Evaluation of genetic markers linked to hemophilia A locus: an Indian experience

Hemophilia A is an X-linked recessive bleeding disorder caused by defects in factor VIII gene (*F8*). Our study examines variations of single nucleotide polymorphism (SNP) in *F8* in the Indian population and establishes the utility of a combination of SNP and microsatellite markers for the successful identification of carriers in the affected families.

Haematologica 2007; 92:1725-1726. DOI: 10.3324/haematol.11545

The disease affects 1 in 5,000-10,000 males worldwide. Due to the large size of F8 (186 Kb), direct mutation analysis is not realistic. Therefore, linkage analysis using informative markers is the most practical approach where family members of the suspect, including an affected member, are available. Single nucleotide polymorphisms (SNPs) are the preferred markers because of their dense coverage of the genome and relatively easy assay format including commonly used restriction fragment length polymorphism (RFLP). SNPs are known to have variable allele frequencies in different populations. Ideally, informative markers should be highly heterozygous in the test population and the set of markers selected for use should not be in linkage disequilibrium (LD) for maximum efficiency of the linkage-based diagnostic test for carrier detection. We recently reported a remarkable variation in the informativeness of RFLP markers linked to hemophilia B locus in several Indian population groups.[1] In this context, we examined several published studies from across different regions of India which have shown variable levels of heterozygosity and informativeness of the F8 markers.[2-4] However, all of them had been conducted on a regional basis and on a relatively small number of samples. Until now, no exhaustive study has been carried out to verify the variability of informativeness of markers in F8 across different population groups in India. To assess the heterozygosity of the SNPs and to evaluate the linkage status among them in the Indian population, 1,486 unrelated individuals recruited as part of the Indian Genome Variation project[5] were included in the present study. These individuals belonged to 54 distinct ethnic groups living in 6 different geographic regions (north, northeast, east, south, west and central) of mainland India. They also included the four major linguistic families of the Indian population i.e. Indo-European (IE), Tibeto-Burman (TB), Austro-Asiatic (AA) and Dravidian (DR). The internal review committee on research using human subjects reviewed and approved the project according to the regula-

