Article

Whole-Exome Sequencing in a Family with an Unexplained Tendency for Venous Thromboembolism: Multicomponent Prediction of Low-Frequency Variant Deleteriousness and of Individual Protein Interaction

Barbara Lunghi [1], Nicole Ziliotto [2], Dario Balestra [1], Lucrezia Rossi [1], Patrizia Della Valle [3], Pasquale Pignatelli [4], Mirko Pinotti [1], Armando D'Angelo [3], Giovanna Marchetti [5] and Francesco Bernardi [1,*]

[1] Department of Life Sciences and Biotechnology, University of Ferrara, 44121 Ferrara, Italy; lngbbr@unife.it (B.L.); blsdra@unife.it (D.B.); lucrezia.rossi@unife.it (L.R.); pnm@unife.it (M.P.)
[2] Department of Pharmacy, University of Pisa, 56126 Pisa, Italy; nicole.ziliotto@unipi.it
[3] Unit of Coagulation Service and Thrombosis Research, IRCCS San Raffaele Hospital, 20132 Milan, Italy; dellavalle.patrizia@hsr.it (P.D.V.); dangelo.armando@hsr.it (A.D.)
[4] Department of Clinical Internal, Anesthesiological, and Cardiovascular Sciences, Sapienza University of Rome, 00185 Rome, Italy; pasquale.pignatelli@uniroma1.it
[5] Department of Neuroscience and Rehabilitation, University of Ferrara, 44121 Ferrara, Italy; mrg@unife.it
*   Correspondence: ber@unife.it; Tel.: +39-0532-974425

Abstract: Whole-exome sequencing (WES) in families with an unexplained tendency for venous thromboembolism (VTE) may favor detection of low-frequency variants in genes with known contribution to hemostasis or associated with VTE-related phenotypes. WES analysis in six family members, three of whom affected by documented VTE, filtered for MAF < 0.04 in 192 candidate genes, revealed 22 heterozygous (16 missense and six synonymous) variants in patients. Functional prediction by multi-component bioinformatics tools, implemented by a database/literature search, including ClinVar annotation and QTL analysis, prioritized 12 missense variants, three of which (*CRP* Leu61Pro, *F2* Asn514Lys and *NQO1* Arg139Trp) were present in all patients, and the frequent functional variants *FGB* Arg478Lys and *IL1A* Ala114Ser. Combinations of prioritized variants in each patient were used to infer functional protein interactions. Different interaction patterns, supported by high-quality evidence, included eight proteins intertwined in the "acute phase" (CRP, F2, SERPINA1 and IL1A) and/or in the "fibrinogen complex" (CRP, F2, PLAT, THBS1, VWF and FGB) significantly enriched terms. In a wide group of candidate genes, this approach highlighted six low-frequency variants (*CRP* Leu61Pro, *F2* Asn514Lys, *SERPINA1* Arg63Cys, *THBS1* Asp901Glu, *VWF* Arg1399His and *PLAT* Arg164Trp), five of which were top ranked for predicted deleteriousness, which in different combinations may contribute to disease susceptibility in members of this family.

Keywords: venous thromboembolism; WES analysis; family studies; low-frequency variants; protein interaction pattern; *CRP*; *F2*; *SERPINA1*; *THBS1*; *VWF*

1. Introduction

Venous thromboembolism (VTE) is a complex multifactorial disorder [1,2], in which the genetic component is estimated to have a major role [3–5]. Historically, susceptibility genes for VTE, mainly codifying for protein of coagulation cascade and its control, were identified in family and association studies [6–11]. Many "private" variants in the anticoagulation genes were found [12], which when combined with each other and/or with specific environmental or lifestyle components greatly increase the risk of developing disease [13,14]. However, the known thrombophilia variants account for a minor fraction of VTE heritability [15].

In genome-wide association studies (GWAS) the associations for part of the candidate VTE genes were successfully replicated and new susceptibility genes were suggested [16–19], although for several variants the biologic impact appear of uncertain significance. Overall, a large proportion of VTE genetic components still remain unexplained.

Part of the "missing" heritability might be due to rare variants [20]. These variants, not captured by a GWAS approach or by genotype imputation but detectable by high through-put sequencing technologies [21] might have even a larger genetic effect than common variants for VTE etiology and could be of clinical importance [21]. Indeed, the whole-exome sequencing (WES) approach in the evaluation of patients with VTE [22–25] has provided multiple novel genetic variants with predicted roles in thrombosis or thrombophilia, thus extending the panel of candidate genes.

More recently, through genomic-transcriptomic-wide analysis [26], single and multi-marker genetic testing [27] and large meta-analyzed GWAS [28], novel genetic risk modifiers for VTE have been suggested to contribute, even with small effects, to VTE susceptibility. Many of the recently identified loci, being outside of known or currently hypothesized pathways for thrombosis, suggest new molecular components belonging to platelet and blood biology, inflammation and immuno-mediated processes, potentially contributing to VTE susceptibility [29,30].

Whether WES could contribute to molecular diagnosis has been investigated in a few families with unexplained VTE and no recognized thrombophilic defects. Novel rare variants responsible for inherited thrombophilia, in the prothrombin gene [31] or outside the coagulation cascade [32,33] have been reported. Some rare variants in genes, not previously reported to be associated with VTE, and likely with an impact on the risk of VTE, were identified in two large pedigrees [34], although their relevance as novel thrombophilic defects was not confirmed.

Prompted by the small number of family studies and by the extended panel of candidate genes, WES analysis was conducted in a family with three subjects experiencing documented VTE, without being carriers of recognized thrombophilic gene variants or of anticoagulant protein deficiencies. WES analysis, combined with multiple bioinformatics approaches and public database mining, was focused on low-frequency missense variants in 192 candidate genes that have been previously suggested for their role in VTE or associated with VTE-related phenotypes.

2. Results

A schematic flow chart of methodology and data analysis is shown in Figure 1.

The family under study (Figure S1) was selected for genetic analysis by WES based on the documented venous thromboembolism in three family members, negative routine thrombophilia testing and absence of conditions (smoking, diabetes, obesity and sedentary lifestyle) that might have favored VTE.

2.1. WES Analysis

The reference list of genes (n = 192, Table S1) for the present study was generated by using as primary resource the PubMed database, for which the search terms "VTE genes" and "VTE GWAS" were inserted.

The WES analysis in the wide panel of candidate genes, performed on six family members (Figure S1), three affected (II2, II3, III2) and three unaffected (I1, II1, III1), did not reveal common thrombophilic variants or point mutations affecting main coagulation inhibitor genes (SERPINC1, PROS, PROC).

Table S2 reports missense changes with MAF 0.04–0.30 in the 1000 Genomes Project (1000G), their zygosity in patients and ClinVar annotation. Among 20 variants present in all affected subjects, the FGB Arg478Lys caused, after recombinant expression, higher clot stiffness and slower fibrinolysis rate [35], was associated with fibrinogen plasma levels and eQTLs, and displayed interaction with the relative amounts of γ' fibrinogen [36]. However, the conferred DVT/PE risk was negligible [37]. The multiple in linkage F5 SNPs have

been associated to decreased VTE risk (OR 0.77, 0.68–0.87) [38]. The LMOD1 rs2820312, in the homozygous condition in all affected subjects, belongs to novel secondary signals in previously established GWAS loci associated with BMI [39].
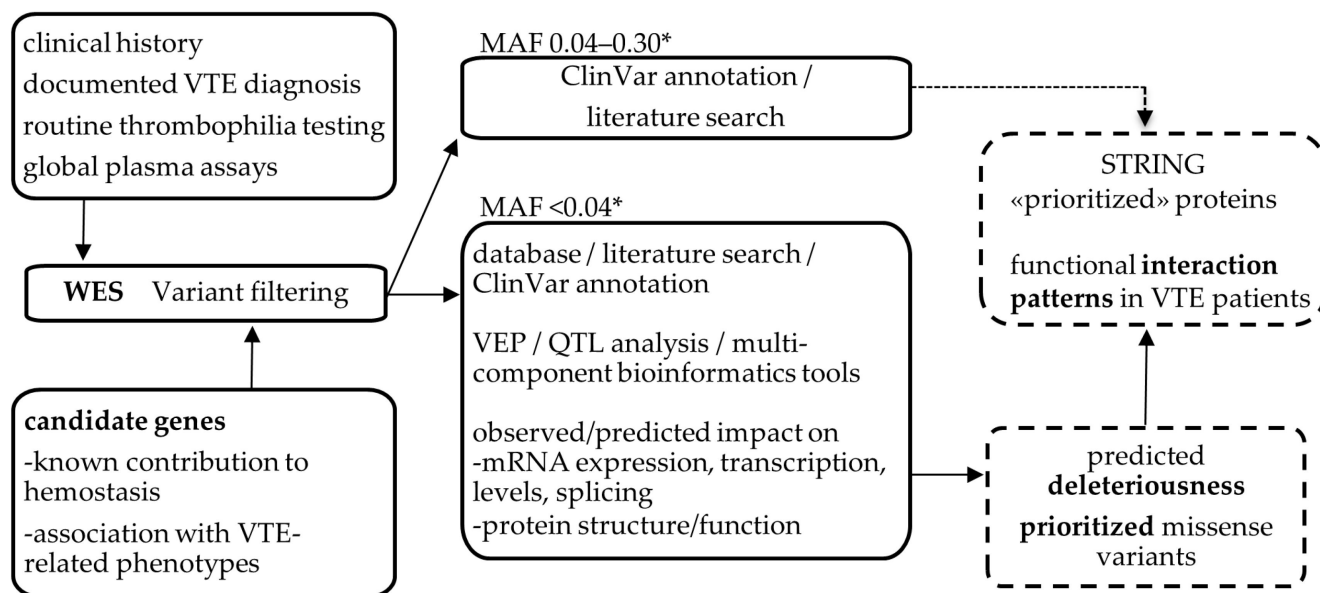

Figure 1. Overview of the family study providing essential information about the research methodology and data analysis. VTE, venous thromboembolism; WES, whole-exome sequencing; MAF, minor allele frequency; VEP, variant effect predictor; QTL, quantitative trait locus. * Allele frequency obtained from gnomAD3 and 1000G databases. The hatched boxes contain the main study output. The hatched arrow indicates the inclusion of two prioritized frequent and functional variants in the STRING analysis. In bold, key terms of the study.

