BLOOD RESEARCH
VOLUME 48 · NUMBER 3
September 2013
ORIGINAL ARTICLE

Sequence variation data of *F8* and *F9* genes in functionally validated control individuals: implications on the molecular diagnosis of hemophilia

Ja Young Seo[1#], Mi-Ae Jang[1], Hee-Jung Kim[1], Ki-O Lee[2], Sun-Hee Kim[1], Hee-Jin Kim[1]

[1]*Department of Laboratory Medicine & Genetics, Samsung Medical Center, Sungkyunkwan University School of Medicine,* [2]*Samsung Biomedical Research Institute, Samsung Medical Center, Seoul, Korea*

p-ISSN 2287-979X / e-ISSN 2288-0011
http://dx.doi.org/10.5045/br.2013.48.3.206
Blood Res 2013;48:206-10.

Received on March 28, 2013
Revised on July 16, 2013
Accepted on July 24, 2013

#Current affiliation: Department of Laboratory Medicine, Gachon University Gil Medical Center, Incheon, Korea

*This study was supported by a grant from the Korea Hemophilia Foundation.

Correspondence to
Hee-Jin Kim, M.D.
Department of Laboratory Medicine & Genetics, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81, Irwon-ro, Gangnam-gu, Seoul 135-710, Korea
Tel: +82-2-3410-2710
Fax: +82-2-3410-2719
E-mail: heejinkim@skku.edu

Background
The *F8* and *F9* genes encode for coagulation factor VIII (FVIII) and FIX, respectively, and mutations in these genes are the genetic basis of hemophilia A/B. To determine whether a sequence variation in *F8*/*F9* is a disease-causing mutation, frequency data from a control population is needed. This study aimed to obtain data on sequence variation in *F8*/*F9* in a set of functionally validated control chromosomes of Korean descent.

Methods
We re-sequenced *F8* and *F9* from DNA samples of 100 Korean male control individuals with normal PT, aPTT, and FVIII activity. PCR and direct sequencing analyses were performed using primer pairs to cover all coding regions and the flanking intronic sequences.

Results
Thirteen individuals (13%) were hemizygous for sequence variations in the coding region of *F8*. Six (6%) had c.3780C>G (p.Asp1260Glu), five (5%) had c.3864A>C (p.Ser1288=). One each individual (1%) had c.4794G>T (p.Glu1598Asp) and c.5069 A>G (p.Glu1690Gly). Asp1260Glu and Ser1288= were known SNPs (rs1800291 and rs1800292, respectively). Glu1598Asp was assigned as a missense mutation in public databases (HGMD and HAMSTeRS), and Glu1690Gly was a novel variation. Based on the normal FVIII activities in control individuals carrying these variations (109% and 148%, respectively), they were considered to be rare SNPs. No variation was observed in *F9* of control individuals.

Conclusion
A significant proportion of control individuals carried sequence variations in *F8*, but not in *F9*. These results can be used as a reference dataset for molecular diagnosis of hemophilia A and B, particularly in Korea.

Key Words Hemophilia, *F8*, *F9*, Sequence variation, Control population, Korea

INTRODUCTION

Hemophilia A and B are X-linked recessive bleeding disorders deficient coagulation factor VIII (FVIII) and IX (FIX), respectively. Mutations in *F8* and *F9* gene are the genetic basis of hemophilia A and B. Affected males suffer from prolonged oozing after injuries or surgery and joint bleeding. The age at diagnosis and the severity of bleeding are closely related to the level of the residual coagulation factor activity [1, 2]. The diagnosis of hemophilia A and B is suspected based on the personal and family history of bleeding episodes and abnormal coagulation tests, which include prolonged activated partial thromboplastin time (aPTT) corrected on mixing test with normal plasma and decreased FVIII or FIX activity. Molecular genetic tests confirm the diagnosis in most patients with hemophilia A and B by identifying disease-causing mutations in *F8* and *F9*, respectively [1, 2]. When a variation is detected, it is critical to gather further information and evidence to label it as a known or novel disease-causing mutation rather than a polymorphism. To date, 105 and 29 coding sequence variants of *F8* and *F9*,