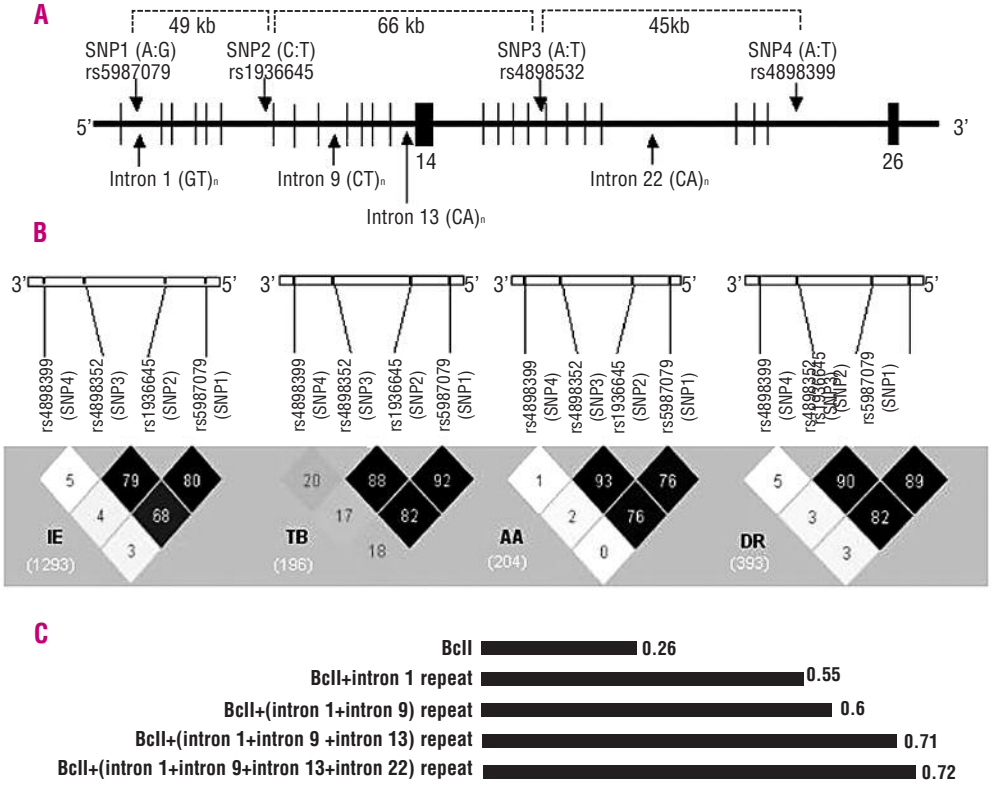

Figure 1. Evaluation of SNP and microsatellite markers at Hemophilia A locus for carrier detection. A. The position of four SNPs (with base change and their rs IDs) and four di-nucleotide repeat (microsatellite) markers at the top and bottom of F8 gene respectively. Relative positions of the markers are indicated in terms of genomic distance and locations with respect to the exons (vertical line and boxes). B. Pairwise Linkage disequilibrium (LD) calculated between the four selected SNPs in the four linguistic groups of the Indian population are shown. Location of the SNPs in F8 is depicted on the top panel. The intensity of the box color is proportional to the strength of the LD (r^2) for the marker pair, which is also indicated in percentage within the boxes. Numbers in the brackets below the population identifier (IE, Indo-European; TB, Tibeto-Burman; AA, Austro-Asiatic; DR, Dravidian) indicate the number of chromosomes analyzed. Genotyping of the SNPs was carried out by allele-specific primer extension followed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) using Sequenom *massarray* system at The Centre for Genomic Application (TCGA, New Delhi, India). Co-efficients of pairwise linkage disequilibrium (r^2) were estimated using the Haploview ver. 3.32 software.[10] C. Increase in informativeness by the combined use of microsatellite and Bcl I RFLP markers in Hemophilia A affected families. Primer sequences for all the markers are available on request.

Table 1. Allele frequency and heterozygosity of microsatellite markers in *F8* gene in normal female population (n=63).

Marker	*Repeat length*	*Allele frequency*	*Heterozygosity*
	$(GT)_{13}$	0.015	
	$(GT)_{14}$	0.015	
Intron 1 $(CA)_n$	$(GT)_{15}$	0.499	0.46
	$(GT)_{16}$	0.373	
	$(GT)_{17}$	0.103	
	$(GT)_{18}$	0.015	
	$(CT)_{22}$	0.052	
Intron 9 $(CT)_n$	$(CT)_{23}$	0.502	0.666
	$(CT)_{24}$	0.278	
	$(CT)_{25}$	0.286	
	$(CA)_{20}$	0.02	
	$(CA)_{21}$	0.048	
	$(CA)_{22}$	0.389	
	$(CA)_{23}$	0.223	
Intron 13 $(CA)_n$	$(CA)_{24}$	0.295	0.46
	$(CA)_{25}$	0.027	
	$(CA)_{26}$	0.014	
	$(CA)_{27}$	0.014	
	$(CA)_{28}$	0.014	
$(CA)_{19}$	0.556		
Intron 22	$(CA)_{18}$	0.383	0.428
$[(CA)_n(CT)_n]$	$(CA)_{20}$	0.029	
	$(CA)_{21}$	0.071	

Allele frequency at each variant site was computed by the allele-counting method. Heterozygosity calculations were made using the following formula:

$$PIC = 1 - \sum_{i=1}^{n} f_i^2$$

(where f_i is the frequency of the i^{th} allele).

tions of the Indian Council of Medical Research. Four SNPS, rs5987079 (SNP1), rs1936645 (SNP2), rs4898352 (SNP3) and rs4898399 (SNP4), were selected (Figure 1, panel A) spanning the entire F8 with almost equal spacing to reduce the possibility of them being in linkage disequilibrium (LD). Three (rs5987079, rs1936645 and rs4898352) of the four SNPs selected were observed to have high minor allele frequencies (MAF) ranging between 0.34-0.42. These could, therefore, be used as markers to distinguish between a mutant and a normal F8 chromosome in the Indian population. However, high LD values (r^2>0.5) between the three SNP markers (i.e. SNP #1-3) in all four linguistic groups suggest that any one of these three SNPs is sufficient to be used successfully as a marker for carrier diagnosis (Figure 1, panel B). The fourth SNP (#4; rs4898399), not in LD with other SNPs, showed low MAFs (0.01-0.1) in all the four linguistic groups and could not be used as a marker. Therefore, we selected the most widely used SNP (#3; rs4898352) having an RFLP assay format using Bcl I restriction enzyme.

