Received: 31 January 2020 | Revised: 17 March 2020 | Accepted: 13 April 2020

DOI: 10.1002/humu.24025

DATABASES

Human Mutation HGVS HUMAN GENOME VARIATION SOCIETY WILEY

The EAHAD blood coagulation factor VII variant database

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

[1]Hematologie Biologique, CHU Montpellier, University of Montpellier, Montpellier, France

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

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

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

[5]Clinical Sciences, Edinburgh Medical School, University of Edinburgh, Edinburgh, UK

[6]Department of Biochemical Sciences, School of Biosciences and Medicine, University of Surrey, Guildford, 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
European Association for Haemophilia and Allied Disorders

Abstract

Hereditary blood coagulation factor VII (FVII) deficiency is a rare autosomal recessive bleeding disorder resulting from variants in the gene encoding FVII (*F7*). Integration of genetic variation with functional consequences on protein function is essential for the interpretation of the pathogenicity of novel variants. Here, we describe the integration of previous locus-specific databases for *F7* into a single curated database with enhanced features. The database provides access to in silico analyses that may be useful in the prediction of variant pathogenicity as well as cross-species sequence alignments, structural information, and functional and clinical severity described for each variant, where appropriate. The variant data is shared with the *F7* Leiden Open Variation Database. The updated database now includes 221 unique variants, representing gene variants identified in 728 individuals. Single nucleotide variants are the most common type (88%) with missense representing 74% of these variants. A number of variants are found with relatively high minor allele frequencies that are not pathogenic but contribute significantly to the likely pathogenicity of coinherited variants due to their effect on FVII plasma levels. This comprehensive collection of curated information significantly aids the assessment of pathogenicity.

KEYWORDS

blood coagulation disorders, factor VII deficiency, genetic variation, hemostasis, LSDB

1 | INTRODUCTION

The initiation of blood coagulation and subsequent wound repair is a fundamental defense mechanism conserved in all vertebrates. Exposure of blood coagulation factor (F) VII/VIIa to cells expressing its cellular receptor and cofactor tissue factor (TF) is both necessary and sufficient to initiate blood coagulation in vivo, leading to the generation of thrombin and a fibrin clot (Figure 1a).

FVII is a zymogen of a vitamin K-dependent serine protease that is synthesized in the liver and circulates in plasma as a single-chain molecule (406 amino acids) at a concentration of approximately 0.5 μg/ml (10 nmol/L). In common with the other serine proteases of the coagulation network (FIX, FX, prothrombin, and protein C) as well as protein S and protein Z, FVII has an N-terminal domain that contains 10 glutamic acid residues that are posttranslationally modified by the addition of a carboxyl group to the γ-carbon by a vitamin K-dependent carboxylase. This γ-carboxyglutamic acid (GLA) domain confers affinity to negatively charged phospholipid membranes such as those of activated platelets, promoting the assembly of functional multiprotein complexes on these surfaces. The primary translation product of FVII (466 amino acids) contains a prepro-leader sequence consisting of a secretory leader (signal) sequence and a highly

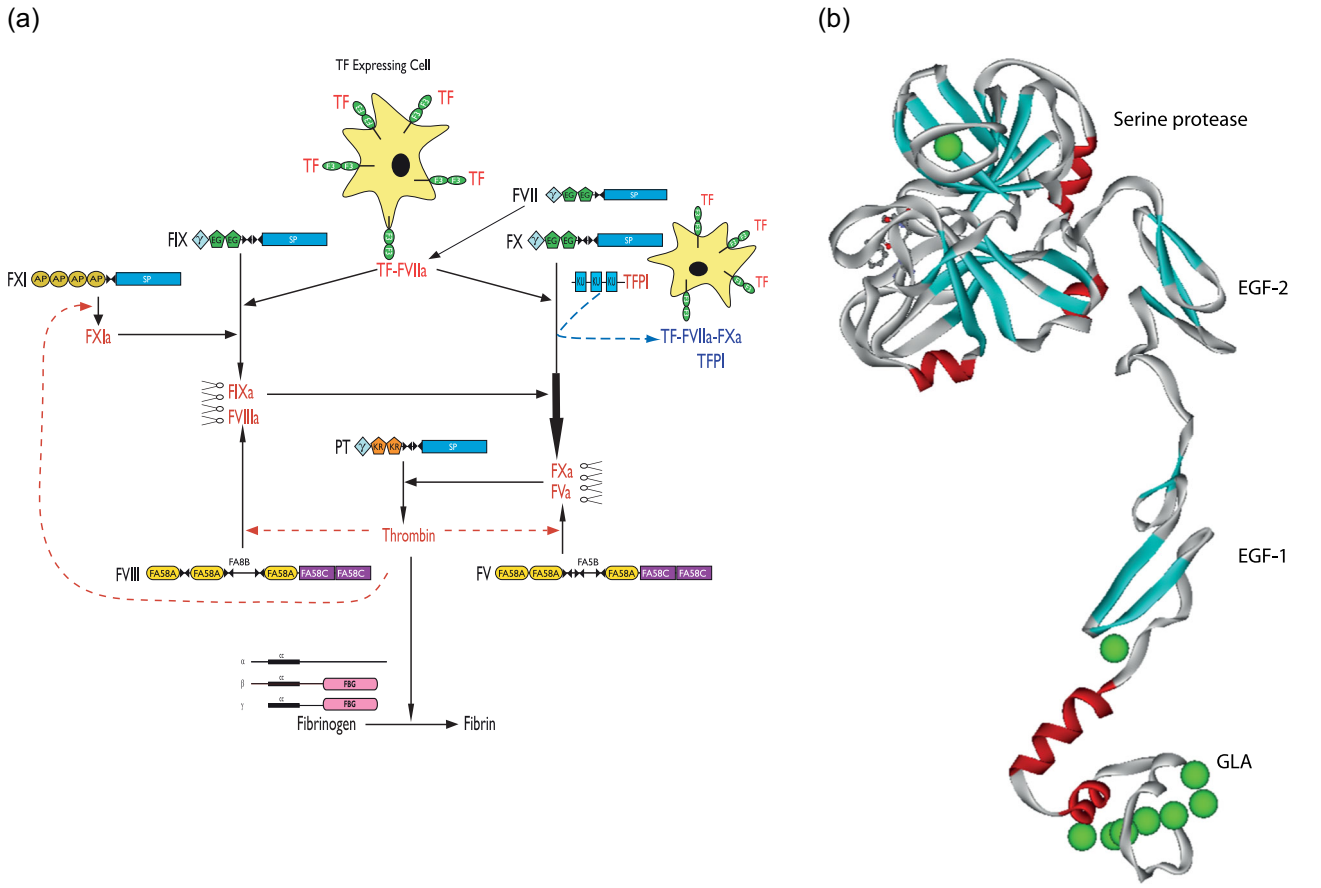


FIGURE 1 Factor VII structure and function. (a) Schematic of TF–FVIIa-initiated thrombin generation. The network includes elements of coagulation initiation via TF–FVIIa interaction, amplification via FIXa–FVIIIa and FXa–FVa complexes leading to thrombin generation, and formation of a fibrin clot. Positive feedback loops are represented by dashed lines colored red and inhibition of the TF–FVIIa–FXa complex by tissue factor pathway inhibitor by a dashed blue line. (b) Molecular model of factor VIIa derived from the complex between active site-inhibited FVIIa and the extracellular domain of tissue factor (Protein Data Bank: 1DAN). FVIIa protein chain is represented by a ribbon. The active site inhibitor is shown in a ball-and-stick model. Calcium ions are represented by green spheres. EGF, epidermal growth factor; GLA, γ-carboxylated Glu; TF, tissue factor

