Wt or variant allele-specific amplimer by nested-PCR (Supplemental Figure 4).

Discussion

We developed a novel approach for haplotype phasing over physical distances up to 30 kb that allowed determination of the gametic phase between DNA markers of different kind directly from the analytical sample (Fig. 2). We provided proof-of-concept of the PI approach by identifying the gametic phase between the Inv22 and an intragenic *F8*Int21 STR *locus* in female carriers bypassing the need of linkage analysis with family relatives. Also, we showed another potentially useful PI application by phasing the Inv22 with the *F8*Int22 STR *locus* taking advantage of the *Pac*I restriction map across the exon 19–exon 23 interval in the Wt *F8* allele.

The potential of these PI applications in discriminating the X-chromosome carrying the Inv22 allele by surrogate markers (i.e., *F8*Int21 and *F8*Int22 STRs), which unlike the Inv22 can be determined even from low-quality DNA samples, represents a significant improvement in genetic counseling in HA, especially for prenatal diagnosis, including PGD. Worth noting, the Inv22 PI application was designed and developed upon request to find *F8* markers *in phase* with the Inv22, suitable for direct genotyping in cases of scarcity and low-quality DNA available for PGD, involving an Inv22 carrier, daughter of a deceased individual with severe HA and without available relatives to achieve linkage analysis.

The estimated combined heterozygosity rates of *F8*Int21 and *F8*Int22 STRs *loci* range from 72 to 91% in different populations [17, 20]. Thus, typing these STRs will resolve the diagnosis of a considerable proportion of carriers in a next-of-kin-independent manner. We also showed that the



E

Marker combination AS::LT	Allele-specific marker	Linked-tracked marker
RFLP::STR	*Bcl*I-iPCR Inv22-RFLP	*F8*Int21(CA)n STR
	*Pac*I-iPCR Inv22-RFLP	*F8*Int22(GT)n STR
SNP::STR	*Pac*I-iPCR c.6118T>C-SNP	*F8*Int22(GT)n STR
RFLP::SNP-RFLP	-	-
SNP::SNP-RFLP	-	-

Fig. 2 Rationale of the PI assay for haplotyping long-distanced marker pairs. **a**–**d** Restriction map showing A–B markers. Enzyme X-restriction of a double-heterozygous DNA sample. Fragment-circularization. Separated characterization of the linked allele of the secondary (tracked) marker B by A-allele-specific inverse-PCR. **e** Table: potentiality and flexibility of the PI approach indicating the characteristics of all possible marker pairs and those exemplified in this study

*F8*Int21[CA]n STR could be genotyped from samples representing few genome equivalents. Since 2005, blastocyst biopsy from a few cells has been widely used for PGD in the clinical practice [21]. The outer trophectoderm biopsy obtained at 5–6 days yields 5–10 cells, which represent more material available for genetic diagnosis and lesser invasively than blastomere biopsy [22].

We note that further modifications in the design of PI approach here depicted will allow the gametic phasing of other markers on the 5′ and 3′ ends on restriction fragments different from the *Bcl*I-RFLP/[CA]n-STR pair validated in this work. For example, allele-specific iPCR amplification based on a RFLP or SNP from a restriction fragment end region in a targeted restriction fragment-ring may form target pairs with other typing SNPs, RFLPs, STRs, or structural variations (Fig. 2).

The PI assay may be adapted to different genomic regions based on the variations of iPCR protocols that include blunt and staggered end-cutting enzymes such *Bgl*II, *Taq*I, *Xba*I [23], *Bam*HI, *Hin*dIII, *Xho*I, and *Pac*I yielding up to 3.3 kb-iPCR products analyzed from circles of up to 30 kb-length on gDNA input of standard quality.

The PI approach based on SNV allele-specific selection (Fig. 2) may be useful to study the effects of *cis*-association of markers in numerous instances: for assessing human leukocyte antigen (HLA) haplotypes associated with more than 100 diseases including drug hypersensitivity induced by specific drugs [24], or *CYP450* haplotypes that define the activity involved in drug metabolism [25–27], for detecting recombination hot spots [28] and for proving that *cis*-regulatory elements influence promoter activity 2–50 kb apart [29, 30].