Among variants present in the propositus (III2), the KNG1 Ile197Met has been associated to decreased kininogen [40] and factor XI [41] levels, and the homozygous IL1A Ala114Ser to ~50% decreased IL-1α release, and in turn with cardiovascular disease [42].

Focusing on low-frequency variants, a MAF < 0.04 was selected as threshold for filtering through the Genome Aggregation (gnomAD3) and 1000 Genomes Project (1000G) Databases. Filtering revealed 22 variants (16 missense and six synonymous) heterozygous in at least one of the affected family members, six with MAF < 0.001 (Table 1).

Twelve out of sixteen missense variants were carried by the propositus' uncle (II3), and nine by the propositus (III2) and his father (II2). All affected family members (II2, II3, III2) carried (Table 1) five missense variants (CRP Leu61Pro, F2 Asn514Lys, JAK2 Leu393Val, NQO1 Arg139Trp and the new PSG8 Cys9Ser). High frequency of the variants was excluded in other populations (Table S3).

Several low and high frequency variants were found in linkage or compound heterozygous condition. The PSG8 Cys9Ser, a newly detected missense variant that may influence signal peptide recognition (http://www.signalpeptide.de/index.php (accessed on 28 April 2023)), was heterozygous in all affected subjects, and compound heterozygous (III2) with PSG8 Gly86Ser and Ile88Arg (Table S2). The THBS1 Asp901Glu, the other newly detected missense variant (II2 and III2), was linked to Thr523Ala and Asn700Ser (Table S2). The SERPINA1 Arg63Cys (II2, III2) was found in linkage with Arg125His and Glu400Asp (Table S2). The VWF Arg1399His and Asp1472His variants were in the compound heterozygosity (II3).

Table 1. Variants (MAF < 0.04) present in at least one affected family member.

Gene Symbol	rsID dbSNP	cDNA	Protein Change	Frequency	Clinical Significance	Family Carriers
Non synonymous						
ARID4A	rs146509016	c.8C > T	p.Ala3Val	4.18819×10^{-5}	NR	II1, III1, **III2**
ARID4A	rs1051029502	c. 2847G > C	p.Met949Ile	6.97876×10^{-6}	NR	I1, **II3**
CRP	rs1376711485	c.182T > C	Leu61Pro	6.98295×10^{-6}	NR	I1, **II2**, **II3**, III1, **III2**
F2	rs199772906	c.1542C > A	p.Asn514Lys	0.000244325	NR	I1, **II2**, **II3**, **III2**
JAK2	rs2230723	c.1177C > G	p.Leu393Val	0.0134624	B	I1, **II2**, **II3**, **III2**
KLK13	rs34089525	c.325C > T	p.His109Tyr	0.0217616	NR	**II2**, **II3**, III1
NCOA1	rs150066931	c.3995A > G	p.Asn1332Ser	0.0015095	NR	I1, **II3**
NQO1	rs1131341	c.415C > T	p.Arg139Trp	0.0255606	NR	I1, **II2**, **II3**, **III2**
PEAR1	rs77795865	c.1142 C > T	p.Ser381Phe	0.0230726	NR	**II2**, **II3**
PLAT	rs2020921	c.490 C > T	p.Arg164Trp	0.012879	NR	**II3**
PSG8	-	c.26G > C	p.Cys9Ser	-	-	I1, **II2**, **II3**, III1, **III2**
SAA2, SAA2-SAA4	rs138605229	c.222A > C	p.Glu74Asp	0.00552043	NR	II1, III1, **III2**
SERPINA1	rs28931570	c.187C > T	p.Arg63Cys	0.00151399	P; LP	**II2**, III1, **III2**
THBS1	-	c.2703T > A	p.Asp901Glu	-	-	I1, **II2**, **III2**
UGT1A3	rs146461519	c.736G > T	p.Val246Leu	0.00043999	NR	**II3**
VWF	rs1800382	c.4196G > A	p.Arg1399His	0.00895051	B; LB; LP; P	**II3**
Synonymous						
KIF26B	rs201717788	c.507C > T	p.Val169=	0.00295131	NR	II1, III1, **III2**
MPL	rs544064034	c.1242G > A	p.Ser414=	6.98227×10^{-6}	LB	**II2**, **II3**, **III2**
PLCG2	rs138637229	c.1146T > C	p.Phe382=	0.0071625	B; LB	**II2**
PTGIS	rs61322884	c.531C > T	p.Tyr177=	0.0221135	NR	**II3**
SERPIND1	rs35646566	c.423G > A	p.Leu141=	0.0173203	NR	I1, **II3**
TMCO1	rs78363884	c.486C > T	p.Leu162=	0.033832	B	**II2**

The gene symbols are reported according to NCBI Gene database (https://www.ncbi.nlm.nih.gov, accessed on 28 April 2023). Gene names: *ARID4A*, AT-rich interaction domain 4; *CRP*, C-reactive protein; *F2*, coagulation factor II (prothrombin); *JAK2*, Janus kinase 2; *KIF26B*, kinesin family member 26B; *KLK13,* kallikrein-related peptidase 13; *MPL*, proto-oncogene, thrombopoietin receptor; *NCOA1*, nuclear receptor coactivator 1; *NQO1*, NAD(P)H quinone dehydrogenase 1; *PEAR1*, platelet endothelial aggregation receptor 1; *PLAT*, tissue plasminogen activator; *PLCG2*, phospholipase C gamma 2; *PSG8*, pregnancy specific beta 1-glycoprotein 8; *PTGIS*, prostaglandin I2 synthase; *SAA2, SAA2-SAA4*, serum amyloid A2, serum amyloid A4; *SERPINA1*, alpha-1 antitrypsin; *SERPIND1*, serpin family D member 1 (heparin cofactor II); *THBS1*, thrombospondin 1; *TMCO1*, transmembrane and coiled coil domains1; *UGT1A3*, UDP glucuronosyltransferase family 1 member A3; *VWF*, von Willebrand factor. Allele frequency as reported in gnomAD3 all; Clinical significance, variant's impact on disease annotated in ClinVar (NCBI resource; accessed April 2023); NR, not reported in ClinVar; B, benign; LB, likely benign; LP, likely pathological; P, pathological; **-**, no data in public databases. Family members affected by thrombosis are reported in bold.

2.2. Observed/Predicted Functional Impact of Variants

Six low-frequency variants were clinically annotated in the ClinVar, a public archive of reports of the relationships among human variations and phenotypes, with supporting evidence. Pathological phenotypes (LP and P) were reported for SERPINA1 Arg63Cys and for VWF Arg1399His, the latter with discrepant pathogenicity assessments (Table 1). Benign or likely benign phenotypes were annotated for JAK2 Leu393Val, and for three synonymous changes (MPL Ser414Ser, PLGC2 Phe382Phe and TMCO1 Leu162Leu).

The functional impact on mRNA expression (transcription and splicing processes) was investigated (Table 2). The ARID4A rs146509016 and NQO1 rs1131341 were located within a splice site region, and the NQO1 rs1131341 also in an open chromatin region. Enhancer sequences overlapped the exons containing the MPL rs544064034 and SERPIND1 rs35646566, and a binding site for the transcription factor CTCF encompassed the exon containing the PLGC2 rs138637229.

Table 2. Functional impact of variants on (m)RNA expression.

