ORIGINAL ARTICLE

Genetics

The European Association for Haemophilia and Allied Disorders (EAHAD) Coagulation Factor Variant Databases: Important resources for haemostasis clinicians and researchers

John H. McVey[1] 🆔 | Pavithra M. Rallapalli[2] 🆔 | Geoffrey Kemball-Cook[3] 🆔 |
Daniel J. Hampshire[4] 🆔 | Muriel Giansily-Blaizot[5] 🆔 | Keith Gomez[3] 🆔 |
Stephen J. Perkins[2] 🆔 | Christopher A. Ludlam[6]

[1]School of Bioscience and Medicine, University of Surrey, Guildford, UK

[2]Department of Structural and Molecular Biology, University College London, London, UK

[3]Katherine Dormandy Haemophilia Centre and Thrombosis Unit, Royal Free London NHS Foundation Trust, London, UK

[4]Department of Biomedical Sciences, University of Hull, Hull, UK

[5]Hematologie Biologique, CHU Montpellier, Montpellier, France

[6]University of Edinburgh, Edinburgh, UK

Correspondence
John H. McVey, Department of Biochemical Sciences, School of Biosciences and Medicine, University of Surrey, Guildford GU2 7XH, UK.
Email: j.mcvey@surrey.ac.uk

Funding information
Pfizer; European Association for Haemophilia and Allied Disorders; KD Trust; CSL Behring

Abstract

Introduction: Advances in genomic sequencing have facilitated the sequencing of genes associated with disorders of haemostasis. The identification of variants within genes and access to curated data incorporating structural, functional, evolutionary as well as phenotypic data has become increasingly important in order to ascribe pathogenicity.

Aim: The European Association for Haemophilia and Allied Disorders (EAHAD) Coagulation Factor Variant Database Project aims to provide a single port of entry to a web-accessible resource for variants in genes involved in clinical bleeding disorders.

Results: New databases have evolved from previously developed single gene variant coagulation database projects, incorporating new data, new analysis tools and a new common database architecture with new interfaces and filters. These new databases currently present information about the genotype, phenotype (laboratory and clinical) and structural and functional effects of variants described in the genes of factor (F) VII (F7), FVIII (F8), FIX (F9) and von Willebrand factor (VWF).

Conclusion: The project has improved the quality and quantity of information available to the haemostasis research and clinical communities, thereby enabling accurate classification of disease severity in order to make assessments of likely pathogenicity.

KEYWORDS
blood coagulation factors, factor VII, genotype, haemophilia A, haemophilia B, von Willebrand factor

1 | INTRODUCTION

The ever-increasing advances in genomic sequencing have facilitated the sequencing of disease-associated genes and in particular, those associated with disorders of haemostasis. The task of reporting and analysing DNA variants is a major challenge. There are currently two types of repositories for these variants. First, central databases contain pooled information on variation across the whole genome. These resources generally contain a list of known genetic variants but less frequently provide detailed information or analyses

that help predict the impact of variations on the structure, function and stability of the specific protein expressed by the gene associated with the disease. None of these central databases contain all the tools required for full clinical interpretation of variants. Second by contrast, gene variant databases, also known as locus-specific databases (LSDBs), that collect and display information on sequence variants (mutations) on a gene-by-gene basis are usually run by a consortium of collaborating researchers with scientific expertise in a particular gene or phenotype. LSDBs provide an invaluable tool for analysing variants, generally promoting submission of data and therefore enabling an accurate and up-to-date resource. Complete and up-to-date variant databases save valuable user time and stimulate effective diagnosis and research by providing access to curated standardized and annotated data, which would otherwise be scattered over many different resources. Importantly, the curators check the data entered into the database, which increases data quality, as there are a significant number of reporting errors in published articles. In addition, LSDBs provide access to variant information that has been collected by direct database submission and would not otherwise be available as it becomes increasingly difficult to publish variant data. With the advent of next-generation sequencing of ever-increasing numbers of individuals and therefore increased identification of variants within genes, access to curated data incorporating structural, functional, evolutionary as well as phenotypic data becomes increasingly important in order to ascribe pathogenicity.

