Integration of molecular and clinical data of 40 unrelated von Willebrand Disease families in a Spanish locus-specific mutation database: first release including 58 mutations

The VWF gene (*VWF*) extends along 178 kb in the genome and comprises a total of 52 exons, whose sizes range from 40 to 1400 bp.[1] Systematic identification of the responsible mutations has been hampered by the large size and complex genomic organization of *VWF*, which is highly polymorphic and has a highly homologous (>96%) partial pseudogene in chromosome 22.[2,3] For several years, our group has been involved in the genetic study of the *VWF*, and this effort led to the design and optimization of a technique for complete sequencing of all exons and the intronic flanking regions of the gene, with very favorable results.[4] Herein, we present the *VWF* Mutation Registry at Hemobase (VWFdb@Hemobase), a freely accessible online database (*http://www.vwf.hemobase.com*) developed to facilitate rapid publication, access to, and interpretation of the *VWF* mutations identified in our laboratory.

All the records in the database contain data from genetic testing carried out at the Molecular Diagnosis and Therapy Unit of the Blood and Tissue Bank in patients clinically diagnosed, mainly at the Hemophilia Unit of Vall d'Hebron Hospital (Barcelona), but also at other Spanish hospitals. All patients were diagnosed as having VWD and met the ISTH criteria.[5,6] Genetic testing of patients and relatives was approved by the Ethics Committee of Vall d'Hebron Hospital, and all participants provided their informed consent. Genomic DNA was extracted and the complete *VWF* of the index case of unrelated families was amplified and sequenced with a previously described method.[4] The numbering and nomenclature used for the mutations identified follow the recommendations of the International Society on Thrombosis and Haemostasis (ISTH) Scientific Subcommittee on von Willebrand Disease.[7] It is the same as that applied in the ISTH SSC VWF Database (http://www.vwf.group.shef.ac.uk/), thus facilitating comparison between the two databases and with previous publications.

The VWFdb@Hemobase is published online (Figure 1) as a subdomain of Hemobase.com, a website that hosts the Spanish hemophilia A and B mutation databases. The content and structure follow several recommendations from the Human Genome Variation Society (HGVS)[8,9] for locus-specific mutation database (LSDB) design and curation: covers many aspects of the variants, provides quality assessment for each entry, and contains information about the disease and resulting clinical phenotype. In this first release of the database, 93 individuals are included: 40 index cases and 53 relatives. A total of 58 mutations identified are distributed over the entire gene (*Online Supplementary Figure S1*): 11 type 1, 15 type 2A, 2 type 2M, 3 type 2N, one type 2 unclassified and 26 type 3. There were 44 substitutions, including 5 nonsense mutations and 6 mutations involved in a single gene conversion event, one indel, 2 insertions, and 11 potential splice site mutations (PSSM). The database contains 19 mutations first described by our group, 12 of which were reported in a previous publication;[4] the remaining 7 are presented herein including molecular and phenotypic data (Table 1). These mutations (4 missense, 2 PSSM, and one insertion) were found in 3 type 1, 3 type 3, and one type 2 patient.

Without the help of functional studies, it is often difficult to decide whether a particular sequence variant is

Table 1. Summary of molecular and clinical data associated to new mutations and their location in the VWF gene. Results of the *in silico* analysis of missense and potential splice site mutations.

IC	VWD type	[VWF:Ag], [VWF:RCo], [FVIII:C],	Blood group	Mutation type	Gene region	Nucleotide change	Amino acid change	Domain or Affected Splice Site	MISSENSE MUTATIONS ANALYSIS (score) PolyPhen[a]	SIFT[b]	PSSM ANALYSIS (Native – mutated splice site) NetGene2[c]	HSF[d]	STRs genotype	Comment
9	2	45, 12, 33	O	Missense	E28	c.4946T>A	p.I1649N	A2	Probably damaging (2.299)	Affects protein function (0.00)	-	-	19-6-12-17 (g) 22-7-14-18	Patient from other hospital. Subtype not established.
115	1	33, 42, 45	O	Missense	E35	c.5890C>A	p.Q1964K	D4	Benign (1.449)	Tolerated (1.00)	-	-	19-10-14-15 (g) 22-10-16-17	Patient with FXII deficiency
135	1	78, 31, 59	B	Missense	E14	c.1672G>T	p.D558Y	D2	Probably damaging (2.288)	Affects protein function (0.00)	-	-	20-6-12-15 (g) 21-7-14-16	Compound heterozygous with mutation 4022G>A
152	1	56, 42, 39	A	PSSM	IVS13	c.1533+15G>A	-	DSS intron 13	-	-	No difference	60.93-89.87 New potential splice site	21-7-16-15* 9-11-14-16	Female patient with a mutation in FIX gene
167	3	6, 9, 3	O	PSSM Missense	IVS5 E14	c.[533-2A>G; 1625C>G]	- p.A542G	ASS intron 6 D2	- Benign (1.265)	- Affects protein function (0.00)	0.71-No ASS Prediction -	96.35-67.4 Broken splice site -	20-7-16-18* 20-7-16-18*	Homozygous mutations. Consanguineous descent
295	3	12, 6, 5	-	Insertion	E19	c.2540_2541insA	p.N847KfsX18	D'	-	-	-	-	19-11-15-15* 19-11-15-15*	Homozygous mutation. Consanguineous descent