Gene	rsID_dbSNP	Variant Region Features	Splicing Process *	QTL eQTL	sQTL
ARID4A	rs146509016	missense; splice region #	ESE disruption	no data	no data
ARID4A	rs1051029502	missense	ESE disruption	no data	no data
F2	rs199772906	missense	ESE disruption	no data	no data
JAK2	rs2230723	missense	new ESS; ESE disruption; new donor	no data	no data
KLK13	rs34089525	missense	new ESS	*IGLON5*	not found
MPL	rs544064034	synonymous; enhancer	no predicted effect	no data	no data
NCOA1	rs150066931	missense	ESE disruption	no data	no data
NQO1	rs1131341	missense; splice region # open chromatin	donor site disruption	*NOB1; COG4; PDXDC2P*	*NQO1; NOB1; NPIPB14P*
PEAR1	rs77795865	missense	no predicted effect	*LRRC71*	not found
PLAT	rs2020921	missense	new ESE, new donor	*PLAT; POLB; AP3M2*	*SLC20A2; POLB*
SAA2-SAA4	rs138605229	missense	ESE disruption	ns	not found
UGT1A3	rs146461519	missense	new donor	no data	no data
PLGC2	rs138637229	synonymous ; CTCF site	new ESS	ns	not found
PTGIS	rs61322884	synonymous	ESE disruption	*SLC9A8*	not found
SERPIND1	rs35646566	synonymous; enhancer	ESE disruption	*AC000089.3*	not found
TMCO1	rs78363884	synonymous	new ESS	*RP11-466F5.10*	not found

The overlapped regulatory features were obtained from VEP, Variant Effect Predictor (Ensembl GRCh38 release 109, February 2023) and were referred to the MANE select transcripts; *, impact on splicing predicted from bioinformatics tools; # variant within the region of the splice site (1–3 bases in the exon or 3–8 bases in the intron); CTCF (CCCTC-binding factor), transcription factor; ESE, exonic splicing enhancer; ESS, exonic splicing silencer; new donor, new splice site donor. QTL, quantitative trait locus, taken from GTEx portal (Released (V8), accessed April 2023); eQTL, expression QTL (mRNA levels); sQTL: splicing QTL; no data, not investigated in GTEx portal; not found, no QTL found in GTEx portal; ns, no significant QTL reported in GTEx portal. *IGLON5*, IgLON Family member 5; *NOB1*, RNA-binding protein NOB1; *COG4*, conserved oligomeric Golgi complex subunit 4; *PDXDC2P*, pyridoxal-dependent decarboxylase domain-containing protein 2, pseudogene; *LRRC71*, leucine-rich repeat-containing protein 71; *POLB*, DNA polymerase beta; *AP3M2*, AP-3 complex subunit mu-2; *SLC20A2*, sodium dependent phosphate transporter 2; *SLC9A8*, sodium/hydrogen exchanger 8; *AC000089.3*, ribosomal protein L7a pseudogene; *RP11-466F5.10*, non-coding gene.

The disruption of a 5′ splice site was predicted for the NQO1 variant, and of an exonic splicing enhancer (ESE) for eight variants. The creation of a new exonic splicing silencer (ESS) was predicted for four variants and of new 5′ splice sites for two (Table 2).

The quantitative trait locus (QTL) analysis (Table 2, GTEx portal) supported a significant functional impact of PLAT rs2020921 on mRNA levels and of NQO1 rs1131341 on splicing [43], and on mRNA level/splicing of other genes (Table 2) encoding several proteins: (i) IGLON5, a member of immunoglobulin superfamily IgLON; (ii) NOB1, a nuclease involved in rRNA processing; (iii) COG4, a protein of the conserved oligomeric Golgi complex; (iv) AP3M2, an AP-3 complex component with a role in protein trafficking to lysosomes and specialized organelles; (v) POLB, the polymerase beta and vi) SLC20A2, a phosphate transporter. The PEAR1 rs77795865 influences mRNA levels of LRRC71, found associated with platelet phenotypes in GWAS catalog.

Concerning variants causing synonymous changes (Table 2), the PTGIS rs61322884 influences mRNA level (eQTL) of SLC9A8, which encodes a Golgi sodium–hydrogen exchanger and has been associated with chronic inflammatory [44] and coronary artery [45] diseases, in turn potentially related to VTE mechanisms. For the other synonymous variants,

analysis provided hints for their involvement in regulatory elements of splicing (ESE and ESS) and of transcription (Table 2), albeit not reflected on recognized QTL.

2.3. Predicted Impact of Variants on Protein Structure and Function

The potential impact of variants on protein structure and function and their pathogenicity were predicted [46] by using 15 multi-component bioinformatics tools, six of which (ClinPred, DANN Coding, MetaSVM, REVEL, VEST4 and FATHMM XF) were based on artificial intelligence. A profile of damaging effects, functional and disease associated, was predicted for each missense variant, referred to the main transcript (Table S4). The degree of predicted deleteriousness is schematically represented in the heat maps (Figure 2) obtained from probability scores (left panel) or from prediction scores, categorized as neutral, moderate or damaging (right panel). The probability scores of REVEL, based on a combination of scores from 13 individual tools, were used to rank variants. Variants detected in at least two thrombotic subjects, classified as damaging by at least four bioinformatics tools, and those associated with eQTL/sQTL and annotated in ClinVar, were highlighted in Figure 2. No evidence for altered post-translational modification (acetylation, phosphorylation or ubiquitination) was found (Phosphosite Plus, https://www.phosphosite.org/homeAction (accessed on 28 April 2023)).
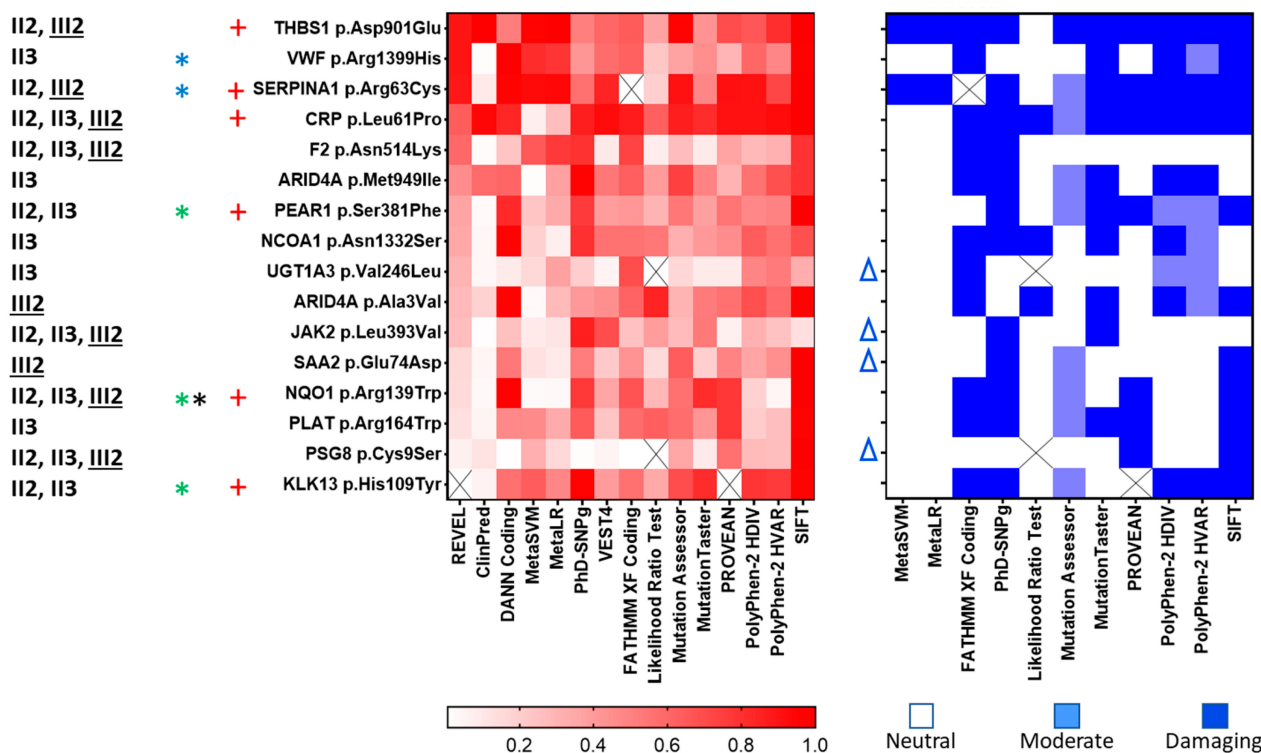

Figure 2. Heatmaps representing the degree of deleteriousness of variants. Full protein names are reported in the legend of Table 1. The 15 multi-component bioinformatics tools used for prediction, six of which exploiting artificial intelligence (ClinPred, DANN Coding, MetaSVM, REVEL, VEST4 and FATHMM XF), are reported. Variants are listed according to REVEL scores. The probability scores are reported in the left panel, according to the color scale below the panel. The right panel reports prediction scores, categorized as neutral, moderate or damaging by cut-off values within each tool. Affected family members, propositus III2 underlined, carrying the selected variants are listed on the left. +, variant detected in ≥2 thrombotic subjects and classified as «damaging» by ≥4 bioinformatics tools. *, variant associated with eQTL (green), sQTL (black), annotated in ClinVar (blue) and classified as «damaging» by ≥4 bioinformatics tools. Δ, variant non included in the interaction analysis (STRING).

Among variants ranking from the ninth to the fifteenth position, those predicted as damaging by only two tools and without additional functional correlates, were not prioritized (UGT1A3, JAK2, SAA2 and PSG8, Figure 2). Overall, 12 missense variants were prioritized, three of which (CRP Leu61Pro, F2 Asn514Lys and NQO1 Arg139Trp) were present in all affected family members.

Among prioritized variants, KLK13 His109Tyr and NQO1 Arg139Trp ranked in the last quartile of predicted deleteriousness, but were characterized by QTL associations (Table 2) with body mass index and immune system through IGLON5 [47] and protein transport and glycosylation through COG4 [48].