conserved (pro) sequence found in other vitamin K-dependent proteins that directs the γ-carboxylation. The GLA domain is followed by two epidermal growth factor (EGF)-like domains, the connecting or activation peptide, and the serine protease domain (Figure 1b).

FVII is converted to its activated form FVIIa as the result of a single proteolytic cleavage between Arg212 and Ile213 (all residues are numbered according to Goodeve, Reitsma, and McVey (2011) using Human Genome Variation Society (HGVS) nomenclature, numbering the initiation methionine of the reference protein sequence (NP_000122.1) as +1, and may differ from legacy numbering +60, producing a disulfide-linked two-chain molecule. In blood, 4% of the total circulating FVII is in the form of FVIIa that has little functional activity in the absence of its cofactor TF. Unlike other members of the trypsin superfamily, the neo-N-terminus generated upon activation of FVII fails to insert into the activation pocket leading to a nonoptimal alignment of the catalytic machinery, rendering the FVIIa "zymogen like" with significantly reduced catalytic activity. Binding of FVIIa to TF allosterically corrects this defect, transforming FVIIa into a catalytically competent enzyme. In addition, TF ensures optimal orientation and positioning of the FVIIa catalytic domain above the membrane for optimal interaction with its substrates, thereby enhancing the proteolytic activity by 10^6-fold. The substrates of the TF–FVIIa complex are blood coagulation factors FIX and FX. A schematic of TF–FVIIa-initiated thrombin generation is shown in Figure 1a.

The FVII gene (*F7*) is located on chromosome 13 (13q34) and spans 14,909 bp. *F7* lies adjacent to the factor X gene (*F10*), 2,118 bp apart, suggesting gene duplication during evolution. Three alternative messenger RNA (mRNA) transcripts have been described. The longest reference transcript (NM_000131.4) has nine exons and encodes the longest isoform composed of 466 amino acid residues including the secretory leader (1–20) and propeptide (21–60) required for appropriate γ-carboxylation of glutamic acid residues in the N-terminal GLA domain. Transcript NM_019616.4 lacks exon 2, which encodes residues 22–43 of the propeptide but encodes an identical mature FVII protein. NM_019616.4 is the most abundant and physiologically relevant mRNA transcript. Transcript NM_001267554.1 is the shortest transcript lacking exons 2, 3, and 4, thus encoding a protein lacking the critically functional GLA and EGF1 domains.

Deficiency of FVII is an autosomal recessive bleeding disorder with a highly variable phenotype that results from variants in *F7*. The clinical heterogeneity ranges from lethal to mild or even asymptomatic forms. In many cases, especially from Southern Europe and African countries, the identification of variants associated with an asymptomatic clinical phenotype arises from the use of nonhuman TF as the trigger for functional coagulation assays. The use of TF from animal sources or recombinant versions thereof may result in a discrepant in vitro result that is inconsistent with FVII activity (FVII:C) assays performed using human TF and the bleeding phenotype of the individual concerned (Bolton-Maggs, Hay, Shanks, Mitchell, & McVey, 2007). Whereas in the more common inherited bleeding disorders hemophilia A and B there is a good correlation between residual clotting activity and the severity of any associated bleeding; in FVII deficiency this relationship is less clear. This might be explained by the lack of sensitivity in assays in differentiating between a total absence of FVII:C activity from extremely low levels. Most individuals experience mild mucous membrane bleeding, menorrhagia, and postsurgical bleeding but more significant events such as life-threatening central nervous system bleeds are well recognized. Neonatal central nervous system bleeds that are often fatal are characteristic of severely affected cases and arise in individuals with variants that result in extremely low or undetectable FVII levels. This mirrors the phenotype observed in mice made null for *F7* by homologous recombination who die either from fatal intra-abdominal hemorrhage in the peripartum period or intracranial hemorrhage before the age of 24 days (Rosen et al., 1997).

The first web-based FVII variant database was established in 2001 (McVey, Boswell, Mumford, Kemball-Cook, & Tuddenham, 2001); however, availability of time and funding impacted the ability to maintain the database as well as continuity of access to a stable URL. A locus-specific database (LSDB) loses its relevance once it is not maintained and updated on a regular basis. Although a replacement *F7* variant database (umd.be/F7; Beroud et al., 2005) was developed, 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 (dbs.eahad.org) to LSDBs for genes in hemostasis (McVey et al., 2020), mirroring the data to Leiden Open Variation Databases (LOVD) thus addressing the concerns of "Yet another database?" (den Dunnen, 2018). The first interactive coagulation factor web database was developed for *F11* variants (Saunders, O'Connell, Lee, Perry, & Perkins, 2005) and the EAHAD databases build on this structure. 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. New databases for FVIII (f8-db.eahad.org) and FIX (f9-db.eahad.org; Rallapalli, Kemball-Cook, Tuddenham, Gomez, & Perkins, 2013) gene variants have evolved from previously developed single-gene variant coagulation database projects (Giannelli et al., 1990; Kemball-Cook, Tuddenham, & Wacey, 1998), incorporating new data, new analysis tools, and a new common database architecture with new interfaces and filters. The project aims to improve the quality and quantity of information available to the hemostasis research and clinical communities, thereby enabling accurate classification of disease severity to make assessments of likely pathogenicity. The databases are curated by international experts in the field who are contributing to the ClinGen (clinicalgenome.org) initiative. We now report the development of the FVII gene (*F7*) Variant Database (f7-db.eahad.org).

2 | MATERIALS AND METHODS

2.1 | Database structure

The database was built on a common architecture developed for blood coagulation variant databases, using a MySQL platform and HTML, CSS, JavaScript, Perl, and PHP interface (McVey et al., 2020). The first EAHAD database using this architecture was for *F9* variants (Rallapalli et al., 2013). The database is available at f7-db.eahad.org. The variant data in the databases are shared with LOVD; databases.lovd.nl/shared/genes/F7, which is a freely available gene-centered collection of DNA variant data and is part of the GEN2-PHEN and Human Variome projects (Fokkema et al., 2011).