Direct (physical) methods of phasing have the unmatched potential to unveil haplotypes using only the sample of interest, without the requirement of genotyping multiple individuals from the same family and the impossibility to phase de novo variations (and eventually new variants associated with disease) in the last generation. Classical methods for direct haplotyping are molecular cloning and allele-specific PCR [31]. Relatively new approaches for direct haplotyping include emulsion haplotype-fusion PCR [32], polony haplotyping [33], drop-phase [34], or whole-genome haplotyping by dilution, amplification, and sequencing [35]. Although some of these approaches can produce accurate long-range haplotypes, their experimental complexity and the lack of expertise for massive data analysis in most genetic testing laboratories still make their implementation difficult in the clinical practice [36]. The PI approach described here for phasing relatively long-distanced DNA markers represents a robust, reliable and cost-effective alternative to the currently available panel of methods for direct haplotyping. The PI approach has proven flexibility to target different marker pair's types and consequently may be useful to associate SNPs, in addition to STRs, as surrogate markers for duplicon-mediated rearrangements. This feature of the PI approach opens the possibility to be useful not only with PGD, which permits typing of both STRs and SNPs, but also with NIPD, which allows detection of only SNPs.

Finally, the PI approach has potential for enhancing its capacity by multiplex characterization of iPCR products by coupling it with massively parallel sequencing technologies.

Beyond the potentiality of the practical approach to be useful in other diagnostic scenarios in human genetics, the

presented PI applications improve hemophilia genetic counseling making possible preimplantation genetic diagnosis in carriers of the prevalent Inv22 bypassing the requirement for family linkage analysis.

Acknowledgements This study was supported by grants from the National Research Council (CONICET), National Agency for Scientific and Technological Promotion (ANPCyT) and the World Federation of Hemophilia. The authors thank Miguel de Tezanos Pinto (National Academy of Medicine and National Hemophilia Foundation) and Judith Mincman and Roberto Cocco (*Fecunditas* Clinic) for their help in different phases of the work.

Authors contributions All authors have contributed to the intellectual content of this paper and have met the following three requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethics statement The Ethics Committee of the Institutes of the National Academy of Medicine of Buenos Aires, Argentina, approved the project (date: 13/06/2012). Peripheral blood samples from participating control, carrier, and hemophilia subjects in nuclear families were collected with written informed consent.

References

1. Morris RW, Kaplan NL. On the advantage of haplotype analysis in the presence of multiple disease susceptibility alleles. Genet Epidemiol. 2002;23:221–33.
2. Fernando MM, Stevens CR, Walsh EC, De Jager PL, Goyette P, Plenge RM, et al. Defining the role of the MHC in autoimmunity: a review and pooled análisis. PLoS Genet. 2008;4:e1000024.
3. Do C, Shearer A, Suzuki M, Terry MB, Gelernter J, Greally JM, et al. Genetic-epigenetic interactions in cis: a major focus in the post-GWAS era. Genome Biol. 2017;18:120.
4. International HapMap Consortium. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007;449:851–61.
5. The International HapMap 3 Consortium. Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, et al. Integrating common and rare genetic variation in diverse human populations. Nature. 2010;467:52–8.
6. Shastry BS. Pharmacogenomics and pharmacoepigenomics in pediatric medicine. Methods Mol Biol. 2014;1175: 671–85.
7. Apellániz-Ruiz M, Inglada-Pérez L, Naranjo ME, Sánchez L, Mancikova V, Currás-Freixes M, et al. High frequency and founder effect of the CYP3A4*20 loss-of-function allele in the Spanish population classifies CYP3A4 as a polymorphic enzyme. Pharm J. 2015;15:288–92.
8. Zheng GX, Lau BT, Schnall-Levin M, Jarosz M, Bell JM, Hindson CM, et al. Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. Nat Biotechnol. 2016;34:303–11.
9. Hui WW, Jiang P, Tong YK, Lee WS, Cheng YK, New MI, et al. Universal haplotype-based noninvasive prenatal testing for single gene diseases. Clin Chem. 2017;63:513–24.
10. Lakich D, Kazazian HH, Antonarakis SE, Gitschier J. Inversions disrupting the factor VIII gene are common cause of severe Haemophilia A. Nat Genet. 1993;5:236–41.
11. Naylor J, Brinke A, Hassock S, Green PM, Giannelli F. Characteristic mRNA abnormality found in half the patients with severe haemophilia A is due to large inversions. Hum Mol Genet. 1993;2:1773–8.
12. Machado FB, Alves Da Silva AF, Rossetti LC, De Brasi CD, Medina-Acosta E. Informativeness of a novel multiallelic marker-set comprising an *F8* intron 21 and three tightly linked loci for haemophilia A carriership analysis. Haemophilia. 2011;17:257–66.
13. 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;6:830–6.
14. Sambrook J, Fritsch EF, Maniatis T. Molecular cloning. A laboratory manual. 2nd ed. New York, USA: Cold Spring Harbor Laboratory Press; 1989.
15. Lahiri DK, Nuremberg JI. A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. Nucl Acids Res. 1991;19:5444.
16. Rossetti LC, Radic P, Larripa IB, De Brasi CD. Genotyping the hemophilia inversion hotspot by use of inverse PCR. Clin Chem. 2005;51:1154–8.
17. Liang Y, Zhao Y, Yan M, Fan XP, Xiao B, Liu JZ. Prenatal diagnosis of haemophilia A in China. Prenat Diagn. 2009;29: 664–7.
18. Zhao M, Chen M, Tan ASC, Cheah FSH, Mathew J, Wong PC, et al. Single-tube tetradecaplex panel of highly polymorphic microsatellite markers <1 Mb from F8 for simplified preimplantation genetic diagnosis of hemophilia A. J Thromb Haemost. 2017;15:1473–83.
19. Liu Q, Nozari G, Sommer SS. Single-tube polymerase chain reaction for rapid diagnosis of the inversion hotspot of mutation in haemophilia A. Blood. 1998;92:1458–9.
20. Machado FB, Medina-Acosta E. High-resolution combined linkage physical map of short tandem repeat loci on human chromosome band Xq28 for indirect haemophilia A carrier detection. Haemophilia. 2009;15:297–308.
21. De Rycke M, Belva F, Goossens V, Moutou C, SenGupta SB, Traeger-Synodinos J, et al. ESHRE PGD Consortium data collection XIII: cycles from January to December 2010 with pregnancy follow-up to October 2011. Hum Reprod. 2015;30:1763–89.
22. Traeger-Synodinos J. Pre-implantation genetic diagnosis. Best Pract Res Clin Obstet Gynaecol. 2017;39:74–88.
23. Citterio CE, Rossetti LC, Souchon PF, Morales C, Thouvard-Viprey M, Salmon-Musial AS, et al. Novel mutational mechanism in the thyroglobulin gene: imperfect DNA inversion as a cause for hereditary hypothyroidism. Mol Cell Endocrinol. 2013;381:220–9.
24. Hosomichi K, Jinam TA, Mitsunaga S, Nakaoka H, Inoue I. Phase-defined complete sequencing of the HLA genes by next-generation sequencing. BMC Genomics. 2013;14:355.
25. Scott SA, Tan Q, Baber U, Yang Y, Martis S, Bander J, et al. An allele-specific PCR system for rapid detection and discrimination