Plasma assays specifically designed for the protein variants might support the predicted functional impact of prioritized missense changes. Global functional assays, thrombin generation and thromboelastometry, which may partially surrogate the protein specific investigation, were performed. In thrombin generation assays induced by low tissue factor concentration, of particular interest for the F2 Asn514Lys variant present in all patients, the extrinsic thrombin potential values in the propositus (1.01 nM/min) and his father (0.9 nM/min) were indistinguishable from normal range (0.88–1.12 nM/min). In the thromboelastometry experiments, to detect "in vitro" hypo-fibrinolysis conditions of interest for VTE, the values of clotting time, maximum clot firmness, lysis onset time and lysis time in patient II3, heterozygous for the PLAT Arg164Trp variant, did not differ from normal.

2.4. Interaction among Proteins Containing Variants in the Affected Family Members

Virtually all substitutions were predicted by modelling to be located on protein surface, were not expected to cause major structural damage, and could perturb domain–domain and protein–protein interactions.

Functional association of proteins containing prioritized low-frequency missense variants were explored in the STRING database. In addition, the frequent and functional variants FGB Arg478Lys, present in all affected family members, and IL1A Ala114Ser, homozygous only in the propositus (Table S2), were also prioritized. Accordingly, Fibrinogen Beta (FGB) and interleukin 1A (IL1A) were included in the interaction analysis. Taking into account the different combinations of variants in the affected subjects, and that several variants with the highest ranking of predicted deleteriousness were not present in all affected subjects, we explored individual protein combinations and interactions in the affected family members (Figure 3).

The affected family members were characterized by five/six proteins' interactions that remarkably differed in the propositus III2 (THBS1, F2, SERPINA1, CRP, FGB and IL1A) and his father II2 (THBS1, F2, SERPINA1, CRP and FGB) as compared with paternal uncle II3 (F2, CRP, VWF, PLAT and FGB). The predicted interactions were characterized for high (FGB with THBS1/SERPINA1/VWF/CRP, F2-SERPINA1, CRP with VWF/IL1A) and very high confidence (F2 with CRP/FGB/VWF and PLAT-VWF), supported by different types of evidence (Figure 3).

Functional enrichment analysis showed that, among Biological Processes (Gene Ontology) the "acute phase response" (GO term:0006953), among Annotated Keywords (Uniprot) the "acute phase" (KW-0011) and among Compartments, the "fibrinogen complex" (GOCC:0005577), displayed highly significant enrichment. In subjects III2 and II2, the interaction pattern exhibited a balanced contribution of proteins in the "acute phase" and "fibrinogen complex", the latter prevailing in the protein interaction pattern of subject II3.
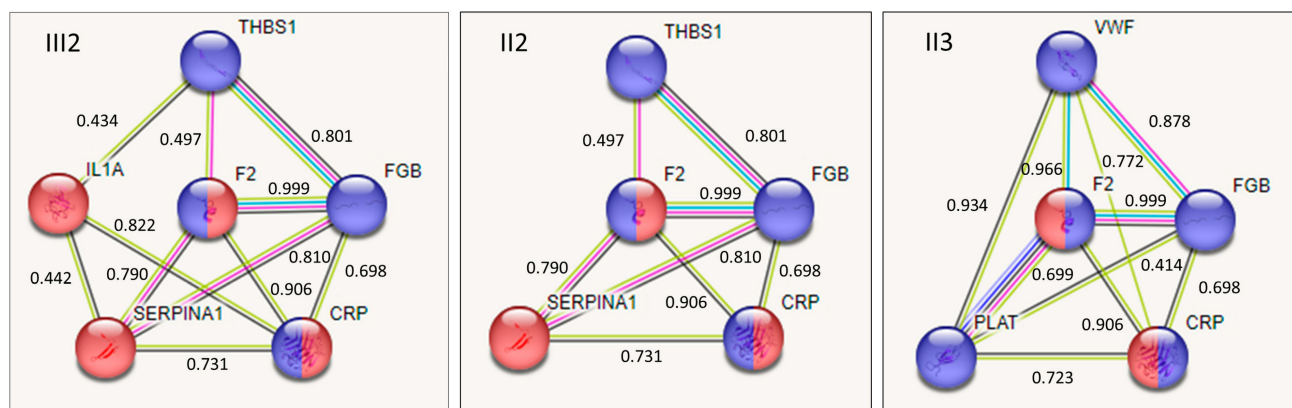
Figure 3. Functional interactions among proteins encoded by genes carrying the prioritized variants in each affected family member. Interactions predicted by STRING database (version 11.5). The protein nodes are reported with the corresponding gene symbol and filled with some known/predicted 3D structure. The proteins involved in the functional interactions are F2, prothrombin; SERPINA1, alpha-1 antitrypsin; CRP, C-reactive protein; THBS1, thrombospondin 1; PLAT, tissue plasminogen activator; VWF, von Willebrand factor; FGB, fibrinogen beta; IL1A, interleukin 1 alpha. The colored lines represent the different types of evidence used in predicting the associations (evidence mode); green line, neighborhood evidence; red line, gene fusion evidence; blue line, cooccurrence evidence; purple line, experimental evidence; light blue line, database evidence; black line, coexpression evidence. The combined score of the predicted interactions is indicated between each couple of proteins. Two significantly enriched terms are reported in the nodes: blue, proteins involved in the compartment «fibrinogen complex»; red, proteins involved in the "acute phase response". Compartment «fibrinogen complex», FDR = 4.88e-07 (III2, II2) and 2.27e-09 (II3), and UniProt Keyword «Acute phase»/GO term "acute phase response", FDR = 6.79e-05/5.09e-05 (III2), 6.79e-05/0.006 (II2) and 0.0097/not significant (II3).

3. Discussion

A family history of VTE indicates important underlying genetic components. In a family with multiple members experiencing VTE without being carriers of recognized thrombophilic defects, WES analysis was focused on low-frequency variants in multiple candidate genes. In relation to biological links to VTE, the genes carrying variants in the affected family members were grouped for belonging mainly to platelet, blood coagulation/fibrinolysis and inflammation processes (Figure 4).

In WES studies, conducted with different filtering/prioritizing strategies in families with an unexplained tendency for VTE (Table 3), nonsense changes/deletions/low-frequency-SNPs homozygosity and genetic conditions supporting contribution to diseases were sporadically detected, as well as new variants. Since missense variants were the main output of all family studies (Table 3), their prioritization is crucial to select those that could play a role in VTE etiology.

The SIFT tool, based on protein sequence homology and physical properties of amino-acids, and PhD-SNPg, a binary classifier trained and tested using >30,000 ClinVar pathogenic and benign SNVs, predicted deleteriousness for the vast majority of variants. Differently, most of the artificial intelligence-based tools, except FATHMM XF, predicted deleteriousness for only some of them. By implementing database/literature observations, functional correlates stemming from QTL analysis, prediction of splicing process involvement, and ClinVar annotation and 12 low-frequency variants, three variants present in all affected family members were prioritized together with two functional frequent variants. Since variants highest ranked for predicted deleteriousness were not present in all affected subjects, who remarkably differed in variant combination, potential protein–protein interaction were investigated in each affected family member (Figure 3). The different protein interaction patterns, supported by several pieces of high-quality evidence, included

proteins intertwined in the "acute phase" (CRP, F2, SERPINA1 and IL1A) and/or in the "fibrinogen complex" (CRP, F2, PLAT, THBS1, VWF and FGB). Interestingly, five proteins carried missense variants top ranked for predicted deleteriousness (CRP Leu61Pro, F2 Asn514Lys, SERPINA1 Arg63Cys, THBS1 Asp901Glu and VWF Arg1399His). For the SERPINA1 Arg63Cys and VWF Arg1399His variants, pathogenicity assessment was also annotated in ClinVar [49,50]. It is worth noting that VWF Arg1399His has been found to be strongly associated with deep-vein thrombosis (OR 3.26, 95% CI 1.18–8.98) [50], at variance with most of the VWF missense changes, which predict bleeding conditions. Moreover, the THBS1 Asp901Glu, a newly detected variant, and the SERPINA1 Arg63Cys, associated with reduced inhibitory activity and increased polymerization susceptibility [49], were linked to missense variants with higher frequency thus resulting in proteins with multiple changes.
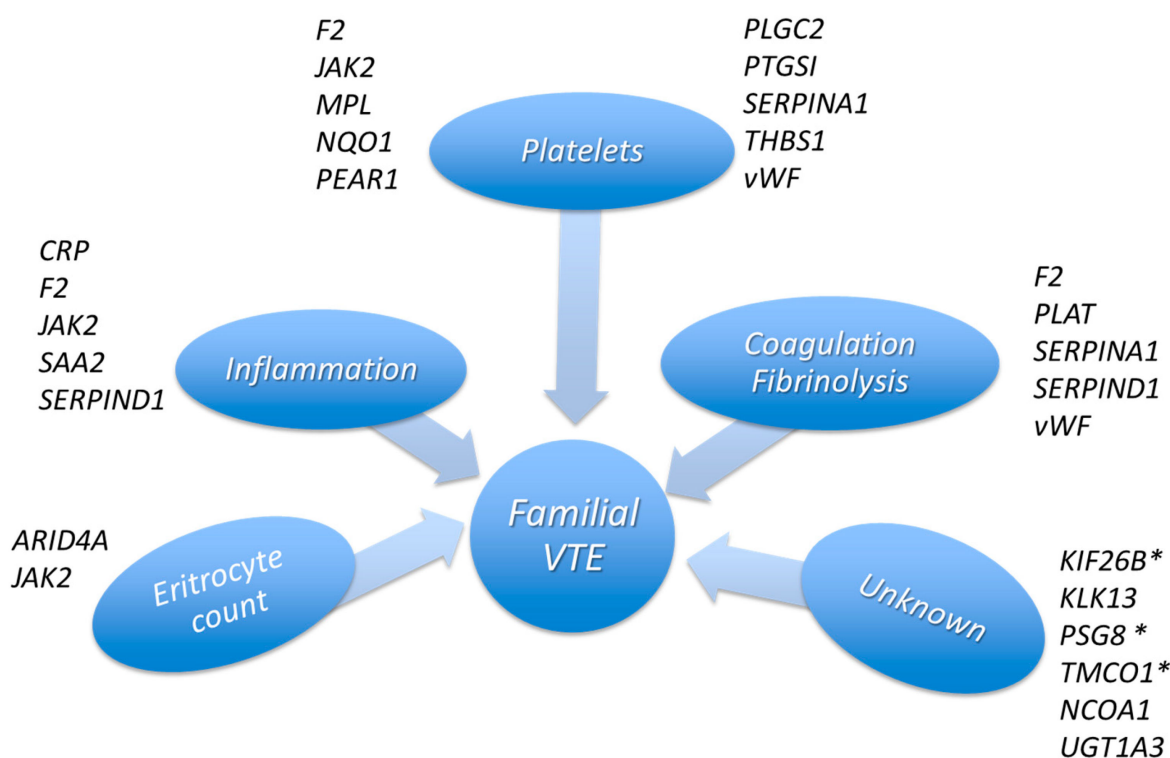