2.2 | Identification of variants

Data was initially imported from the original MRC FVII mutation database (McVey et al., 2001) and the UMD-F7 mutation database (umd.be/F7; Beroud et al., 2005). Subsequently, additional variants were identified in the published literature. All variants incorporated into the new database were verified for accuracy and HGVS nomenclature was generated and checked with Mutalyzer (mutalyzer.nl/). All data referring to individual cases with variants in the *F7* database is pseudo-anonymized and no information is provided on the site that identifies individuals.

2.3 | 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, before 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. To reduce confusion, an International Society on Thrombosis and Hemostasis 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 hemostasis (Goodeve et al., 2011). All variants described in the factor VII gene (*F7*) variant database, therefore, conform to HGVS guidelines and are reported in relation to reference sequences. The longest

transcript NM_000131.4 is used to describe DNA variants with the A of the ATG initiator methionine utilized as the sequence start point +1 and similarly, the first methionine is numbered +1 in the protein reference NP_000121.1. This differs from the legacy numbering of FVII protein amino acid variants where the signal peptide and propeptide were numbered negatively, and amino acid numbering started from the beginning of the mature protein. However, to aid comparison with previous variant descriptions, both HGVS and legacy numbering is provided for each variant. Variants are described using complementary DNA (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 LOVD (lovd.nl) to which we mirror all variants described in the EAHAD database.

2.4 | In silico analyses

The database provides access to in silico analyses that may be useful in the prediction of variant pathogenicity, using open-access software packages (Align-GVGD, PolyPhen-2, SIFT, and PROVEAN). The minor allele frequency (MAF) from the Genome Aggregation Database (gnomAD; gnomad.broadinstitute.org) is also provided. Where a MAF is not presented, it 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 position of a missense variant can also be visualized if the variant appears in the molecular model based on the structure of active site-inhibited FVIIa in complex with human soluble TF (Protein Data Bank: 2A2Q). Finally, the evolutionary conservation of the amino acid sequence of the protein at the variant residue in closely related species chimpanzee, gorilla, gibbon, bushbaby, and marmoset can be inspected. Further multiple sequence alignments from more distantly related species and multiple alignments of human vitamin K-dependent coagulation factor protease domains are also available from the *AA Alignments* tab.

3 | DATABASE CONTENT

A total of 728 individual cases with plasma FVII:C levels outwith the normal range, identified by in vitro functional coagulation assays, and variants identified in their *F7* gene(s) are compiled in the FVII gene (*F7*) variant database (f7-db.eahad.org). Two hundred and twenty-one unique variants have been identified in the cohort. The database allows 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, and severity (where known). A search returns the unique variants at the selected position in the gene or protein as a list with links, where appropriate, for in-depth analysis of the variation and anonymized patient information.

3.1 | Mutational spectrum

The majority of the *F7* variants reported are small lesions including deletions (7.7%), duplications (1.8%), insertions (0.5%), indel rearrangements (1.8%) all smaller than 20 nucleotides with the exception of two large deletions; and single nucleotide substitutions (88.2%) with the majority of these being missense variants (74.3%; Figure 2a; f7-db.eahad.org/statistics.html.php). *F7* is located 2.8 kb upstream of *F10* and therefore large rearrangements often involve both genes leading to combined FVII and FX deficiency (Pavlova et al., 2015).

Variants have been identified throughout the gene with 14 identified in the 5′ untranslated and flanking region, 184 in exons, and 23 in introns. Similarly, variants have been identified in the coding sequence for all the protein domains: signal peptide (9), propeptide (9), GLA (18), EGF1 (13), EGF2 (17), activation peptide (7), and protease (116) domains (Figure 3). The 66 bp exon 2 is an alternatively used exon present in the longest reference transcript (NM_000131.4) but absent from the most abundant and

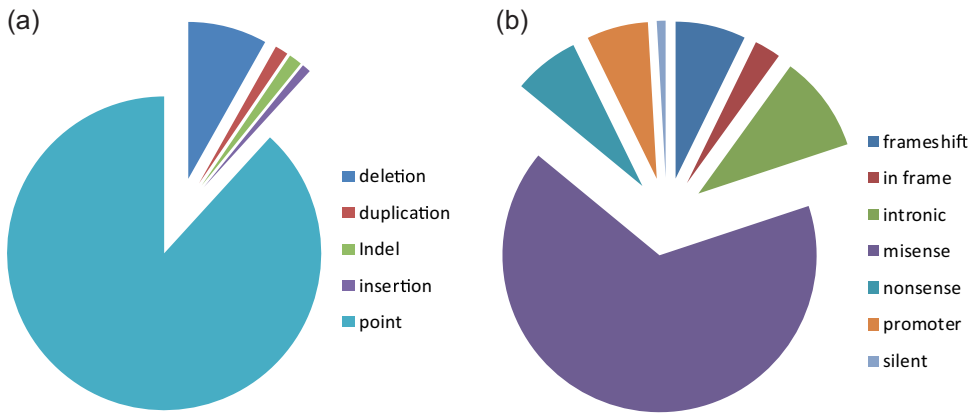
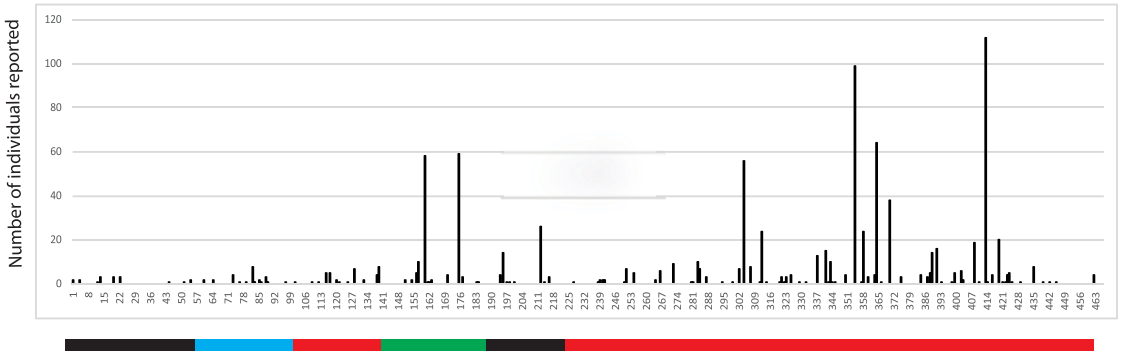


FIGURE 2 Pie charts indicating the variant type and clinical severity. (a) Variant type of the 221 unique variants in the database. (b) Classification of the 728 individual cases in the database according to clinical severity



	Pre-pro leader	Gla	EGF1	EGF2	Activation peptide	Serine protease
Number of Unique variants	18	18	13	17		116
% Unique variants	9.2	9.1	6.6	8.6		58.9

FIGURE 3 Frequency and location of the 221 unique variants. The number of individuals reported with variants in the amino acid sequence of the protein is plotted. The correspondence of the primary amino acid sequence with the domain structure of the protein is shown below with the number and percentage of unique variants reported in each domain

physiologically relevant mRNA transcript (NM_019616.3). It encodes additional amino acids in the propeptide sequence. Only two variants are described in this exon, the synonymous c.66C>T p.(Gly22=) and the c.130G>A p.(Val44Ile) variant. The increased number identified in the protease domain reflects the relative size of the coding sequence and not a higher variant rate.