of the CYP2C19*4A, *4B, and *17 alleles: implications for clopidogrel response testing. J Mol Diagn. 2013;15:783–9.
26. Sim SC, Ingelman-Sundberg M. The Human Cytochrome P450 (CYP) Allele Nomenclature website: a peer-reviewed database of CYP variants and their associated effects. Hum Genomics. 2010;4:278–81.
27. Zanger UM, Klein K. Pharmacogenetics of cytochrome P450 2B6 (CYP2B6): advances on polymorphisms, mechanisms, and clinical relevance. Front Genet. 2013;4:24.
28. Jeffreys AJ, Kauppi L, Neumann R. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. Nat Genet. 2001;29:217–22.
29. Kolovos P, Knoch TA, Grosveld FG, Cook PR, Papantonis A. Enhancers and silencers: an integrated and simple model for their function. Epigenetics Chromatin. 2012;5:1.
30. Mora A, Sandve GK, Gabrielsen OS, Eskeland R. In the loop: promoter-enhancer interactions and bioinformatics. Brief Bioinformatics. 2016;17:980–95.
31. Jeffreys AJ, Neumann R. Factors influencing recombination frequency and distribution in a human meiotic crossover hotspot. Hum Mol Genet. 2005;14:2277–87.
32. Turner DJ, Tyler-Smith C, Hurles ME. Long-range, high-throughput haplotype determination via haplotype-fusion PCR and ligation haplotyping. Nucleic Acids Res. 2008;36:e82.
33. Zhang K, Zhu J, Shendure J, Porreca GJ, Aach JD, Mitra RD, et al. Long-range polony haplotyping of individual human chromosome molecules. Nat Genet. 2006;38:382–7.
34. Regan JF, Kamitaki N, Legler T, Cooper S, Klitgord N, Karlin-Neumann G, et al. A rapid molecular approach for chromosomal phasing. PLoS ONE 2015;10:e0118270.
35. Kaper F, Swamy S, Klotzle B, Munchel S, Cottrell J, Bibikova M, et al. Whole-genome haplotyping by dilution, amplification, and sequencing. Proc Natl Acad Sci USA. 2013;110:5552–7.
36. Snyder MW, Adey A, Kitzman JO, Shendure J. Haplotype-resolved genome sequencing: experimental methods and applications. Nat Rev Genet. 2015;16:344–58.