respectively, are enlisted in the National Center for Biotechnology Information (NIH) single nucleotide polymorphism database (dbSNP) (last accessed March 20, 2013) [3]. Since the frequency and distribution of sequence variants vary across different populations, particularly in *F8* [4], sequence variation data should be obtained from a panel of DNA samples of a control population from a given ethnic origin. In particular, variation data from control samples from functionally validated individuals are critical for reference in molecular diagnostics. In this study, we aimed to obtain sequence variation data of *F8* and *F9* in a set of control chromosomes from individuals of Korean descent and establish a set of sequence variation data to serve as a reference for the molecular diagnosis of hemophilia in Korean patients.

MATERIALS AND METHODS

Re-sequencing of *F8* and *F9* genes

The DNA samples were obtained from 100 male control individuals of Korean descent (total, 100 X chromosomes). They had normal prothrombin time (PT), aPTT, FVIII activity (%), and liver function tests. Since FVIII is an acute-phase reactant, we confirmed that the study subjects had a normal level of C-reactive protein. Genomic DNA was isolated from the peripheral blood leukocytes of study subjects using the Wizard Genomic DNA Purification Kit, according to the manufacturer's instructions (Promega, Madison, WI, USA). All exons and their flanking intronic sequences of *F8* and *F9* were amplified by PCR using the primer pairs designed by the authors (available on request) on a thermal cycler (Model 9700; Applied Biosystems, Foster City, California, USA). Direct sequencing was performed using the same primer sets using the BigDye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems, Rotkreuz, Switzerland) on the ABI Prism 3130 Genetic Analyzer (Applied Biosystems). To identify sequence variations, the patient's sequences were compared with the reference sequence (GenBank accession number NM_000132.3 for *F8*; NM_000133.3 for *F9*) using the DNA sequence assembly software, Sequencher 4.10.1 (Gene Codes Corporation, Ann Arbor, Michigan, USA). Any variations detected were described according to the guidelines by the Human Genome Variation Society (HGVS; http://www.hgvs.org/). For each cDNA sequence, +1 signified the first base (A of ATG) for translational initiation, while at the protein level, +1 signified the first methionine, as indicated by the HGVS guideline. The conventional numbering system based on the mature protein (the first methionine numbered as -19) was given in parentheses or brackets. We referred to the dbSNP, the Human Gene Mutation Database (HGMD; http://www.hgmd.cf.ac.uk/ac/), the Haemophilia A Mutation, Structure, Test and Resource Site (HAMSTeRS), 1000 Genomes (http://www.1000genomes.org/), the Exome Variant Server (http://evs.gs.washington.edu/EVS/), as well as published re-sequencing studies on a panel of non-hemophiliac individuals from several population groups to decide whether an identified missense variation was previously reported or novel [4, 5]. For a novel amino acid-changing variation, cross-species amino acid alignment was performed using ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Bioinformatics analyses were performed to predict the effect of the amino acid substitution by using SIFT (http://sift.jcvi.org/) and PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/). This study was approved by the institutional review board.

Statistical analyses

The frequency of coding sequence variations were described using the observed frequency and standard error. The comparison of FVIII activities between 2 groups with and without a variation of interest was performed by Mann-Whitney test. A P value less than 0.05 was considered statistically significant. All statistical analyses were performed using MedCalc statistical software, Version 11.5.1 (MedCalc Software, Mariakerke, Belgium).

RESULTS

Thirteen individuals (13%) were shown to be hemizygous for a sequence variation in the coding sequence of *F8* (Table 1). Six individuals (6%) had c.3780C>G (p.Asp1260Glu)

Table 1. Four sequence variations of the *F8* gene observed in functionally validated male control individuals of Korean descent.