Locus-specific databases for haemostatic genes were first established with the publication of collated data for variants in F9,[1] F8[2] and VWF[3] in the early 1990s. The first web-based database for F8 appeared in 1996.[4] Over the years, several additional coagulation factor variant databases have also been developed (F7,[5] F10, F11[6] and fibrinogen [FGA, FGB, FGG][7]). Some allowed database searching and provided a number of additional resources, whereas others offered data in a downloadable Excel spreadsheet. These resources provided an invaluable resource to the haemostasis community and were highly cited. However, many of these databases were established and maintained by individuals and availability of time and funding had a significant impact on their ability to maintain the database as well as continuity of access to a stable URL. An LSDB loses its relevance once it is not maintained and updated on a regular basis.

In response to these concerns, the European Association for Haemophilia and Allied Disorders (EAHAD) initiated a Coagulation Factor Variant Database Project with the aim of gathering together single gene variant databases involved in clinical bleeding disorders that would provide a single web portal to LSDBs for genes in haemostasis. The LSDBs would share a common architecture making navigation of the database(s) easier as well as providing greater support for maintenance of the sites.

2 | CURATION AND GOVERNANCE

The EAHAD Variant Database Project was established in 2011 and is administered by a Steering Group, a Curator Group and individual Variant Database Advisory Committees; further details of the composition of these groups can be found on the EAHAD Coagulation Factor Variant Database Project home page (dbs.eahad.org). Each individual variant database has a lead curator and a curator who are supported by further experts in the field. One of the important roles of the curators is to ensure that variants are correctly described with the correct nomenclature relative to the correct reference sequence. An increasingly important task of the curator(s) is also to suggest where possible the potential functional consequences of a variant. Current EAHAD curators are working as co-chairs or members of relevant ClinGen Variant Curation Expert Panels (clinicalgenome.org). ClinGen is an initiative to build a knowledge base on genes associated with disease and to assess likely pathogenicity of variants within a given gene. The variant curation process combines clinical, genetic, population and functional evidence with expert review to classify variants according to ACMG/AMP guidelines. The results are deposited in an open access database (ClinVar; ncbi.nlm.nih.gov/clinvar/).

The availability of an up-to-date open access variant database is a crucial tool for diagnostic and research laboratories; therefore, there is free and unrestricted access to all the online EAHAD variant databases. However, in order to honour the legal requirements of medical/health information and data protection ensuring that no individuals can be identified from the data stored and distributed in the databases, no patient IDs or the geographical location of the submitting laboratory are publicly available (unless they are already in the public domain through published articles). However, these details are maintained off-line in a secure database to make it possible to identify individual patient records and their source if required. Further details of the EAHAD variant database policy can be found on the home page (dbs.eahad.org).

3 | NOMENCLATURE

It is particularly important in molecular genetic analysis that there is no confusion resulting from differences in variant nomenclature between laboratories/publications. Many coagulation genes were cloned and initially sequenced during the 1980s, prior to the introduction of standardized nomenclature. As a result, genes and proteins have their own idiosyncrasies of naming and numbering. This can lead to confusion in the laboratory, literature and diagnostic setting. The Human Genome Variation Society (HGVS) has devised an extensive scheme for nomenclature of sequence variants (varnomen.hgvs.org[8]). To reduce confusion, an International Society of Thrombosis and Haemostasis (ISTH) working group published recommendations that full adoption of standard gene names and of DNA and protein sequence variant numbering according to HGVS guidelines should be adopted for all genes/proteins in haemostasis.[9] All variants described in the EAHAD databases therefore conform to HGVS guidelines and are reported in relation to reference sequences (RefSeq) maintained by the American National Centre for Biotechnology Information (NCBI). Variants are described using cDNA and protein reference sequences; however in future releases,

variants will also be described using a locus genomic reference (LRG) sequence. This description is currently available from the Leiden Open Variation Database (LOVD; lovd.nl) to which we mirror all variants described in the EAHAD database. Sequence variants differing from the reference sequence are denoted by a prefix, c. for cDNA and p. for amino acid variants. The A of the ATG initiator methionine is utilized as the sequence start point +1 and similarly the first methionine is numbered +1. This differs from some legacy numbering of haemostasis proteins where the signal peptide (and sometimes pro-peptide) was numbered negatively and amino acid numbering started from the beginning of the mature protein. Both HGVS and 'Legacy' numbering are provided for the *F7*, *F8* and *F9* databases. The CoagBase web page (coagbase.org) provides information for coagulation factor genes including gene names, RefSeq as well as an indication of the difference between legacy and HGVS amino acid numbering schemes.