*IC: index case; PSSM: potential splice site mutation; ASS: acceptor splice site; DSS: donor splice site; (g): haplotype not established; *: allele carrying the mutation. [a]Large values of PolyPhen Score indicate that the substitution is rarely or never observed in the protein family, suggesting likelihood that the amino acid replacement will be deleterious. [b]SIFT Score use the comparison to experimental data as a cut off for prediction and considers that substitutions with less than 0.05 are deleterious. [c]NetGene2 values close to 0.0 indicate intron region, while values close to 1.0 indicate exon and it is used to select donor and acceptor site predictions. [d]HSF define consensus values (CV) to analyze the difference between wild-type active sites and mutant inactive sites. Strong sites presented a CV higher than 80 and less strong sites a CV ranging between 70 and 80, and only a minor fraction of active sites showed a CV between 65 and 70.*

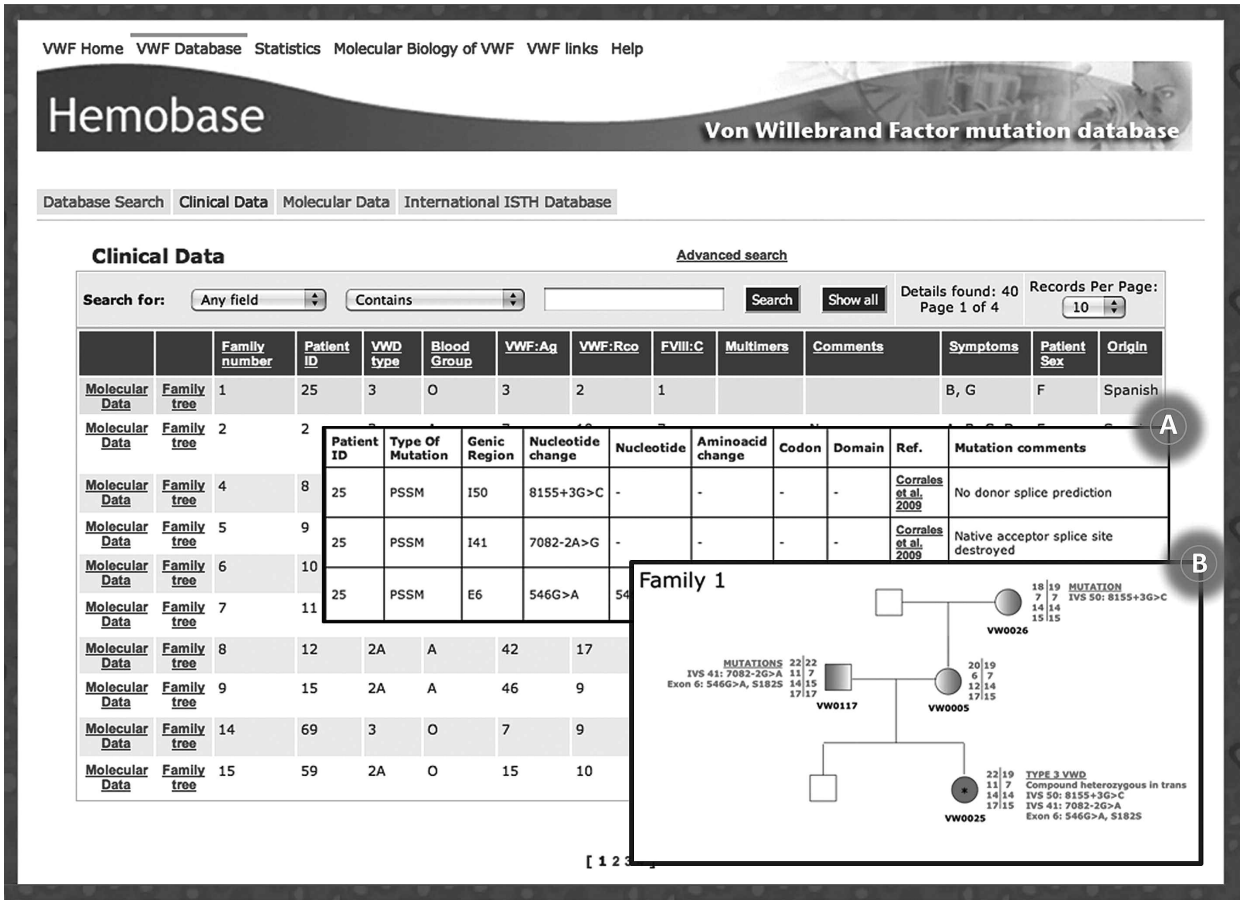

Figure 1. Screenshot of the web page for clinical data (*http://www.vwf.hemobase.com*), showing the patient-specific clinical and coagulation variables (blood group, laboratory coagulation parameters, symptoms, sex, geographical origin, etc.). The related molecular data (A) and family tree with inheritance information (B) can be easily consulted from this page by direct access with pop-up windows. The emergent molecular data table summarizes the significant information about the mutation(s) found in the patient. In addition, there is a link to PubMed entries of the articles reporting this variant. The website was built using the template-based website design tool RapidWeaver 4.1 (Realmac Software). The VWF database was developed using MS Access software (Microsoft Corporation), and ASPRunnerPro (Xlinesoft), a database management tool, was used to create a set of Active Server Pages to access, modify, display, and navigate the data on the web.

associated with disease or represents a rare polymorphism with no deleterious effect. In this regard, molecular studies are being conducted at the mRNA level to determine the true impact of PSSM (I Corrales, L Ramírez, R Parra and F Vidal, manuscript in preparation). However, software tools can provide useful information before in-depth mutation analysis is