Some sites, however, have multiple variants in a single codon (29 codons with 2 variants and 8 codons with 3 or more variants), for example, p.Ala251 has four different amino acid substitutions associated with variants in both c.751G and c.752C.

There are 195 single nucleotide substitutions: 144 missense, 15 nonsense, 4 silent (synonymous) within the coding sequence and an additional 10 within the 5′ flanking region located within transcription factor binding sites, 2 in the 5′ untranslated region and 20 intronic. Tools for assessing the pathogenicity of these variants are discussed below but many missense variants are predicted to result in a variety of consequences including functional and misfolding defects. The variants in transcription factor binding sites have been extensively studied and shown to result in either increased or decreased transcription. The intronic variants are predicted to result in splice variants as a consequence of various mechanisms: 10 are located in the canonical "−1, +1, −2, +2" splice nucleotides, 4 in the critical intronic "+5" position, 2 are located at less conserved positions namely, −11 and −12. In addition, a small four-nucleotide deletion within intron 8 (c.805+3_805+6del) and a synonymous variant located at the last nucleotide of exon 3 (c.291G>C or p.(Thr97=)) are also predicted to result in aberrant splicing (Millar et al., 2000; Pinotti et al., 1998).

3.2 | Impact of common variants known to modulate FVII levels

A number of common variants have been identified in *F7* (Table 1). Some of these variants, which are in strong disequilibrium with each other, have been extensively studied in relation to FVII:C and antigen (FVII:Ag) levels. The *F7* promoter haplotype [c.-401G>T, c.-325_-324insCCTATATCCT, c.-122T>C], and the missense variant c.1238G>A p.(Arg413Gln) are associated with decreased levels of FVII:C and FVII:Ag of approximately 30% and 23%, respectively (Bernardi et al., 1996) in patients and confirmed by in vitro studies (Hunault, Arbini, Lopaciuk, Carew, & Bauer, 1997; Pollak, Hung, Godin, Overton, & High, 1996). In contrast, the rare c.-402A allele is associated with increased transcriptional activity (van 't Hooft et al., 1999) resulting in significantly higher FVII:C and FVII:Ag levels than individuals homozygous for the common c.-402G allele (Lindman, Pedersen, Arnesen, Hjerkinn, & Seljeflot, 2005). Intron 8 is the location of a variable number tandem repeat (VNTR) of 37 bp with its own modulating effect. A quantitative analysis of transcripts indicated a parallel decrease of the VNTR repeat number and mRNA relative expression. Both c.795_805+26[8] and c.795_805+26[7] repeats showed higher values than the reference c.795_805+26[6], whereas c.795_805+26[5] repeats showed lower values (Pinotti et al., 2000). The in vitro results were confirmed in genetic association studies for the c.795_805+26[7] allele contributing to a 17% variance of FVII:C levels (Bernardi et al., 1996).

The impact of the common variants on the combined expression of function FVII:C activity is complex as they may be either inherited in cis or trans with the rare variant being analyzed. Often it will not be possible to assign the linkage because family studies are usually not reported. Nevertheless, it may be possible to draw conclusions from the impact of the common variants when they are present in a homozygous state. To illustrate this point, eight patients heterozygous for the p.(Ala304Val) missense variant but with various combinations of the p.(Arg413Gln) common variant were compared with each other. Four of them were homozygous for the p.(Arg413Gln) allele resulting in the p.[(Ala304Val);Arg413Gln)];[(Arg413Gln)] genotype whereas the remaining four were classified into the other possible combinations of

TABLE 1 *F7* common variants

ID	Type	Effect	Location	cDNA change	Amino acid change	MAF	Legacy nomenclature
rs510317	Point	Promoter	Flanking (5′)	c.-402G>A		0.2326	NA
rs510335	Point	Promoter	Flanking (5′)	c.-401G>T		0.2041	NA
rs5742910	Insertion	Promoter	Flanking (5′)	c.-325_-324insCCTATATCCT		0.23	WT = A1: Var = A2
rs561241	Point	Promoter	Flanking (5′)	c.-122T>C		0.1436	WT = P1: Var = P2
rs6039	Point	Intronic	Intron (1)	c.64+9G>A		0.1403992	G73A (DNA numbering) WT = G1: Var = G2
rs6042	Point	Silent	Exon (6)	c.525C>T	p.(His175=) (H115=)	0.1345885	WT = H1 (or C1): Var = H2 (or C2)
VNTR[8]	Indel	Intronic	Exon (8)	c.795_805+26[8]			NA
VNTR[5]	Indel	Intronic	Exon (8)	c.795_805+26[5]			NA
VNTR[7]	Indel	Intronic	Exon (8)	c.795_805+26[7]		0.31	Var = a (7) or h7
rs6041	Point	Intronic	Intron (8)	c.806-20G>A		0.1274271	WT = I1: Var = I2
rs6046	Point	Missense	Exon (9)	c.1238G>A	p.(Arg413Gln) (R353Q)	0.1265056	WT = M1: Var = M2

Abbreviations: cDNA, complementary DNA; MAF, minor allele frequency.

genotypes: p.[(Ala304Val;Arg413Gln)];[(Arg413=)], p.[(Ala403Val;Arg413=)];[(Arg413=)] or p.[(Ala403Val;Arg413=)];[(Arg413Gln)] according to the provided pedigree (Alshinawi, Scerri, Galdies, Aquilina, & Felice, 1998). Interestingly, both p.[(Ala304Val;Arg413Gln)];[(Arg413Gln)] and p.[(Ala403Val;Arg413=)];[(Arg413Gln)] patients with no wild-type allele remaining, displayed the lowest FVII:C levels ranging from 25% to 32% (Tamary et al., 1996) whereas the p.[(Ala304Val;Arg413Gln)];[(Arg413=)] and p.[(Ala403Val;Arg413=)];[(Arg413=)] had the highest FVII:C levels of 38–45% and 46%, respectively. The location in cis or in trans of the modulating common variants will, therefore, be provided when available for future submissions to this database.