4 | DATABASE STRUCTURE

The databases (Table 1) are built on a common architecture using a MySQL platform, and their interface was developed using HTML, CSS, JavaScript, Perl and PHP. They are available at dbs.eahad.org. In order to encourage the ethos of data sharing and open access to resources, the variant data in the databases are shared with LOVD (lovd.nl[10]), which is a freely available gene-centred collection of DNA variant data.

The databases allow simple searches of variants based on nucleotide or amino acid numbering (HGVS and legacy), variant type or location within the gene. In addition, advanced search options allow further refinement of searches based on variant effect, protein domain, severity and inhibitor status (where known). A search returns the unique variants at the selected position in the gene or protein as a list with links where appropriate to in-depth analysis of the variation and anonymized patient information (Figure 1). Unlike other variant DBs, the EAHAD databases store and display all cases reported regardless of how many individual reports of the identical genetic variant have been reported, allowing users to survey phenotypic variability among dozens or even hundreds of cases with the same variant.

It is possible to visualize/export the data either as list of unique variants at the selected position, or as a multiple patient list showing all individuals with a particular variant at that position, allowing between-case comparisons to assess variability of presentation. This also allows the assessment of the frequency of reporting a particular variant; however, it should be noted that this frequency may be biased by extensive family studies, founder effects in certain ethnic groups and does not represent the frequency within the affected population. In addition, the database presents statistics and graphics on all the variants in the database by specific type of variant, by protein domain and by disease severity (Figure 2). There is also a variant map of all missense variants mapped onto the amino acid sequence of the protein.

Predicting whether or not a variant is pathogenic may not be straightforward, especially for synonymous and missense variants or for those occurring in intronic sequences or promoter regions of the gene. The databases provide access to in silico analyses that may be useful in the prediction of variant pathogenicity of missense variants using open access software packages (Table 2). The minor allele frequency (MAF) of variants in large population studies such as the Genome Aggregation Database (gnomAD; gnomad.broad institute.org[11]) is also provided and can be useful in assessing the predicted variant pathogenicity. Where a MAF is not presented, this indicates that this variant has not been identified in the data set and is a rare variant or in a region that is not read to an adequate depth in the control data sets. The structural consequence of a missense variant can also be visualized if an appropriate molecular model of the relevant protein structure is available. Finally, the evolutionary conservation of the amino acid sequence of the protein at the variant residue in closely related species (eg, chimpanzee, gorilla, gibbon, bushbaby and marmoset), or of closely related human proteins, can be inspected. While the use of in silico methods such as these can be extremely useful in predicting the severity of the amino acid change, these methods should not be relied on solely to assign pathogenicity.

5 | IDENTIFICATION OF VARIANTS

The databases have been built incorporating variants previously reported in databases for each of the relevant genes (Table 1). Additional variants were identified by searches of the Human Gene

TABLE 1 European Association for Haemophilia and Allied Disorders (EAHAD) variant databases

Gene	Disorder	EAHAD DB website	LOVD DB website	Contributing databases
F7	Factor VII deficiency	f7-db.eahad.org	databases.lovd.nl/shared/genes/F7	MRC FVII mutation database[5]; UMD-F7 database
F8	Haemophilia A	f8-db.eahad.org	databases.lovd.nl/shared/genes/F8	The factor VIII structure and mutation resource site: HAMSTeRs; HADB[2,4,18]
F9	Haemophilia B	f9-db.eahad.org	lovd.nl/shared/genes/F9	Haemophilia B: database of point mutations and short additions and deletions[19]
VWF	von Willebrand disease	vwf-db.eahad.org	databases.lovd.nl/shared/genes/VWF	Sheffield VWFdb[20]

In Depth Variant Analysis: c.1244C>T (p.Ala415Val)

c.1244C>T	p.Ala415Val (Legacy AA No. 396)

Variant Type: Point	**Domain:** A2	**Sequence Context:** GCT > GTT
Variant Effect: Missense	**Location:** Exon(8)	**No of bases:** 1
No. of cases reported: 8	**Allele Frequency (MAF):**	