Figure 4. Genes, carrying low-frequency variants present in at least one affected family member, grouped by biological links. Genes are grouped according to biological processes or traits known to be linked to VTE. * Reported in Herrera-Rivero et al. [27] as involved in vascular repair (*KIF26B*) and platelet activation (*PSG8* and *TMCO1*).

Table 3. WES studies in families with unexplained tendency for VTE.

Reference/*Strategy*	Population	Gene	rsID_dbSNP	MAF	Variant
Cunha MLR et al., 2017 [34] - *Candidate genes (n = 126)* - *Variant MAF < 5%*	Dutch 2 Families n_a = 5 + 5	*STX2* *ITGB3* *APOH* *KLK8* *KLK11*	rs137928907 rs5918 rs4581 rs16988799 rs3745539	0.014 0.121 0.353 0.046 0.063	Phe32Val Leu59Pro Val266Leu Val154Ile Gly17Glu
Chang WA et al., 2018 [32] - *Variant MAF < 1%* - *ClinVar annotation*	Asian n_a = 3	*SLC4A1* *GP1BA*	rs121912749 rs770089708	0.000135 0.109	Gly130Arg Ser441fs

Table 3. *Cont.*

Reference/*Strategy*	Population	Gene	rsID_dbSNP	MAF	Variant
Mulder R et al., 2020 [31] *- GWLA—Pathogenicity prediction* *- Rec. proteins assays—Molec. dynamics*	Dutch $n_a = 5$	*F2*	rs886048338	0.000004	Arg541Trp
Morange PE et al., 2021 [33] *- MAF < 0.1%—Deleteriousness prediction-RNA seq in siRNAs targeted EC*	French $n_a = 4$	*MAST2*	rs1387081220	0.000004	Arg89Gln
Present study *- Candidate genes (n = 192)* *- MAF < 4%—Deleteriousness prediction—QTL analysis* *- Protein–protein interaction analysis*	Italian $n_a = 3$	*THBS1* *VWF* *SERPINA1* *CRP* *F2* *PLAT*	-- rs1800382 rs28931570 rs1376711485 rs199772906 rs2020921	-- 0.008950 0.001513 0.000007 0.000244 0.013	Asp901Glu Arg1399His Arg63Cys Leu61Pro Asn514Lys Arg164Trp

Strategies combined with WES are reported. GWLA, Genome-wide linkage analysis; Rec., recombinant; Molec., molecular; EC, endothelial cell. n_a = number of affected family members; -- no ID number. MAF, gnomAD-Exomes Global.

We can tentatively speculate that partially different mutational backgrounds and remarkably different functional interactions predisposed the brothers II2 and II3 to VTE episodes at around 60 years of age for both of them. On the other hand, these subjects were both carriers of other prioritized variants that may be associated with platelet phenotypes either directly (*NQO1* and *PEAR1*, Figure 3) [51,52], or indirectly (*PEAR1*) through mRNA level of *LRRC71* [53]. In addition, the rare *F2* Asn514Lys variant is located in thrombin, a protein with a crucial role in platelet activation. Furthermore, it can be hypothesized that trauma, a strong acquired condition, could have substantially contributed to early onset of VTE in the propositus III2.

With an approach aimed at highlighting genetic components that may confer VTE susceptibility by interaction [14], six low-frequency variants (*CRP* Leu61Pro, *F2* Asn514Lys, *SERPINA1* Arg63Cys, *THBS1* Asp901Glu, *VWF* Arg1399His and *PLAT* Arg164Trp), the first two present in all affected family members, and the first five top ranked for deleteriousness, were the main findings of this family-based WES analysis (Table 3). The frequent functional variants *FGB* Arg478Lys and *IL1A* Ala114Ser have further boosted "acute phase" and "fibrinogen complex"-related interactions, which were highly significant also without fibrinogen beta and interleukin 1 alpha proteins inclusion. Taking into account that experimental investigation of protein variant combinations is still a very difficult task, this study has several limitations: (i) The low number of family members affected by VTE and the presence of only one elderly member not affected by VTE (I1). It is worth noting that the grandfather shares with the affected sons and nephew only three out of the six low frequency variants determining the protein interaction patterns; (ii) The contribution of synonymous variants was not further explored, despite their potential involvement in regulatory elements of splicing and of transcription. However, synonymous variant predictions were not reflected on recognized QTL; (iii) Exploring around 200 candidate genes favors prediction of interactions as new combinations of small size genetic effects. However, the low number of proteins with prioritized missense variants displayed multiple interactions with high confidence scores and identified statistically significant enriched terms; (iv) The functional implication of variants was not experimentally explored. It is worth noting that the amino acid substitutions affect exposed residues and do not suggest quantitative defects of proteins. Moreover, the variants were present in the heterozygous condition, which does not favor their functional investigation and (v) Although several variants were located in genes involved in platelet biology, the study design did not include platelet functional investigation.

4. Materials and Methods

4.1. Clinical History of the Family and Laboratory Assays

Clinical history of the family members was obtained by the hospital discharge letters and by validated questionnaire [54]. The following criteria supported deep genetic analysis in the family: (i) documented pulmonary embolism episodes in the three family members associated with vein thrombosis in the three family members, which clearly define the phenotypes (see following description for each patient) under study; (ii) routine thrombophilia testing excluding the deficiencies of antithrombin, protein C, protein S, APC resistance, and high factor IX and factor VIII levels; (iii) absence of lupus anticoagulant, anti-cardiolipin-, anti-beta2GPI-(IgM, IgG) antibodies, extractable nuclear antigens and antinuclear antibodies and (iv) genetic assays excluding the factor V Leiden and prothrombin G20210A mutations.

Routine assays revealed normal platelet number in the patients. The risks conferred by smoking, diabetes, obesity and sedentary lifestyle were excluded.

The propositus (III2, Figure S1) at the age of 18 developed deep-vein thrombosis (DVT) of the right popliteal vein after trauma of the ankle during sports activity. Twenty days after trauma, a CT scan after dyspnea revealed the presence of PE at the level of the pulmonary artery extended to the segmental branches.

After hospitalization for dyspnea at the age of 59, the propositus' father (II2) was found to be affected by PE, confirmed by CT scan, and by superficial vein thrombosis at the small saphenous veins.

After hospitalization for thoracalgia at the age of 58, the propositus's paternal uncle (II3) was found to be affected by acute PE confirmed by CT scan, and by femoral vein thrombosis. The patient reported recent pneumonia.

No additional family member has been affected by VTE episodes, as derived from validated questionnaire.

4.2. Global Functional Assays

Thrombin generation induced in plasma by 3 pM of tissue factor was performed as described [55]. Thromboelastometry (ROTEM) in plasma was performed as previously described to assay exogenous activation (recombinant tissue plasminogen activator) of fibrinolysis [56].

4.3. Selection of Candidates Genes

The primary source was the PubMed database, for which search terms "venous thromboembolism (VTE)" and "GWAS" were used, and a list of 192 genes was generated for the present study.

4.4. Whole-Exome Sequencing and Analysis

Genomic DNA was extracted from peripheral blood using the Wizard Genomic DNA Purification Kit (Promega, Madison, WI, USA). WES [57,58] was performed on six individuals, three diagnosed with VTE, from an Italian family by using the SureSelect Human Exon 6 exome capture kit (Agilent Technologies) and the NovaSeq platform (Illumina, San Diego, CA, USA) with 150-bp paired-end reads. Reads were mapped against the hg19 human reference sequence using SOAPaligner. Variants calling was performed by the Complete Genomics Small Variant Caller.