3.3 | Individual data

In common with the other EAHAD coagulation factor databases, the *F7* database stores and displays all cases reported regardless of how many individual reports of the identical genetic variant have been reported, allowing users to survey phenotypic variability among cases with the same genetic variant. Where appropriate additional variants within an individual case are reported, including the common *F7* variants allowing analysis of combined genotypes of a rare variant and a common variant. It should be noted that in some cases, the frequency of reporting variants is biased by extensive family studies or founder effects in certain ethnic groups, for example, p.(Ala354Val) in North and North-East Europe (99 individuals) and p.(Ala304Val) in Jewish populations (38 individuals). It is possible to visualize/export the data either as a 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 variant frequency and variability of presentation. Thus, despite the wide heterogeneity, 17 variants are highly frequent. Four missense variants are found more than 50 times, namely p.(Gln160Arg), p.(Ala304Val), p.(Ala354Val) combined with p.(Pro464Hisfs*32) and p.(Arg364Gln) (also known as FVII Padua). Five missense variants are found between 20 and 50 times and nine variants between 10 and 19 times (Figure 3).

3.4 | Impact of thromboplastin source for FVII:C measurement

It is well-documented that some FVII protein variants display variations in the FVII:C measurement according to the species of the thromboplastin reagent (TF) that has been used to trigger the in vitro measurement. Historically bovine and rabbit sources were used before the introduction of recombinant human thromboplastin. The FVII:C value has, therefore, been recorded in the database with the species of the corresponding thromboplastin reagent where this information was available in the original report. Thirty-nine individuals were reported with FVII:C levels determined with thromboplastin of both human and rabbit origin. Seventeen displayed significant variations between FVII:C levels when measured with the different thromboplastins. The corresponding variants are p.(Gly391Asp), p.(Arg364Gln), p.(Arg364Trp), p.(Arg337His), p.(Arg139Gln) (Matsushita, Kojima, Emi, Takahashi, & Saito, 1994; Mourey et al., 2014; O'Brien et al., 1991; Takamiya & Takeuchi, 1998; Zheng, Shurafa, & James, 1996). Importantly, the impact of the residual FVII:C levels measured in vitro on the potential bleeding phenotype of the individual should only be considered when assayed with human thromboplastin.

3.5 | Clinical phenotype and genotype

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, available from the *Variants* tab allowing users to analyze relationships between clinical phenotype and genotype. Mariani et al. (2005) proposed a standard classification of clinical severity for FVII deficiency which we would encourage all reporting of FVII deficiency to use. However, the lack of a standard classification of the observed clinical phenotype before 2005 requires users to refer back to the original publication to ascertain the scheme used if the phenotypic severity is to be used in the interpretation of the pathogenicity of the variant.

The 728 patients present a wide range of clinical phenotypes, from asymptomatic (34.6%; often identified in preoperative coagulation screening) to individuals suffering from severe bleeds (21.2%; Figure 2b). The large majority of the 154 patients (92%) with severe bleeding tendency displayed FVII:C levels below 5%. Of those determined with human thromboplastin, the majority displayed FVII:C levels below 2.9%. Genotypically, severe patients are either homozygous or compound heterozygous for nonsense, canonical splice site, frameshift, missense variants resulting in the substitution of critical residues for the FVII function (viz., arginine residue at the proteolytic activation site (Arg212), the catalytic triad (Asp302), or residues involved in the disulfide bond (Cys195), between the light and heavy chains of the activated form of FVII), variants at transcription factor binding sites within the *F7* promoter or variants predicted to dramatically alter the protein folding namely, p.(Gln160Arg) or p.(Thr419Met). However, the converse was not always true. Of the 65 individuals homozygous for variants predicted to be pathogenic (Table 2) there were 27 reported not to have a severe bleeding tendency. Two asymptomatic patients were homozygous for the p.(Arg462*) nonsense variant occurring only four residues before the natural termination signal. The 4-residue carboxy-terminal-truncated protein has been reported to have increased specific activity and gain-of-function features (Branchini et al., 2012) explaining the observed asymptomatic phenotype. Conversely, it was unexpected that an individual homozygous for the consensus splice site c.572-1G>A was reported with only epistaxis at the age of three (Kwon, Yoo, Lee, Kim, & Kim, 2011) or an individual homozygous for the critical variant p.(Arg212Gln) reportedly with a mild bleeding phenotype despite FVII:C levels of <1% and four other cases homozygous for this variant reported as having severe hemorrhagic phenotypes. As for other clotting factor deficiencies, homozygotes for frameshift variants displayed the largest clinical heterogeneity. One explanation could be a partial correction by ribosomal slippage and DNA replication/RNA transcription errors as suggested for *F8* (Young et al., 1997).

3.6 | Assessing pathogenicity of F7 variants

Variant classification is central to the utility of molecular genetic diagnostics in clinical practice; however, predicting whether gene variants are likely to be pathogenic may not be straightforward. The *F7* variant database currently does not assign pathogenicity to an individual variant but rather provides access to a number of tools to allow assessment of the variant according to published guidelines that establish a framework for variant classification (Nykamp et al., 2017; Richards et al., 2015). In future releases, the database 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.

Evidence for pathogenicity can be obtained from the evaluation of variant frequencies from large population data sets, variant type, clinical observations, experimental studies, and computational analyses. Each variant in the *F7* database is assigned a variant ID and the variant ID view (Figure 4a) displays the cDNA change, the amino acid change, the type of variant (deletion, duplication, indel, insertion or point), the effect (frameshift, in-frame, intronic, missense, nonsense, promoter or silent), the location within the gene (untranslated region, flanking sequence, exon or intron) as well as within the protein (domain) and finally, the MAF is obtained from the gnomAD database that calculates the frequency from 125,748 exomes and 15,708 genomes.

The MAF value can provide strong evidence that a variant is benign or that it is sufficiently rare to be considered a candidate pathogenic variant. However, MAF values can sometimes be misleading. For example, two variants with high MAFs of 0.1419 and 0.1341, respectively, are p.(His175=) and p.(Arg413Gln). The high frequency and variant type would lead to the assignment of p.(His175=) as nonpathogenic. Whereas p.(Arg413Gln) is a frequent missense variant associated with decreased levels of FVII and although not directly pathogenic it can contribute to the consequence of coinherited variants.

To assess the clinical observations each variant is associated with the number of individuals reported, laboratory measurements of FVII:C (TF: unknown, human or rabbit) and FVII:Ag plasma levels, reported clinical severity as well as whether the variant is homozygous or heterozygous. Additional variants identified in each individual are also collated. For each missense variant, computational analysis using in silico tools (Align-GVGD, Polyphen-2, SIFT, and PROVEAN) to predict the likely impact on the resultant protein is provided. It is also possible to visualize the potential structural implications by viewing the variant in a molecular model based on the X-ray crystal structure of the TF–FVIIa complex or in multiple sequence alignments of the FVII amino acid sequence across primates or the serine protease domains of the human vitamin K-dependent coagulation factors (FVII, FIX, FX, protein C; Figure 4b).