initiated. The results of *in silico* analysis of new mutations are shown in Table 1. The predicted impact of the novel missense mutations described was evaluated with PolyPhen (*http://genetics.bwh.harvard.edu/pph/*) and the PSSM with NetGene2 (*http://www.cbs.dtu.dk/services/NetGene2/*). Because of the limitations of *in silico* tools[10] a second program was used to perform the analysis: Sort Intolerant from Tolerant (SIFT; *http://blocks.fhcrc.org/sift/SIFT.html*) for missense mutations and Human Splicing Finder (HSF; *http://www.umd.be/SSF*) for PSSM.

An excellent International database of VWF mutations has been assembled by the University of Sheffield, UK.[11] All the mutations contained in the database have been sent (and will be submitted in the future) to the ISTH VWF database. There had been only 18 entries from Spanish laboratories before our participation with 41 different (58 total) mutations, which made us a significant single-laboratory contributor to the database. The ISTH VWF database provides users with a rich repository of mutations from worldwide contributing researchers, but its aim differs somewhat from ours. The VWFdb@Hemobase provides additional supplementary data that is not compiled by the international database, such as tabulated laboratory studies, STR gene tracking and pedigree representation of mutation and disease inheritance in the families recorded (Figure 1).

A major feature in designing the database user interface was the capability to browse and search for mutations in an intuitive and straightforward manner. Users of the VWFdb@Hemobase can view the information via 4 different main web pages: the advanced search page, a fast route to find mutations; the clinical data page, showing relevant patient-specific clinical and laboratory parameters; the molecular data page, which includes all mutation information in an individualized manner; and the international VWF database page, which corresponds to an up-to-date copy of the ISTH SSC VWF Database.[11] An advanced search system allows easy retrieval of any mutation in the database. The VWFdb@Hemobase also

includes general facts about VWD, the classification system into subtypes, clinical features of the disease, diagnostic difficulties involved, and the biochemical and molecular characteristics of VWF.