Genetic variations were verified in the database of Single-Nucleotide Polymorphisms (dbSNP, http://www.ncbi.mln.nih.gov/snp (accessed on 28 April 2023)) and their frequency was verified in the 1000 Genomes Project (1000G, all) and the Genome Aggregation database (gnomAD v3.1, all).

Base calling accuracy, measured by the Phred quality score (Q score), was 98.4% for $Q > 20$ and 95.5% for $Q > 30$. Filtering of variants was based on a quality score >30 and a minor allele frequency (MAF) < 0.04 in the three affected family member.

4.5. Analysis of the Functional Impact of Variants

The annotated variant's impact on disease was verified in ClinVar (NCBI resource; accessed April 2023). The overlapped regulatory features were obtained from VEP, Variant Effect Predictor (Ensembl GRCh38 release 109, February 2023). The quantitative trait loci (QTL) analysis was obtained from GTEx portal (release V8, https://gtexportal.org/home/accessed (accessed on 28 April 2023)).

The functional impact of variants on splicing was investigated by the HOT-SKIP (https://hot-skip.img.cas.cz/ (accessed on 28 April 2023)) for the assessment of the impact on ESE and ESS, by the SpliceAI (https://spliceailookup.broadinstitute.org/# (accessed on 28 April 2023)) and SpliceRover (http://bioit2.irc.ugent.be/rover/splicerover/ (accessed on 28 April 2023)) for the impact on canonical sites.

Information on gene function, and on SNPs—associated phenotypes were obtained from Gene Cards database (https://www.genecards.org (accessed on 28 April 2023)) and GWAS catalog (https://www.ebi.ac.uk/gwas/ (accessed on 28 April 2023)), accessed April 2023.

The potential impact of variants on secondary protein structure was explored by prediction of the structural changes introduced by an amino acid substitution, conducted by exploiting the Missense3D bioinformatic tool (http://missense3d.bc.ic.ac.uk/~missense3d/ (accessed on 28 April 2023)) [59].

The potential impact of variants on protein function and their pathogenicity were predicted by 15 algorithms. All analyses were automated by exploiting the OpenCRAVAT server (https://opencravat.org/ (accessed on 28 April 2023)) [46]. The algorithms are reported with a brief description: ClinPred is an efficient tool for identifying disease-relevant nonsynonymous variants. It is based on two machine learning algorithms that use existing pathogenicity scores and, notably, benefits from inclusion of normal population allele frequency from the gnomAD database as an input feature. DANNCoding is a deep learning approach for annotating the pathogenicity of genetic variants which uses the same feature set and training data as CADD to train a deep neural network (DNN). MetaSVM is an ensemble-based prediction algorithm developed by integrating 10 component scores (SIFT, PolyPhen-2 HDIV, PolyPhen-2 HVAR, GERP++, MutationTaster, Mutation Assessor, FATHMM, LRT, SiPhy, PhyloP) and the maximum frequency observed in the 1000 genomes populations, using a support vector machine model. REVEL is an ensemble method for predicting the pathogenicity of missense variants based on a combination of scores from 13 individual tools: MutPred, FATHMM v2.3, VEST 3.0, PolyPhen-2, SIFT, PROVEAN, MutationAssessor, MutationTaster, LRT, GERP++, SiPhy, phyloP and phastCons. VEST is a machine learning method that predicts the functional significance of missense mutations based on the probability that they are pathogenic. FATHMM-XF (FATHMM with eXtended Features) is an improvement of previous predictor, FATHMM-MKL, and predicts whether single nucleotide variants (SNVs) in the human genome are likely to be functional or non-functional in inherited diseases. MetaLR is an ensemble-based prediction algorithm developed by integrating 10 component scores (SIFT, PolyPhen-2 HDIV, PolyPhen-2 HVAR, GERP++, MutationTaster, Mutation Assessor, FATHMM, LRT, SiPhy and PhyloP) and the maximum frequency observed in the 1000 genomes populations, using a logistic regression model. PhD-SNPg is a binary classifier that implements Gradient Boosting-based algorithm for predicting pathogenic variants in coding and non-coding regions. Likelihood Ratio Test can accurately identify a subset of deleterious mutations that disrupt highly conserved amino acids within protein-coding sequences by using a comparative genomics data set of 32 vertebrate species. Mutation Assessor is a database providing prediction of the functional impact of amino-acid substitutions in proteins. Functional impact is calculated based on evolutionary conservation of the affected amino acid in protein homologs.

MutationTaster evaluates disease-causing potential of sequence alterations. PROVEAN (Protein Variant Effect Analyzer, Version 1.1.5) is a software tool which predicts whether an amino acid substitution or indel has an impact on the biological function of a protein. PolyPhen-2 (Polymorphism Phenotyping v2) is a tool which predicts possible impact

of an amino acid substitution on the structure and function of a human protein using straightforward physical and comparative considerations. SIFT predicts whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids.

Protein–protein interaction was predicted by STRING database (v.11.5, https://string-db.org/, accessed on June 2023) which collects and integrates protein–protein interactions, both physical interactions as well as functional associations. In the STRING setting the network edges were reported with the evidence mode, in which the types of evidence used in predicting the associations are shown as differently colored lines. Each protein–protein interaction is annotated with a score, which is indicator of confidence, the approximate probability that a predicted link exists between two proteins. Confidence scores are first computed separately per evidence type, and then integrated into a final, "combined" confidence score. Confidence limits are as follows: low confidence, 0.15; medium confidence, 0.4; high confidence, 0.7 and highest confidence, 0.9.

In parallel with the network prediction, functional enrichment analysis was performed by STRING, which imports knowledge from the three Gene Ontology branches (Biological process, Molecular Function and Cellular Component), KEGG pathways, UniProtKB Keywords, COMPARTMENTS and TISSUES.

5. Conclusions

We believe that our approach has the potential to detect in wide groups of candidate genes and gene variations the most plausible combinations of small-size-effects variants supporting disease susceptibility in families with an unexplained tendency for VTE. Whether this strategy could be of clinical and diagnostic importance deserves further investigation in other families.