Individual Case Information : Hide

Case ID	FVIII:C% 1-st (presumed)	FVIII:C% (2-st/Chr)	Assay Ratio (1-st/2-st)	FVIII:Ag(%)	Type I/II	Severity	Comments	Inhibitors	Reference	Reporting Centre
1050	5			14	II	Moderate		No	Casana et al (conference abstract)	Spain
1051	4			12	II	Moderate		No	Casana et al (2008)	Spain
1052	4			4	I	Moderate		No	Casana et al (2008)	Spain
1053	2					Moderate		No	Centre A11 (unpublished)	
1054	<1			3		Severe		Yes	Vinciguerra et al (2006)	France
1055	<1					Severe		No	Vinciguerra et al (2006)	France
1056	<1					Severe		No	Vinciguerra et al (2006)	France
1057	4.2					Moderate			Lin et al (2008)	Taiwan

Residue Information :

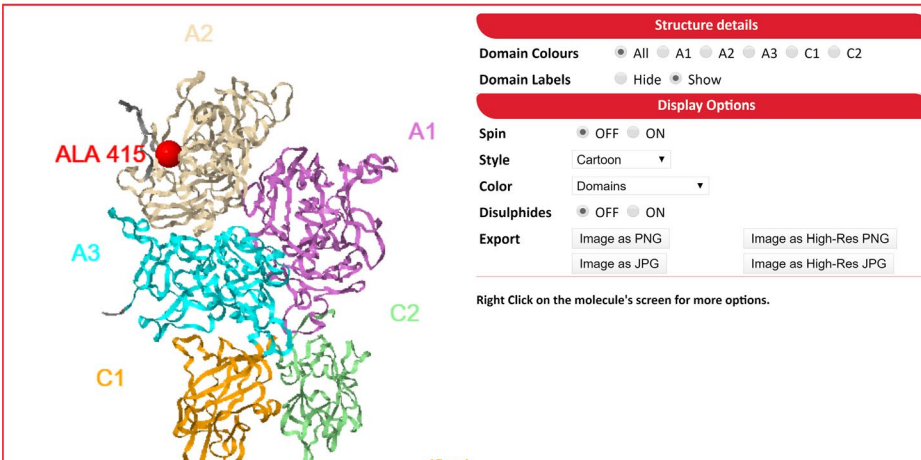|	Name	Type	Cyclic	Size	Hydrophobicity	Charge
Wild Type	Ala	aliphatic	acyclic	small	hydrophobic	neutral
Variant	Val	aliphatic	acyclic	medium	hydrophobic	neutral

Substitution Analysis :

- **Grantham Score : 64**
- **PolyPhen-2 Prediction : possibly damaging (Probability: 0.911)**
- **SIFT Prediction :Damaging (SCORE: 0.019)**
- **PROVEAN (Protein Variation Effect Analyzer) Prediction : Deleterious (SCORE: -3.03)**

Structural Implications :

- **Ala415 is shown below as a red sphere.**
- **Ala415 is a buried residue (the surface accessibility from the FVIII structure is 0).**
- **Ala415 is in a random coil area of the FVIII structure (the DSSP assignment from the modelled FVIII domains is C).**

Hint: Left click mouse | To rotate the structure
Hint: Rotate mousewheel | To zoom in/out the structure
Hint: Right click mouse | to use applet control options

Multiple Sequence Alignment of Factor VIII protein sequence across Primates.

Note: Clicking and Dragging the mouse over the sequences will take you across the sequence limits
Note: All numbering is for human FVIII and is HGVS