4 | CONCLUSIONS

DNA sequence analysis (either classical Sanger DNA sequencing or more commonly exome/complete genome sequencing) is now

TABLE 2 Characteristics of the 65 cases homozygous for a pathogenic variant

	Variant[a]		Phenotype		
Type	cDNA numbering	Protein numbering	FVII:C (FVII:Ag)	Bleeding tendency	Number of cases
Nonsense	c.335C>G	p.(Ser112*)	<1 (ND)	Severe	1
	c.396C>A	p.(Cys132*)	<1 (ND)	Severe	1
	c.859C>T	p.(Gln287*)	<1 (ND)	Severe	3
	c.1324C>T	p.(Gln442*)	<1 (ND)	Severe	1
	c.1384C>T	p.(Arg462*)	2–3 (<1–ND)	Asymptomatic	2
Consensus splice site	c.291G>C	p.(Thr97=)	<1 (19)	Severe	1
	c.291+1G>A	NA	<1 (<1)	Severe	1
	c.291+1G>C	NA	<1 (<1)	Severe	2
	c.292-2A>G	NA	ND	Severe	3
	c.430+1G>A	NA	<1 (4 for one patient)	Severe	1
	c.572-1G>A	NA	<5 (ND)	1 Severe	2
				1 Mild	
	c.681+1G>T	NA	<1 (ND)	Severe	1
Frameshift	c.16delC	p.(Leu6Serfs*41)	<1	1 Severe	2
				1 Moderate	
	c.27_28delCT	p.(Cys10Profs*16)	<1	2 Severe	4
				2 Mild	
	c.1061C>T+c.1391delC	p.(Ala354Val) + p.(Pro464Hisfs*32)	<2 (<2–3)	5 Severe	25
				15 Moderate	
				1 Mild	
				4 Asymptomatic	
Missense variants resulting in substitution of critical residues	c.583T>C	p.(Cys195Arg)	1(<5)	Severe	2
	c.635G>A	p.(Arg212Gln)	<1 (>75)	4 Severe	5
				1 Mild	
	c.635G>A + c.416G>A	p.(Arg212Gln) + p.(Arg139Gln)	<1 (>75)	Severe	1
	c.904G>A	p.(Asp302Asn)	<1 (9)	Severe	2
Promoter transcription factor binding site	c.-94C>G	NA	<1 (<1)	Severe	1
	c.-65G>C	NA	<1 (<1)	Severe	1
	c.-61T>G	NA	<1 (<1)	Severe	1
	c.-60_-59delTT	NA	ND	Severe	1
	c.-55C>T	NA	<1 (<1)	Severe	1

Abbreviation: cDNA, complementary DNA.

[a]The variants classified as pathogenic are nonsense, canonical splice site, frameshift, missense variants resulting in the substitution of critical residues for FVII function namely, residues at the Arg212–Ile213 bond for proteolytic activation, at the catalytic site (His253, Asp302, Ser404), or residues involved in the unique disulfide bond between the light and heavy chains of the activated form of FVII (Cys195, Cys322), or variants at transcription factor binding sites within the F7 promoter (HNF4 and Sp1).

standard in the characterization of inherited disease. Most variants identified are single nucleotide variants and predicting the pathogenicity of these variants may be difficult. Deficiency of FVII is an autosomal recessive bleeding disorder with a highly variable phenotype, ranging from lethal to mild to even asymptomatic. The correlation between residual clotting activity and the clinical severity of associated bleeding is less clear than in the more common inherited bleeding disorders hemophilia A and B. This is further complicated by the use of nonhuman TF as the trigger for functional coagulation assays resulting in discrepant in vitro results that are inconsistent with FVII:C assays performed using human TF and the bleeding phenotype of the individual concerned. This has often resulted in the identification of individuals with variants in their *F7* gene who are asymptomatic. The coinheritance of common variants that can affect the expression levels of plasma FVII:Ag further complicates the analysis of potential pathogenicity.

An integrated comprehensive analysis is, therefore, key to understanding the consequence of a variant. In the *F7* variant database for any variant, it is possible to obtain the description at cDNA and amino acid (both HGVS and legacy) levels, the type of variant, and the location of the variant within the gene and protein. Each individual reported to carry the variant is displayed along with the reported clinical severity, whether the individual is heterozygous or homozygous for the variant and any other variants identified in each