Serial number	Description	Observed frequency (%)	Standard error	Exon	Domain	Database	Functional prediction (SIFT/PolyPhen-2)	FVIII activity level[a] (mean±SD)
1	c.3780C>G (p.Asp1260Glu)	6	0.024	14	B	dbSNP (MAF 24.6%)	Tolerated/ Benign	104%±25%
2	c.3864A>C (p.Ser1288=)	5	0.022	14	B	dbSNP (MAF 10.5%)	NA	121%±47%
3	c.4794G>T (p.Glu1598Asp)	1	0.010	14	B	HGMD/HAMSTeRS	Tolerated/Benign	109%
4	c.5069A>G (p.Glu1690Gly)	1	0.010	14	B	None	Tolerated/Probably damaging	148%

[a]Reference interval of FVIII activity: 55–150% for blood type O, 77–205% for blood type non-O.
Abbreviations: MAF, minor allele frequency; NA, not applicable; FVIII, coagulation factor VIII; SD, standard deviation.

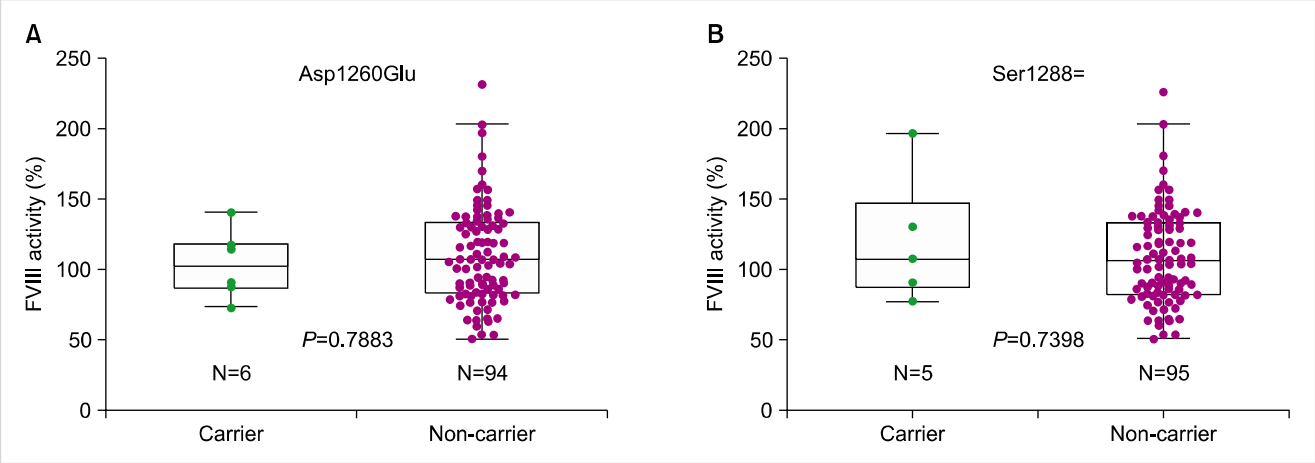

Fig. 1. The distribution of factor VIII activity levels of the carriers and non-carriers of 2 common polymorphisms Asp1260Glu **(A)** and Ser1288= **(B)** among the control individuals. The upper/lower ends and the inner lines of the boxes correspond to the upper/lower quartiles and median values, respectively. Circles represent factor VIII activity values. The lines from each box extend from the minimum to the maximum values, excluding outlying values displayed as a separate circle.