In conclusion, we present an online VWD LSDB containing the putative mutations responsible for VWD identified in several families after complete *VWF* sequencing of the index case. This database endeavors to assure that the relevant data obtained from molecular study of VWD patients will contribute to a better understanding of the mechanisms involved in the pathophysiology of this disease, providing a dynamic view of the molecular epidemiology of VWD in our population. Furthermore, it is a practical instrument to maintain updated information for clinicians and researchers,[8] and to offer fast and simple access to new data that will grow rapidly with the introduction of the next generation sequencing platforms.[12]

Irene Corrales,[1] Lorena Ramírez,[1] Júlia Ayats,[2] Carme Altisent,[2] Rafael Parra,[1,2] and Francisco Vidal[1]

[1]Unitat de Diagnòstic i Teràpia Molecular. Banc de Sang i Teixits, Barcelona; and [2]Unitat d'Hemofília de l'Hospital General Vall d'Hebron, Barcelona, Spain

Correspondence: Francisco Vidal, Unitat de Diagnòstic i Teràpia Molecular, Banc de Sang i Teixits Passeig Vall d'Hebron 119-129, 08035 Barcelona, Spain. Phone: international + 34.93 2746725. Fax: international + 34.93.2746727. E-mail: fvidal@bstcat.net

Funding: this work was supported in part with a grant (FIS, PI080385) from the Fondo de Investigaciones Sanitarias, Spanish Ministry of Science and Innovation, and by the Fundació Privada Catalana de l'Hemofília.

Citation: Corrales I, Ramírez L, Ayats J, Altisent C, Parra R and Vidal F. Integration of molecular and clinical data of 40 unrelated VWD families in a Spanish locus-specific mutation database: first release including 58 mutations. Haematologica 2010; 95(11):1982-1984. doi:10.3324/haematol.2010.028977

References

1. Mancuso DJ, Tuley EA, Westfield LA, Worrall NK, Shelton-Inloes BB, Sorace JM, et al. Structure of the gene for human von Willebrand factor. J Biol Chem. 1989;264(33):19514-27.
2. Federici AB, Canciani MT. Clinical and laboratory versus molecular markers for a correct classification of von Willebrand disease. Haematologica. 2009;94(5):610-5.
3. Bernardi F, Marchetti G, Casonato A, Gemmati D, Patracchini P, Legnani C, et al. Characterization of polymorphic markers in the von Willebrand factor gene and pseudogene. Br J Haematol. 1990;74(3):282-9.
4. Corrales I, Ramirez L, Altisent C, Parra R, Vidal F. Rapid molecular diagnosis of von Willebrand disease by direct sequencing. Detection of 12 novel putative mutations in VWF gene. Thromb Haemost. 2009;101(3):570-6.
5. Sadler JE. A revised classification of von Willebrand disease. For the Subcommittee on von Willebrand Factor of the Scientific and Standardization Committee of the International Society on Thrombosis and Haemostasis. Thromb Haemost. 1994;71(4):520-5.
6. Sadler JE, Budde U, Eikenboom JC, Favaloro EJ, Hill FG, Holmberg L, et al. Update on the pathophysiology and classification of von Willebrand disease: a report of the Subcommittee on von Willebrand Factor. J Thromb Haemost. 2006;4(10):2103-14.
7. Goodeve AC, Eikenboom JC, Ginsburg D, Hilbert L, Mazurier C, Peake IR, et al. A standard nomenclature for von Willebrand factor gene mutations and polymorphisms. On behalf of the ISTH SSC Subcommittee on von Willebrand factor. Thromb Haemost. 2001;85(5):929-31.
8. Cotton RG, Auerbach AD, Beckmann JS, Blumenfeld OO, Brookes AJ, Brown AF, et al. Recommendations for locus-specific databases and their curation. Hum Mutat. 2008;29(1):2-5.
9. Claustres M, Horaitis O, Vanevski M, Cotton RG. Time for a unified system of mutation description and reporting: a review of locus-specific mutation databases. Genome Res. 2002;12(5):680-8.
10. Houdayer C, Dehainault C, Mattler C, Michaux D, Caux-Moncoutier V, Pages-Berhouet S, et al. Evaluation of in silico splice tools for decision-making in molecular diagnosis. Hum Mutat. 2008;29(7):975-82.
11. Ginsburg D, Sadler JE. von Willebrand disease: a database of point mutations, insertions, and deletions. For the Consortium on von Willebrand Factor Mutations and Polymorphisms, and the Subcommittee on von Willebrand Factor of the Scientific and Standardization Committee of the International Society on Thrombosis and Haemostasis. Thromb Haemost. 1993;69(2):177-84.
12. Mardis ER. The impact of next-generation sequencing technology on genetics. Trends Genet. 2008;24(3):133-41.