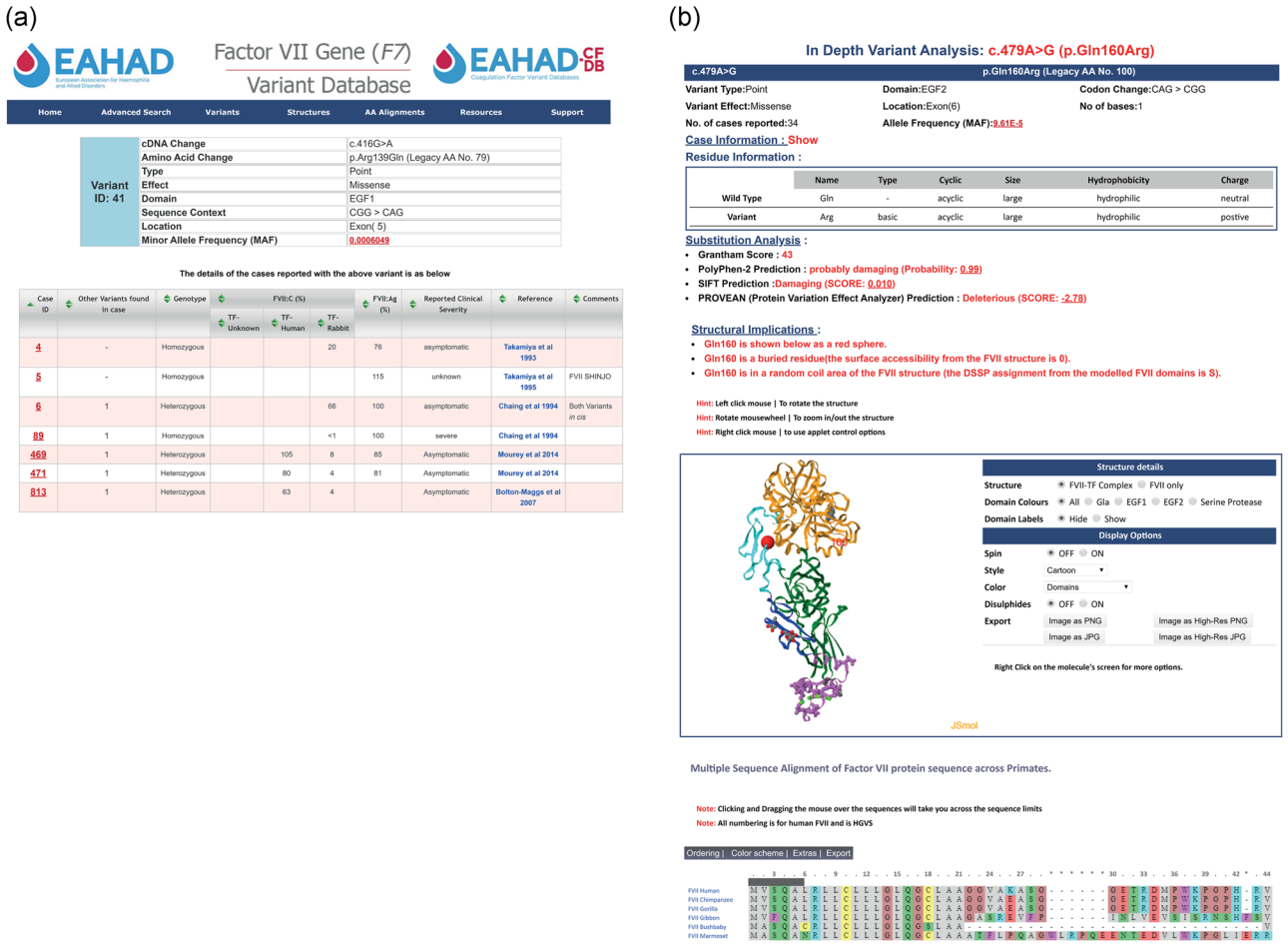


FIGURE 4 Representative screenshots of the EAHAD factor VII gene (*F7*) variant database. (a) An example of a *Variant ID* page; shown for *F7*:p.(Arg139Gln). The molecular details of the variant are listed. All cases are reported to carry the variant on at least one allele with links to the individual case details and reference. (b) An example of an *In depth variant analysis* view; shown for *F7*:p.(Gln160Arg). Substitution analysis scores are provided for Grantham (Align-GVGD), PolyPhen-2, SIFT, and PROVEAN. The location of the missense variant is indicated by a red sphere in a molecular model derived for the X-ray structure of the TF–FVIIa complex. EAHAD, European Association for Haemophilia and Allied Disorders

individual. The generation and maintenance of an online open-access database by a team of specialized curators that incorporates all described variants in *F7* and their associated properties and functional activities in addition to in-depth computational analysis of the consequence of the variant significantly aids the assessment of pathogenicity.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support of the UMD database (Beroud et al., 2005) in providing *F7* variants data from the UMD-F7 database (umd.be/F7/). The authors also acknowledge the support of the EAHAD for this project.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

DATA AVAILABILITY STATEMENT

The database is available at f7-db.eahad.org.

ORCID

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

REFERENCES

Alshinawi, C., Scerri, C., Galdies, R., Aquilina, A., & Felice, A. E. (1998). Two new missense mutations (P134T and A244V) in the coagulation factor VII gene. *Human Mutation*, *11*(Suppl 1), S189–S191.

Bernardi, F., Marchetti, G., Pinotti, M., Arcieri, P., Baroncini, C., Papacchini, M., … Mariani, G. (1996). Factor VII gene polymorphisms contribute about one third of the factor VII level variation in plasma. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *16*(1), 72–76.

Beroud, C., Hamroun, D., Collod-Beroud, G., Boileau, C., Soussi, T., & Claustres, M. (2005). UMD (Universal Mutation Database): 2005

update. *Human Mutation*, *26*(3), 184–191. https://doi.org/10.1002/humu.20210

Bolton-Maggs, P. H., Hay, C. R., Shanks, D., Mitchell, M. J., & McVey, J. H. (2007). The importance of tissue factor source in the management of factor VII deficiency. *Thrombosis and Haemostasis*, *97*(1), 151–152.

Branchini, A., Rizzotto, L., Mariani, G., Napolitano, M., Lapecorella, M., Giansily-Blaizot, M., … Bernardi, F. (2012). Natural and engineered carboxy-terminal variants: Decreased secretion and gain-of-function result in asymptomatic coagulation factor VII deficiency. *Haematologica*, *97*(5), 705–709. https://doi.org/10.3324/haematol.2011.049403

den Dunnen, J. T. (2018). Yet another database? *Human Mutation*, *39*(6), 755. https://doi.org/10.1002/humu.23429

Fokkema, I. F., Taschner, P. E., Schaafsma, G. C., Celli, J., Laros, J. F., & den Dunnen, J. T. (2011). LOVD v.2.0: The next generation in gene variant databases. *Human Mutation*, *32*(5), 557–563. https://doi.org/10.1002/humu.21438

Giannelli, F., Green, P. M., High, K. A., Lozier, J. N., Lillicrap, D. P., Ludwig, M., … Brownlee, G. G. (1990). Haemophilia B: Database of point mutations and short additions and deletions. *Nucleic Acids Research*, *18*(14), 4053–4059.

Goodeve, A. C., Reitsma, P. H., & McVey, J. H. (2011). Nomenclature of genetic variants in hemostasis. *Journal of Thrombosis and Haemostasis*, *9*(4), 852–855.

Hunault, M., Arbini, A. A., Lopaciuk, S., Carew, J. A., & Bauer, K. A. (1997). The Arg353Gln polymorphism reduces the level of coagulation factor VII. In vivo and in vitro studies. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *17*(11), 2825–2829. https://doi.org/10.1161/01.atv.17.11.2825

Kemball-Cook, G., Tuddenham, E. G., & Wacey, A. I. (1998). The factor VIII structure and mutation resource site: HAMSTeRS version 4. *Nucleic Acids Research*, *26*(1), 216–219.

Kwon, M. J., Yoo, K. Y., Lee, K. O., Kim, S. H., & Kim, H. J. (2011). Recurrent mutations and genotype-phenotype correlations in hereditary factor VII deficiency in Korea. *Blood Coagulation & Fibrinolysis*, *22*(2), 102–105. https://doi.org/10.1097/MBC.0b013e328343641a

Lindman, A. S., Pedersen, J. I., Arnesen, H., Hjerkinn, E. M., & Seljeflot, I. (2005). Association between the factor VII haplotype containing the -402A allele and myocardial infarction in a population of elderly men at high risk for coronary heart disease. *Thrombosis and Haemostasis*, *94*(1), 226–227.