Supplemental Table 1: Information of the short tandem repeats *F8*Int21 and *F8*Int22.

Marker Name	Sequence Repeat Unit	Number of Repeat Unit	[a]STR nomenclature	Reference / Database
*F8*Int21	AC	26	g.154896506AC[26]	[12] / [b]HEMApSTR
HSC271 (M4) *F8*Int22	GT	19	g.154875736GT[19]	[17] / [b]HEMApSTR

[a]NCBI nucleotide sequence file NC_000023.11 genomic build GRCh38.p12 accessed 26/03/2018.
[b]URL: http://www.uenf.br/Uenf/Pages/CBB/LBT/HEMApSTR.html.

Supplemental Table 2: Genotype frequency distribution for the *F8*Int21[CA]n STR *locus* in unrelated women from Argentina.

Genotype	Inv22+/- observed frequency [a]n= 49	Inv22+/- HWE-Expected frequency n= 49	Inv22+NC females observed frequency [b]n= 64	Inv22+NC females HWE-Expected frequency n=64
161/161	0	0	0	0
161/163	0.041 (2)	0.014 (0.7)	0.031 (2)	0.011 (0.7)
163/163	0.041 (2)	0.114 (5.6)	0.078 (5)	0.108 (6.9)
161/165	0	0.002 (0.1)	0	0.001 (0)
163/165	0.041 (2)	0.02 (1.0)	0.031 (2)	0.015 (1)
165/165	0	0.001 (0)	0	0
161/167	0	0.024 (1.2)	0	0.020 (1.3)
163/167	0.510 (25)	0.412 (20.2)	0.438 (28)	0.416 (26.6)
165/167	0.020 (1)	0.037 (1.8)	0.016 (1)	0.030 (1.9)
167/167	0.347 (17)	0.376 (18.4)	0.406 (26)	0.400 (25.6)
[c]H_O	0.61		0.52	
[d]H_E	0.511		0.4934	
[e]PIC	0.42370174		0.40614558	
[f]χ^2	9.802; 6df		6.470; 6df	
[g]P value	0.133237		0.372672	

a Inv22+/-: women carriers of the *F8* intron 22 inversion

b Inv22+/-: women carriers of the *F8* intron 22 inversion + NC: normal control women (n = 15)