FIGURE 1 Screenshot of the in-depth variant analysis following a search of the Factor VIII Gene Variant Database for variants in codon 415 (legacy 396) showing the predicted consequence of variant c.1244C>T (p.Ala415Val) using in silico prediction programmes, structural implications and location of the variant within the FVIII structure and finally multiple amino acid sequence alignments across primates. Further multiple sequence alignments are available from the AA Alignments tab at the top of the page. The case information for this variant is also shown demonstrating the range of laboratory data and ascribed bleeding severity in eight reported individuals
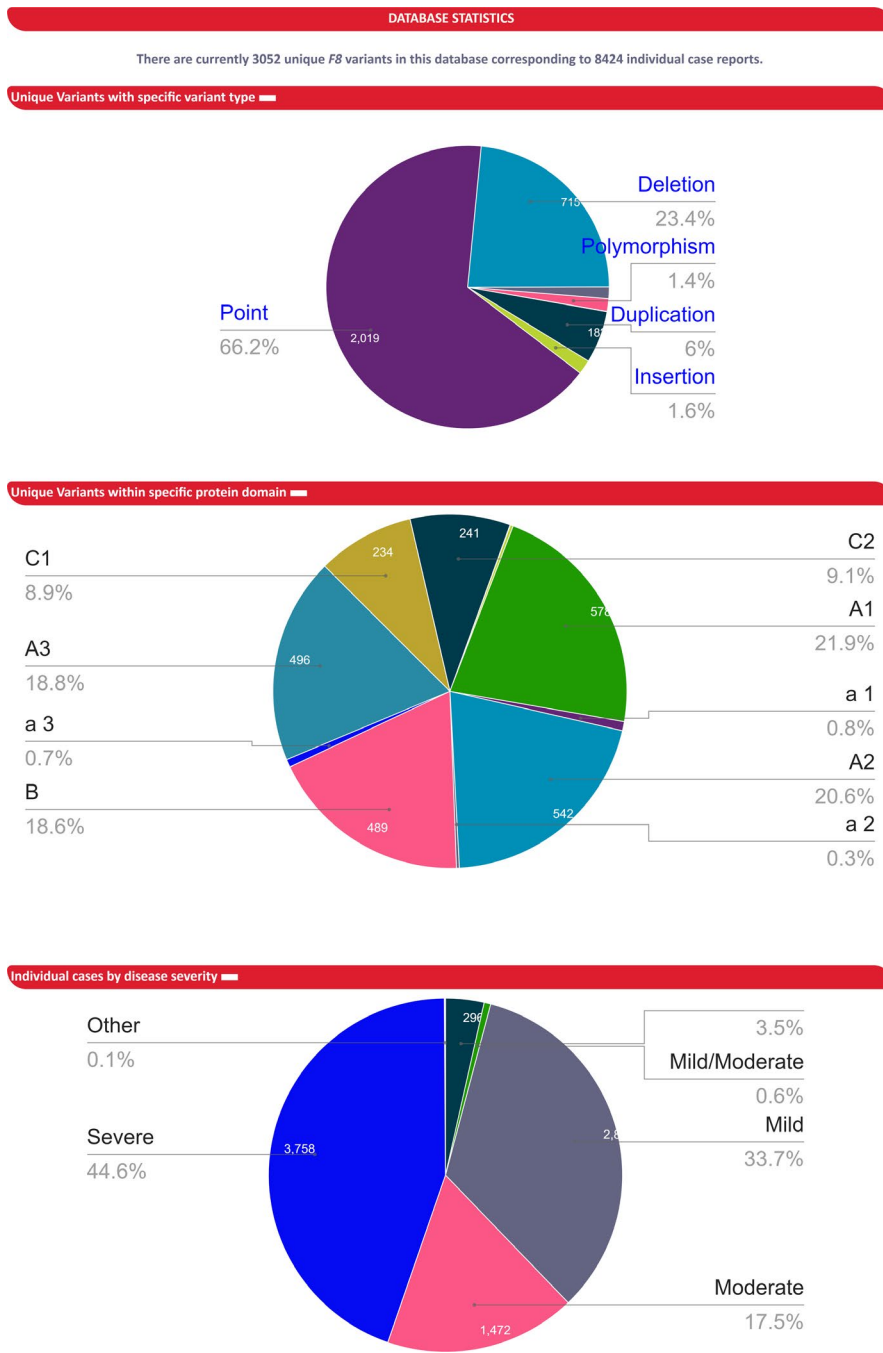

FIGURE 2 Screenshot of the database statistics available on the Factor VIII Gene Variant Database f8-db.eahad. org/statistics.html.php illustrating the percentage of: variants by type; variants within specific protein domains; and cases by disease severity. Similar information is available on the factor IX and VII websites

Mutation Database (hgmd.cf.ac.uk[12]), literature searches and direct submission by reporting laboratories. All variants incorporated into the new databases were verified for accuracy, and HGVS nomenclature was generated and checked with Mutalyzer (mutalyzer.nl[13]). Each variant was linked to the original publication describing the variant(s). Finally, any duplicated cases were removed. Identifying duplicate accounts of the same patient can be problematic as multiple published accounts of the same cases can occur and there may be no explicit reference to previous publications and/or the use of a unique laboratory patient identifier. If there was any doubt over duplication of reporting an individual, then both records were maintained in the database. Future updates to the databases will be made