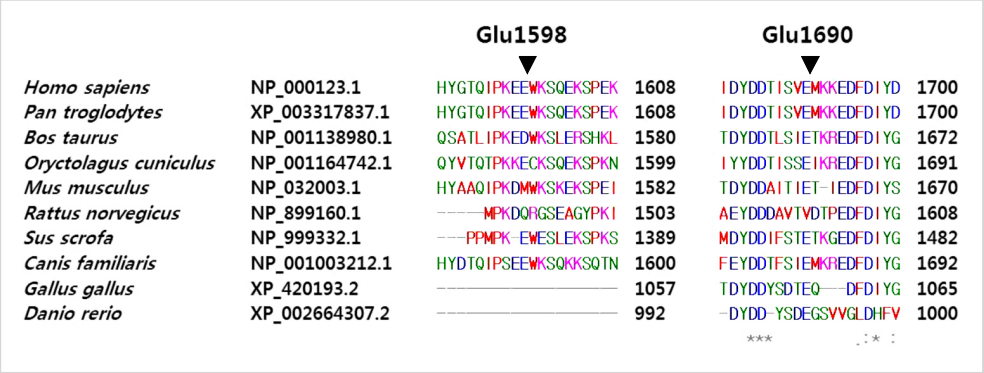

Fig. 2. Alignment of the F8 peptide sequences in human and other species using ClustalW2. Note that the Glu1598 and Glu1690 residues are not conserved across mammalian and non-mammalian species.

[p.Asp1241Glu] and five individuals (5%) had c.3864A>C (p.Ser1288=) [p.Ser1269=]. One each individual had c.4794 G>T (p.Glu1598Asp) [p.Glu1579Asp] and c.5069A>G (p.Glu1690Gly) [p.Glu1671Gly] (Table 1). According to the 1000 Genomes database, c.3780C>G (Asp1260Glu) and c.3864A>C (Ser1288=) were known common SNPs (rs1800291 and rs1800292, respectively) with minor allele frequencies (MAF) of 24.6% and 10.5%, respectively. Since Asp1260Glu was previously described to be associated with the FVIII activity [4, 6, 7], we compared the FVIII activities of individuals carrying Asp1260Glu to those individuals without Asp1260Glu. The results showed that there was no significant difference in FVIII activities between the 2 groups by a univariate analysis (104% vs. 110% in individuals with Asp1260Glu and without Asp1260Glu, respectively, P=0.7883, Fig. 1). We also found no significant difference in FVIII activities between the carriers of Asp1260Glu and individuals without any variants (104% vs. 95%, P=0.8328). Glu1598Asp (Glu1579Asp) was enlisted in both HGMD and HAMSTeRS as a disease-causing mutation in a patient with severe hemophilia A, but not in dbSNP [8]. Glu1690Gly (Glu1671Gly) was a novel missense variation enlisted neither in public database nor in the literature [4, 5]. Comparative genomic analyses demonstrated neither Glu1598 nor Glu1690 residues were conserved among mammalian and non-mammalian species (Fig. 2). Functional predictions using SIFT and PolyPhen-2 demonstrated that Glu1598Asp was predicted to be tolerated and benign. Glu1690Gly was predicted to be tolerated by SIFT, but to be probably damaging by PolyPhen-2. Based on the normal FVIII activities in the control individuals carrying these variations (109% in individuals carrying Glu1598Asp and 148% in individuals carrying Glu1690Gly), both were considered to be rare SNPs rather than a causative mutation of hemophilia A. No sequence variation was observed in the *F9* gene in the 100 control chromosomes.

DISCUSSION

Historically, linkage analysis was used to determine the mutation status in individuals with hemophilia [9-11]. In recent years, direct mutation detection techniques such as direct sequencing analysis, targeted mutation analysis (intron 22 and intron 1 inversion mutations of *F8*), and deletion/duplication analysis have replaced linkage analysis, and these direct techniques have detected disease-causing mutations in >95% of patients with hemophilia [12-16]. Given