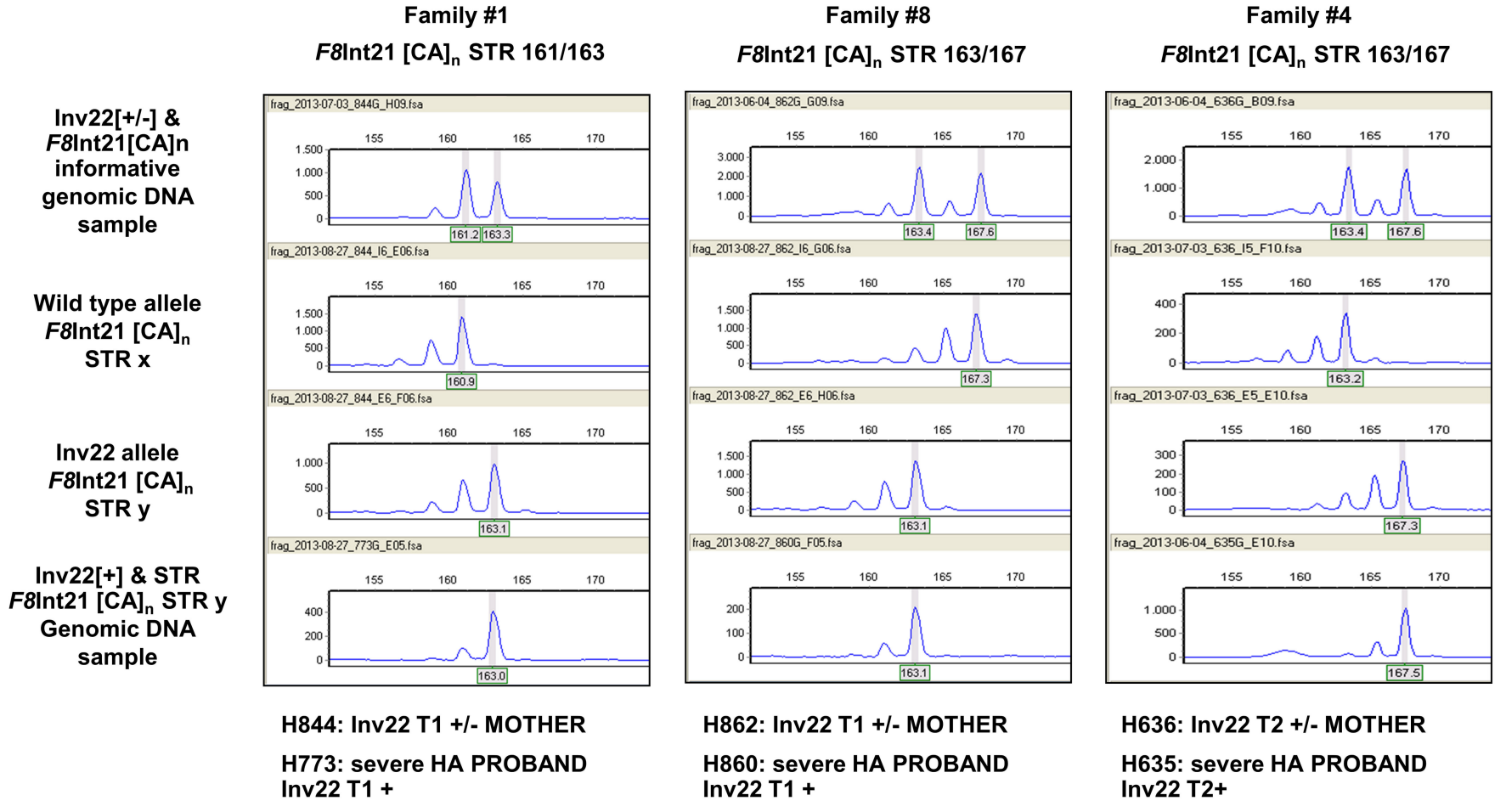
c Observed heterozigosity

d Expected heterozigosity

e PIC: Polymorphic information content

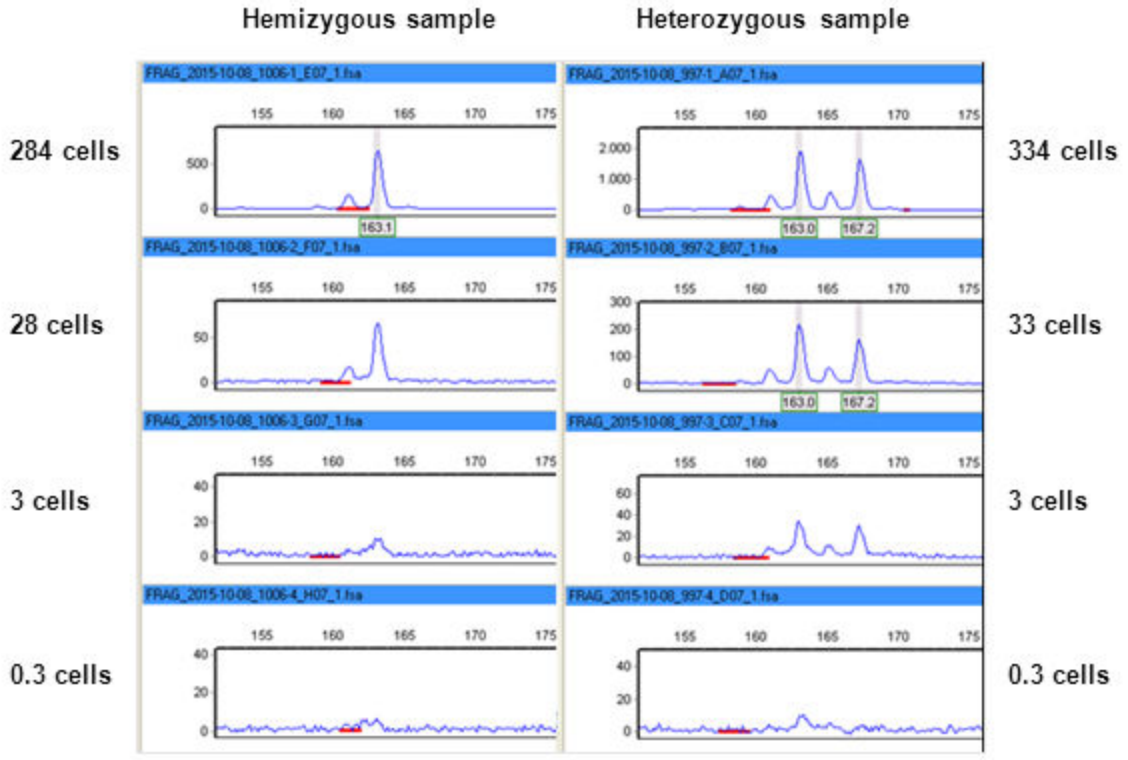
f Pearson´s chi-square test (6 degrees of freedom -df)
Hardy-Weinberg Equilibrium estimations for codominant locus systems were using the Statistical Genetics Utility programs available at http://www.jurgott.org/linkage/util.htm (Access data: January, 2017)