TABLE 2 In silico tools used to predict the consequence of missense variants

Program name	URL and key reference	Operating principle
ALIGN-GVGD (Grantham)	agvgd.hci.utah.edu;[21]	Combines the biophysical characteristics of amino acids and protein multiple sequence alignments to predict where missense substitutions in genes of interest fall in a spectrum from enriched deleterious to enriched neutral
POLYPHEN-2	genetics.bwh.harvard.edu;[22]	Predicts possible impact of an amino acid substitution on the structure and function of the protein based on physical and evolutionary conservation
SIFT	sift.jcvi.org/;[23]	Uses sequence alignments to score the probability that the substitution would be evolutionarily tolerated
PROVEAN	provean.jcvi.org;[24]	Uses pairwise sequence alignment scores to generate predictions for amino acid substitutions in protein sequence homolog

on a yearly basis by continued literature searches. Direct submission of variant cases to EAHAD database curators is highly encouraged, and electronic submission forms are available on the EAHAD Variant Database Project home page (dbs.eahad.org). All new submissions are validated by the expert curator(s) before they are added to the relevant database.

6 | UTILITY OF EAHAD DATABASES

It is recognized that the availability of standardized, validated DNA variant and phenotypic data is a crucial tool for diagnostic and research laboratories in the accurate classification of disease severity, and in order to make assessments of likely pathogenicity. The establishment of the EAHAD Coagulation Factor Variant Database Project has enabled the development of a single platform for gene-specific variant databases for the rare bleeding disorders. It has allowed the development of three new online databases for variants in the genes encoding FVII, FVIII and FIX and links with the existing database for variants in the gene for VWF. One of the important advances has been the conversion of legacy descriptions of variant data to a standardized nomenclature as recommended by the HGVS and the use of reference sequences to allow unambiguous mapping of the positions of variants. The new databases have been updated substantially with new variants. The databases for *F7* (f7-db.eahad.org), *F8* (f8-db.eahad.org) and *F9* (f9-db.eahad.org) currently have 221, 2015 and 2591 unique variants in the respective genes corresponding to 728, 8424 and 3713 individual cases (accessed 18 July 2019).

6.1 | Assigning pathogenicity

Variant classification is central to the utility of molecular genetic diagnostics in clinical practice; however, predicting whether gene variants are pathogenic is not straightforward. While nonsense and frameshifting small and large rearrangements are accepted as pathogenic in most genes, the prediction of the effects of synonymous and missense variants or those located in introns or in promoter regions of genes is problematic. Classification should be evidence-based, objective and systematic. Guidelines to establish a framework for variant classification have recently been published[14,15] with the aim of evaluating evidence in a systematic hierarchical manner evaluating as follows: population data; expected effect of the variant on the gene product; clinical data; and predictive data derived from functional experiments or mutagenesis. The databases do not currently assign pathogenicity classification for individual variants but in future will link directly to ClinVar (ncbi.nlm.nih.gov/clinvar) which is an open access database that reports curated information and likely pathogenicity on gene variants. EAHAD curators are working as co-chairs or members of relevant curation panels that input into ClinVar.

Evaluation of variant frequencies from large population data sets can provide strong evidence that a variant is benign or can identify one that is sufficiently rare to be considered a candidate pathogenic variant. The EAHAD variant databases link to the MAF of a variant in gnomAD database (based on exome and genomes of 141 456 individuals[11]). MAF values >0.05 in a control population may be considered stand-alone support for a benign interpretation. However, many benign variants are private and therefore absence or a low allele frequency is not considered sufficient evidence for pathogenicity. The expected allele frequency for haemophilia A would be between 0.0001 (one in 10 000) and 0.00005 (one in 20 000). The most frequent allele associated with reduced factor VIII antigen and activity levels (20%) and a mild bleeding phenotype is p.Val64Gly with a MAF of 0.000228. For *F9*, the most frequent allele associated with reduced FIX activity and a mild bleeding phenotype is p.Glu323Lys with a MAF of 0.0009748. However, variation in *F7* is more complex with common variants (p.Arg314Gln, MAF = 0.1341; c.-325_-324insCCTATATCCT,

MAF = 0.1632) that are not pathogenic in themselves but are associated with substantially reduced levels of FVII and therefore when co-inherited with other variants can have a significant impact on the overall phenotype.