the scarcity of mutation hotspots within these genes and the large size of *F8*, a significant proportion of the hemophilia mutations in recent reports are still novel, and it is critical to determine whether the novel variation is a disease causing mutation or polymorphism. In particular, when the variation is neither apparently deleterious nor silent but is predicted to change the amino acid, deciding whether the variation is a disease-causing mutation or a polymorphism becomes highly challenging. For these reasons, variation dataset from functionally validated (factor levels within reference ranges) control samples are crucial for molecular diagnosis of hemophilia. In particular, male control samples from ethnic origin of interest are relevant to hemophilia diagnostics because carrier females of hemophilia A/B may have a normal factor level and the SNP frequencies throughout the entire *F8* gene differ widely across different ethnic groups [4, 17]. According to the validation of mutations approved by the Human Genome Organization (HUGO) Mutation Database Initiative/Human Genome Variation Society (HGVS), the determination of disease-causativeness needs to be supported by observation of 0% frequency of the variation in 100 normal chromosomes.

In the present study, we obtained a set of variation data for the *F8* gene in 100 control male individuals (100 X chromosomes) of Korean descent with normal FVIII activities. A significant proportion (13%; 13/100) of the control chromosomes carried sequence variations in *F8*. Asp1260Glu (Asp1241Glu) and Ser1288= (Ser1269=) are common SNPs previously enlisted in the dbSNP (rs1800291 and rs1800292, respectively). The frequencies of both Asp1260Glu and Ser1288= in Korean were lower than the minor allele frequencies in the dbSNP (6% vs. 24.5%, 5% vs. 10.5%, respectively). c.3780C>G (p.Asp1260Glu) demonstrated a large difference in the frequency of the variant G allele. The frequency of the variant G allele was reported at 3.4% in Japanese, 7.8% in Han Chinese, 15.3% in European, and 69.2% in Sub-Saharan African individuals [3]. Thus, the frequency of c.3780C>G (Asp1260Glu) in Korean control individuals (6%) was between those reported for Japanese and Han Chinese individuals. c.3864A>C (p.Ser1288=) is a common SNP across population and the frequency of the variant C allele in this study was 5%, which fell between those reported for Japanese (2.2%) and Han Chinese (7.8%) individuals [3]. The frequency of the variant allele in the Korean individuals was similar to that of sub-Saharan African individuals (5.8%) and lower than that of European individuals (8.3%). The aspartate residue at 1260 is located in the B domain of the FVIII protein, and Asp1260Glu was reported to be associated with decreased FVIII activity (by −10%) [4, 6, 7]. However, the difference of FVIII activities was not significant between individuals with or without Asp1260Glu in our study, possibly due to the limited number of samples or other factors such as age, ABO blood type, smoking, and plasma level of the von Willebrand factor [4]. Glu1598Asp (Glu1579Asp) was registered in LSDB (HAMSTeRS) and HGMD as a disease-causing mutation in a patient with severe hemophilia A, but was not registered in dbSNP and other polymorphism databases [8]. However, our data indicate that Glu1598Asp is a rare SNP rather than a mutation associated with hemophilia A because it was found in our control chromosomes. Glu1690Gly (Glu1671Gly) was a novel amino acid-changing variation and was found in 1%(1/100) of control chromosomes. Based on the normal FVIII activity in control individual carrying this variation (148%), it was considered a rare SNP, also. Unlike *F8*, no variation was observed in coding sequences of *F9* in all 100 of the control chromosomes. Although our control sample set was not tested for FIX activity, the result implicates that *F9* is less polymorphic in Korean descendants compared to *F8*. So, when a novel amino acid-changing variation is found in the *F9* gene in a patient suspected to have hemophilia B, one should search for other evidence to prove disease causativeness of the variation.

The results obtained in this study can be used as a reference dataset for molecular diagnosis of hemophilia A and B in Korea. In addition, the data we obtained revealed rare SNPs, including one enlisted as a missense mutation causing hemophilia A. The data presented here indicate that interpretation of sequencing data using public databases to diagnose hemophilia A and B should be approached cautiously. It should also be noted that the number of chromosomes (100) used in this study has a limitation in identifying other rare SNPs, especially those with a minor allele frequency <1% [18].