References

1. Rosendaal, F.R. Venous Thrombosis: A Multicausal Disease. *Lancet* **1999**, *353*, 1167–1173. [CrossRef] [PubMed]
2. Pastori, D.; Cormaci, V.M.; Marucci, S.; Franchino, G.; Del Sole, F.; Capozza, A.; Fallarino, A.; Corso, C.; Valeriani, E.; Menichelli, D.; et al. A Comprehensive Review of Risk Factors for Venous Thromboembolism: From Epidemiology to Pathophysiology. *Int. J. Mol. Sci.* **2023**, *24*, 3169. [CrossRef] [PubMed]
3. Zöller, B.; García de Frutos, P.; Hillarp, A.; Dahlbäck, B. Thrombophilia as a Multigenic Disease. *Haematologica* **1999**, *84*, 59–70. [PubMed]
4. Souto, J.C.; Almasy, L.; Borrell, M.; Blanco-Vaca, F.; Mateo, J.; Soria, J.M.; Coll, I.; Felices, R.; Stone, W.; Fontcuberta, J.; et al. Genetic Susceptibility to Thrombosis and Its Relationship to Physiological Risk Factors: The GAIT Study. Genetic Analysis of Idiopathic Thrombophilia. *Am. J. Hum. Genet.* **2000**, *67*, 1452–1459. [CrossRef]
5. Seligsohn, U.; Lubetsky, A. Genetic Susceptibility to Venous Thrombosis. *N. Engl. J. Med.* **2001**, *344*, 1222–1231. [CrossRef]
6. Egeberg, O. Inherited Antithrombin Deficiency Causing Thrombophilia. *Thromb. Diath. Haemorrh.* **1965**, *13*, 516–530. [CrossRef]
7. Griffin, J.H.; Evatt, B.; Zimmerman, T.S.; Kleiss, A.J.; Wideman, C. Deficiency of Protein C in Congenital Thrombotic Disease. *J. Clin. Invest.* **1981**, *68*, 1370–1373. [CrossRef]
8. Dahlbäck, B.; Carlsson, M.; Svensson, P.J. Familial Thrombophilia Due to a Previously Unrecognized Mechanism Characterized by Poor Anticoagulant Response to Activated Protein C: Prediction of a Cofactor to Activated Protein C. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 1004–1008. [CrossRef]
9. Bertina, R.M.; Koeleman, B.P.; Koster, T.; Rosendaal, F.R.; Dirven, R.J.; de Ronde, H.; van der Velden, P.A.; Reitsma, P.H. Mutation in Blood Coagulation Factor V Associated with Resistance to Activated Protein C. *Nature* **1994**, *369*, 64–67. [CrossRef]
10. Poort, S.R.; Rosendaal, F.R.; Reitsma, P.H.; Bertina, R.M. A Common Genetic Variation in the 3′-Untranslated Region of the Prothrombin Gene Is Associated with Elevated Plasma Prothrombin Levels and an Increase in Venous Thrombosis. *Blood* **1996**, *88*, 3698–3703. [CrossRef]
11. Bernardi, F.; Faioni, E.M.; Castoldi, E.; Lunghi, B.; Castaman, G.; Sacchi, E.; Mannucci, P.M. A Factor V Genetic Component Differing from Factor V R506Q Contributes to the Activated Protein C Resistance Phenotype. *Blood* **1997**, *90*, 1552–1557. [CrossRef] [PubMed]
12. Reitsma, P.H.; Rosendaal, F.R. Past and Future of Genetic Research in Thrombosis. *J. Thromb. Haemost.* **2007**, *5* (Suppl. S1), 264–269. [CrossRef]
13. De Stefano, V.; Martinelli, I.; Mannucci, P.M.; Paciaroni, K.; Chiusolo, P.; Casorelli, I.; Rossi, E.; Leone, G. The Risk of Recurrent Deep Venous Thrombosis among Heterozygous Carriers of Both Factor V Leiden and the G20210A Prothrombin Mutation. *N. Engl. J. Med.* **1999**, *341*, 801–806. [CrossRef] [PubMed]
14. Castoldi, E.; Simioni, P.; Kalafatis, M.; Lunghi, B.; Tormene, D.; Girelli, D.; Girolami, A.; Bernardi, F. Combinations of 4 Mutations (FV R506Q, FV H1299R, FV Y1702C, PT 20210G/A) Affecting the Prothrombinase Complex in a Thrombophilic Family. *Blood* **2000**, *96*, 1443–1448. [CrossRef] [PubMed]
15. Morange, P.-E.; Tregouet, D.-A. Deciphering the Molecular Basis of Venous Thromboembolism: Where Are We and Where Should We Go? *Br. J. Haematol.* **2010**, *148*, 495–506. [CrossRef] [PubMed]
16. Trégouët, D.-A.; Heath, S.; Saut, N.; Biron-Andreani, C.; Schved, J.-F.; Pernod, G.; Galan, P.; Drouet, L.; Zelenika, D.; Juhan-Vague, I.; et al. Common Susceptibility Alleles Are Unlikely to Contribute as Strongly as the FV and ABO Loci to VTE Risk: Results from a GWAS Approach. *Blood* **2009**, *113*, 5298–5303. [CrossRef] [PubMed]
17. Tang, W.; Basu, S.; Kong, X.; Pankow, J.S.; Aleksic, N.; Tan, A.; Cushman, M.; Boerwinkle, E.; Folsom, A.R. Genome-Wide Association Study Identifies Novel Loci for Plasma Levels of Protein C: The ARIC Study. *Blood* **2010**, *116*, 5032–5036. [CrossRef]
18. Heit, J.A.; Armasu, S.M.; Asmann, Y.W.; Cunningham, J.M.; Matsumoto, M.E.; Petterson, T.M.; De Andrade, M. A Genome-Wide Association Study of Venous Thromboembolism Identifies Risk Variants in Chromosomes 1q24.2 and 9q. *J. Thromb. Haemost.* **2012**, *10*, 1521–1531. [CrossRef]
19. Germain, M.; Chasman, D.I.; de Haan, H.; Tang, W.; Lindström, S.; Weng, L.-C.; de Andrade, M.; de Visser, M.C.H.; Wiggins, K.L.; Suchon, P.; et al. Meta-Analysis of 65,734 Individuals Identifies TSPAN15 and SLC44A2 as Two Susceptibility Loci for Venous Thromboembolism. *Am. J. Hum. Genet.* **2015**, *96*, 532–542. [CrossRef] [PubMed]
20. McCarthy, M.I.; Abecasis, G.R.; Cardon, L.R.; Goldstein, D.B.; Little, J.; Ioannidis, J.P.A.; Hirschhorn, J.N. Genome-Wide Association Studies for Complex Traits: Consensus, Uncertainty and Challenges. *Nat. Rev. Genet.* **2008**, *9*, 356–369. [CrossRef]
21. Trégouët, D.-A.; Morange, P.-E. What Is Currently Known about the Genetics of Venous Thromboembolism at the Dawn of next Generation Sequencing Technologies. *Br. J. Haematol.* **2018**, *180*, 335–345. [CrossRef]
22. Lotta, L.A.; Wang, M.; Yu, J.; Martinelli, I.; Yu, F.; Passamonti, S.M.; Consonni, D.; Pappalardo, E.; Menegatti, M.; Scherer, S.E.; et al. Identification of Genetic Risk Variants for Deep Vein Thrombosis by Multiplexed Next-Generation Sequencing of 186 Hemostatic/pro-Inflammatory Genes. *BMC Med. Genom.* **2012**, *5*, 7. [CrossRef]
23. Lee, E.-J.; Dykas, D.J.; Leavitt, A.D.; Camire, R.M.; Ebberink, E.; García de Frutos, P.; Gnanasambandan, K.; Gu, S.X.; Huntington, J.A.; Lentz, S.R.; et al. Whole-Exome Sequencing in Evaluation of Patients with Venous Thromboembolism. *Blood Adv.* **2017**, *1*, 1224–1237. [CrossRef] [PubMed]
24. Lindström, S.; Brody, J.A.; Turman, C.; Germain, M.; Bartz, T.M.; Smith, E.N.; Chen, M.-H.; Puurunen, M.; Chasman, D.; Hassler, J.; et al. A Large-Scale Exome Array Analysis of Venous Thromboembolism. *Genet. Epidemiol.* **2019**, *43*, 449–457. [CrossRef]

25. Desch, K.C.; Ozel, A.B.; Halvorsen, M.; Jacobi, P.M.; Golden, K.; Underwood, M.; Germain, M.; Tregouet, D.-A.; Reitsma, P.H.; Kearon, C.; et al. Whole-Exome Sequencing Identifies Rare Variants in STAB2 Associated with Venous Thromboembolic Disease. *Blood* **2020**, *136*, 533–541. [CrossRef]

26. Lindström, S.; Wang, L.; Smith, E.N.; Gordon, W.; van Hylckama Vlieg, A.; de Andrade, M.; Brody, J.A.; Pattee, J.W.; Haessler, J.; Brumpton, B.M.; et al. Genomic and Transcriptomic Association Studies Identify 16 Novel Susceptibility Loci for Venous Thromboembolism. *Blood* **2019**, *134*, 1645–1657. [CrossRef] [PubMed]

27. Herrera-Rivero, M.; Stoll, M.; Hegenbarth, J.-C.; Rühle, F.; Limperger, V.; Junker, R.; Franke, A.; Hoffmann, P.; Shneyder, M.; Stach, M.; et al. Single- and Multimarker Genome-Wide Scans Evidence Novel Genetic Risk Modifiers for Venous Thromboembolism. *Thromb. Haemost.* **2021**, *121*, 1169–1180. [CrossRef] [PubMed]

28. Thibord, F.; Klarin, D.; Brody, J.A.; Chen, M.-H.; Levin, M.G.; Chasman, D.I.; Goode, E.L.; Hveem, K.; Teder-Laving, M.; Martinez-Perez, A.; et al. Cross-Ancestry Investigation of Venous Thromboembolism Genomic Predictors. *Circulation* **2022**, *146*, 1225–1242. [CrossRef] [PubMed]

29. Zöller, B. Genetics of Venous Thromboembolism Revised. *Blood* **2019**, *134*, 1568–1570. [CrossRef] [PubMed]

30. D'Andrea, G.; Margaglione, M. Rare Defects: Looking at the Dark Face of the Thrombosis. *Int. J. Environ. Res. Public Health* **2021**, *18*, 9146. [CrossRef] [PubMed]

31. Mulder, R.; Lisman, T.; Meijers, J.C.M.; Huntington, J.A.; Mulder, A.B.; Meijer, K. Linkage Analysis Combined with Whole-Exome Sequencing Identifies a Novel Prothrombin (F2) Gene Mutation in a Dutch Caucasian Family with Unexplained Thrombosis. *Haematologica* **2020**, *105*, e370–e372. [CrossRef] [PubMed]

32. Chang, W.-A.; Sheu, C.-C.; Liu, K.-T.; Shen, J.-H.; Yen, M.-C.; Kuo, P.-L. Identification of Mutations in SLC4A1, GP1BA and HFE in a Family with Venous Thrombosis of Unknown Cause by next-Generation Sequencing. *Exp. Ther. Med.* **2018**, *16*, 4172–4180. [CrossRef] [PubMed]

33. Morange, P.-E.; Peiretti, F.; Gourhant, L.; Proust, C.; Soukarieh, O.; Pulcrano-Nicolas, A.-S.; Saripella, G.-V.; Stefanucci, L.; Lacroix, R.; Ibrahim-Kosta, M.; et al. A Rare Coding Mutation in the MAST2 Gene Causes Venous Thrombosis in a French Family with Unexplained Thrombophilia: The Breizh MAST2 Arg89Gln Variant. *PLoS Genet.* **2021**, *17*, e1009284. [CrossRef]

34. Cunha, M.L.R.; Meijers, J.C.M.; Rosendaal, F.R.; Vlieg, A.V.; Reitsma, P.H.; Middeldorp, S. Whole Exome Sequencing in Thrombophilic Pedigrees to Identify Genetic Risk Factors for Venous Thromboembolism. *PLoS ONE* **2017**, *12*, e0187699. [CrossRef]