g *p* value: **p* < 0.05 significant; ***p* < 0.01 highly significant
d and **e** were calculated using the online PICcalc program available at http://w3.georgikon.hu/pic/english/default.aspx.



Supplemental Figure 1. PI validation assay in three families (#1, #4 and #8). The figure depicts an example of a blind experiment comparing the result of the *F8*Int21[CA]n allele in phase with the Inv22 allele obtained by the PI in three of the 15 informative duos (carrier mother-child affected) analyzed. The Wild type allele (primers 22B+ID) and the Inv22 allele (primers 22B+ED) were separately amplified by allele-specific iPCR, both products includes each STR *F8*Int21[CA]n allele in phase (haplotype), followed by nested-PCR and capillary electrophoresis for genotyping *F8*Int21[CA]n linked allele. For example, in the Family#4, the mother has the *F8*Int21[CA]n 167 pb allele in phase with the Inv22 allele and the *F8*Int21[CA]n 163 pb allele in phase with the Wild type allele. The affected Inv22 child bears the *F8*Int21[CA]n 167 pb allele segregating in block with the Inv22 from his carrier mother.

A



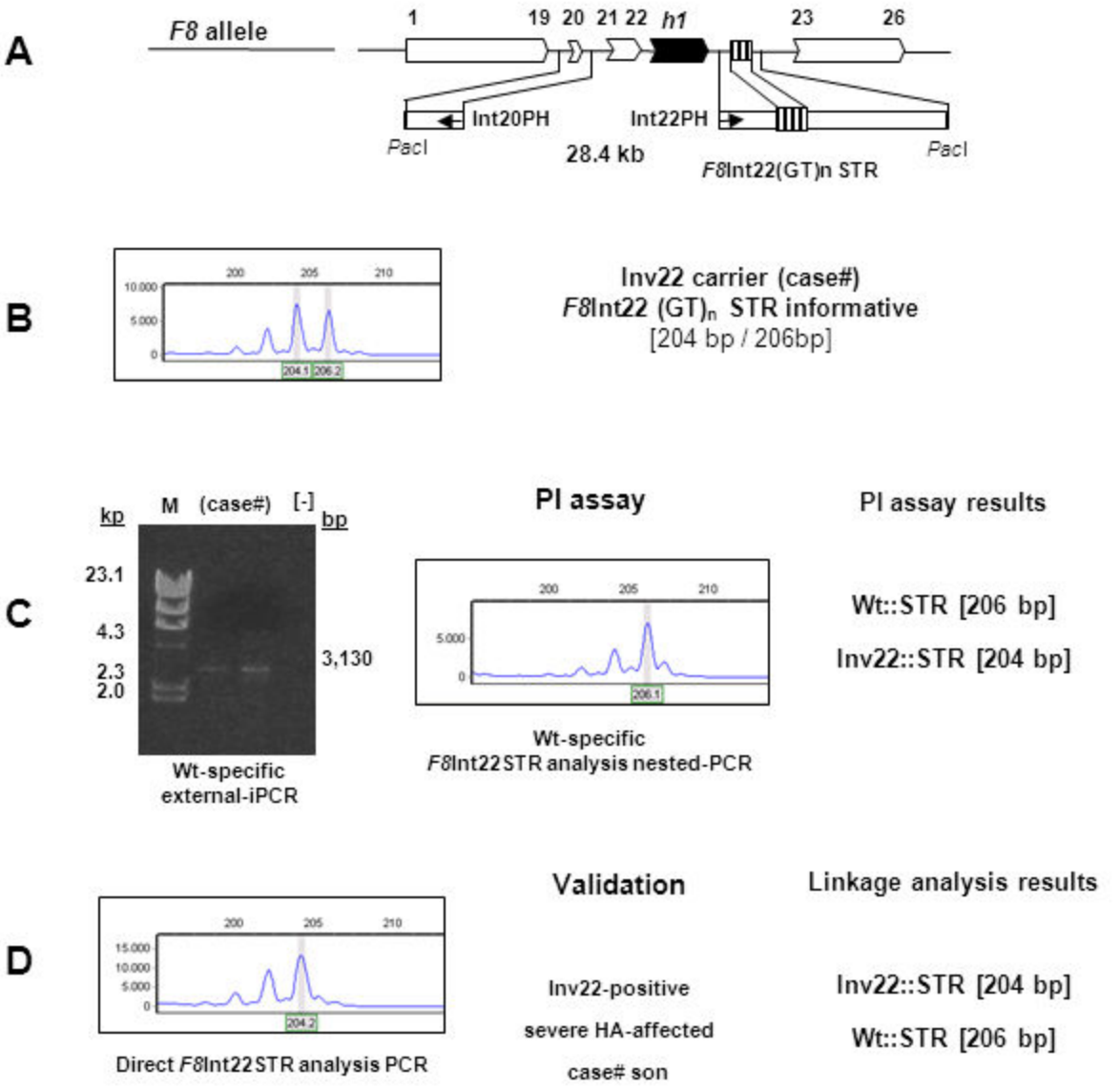
B

Sample ID & dilution	DNA (ng/µL)	Estimated number of diploid cells [a]	Amplification readout
#997 E ♀ [a]	19.36	3340	ND
1/10	2	334	+
1/100	0.2	33	+
1/1000	0.02	3	+/-
1/10000	0.002	<1	-
#1006 E ♂ [b]	17.35	2840	ND
1/10	1.7	284	+
1/100	0.17	28	+/-
1/1000	0.017	3	+/--
1/10000	0.0017	<1	-

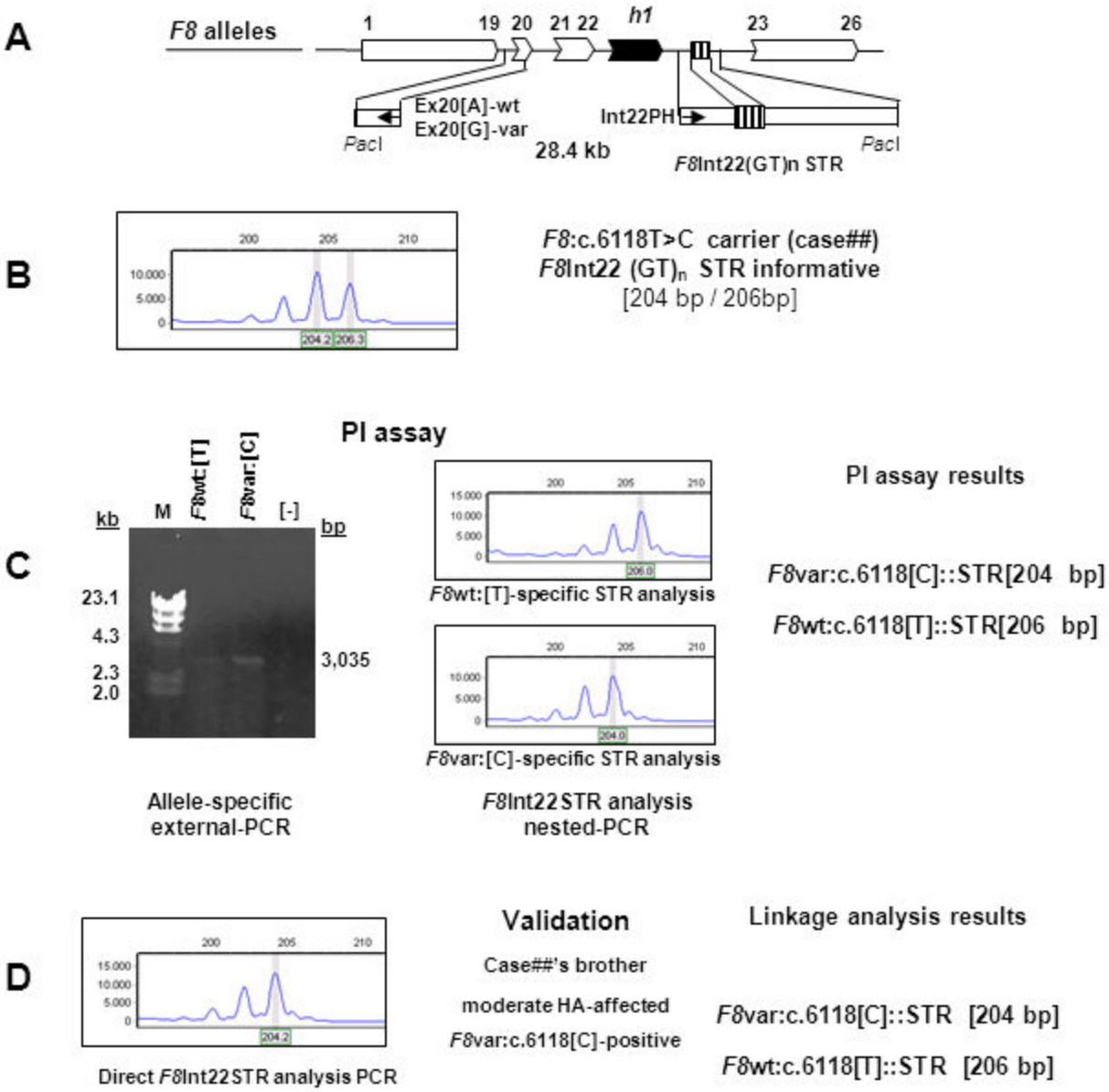
[a] The estimation of the number of cells represented in 1ng = 167 diploid cells is based on the approximation that a single human diploid cell contains about 6 pg of DNA ($6x10^9$ pb).