Authors' Disclosures of Potential Conflict of Interest

No potential conflicts of interest relevant to this article were reported.

REFERENCES

1. GeneReviews: Hemophilia B. Seattle, WA: University of Washington, 2011. (Accessed February 25, 2013, at http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=gene&part=hemo-b)
2. GeneReviews: Hemophilia A. Seattle, WA: University of Washington, 2011. (Accessed February 25, 2013, at http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=gene&part=hemo-a)
3. dbSNP Short genetic variations. The NCBI Single Nucleotide Polymorphism database. Bethesda, MD: NCBI, 2012. (Accessed February 25, 2013, at http://www.ncbi.nlm.nih.gov/SNP/)
4. Viel KR, Machiah DK, Warren DM, et al. A sequence variation scan of the coagulation factor VIII (FVIII) structural gene and associations with plasma FVIII activity levels. Blood 2007;109:3713-24.
5. Singer H, Ahmed R, Ivaskevicius V, et al. Genetic variability of the factor VIII gene in the normal population. 34th Hemophilia Symposium 2005:348-50.
6. Machiah D, Viel K, Almasy L, et al. A common SNP in the factor VIII (f-VIII) gene encodes a conservative aspartate to glutamate substitution (Asp1241Glu) in the B-domain that influences f-VIII activity levels. Blood 2003;102(Suppl):abst 181.
7. Scanavini D, Legnani C, Lunghi B, Mingozzi F, Palareti G,

Bernardi F. The factor VIII D1241E polymorphism is associated with decreased factor VIII activity and not with activated protein C resistance levels. Thromb Haemost 2005;93:453-6.
8. HAMSTeRS. The Haemophilia A Mutation, Structure, Test and Resource Site. London, UK: MRC Clinical Sciences Centre, 2012. (Accessed February 25, 2013, at http://hadb.org.uk/)
9. Grunebaum L, Cazenave JP, Camerino G, et al. Carrier detection of Hemophilia B by using a restriction site polymorphism associated with the coagulation Factor IX gene. J Clin Invest 1984;73:1491-5.
10. McCurdy PR. Use of genetic linkage for the detection of female carriers of hemophilia. N Engl J Med 1971;285:218-9.
11. Oberle I, Camerino G, Heilig R, et al. Genetic screening for hemophilia A (classic hemophilia) with a polymorphic DNA probe. N Engl J Med 1985;312:682-6.
12. Liu Q, Nozari G, Sommer SS. Single-tube polymerase chain reaction for rapid diagnosis of the inversion hotspot of mutation in hemophilia A. Blood 1998;92:1458-9.
13. Keeney S, Watson P, Hay C, Cumming A. Rapid turnaround for full mutation analysis of larger genes: robotic processing and automated DNA sequencing of the essential regions of the FVIII (F8) gene for mutation identification in haemophilia A. J Med Genet 2004;41(Suppl 1):S67.
14. Rossetti LC, Radic CP, Larripa IB, De Brasi CD. Genotyping the hemophilia inversion hotspot by use of inverse PCR. Clin Chem 2005;51:1154-8.
15. Kwon MJ, Yoo KY, Kim HJ, Kim SH. Identification of mutations in the F9 gene including exon deletion by multiplex ligation-dependent probe amplification in 33 unrelated Korean patients with haemophilia B. Haemophilia 2008;14:1069-75.
16. Kim HJ, Chung HS, Kim SK, et al. Mutation spectrum and inhibitor risk in 100 Korean patients with severe haemophilia A. Haemophilia 2012;18:1008-13.
17. Seattle SNPs. National Heart, Lung, and Blood Institute Program for genomic applications. Bethesda, MD: NHLBI, 2009. (Accessed February 25, 2013, at http://pga.gs.washington.edu)
18. Collins JS, Schwartz CE. Detecting polymorphisms and mutations in candidate genes. Am J Hum Genet 2002;71:1251-2.