35. Ajjan, R.; Lim, B.C.B.; Standeven, K.F.; Harrand, R.; Dolling, S.; Phoenix, F.; Greaves, R.; Abou-Saleh, R.H.; Connell, S.; Smith, D.A.M.; et al. Common Variation in the C-Terminal Region of the Fibrinogen Beta-Chain: Effects on Fibrin Structure, Fibrinolysis and Clot Rigidity. *Blood* **2008**, *111*, 643–650. [CrossRef]

36. Kotzé, R.C.; Nienaber-Rousseau, C.; De Lange, Z.; De Maat, M.P.; Hoekstra, T.; Pieters, M. Genetic Polymorphisms Influencing Total and γ' Fibrinogen Levels and Fibrin Clot Properties in Africans. *Br. J. Haematol.* **2015**, *168*, 102–112. [CrossRef]

37. Klovaite, J.; Nordestgaard, B.G.; Tybjærg-Hansen, A.; Benn, M. Elevated Fibrinogen Levels Are Associated with Risk of Pulmonary Embolism, but Not with Deep Venous Thrombosis. *Am. J. Respir. Crit. Care Med.* **2013**, *187*, 286–293. [CrossRef] [PubMed]

38. Heit, J.A.; Cunningham, J.M.; Petterson, T.M.; Armasu, S.M.; Rider, D.N.; DE Andrade, M. Genetic Variation within the Anticoagulant, Procoagulant, Fibrinolytic and Innate Immunity Pathways as Risk Factors for Venous Thromboembolism. *J. Thromb. Haemost.* **2011**, *9*, 1133–1142. [CrossRef]

39. Turcot, V.; Lu, Y.; Highland, H.M.; Schurmann, C.; Justice, A.E.; Fine, R.S.; Bradfield, J.P.; Esko, T.; Giri, A.; Graff, M.; et al. Publisher Correction: Protein-Altering Variants Associated with Body Mass Index Implicate Pathways That Control Energy Intake and Expenditure in Obesity. *Nat. Genet.* **2018**, *50*, 766–767. [CrossRef]

40. Rohmann, J.L.; de Haan, H.G.; Algra, A.; Vossen, C.Y.; Rosendaal, F.R.; Siegerink, B. Genetic Determinants of Activity and Antigen Levels of Contact System Factors. *J. Thromb. Haemost.* **2019**, *17*, 157–168. [CrossRef]

41. Sabater-Lleal, M.; Martinez-Perez, A.; Buil, A.; Folkersen, L.; Souto, J.C.; Bruzelius, M.; Borrell, M.; Odeberg, J.; Silveira, A.; Eriksson, P.; et al. A Genome-Wide Association Study Identifies KNG1 as a Genetic Determinant of Plasma Factor XI Level and Activated Partial Thromboplastin Time. *Arter. Thromb. Vasc. Biol.* **2012**, *32*, 2008–2016. [CrossRef] [PubMed]

42. Wiggins, K.A.; Pyrillou, K.; Humphry, M.; Butterworth, A.S.; Clarke, M.C. The Common IL1A Single Nucleotide Polymorphism Rs17561 Is a Hypomorphic Mutation That Significantly Reduces Interleukin-1α Release from Human Blood Cells. *Immunology* **2023**, *168*, 459–472. [CrossRef] [PubMed]

43. Lienhart, W.-D.; Strandback, E.; Gudipati, V.; Koch, K.; Binter, A.; Uhl, M.K.; Rantasa, D.M.; Bourgeois, B.; Madl, T.; Zangger, K.; et al. Catalytic Competence, Structure and Stability of the Cancer-Associated R139W Variant of the Human NAD(P)H:Quinone Oxidoreductase 1 (NQO1). *FEBS J.* **2017**, *284*, 1233–1245. [CrossRef] [PubMed]

44. Ellinghaus, D.; Jostins, L.; Spain, S.L.; Cortes, A.; Bethune, J.; Han, B.; Park, Y.R.; Raychaudhuri, S.; Pouget, J.G.; Hübenthal, M.; et al. Analysis of Five Chronic Inflammatory Diseases Identifies 27 New Associations and Highlights Disease-Specific Patterns at Shared Loci. *Nat. Genet.* **2016**, *48*, 510–518. [CrossRef]

45. van der Harst, P.; Verweij, N. Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. *Circ. Res.* **2018**, *122*, 433–443. [CrossRef]

46. Pagel, K.A.; Kim, R.; Moad, K.; Busby, B.; Zheng, L.; Tokheim, C.; Ryan, M.; Karchin, R. Integrated Informatics Analysis of Cancer-Related Variants. *JCO Clin. Cancer Inform.* **2020**, *4*, 310–317. [CrossRef]

47. Huang, J.; Huffman, J.E.; Huang, Y.; Do Valle, Í.; Assimes, T.L.; Raghavan, S.; Voight, B.F.; Liu, C.; Barabási, A.-L.; Huang, R.D.L.; et al. Genomics and Phenomics of Body Mass Index Reveals a Complex Disease Network. *Nat. Commun.* **2022**, *13*, 7973. [CrossRef]

48. Richardson, B.C.; Smith, R.D.; Ungar, D.; Nakamura, A.; Jeffrey, P.D.; Lupashin, V.V.; Hughson, F.M. Structural Basis for a Human Glycosylation Disorder Caused by Mutation of the COG4 Gene. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 13329–13334. [CrossRef]

49. Seixas, S.; Marques, P.I. Known Mutations at the Cause of Alpha-1 Antitrypsin Deficiency an Updated Overview of SERPINA1 Variation Spectrum. *Appl. Clin. Genet.* **2021**, *14*, 173–194. [CrossRef]

50. Pagliari, M.T.; Cairo, A.; Boscarino, M.; Mancini, I.; Pappalardo, E.; Bucciarelli, P.; Martinelli, I.; Rosendaal, F.R.; Peyvandi, F. Role of ADAMTS13, VWF and F8 Genes in Deep Vein Thrombosis. *PLoS ONE* **2021**, *16*, e0258675. [CrossRef]

51. Ansari, N.; Najafi, S.; Shahrabi, S.; Saki, N. PEAR1 Polymorphisms as a Prognostic Factor in Hemostasis and Cardiovascular Diseases. *J. Thromb. Thrombolysis* **2021**, *51*, 89–95. [CrossRef]

52. Chen, M.-H.; Raffield, L.M.; Mousas, A.; Sakaue, S.; Huffman, J.E.; Moscati, A.; Trivedi, B.; Jiang, T.; Akbari, P.; Vuckovic, D.; et al. Trans-Ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. *Cell* **2020**, *182*, 1198–1213.e14. [CrossRef]

53. Vuckovic, D.; Bao, E.L.; Akbari, P.; Lareau, C.A.; Mousas, A.; Jiang, T.; Chen, M.-H.; Raffield, L.M.; Tardaguila, M.; Huffman, J.E.; et al. The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell* **2020**, *182*, 1214–1231. [CrossRef] [PubMed]

54. Frezzato, M.; Tosetto, A.; Rodeghiero, F. Validated Questionnaire for the Identification of Previous Personal or Familial Venous Thromboembolism. *Am. J. Epidemiol.* **1996**, *143*, 1257–1265. [CrossRef] [PubMed]

55. Consolo, F.; Pozzi, L.; Pieri, M.; Della Valle, P.; Redaelli, A.; D'Angelo, A.; Pappalardo, F. Influence of Different Antithrombotic Regimens on Platelet-Mediated Thrombin Generation in Patients with Left Ventricular Assist Devices. *ASAIO J.* **2020**, *66*, 415–422. [CrossRef] [PubMed]

56. Panigada, M.; Zacchetti, L.; L'Acqua, C.; Cressoni, M.; Anzoletti, M.B.; Bader, R.; Protti, A.; Consonni, D.; D'Angelo, A.; Gattinoni, L. Assessment of Fibrinolysis in Sepsis Patients with Urokinase Modified Thromboelastography. *PLoS ONE* **2015**, *10*, e0136463. [CrossRef]

57. Ziliotto, N.; Marchetti, G.; Scapoli, C.; Bovolenta, M.; Meneghetti, S.; Benazzo, A.; Lunghi, B.; Balestra, D.; Laino, L.A.; Bozzini, N.; et al. C6orf10 Low-Frequency and Rare Variants in Italian Multiple Sclerosis Patients. *Front. Genet.* **2019**, *10*, 573. [CrossRef] [PubMed]

58. Scapoli, C.; Ziliotto, N.; Lunghi, B.; Menegatti, E.; Salvi, F.; Zamboni, P.; Baroni, M.; Mascoli, F.; Bernardi, F.; Marchetti, G. Combination of Genomic and Transcriptomic Approaches Highlights Vascular and Circadian Clock Components in Multiple Sclerosis. *Int. J. Mol. Sci.* **2021**, *23*, 310. [CrossRef]

59. Ittisoponpisan, S.; Islam, S.A.; Khanna, T.; Alhuzimi, E.; David, A.; Sternberg, M.J.E. Can Predicted Protein 3D Structures Provide Reliable Insights into Whether Missense Variants Are Disease Associated? *J. Mol. Biol.* **2019**, *431*, 2197–2212. [CrossRef]