[b] ♀ female sample DNA (two X-chromosomes); ♂ male sample DNA (one X-chromosome). Both samples predigested with *EcoR*I (**E**) before PCR amplification.

Supplemental Figure 2. PI detection limit assay. (**A**) Allele profiles for the *F8*Int21[CA]n STR *locus* using serial dilutions of DNA from either a male (hemizygous) or an informative (heterozygous) female donors. (**B**) Observed effect of DNA template quantity and cell number equivalent on the performance of the assay. ND: not done.



Supplemental Figure 3. PI application assay for haplotyping the Inv22 and the *F8Int22[GT]n* STR *locus*. This PI application is based on *Pac*I restriction and circularization, normal (NoInv22) specific iPCR amplification followed by nested PCR analysis of the linked *F8*Int22[GT]n allele by capillary fluorescent electrophoresis (Materials and Methods). The Inv22 allele was not analyzed due to the DNA quality precluding the circle formation spanning ≈70 kb. **A**. Schematic map showing the relevant details to understand this PI application to determine the phase Inv22-*F8*Int22[GT]n STR *locus*. Relevant iPCR primer target sites, *F8* exons, *int22h*-1 (*h1*) and *Pac*I restriction sites are shown. **B.** Analyzed case#: Inv22-type1 carrier mother of a hemophilia A-affected son, showing the heterozygosity for *F8*Int22 STR [204 bp / 206 bp]. **C.** Sequential molecular steps of the PI assay including the wildtype (Wt) specific external-iPCR on the left (3,130 bp-product), the Wt-specific analysis of the *F8*Int22 STR showing the identification of the allele in phase (206 bp) and the final integrated results of the PI assay on the right. **D**. Validation of the PI approach results by linkage analysis using the DNA sample of the affected son of case# with severe HA and the Inv22 determined in phase with the 204 bp-*F8*Int22 allele.



Supplemental Figure 4. PI application assay for haplotyping an exon 20 single nucleotide variant *F8:c.6118T>C* and the *F8*Int22[GT]n STR *locus*. This PI application is based on *Pac*I restriction and circularization, separated *F8*wt:[T] and *F8*var:[C] allele-specific iPCR amplification followed by nested PCR analysis of each linked *F8*Int22[GT]n allele by capillary fluorescent electrophoresis (Materials and Methods). **A**. Schematic map showing the relevant details to understand this PI application to determine the phase between the SNV c.6118T>C and *F8*Int22[GT]n STR alleles. Relevant allele-specific iPCR primer target sites, *F8* exons, *int22h*-1 (*h1*) and *Pac*I restriction sites are shown. **B.** Analyzed case## characteristics: c.6118T>C carrier sister of a moderate hemophilia A-affected proband, showing heterozygosity for *F8*Int22 STR [204 bp / 206 bp]. **C.** Sequential molecular steps of the PI assay including *F8*wt:[T] and *F8*var:[C] allele-specific iPCR amplification (3,035 bp-product) followed by nested PCR analysis of each linked *F8*Int22[GT]n STR allele and the resulting phase of case## on the right panel *F8*var::STR[204 bp] and *F8*wt::STR[206 bp]. **D**. Validation of the results obtained by this PI application by familiar linkage analysis using the DNA sample of the affected brother of case## with variant c.6118T>C determined in phase with the 204 bp-*F8*Int22 allele.