The databases do not contain detailed clinical information but provide information on the ascribed severity of bleeding based on the submitter's observations when available. Where available, laboratory data from individual cases have been included. There are a number of validated assays that measure functional activity or quantify plasma concentrations of coagulation factors. This type of data straddles the boundary between clinical phenotype and in vitro functional data but must be interpreted in the context of the pathophysiological levels associated with particular diseases. For example, in the case of FVII activity it is essential to use only data obtained using human tissue factor (TF; thromboplastin) as the trigger for coagulation. The use of TF from animal sources or recombinant versions thereof may result in discrepant and misleading in vitro results that are inconsistent with assays performed with human TF and the bleeding phenotype of the individual concerned.[16] Thus, the species of TF (unknown, rabbit or human) used to generate the FVII activity values (FVII:C) presented in the *F7* variant database are indicated where known. Similarly, it is well recognized that certain missense variants in *F8* are associated with 'assay discrepancy' such that the values obtained with two-stage or chromogenic assays are substantially lower than the result with the one-stage assay. The opposite effect with a lower one-stage result is much less common but may also occur. This is of clinical significance as reliance on one assay might lead to a diagnosis of mild haemophilia A being missed. Variants known to be associated with 'assay discrepancy' are labelled as such in the database and on clinical reports. Furthermore, by collecting these variants together in the database insights into their localization in the FVIII tertiary structure have been gained which allows prediction of whether a novel variant might have a similar effect.

It should however be noted that discrepancies in the laboratory value or ascribed clinical severity between cases can arise due to differences in classification schemes used in different centres, different assays and standards used in reporting residual activity and finally the reported level may reflect a trough level rather than the true baseline in patients on prophylaxis. A variant is more likely to be pathogenic if it has an impact on the expression or function of the gene product. We have used in silico programmes (Table 2) to predict the likely impact of missense variants on the resultant protein based on the biophysical properties of the variant and the evolutionary conservation of the amino acid residue. Prediction of non-coding variants is more problematic. Whereas variants in the invariant AG/GT dinucleotides located at the 5′ and 3′ intron-exon boundaries are highly likely to result in a pathogenic phenotype, it is much more difficult to predict the consequence of variants in promoter or intronic sequences. Multiple software programs have been developed to predict the possible effect of a variant on RNA splicing. Although not sufficient to establish clinical relevance, splice-site prediction can be very helpful to predict those variants that may merit further investigation. These programs have been shown to be accurate in predicting whether an effect on splicing is likely.[17]

Although predictions from in silico tools and biochemical/functional evidence are useful in predicting the severity of amino acid substitutions, they should not be used to assign pathogenicity in the absence of other supporting data. In particular, supporting clinical observations should have primacy as clinical data describe human disease directly, whereas quantitative and qualitative effects are only relevant to the disease as it correlates with disease pathophysiology.

6.2 | Future developments

We support the ethos of open access and data sharing, and we encourage researchers and clinicians to contribute to the project by submitting their variant data to the relevant databases via the submission forms on the EAHAD Coagulation Factor Variant Project home page (dbs.eahad.org). We have made all reasonable efforts to ensure that the information contained in the databases and displayed on the web pages is accurate; however, the database may contain errors and we would be grateful if users could inform us of any that they find.

The development of a common database architecture will allow other coagulation factor variant databases to be developed. Databases for *F10* and *F11* are planned as well as FGA, FGB and FGG, responsible for FX, FXI and fibrinogen deficiency, respectively. If you would like to contribute to the project by either developing future databases in the field or by offering your services as a curator, please contact us.

7 | CONCLUSION

The analysis of genotype-phenotype relationships in rare diseases such as the bleeding disorders associated with deficiencies in FVII, FVIII, FIX and VWF requires international collaboration and standardized reporting. The establishment of the EAHAD Coagulation Factor Variant Database Project has generated a common database architecture and web-based platform that includes analytical tools to assess the structure/function consequence of variants. This has allowed the implementation of updated variant databases for *F7*, *F8*, *F9* and *VWF*. The variant databases will be an invaluable resource for scientists and clinicians in the field of haemostasis.

DISCLOSURES

The authors stated that they had no interests which might be perceived as posing a conflict or bias.