To increase the number of testable markers in our repertoire, we selected four microsatellite markers[6-9] within F8 gene (Figure 1, panel A) and tested for informativeness in a pool of 63 normal females with no bleeding history in their family and recruited from different parts of India. The allele frequency and heterozygosity of the microsatellite markers are shown in Table 1. Finally, by using the Bcl I RFLP and the microsatellite markers we could experimentally determine carrier status in 74% (31/42) of obligate carriers for hemophilia A (Figure 1, panel C) who enrolled at the Hemophilia Federation of India (Kolkata Chapter) and at the Center for Cellular and Molecular Biology, Hyderabad, India. Our studies on the Indian population and the available data on International HAPMAP population groups suggest that microsatellite markers, and not SNPs, should preferably be used for carrier detection of hemophilia A. Among these multi-allelic markers, the intron 9 CT-repeat showing higher heterozygosity should be tested across different population groups for wide application for carrier detection.

Atreyee Saha, Saibal Mukherjee,*,+ Mahua Maulik,*,§ Giriraj Ratan Chandak,° The Indian Genome Variation Consortium,# Kunal Ray**

**Molecular & Human Genetics Division, Indian Institute of Chemical Biology, Kolkata, India; °Genome Research Group, Centre for Cellular and Molecular Biology, Hyderabad, India; #Nodal Laboratory: Institute of Genomics & Integrative Biology, New Delhi, India*

+Current address: the University of New Mexico, Albuquerque, USA; §Current address: Centre for Neuroscience, University of Alberta, Canada.

Acknowledgments: the authors thank all the donors who participated in the study and Dr. Amal Saha and Mr. Ujjal Roy of the Haemophilia Federation of India, Kolkata Chapter, Kolkata.

Funding: the study was supported from Indian Government Council of Scientific and Industrial Research (CSIR), through a grant (CMM0016). AS is a recipient of Lady Tata Memorial Senior Scholarship.

Key words: hemophilia A, F8, SNP markers, microsatellite markers.

Correspondence: Kunal Ray, Molecular & Human Genetics Division, Indian Institute of Chemical Biology, 4 Raja S. C. Mullick Road, Jadavpur, Kolkata, 700 032, India. Phone: international +91.33.24733491 (or 0492/6793). Fax: international 91.33.24735197 (or 2472/3967). E-mail: kunalray@gmail.com/thisisker@rediffmail.com

References

1. Mukherjee S, Saha A, Kumar PS, Chandak GR, Majumder PP, Ray K. Remarkable variation in the informativeness of RFLP markers linked to hemophilia B locus in Indian population groups: implication in the strategy for carrier detection. Dis Markers 2006;22:327-34.
2. Shetty S, Ghosh K, Pathare A, Colah R, Badakare S, Mohanty D. Factor VIII and IX gene polymorphisms and carrier analysis in Indian population. Am J Hematol 1997; 54:271-5.
3. Jayandharan G, Shaji RV, George B, Chandy M, Srivastava A. Informativeness of linkage analysis for genetic diagnosis of haemophilia A in India. Haemophilia 2004;10:553-9.
4. Ahmed R, Kannan M, Biswas A, Ranjan R, Choudhry VP, Saxena R. Use of Intron 1 and 22 inversions and linkage analysis in carrier detection of hemophilia A in Indians. Clin Chim Acta 2006;365:109-12.
5. The Indian Genome Variation Consortium. The Indian Genome Variation database (IGVdb): A project overview, Hum Gen 2005;118:1-11.
6. Tizzano E, Vencesla A, Cornet M, Baena M, Baiget M. Utility of a (GT) dinucleotide repeat in intron 1 of the factor 8 gene for haemophilia A carrier diagnosis. Haemophilia 2005;11: 142-4.
7. Dardik R, Kenet G, Kornbrot N, Rosenberg N, Martinowitz U. Two novel polymorphic microsatellite markers in the FVIII gene useful in haemophilia A diagnosis. Haemophilia 2002;8: 567-8.
8. Lalloz MR, McVey JH, Pattinson JK, Tuddenham EG. Haemophilia A diagnosis by analysis of a hypervariable dinucleotide repeat within the factor VIII gene. Lancet 1991;338: 207-11.
9. Lalloz MR, Schwaab R, McVey JH, Michaelides K, Tuddenham EG. Haemophilia A diagnosis by simultaneous analysis of two variable dinucleotide tandem repeats within the factor VIII gene. Br J Haematol 1994;86:804-9.
10. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263-5.