Mariani, G., Herrmann, F. H., Dolce, A., Batorova, A., Etro, D., Peyvandi, F., … International Factor VII Deficiency Study Group. (2005). Clinical phenotypes and factor VII genotype in congenital factor VII deficiency. *Thrombosis and Haemostasis*, *93*(3), 481–487. https://doi.org/10.1160/TH04-10-0650

Matsushita, T., Kojima, T., Emi, N., Takahashi, I., & Saito, H. (1994). Impaired human tissue factor-mediated activity in blood clotting factor VIINagoya (Arg304 → Trp). Evidence that a region in the catalytic domain of factor VII is important for the association with tissue factor. *Journal of Biological Chemistry*, *269*(10), 7355–7363.

McVey, J. H., Boswell, E., Mumford, A. D., Kemball-Cook, G., & Tuddenham, E. G. (2001). Factor VII deficiency and the FVII mutation database. *Human Mutation*, *17*(1), 3–17. https://doi.org/10.1002/1098-1004(2001)17:1<3::aid-humu2>3.0.co;2-v

McVey, J. H., Rallapalli, P. M., Kemball-Cook, G., Hampshire, D. J., Giansily-Blaizot, M., Gomez, K., … Ludlam, C. A. (2020). The European association for haemophilia and allied disorders (EAHAD) coagulation factor variant databases: Important resources for haemostasis clinicians and researchers. *Haemophilia*, *26*(2), https://doi.org/10.1111/hae.13947

Millar, D. S., Kemball-Cook, G., McVey, J. H., Tuddenham, E. G., Mumford, A. D., Attock, G. B., … Cooper, D. N. (2000). Molecular analysis of the genotype-phenotype relationship in factor VII deficiency. *Human Genetics*, *107*(4), 327–342.

Mourey, G., Tachon, G., Pellequer, J. L., Zawadzki, C., Trossaert, M., Bertrand, M. A., … Giansily-Blaizot, M. (2014). A Haut-Doubs FVII variant depending on species-derived-thromboplastin reagent (F7:P.Arg337His). *Haemophilia*, *20*(4), e347–e349. https://doi.org/10.1111/hae.12455

Nykamp, K., Anderson, M., Powers, M., Garcia, J., Herrera, B., Ho, Y. Y., … Topper, S. (2017). Sherloc: A comprehensive refinement of the ACMG-AMP variant classification criteria. *Genetics in Medicine*, *19*(10), 1105–1117. https://doi.org/10.1038/gim.2017.37

O'Brien, D. P., Gale, K. M., Anderson, J. S., McVey, J. H., Miller, G. J., Meade, T. W., & Tuddenham, E. G. (1991). Purification and characterization of factor VII 304-Gln: A variant molecule with reduced activity isolated from a clinically unaffected male. *Blood*, *78*(1), 132–140.

Pavlova, A., Preisler, B., Driesen, J., de Moerloose, P., Zieger, B., Hutker, S., … Oldenburg, J. (2015). Congenital combined deficiency of coagulation factors VII and X—Different genetic mechanisms. *Haemophilia*, *21*(3), 386–391. https://doi.org/10.1111/hae.12604

Pinotti, M., Toso, R., Girelli, D., Bindini, D., Ferraresi, P., Papa, M. L., … Bernardi, F. (2000). Modulation of factor VII levels by intron 7 polymorphisms: Population and in vitro studies. *Blood*, *95*(11), 3423–3428.

Pinotti, M., Toso, R., Redaelli, R., Berrettini, M., Marchetti, G., & Bernardi, F. (1998). Molecular mechanisms of FVII deficiency: Expression of mutations clustered in the IVS7 donor splice site of factor VII gene. *Blood*, *92*(5), 1646–1651.

Pollak, E. S., Hung, H. L., Godin, W., Overton, G. C., & High, K. A. (1996). Functional characterization of the human factor VII 5′-flanking region. *Journal of Biological Chemistry*, *271*(3), 1738–1747. https://doi.org/10.1074/jbc.271.3.1738

Rallapalli, P. M., Kemball-Cook, G., Tuddenham, E. G., Gomez, K., & Perkins, S. J. (2013). An interactive mutation database for human coagulation factor IX provides novel insights into the phenotypes and genetics of hemophilia B. *Journal of Thrombosis and Haemostasis*, *11*(7), 1329–1340. https://doi.org/10.1111/jth.12276

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., … ACMG Laboratory Quality Assurance Committee. (2015). 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. *Genetics in Medicine*, *17*(5), 405–424. https://doi.org/10.1038/gim.2015.30

Rosen, E. D., Chan, J. C., Idusogie, E., Clotman, F., Vlasuk, G., Luther, T., … Carmeliet, P. (1997). Mice lacking factor VII develop normally but suffer fatal perinatal bleeding. *Nature*, *390*(6657), 290–294. https://doi.org/10.1038/36862

Saunders, R. E., O'Connell, N. M., Lee, C. A., Perry, D. J., & Perkins, S. J. (2005). Factor XI deficiency database: An interactive web database of mutations, phenotypes, and structural analysis tools. *Human Mutation*, *26*(3), 192–198. https://doi.org/10.1002/humu.20214

Takamiya, O., & Takeuchi, S. (1998). Dysfunctional factor VII variant (FVII Tondabayashi) with R79Q: Determination of mutated site with monoclonal anti-human factor VII antibody (B101/B1). *Clinical Chemistry*, *44*(9), 1993–1995.

Tamary, H., Fromovich, Y., Shalmon, L., Reich, Z., Dym, O., Lanir, N., … Seligsohn, U. (1996). Ala244Val is a common, probably ancient mutation causing factor VII deficiency in Moroccan and Iranian Jews. *Thrombosis and Haemostasis*, *76*(3), 283–291.

van 't Hooft, F. M., Silveira, A., Tornvall, P., Iliadou, A., Ehrenborg, E., Eriksson, P., & Hamsten, A. (1999). Two common functional polymorphisms in the promoter region of the coagulation factor VII gene determining plasma factor VII activity and mass concentration. *Blood*, *93*(10), 3432–3441.

Young, M., Inaba, H., Hoyer, L. W., Higuchi, M., Kazazian, H. H., Jr., & Antonarakis, S. E. (1997). Partial correction of a severe molecular defect in hemophilia A, because of errors during expression of the factor VIII gene. *American Journal of Human Genetics*, *60*(3), 565–573.

Zheng, D. Q., Shurafa, M., & James, H. L. (1996). Factor VII G331D: A variant molecule involving replacement of a residue in the substrate-binding region of the catalytic domain. *Blood Coagulation & Fibrinolysis*, *7*(1), 93–96.

How to cite this article: Giansily-Blaizot M, Rallapalli PM, Perkins SJ, et al. The EAHAD blood coagulation factor VII variant database. *Human Mutation*. 2020;1–11. https://doi.org/10.1002/humu.24025