ORCID

John H. McVey https://orcid.org/0000-0002-7416-533X
Pavithra M. Rallapalli https://orcid.org/0000-0002-8261-6209
Geoffrey Kemball-Cook https://orcid.org/0000-0002-2649-9503
Daniel J. Hampshire https://orcid.org/0000-0002-1387-8926
Muriel Giansily-Blaizot https://orcid.org/0000-0002-3705-6177
Keith Gomez https://orcid.org/0000-0002-8934-0700
Stephen J. Perkins https://orcid.org/0000-0001-9218-9805

REFERENCES

1. Giannelli F, Green PM, High KA, et al. Haemophilia B: database of point mutations and short additions and deletions. *Nucleic Acids Res*. 1990;18(14):4053-4059.
2. Tuddenham EG, Cooper DN, Gitschier J, et al. Haemophilia A: database of nucleotide substitutions, deletions, insertions and rearrangements of the factor VIII gene. *Nucleic Acids Res*. 1991;19(18):4821-4833.
3. Ginsburg D, von Sadler JE. 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):177-184.
4. Wacey AI, Kemball-Cook G, Kazazian HH, et al. The haemophilia A mutation search test and resource site, home page of the factor VIII mutation database: HAMSTeRS. *Nucleic Acids Res*. 1996;24(1):100-102.
5. McVey JH, Boswell E, Mumford AD, Kemball-Cook G, Tuddenham EG. Factor VII deficiency and the FVII mutation database. *Hum Mutat*. 2001;17(1):3-17.
6. Saunders RE, O'Connell NM, Lee CA, Perry DJ, Perkins SJ. Factor XI deficiency database: an interactive web database of mutations, phenotypes, and structural analysis tools. *Hum Mutat*. 2005;26(3):192-198.
7. Hanss M, Biot F. A database for human fibrinogen variants. *Ann N Y Acad Sci*. 2001;936:89-90.
8. den Dunnen JT, Dalgleish R, Maglott DR, et al. HGVS recommendations for the description of sequence variants: 2016 update. *Hum Mutat*. 2016;37(6):564-569.
9. Goodeve AC, Reitsma PH, McVey JH. Nomenclature of genetic variants in hemostasis. *J Thromb Haemost*. 2011;9(4):852-855.
10. Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT. LOVD vol 2.0: the next generation in gene variant databases. *Hum Mutat*. 2011;32(5):557-563.
11. Karczewski KJ, Francioli LC, Tiao G, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*. 2019:531210.
12. Stenson PD, Mort M, Ball EV, et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet*. 2017;136(6):665-677.
13. Wildeman M, van Ophuizen E, den Dunnen JT, Taschner PE. Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker. *Hum Mutat*. 2008;29(1):6-13.
14. Nykamp K, Anderson M, Powers M, et al. Sherloc: a comprehensive refinement of the ACMG-AMP variant classification criteria. *Genet Med*. 2017;19(10):1105-1117.
15. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-424.
16. Bolton-Maggs PH, Hay CR, Shanks D, Mitchell MJ, McVey JH. The importance of tissue factor source in the management of factor VII deficiency. *Thromb Haemost*. 2007;97(1):151-152.
17. Vreeswijk MP, Kraan JN, van der Klift HM, et al. Intronic variants in BRCA1 and BRCA2 that affect RNA splicing can be reliably selected by splice-site prediction programs. *Hum Mutat*. 2009;30(1):107-114.
18. Kemball-Cook G, Tuddenham EG, Wacey AI. The factor VIII structure and mutation resource site: HAMSTeRS version 4. *Nucleic Acids Res*. 1998;26(1):216-219.
19. Rallapalli PM, Kemball-Cook G, Tuddenham EG, Gomez K, Perkins SJ. An interactive mutation database for human coagulation factor IX provides novel insights into the phenotypes and genetics of hemophilia B. *J Thromb Haemost*. 2013;11(7):1329-1340.
20. Hampshire DJ, Goodeve AC. The international society on thrombosis and haematosis von Willebrand disease database: an update. *Semin Thromb Hemost*. 2011;37(5):470-479.
21. Tavtigian SV, Deffenbaugh AM, Yin L, et al. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J Med Genet*. 2006;43(4):295-305.
22. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4):248-249.
23. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4(7):1073-1081.
24. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One*. 2012;7(10):e46688.