CHARLES UNIVERSITY

FACULTY OF SCIENCE

Anna Gregušová

EVALUATION OF PROGRAMS FOR PREDICTING THE CLINICAL SIGNIFICANCE OF MUTATIONS IN FIBRINOGEN

Bachelor thesis

Supervisor of the bachelor thesis: Mgr. Žofie Sovová, PhD.

Study programme: Bioinformatics

Praha 2024

Dedication

I would like to thank my supervisor Žofie Sovová for guidance, feedback, and invaluable advice during the writing of this thesis. I would also like to thank my mum, dad, and brother for the support, and encouragement, not only throughout my three years of study. I am also thankful to my classmates namely Anetka, Mara, Tomáš and Tom for their help and motivation. Last but not least, I would like to thank my best friends, Justýnka, Evička, and Terezka for enduring the friendship during the hardest times.

Prohlašuji, že jsem předkládanou bakalářskou práci vypracovala samostatně za využití zdrojů uvedených v seznamu použité literatury a na základě konzultací se svým vedoucím práce.

V Praze dne 28. 4. 2024 Anna Gregušová

Abstract

Keywords: Mutations in proteins, missense mutations, prediction reliability, Fibrinogen, *in silico* prediction of pathogenicity

This bachelor's thesis compares the reliability of predicting the pathogenicity of mutations in the γ chain of fibrinogen using various predictive algorithms. The work is designed as a "blind study," where we investigate the potential pathogenicity of 70 missense mutations described in the literature. To analyse the reliability of the individual programs, we used statistical metrics, specifically sensitivity, specificity, accuracy, and Matthews correlation coefficient. We tested the following programs PANTHER-PSEP, PMut, SNPs&GO, PhD-SNP, SIFT, Mutation Taster, PolyPhen2, and Provean. These are introduced in the introductory theoretical part, which also describes fibrinogen, its role in blood clotting, and diseases related to mutations in fibrinogen. The comparison showed that the quality of predictions by the various programs differs significantly. Programs more reliable predict pathogenic than benign mutations.

Abstrakt

Klíčová slova: Mutace v proteinech, Missense mutace, spolehlivost předpovědi, Fibrinogen, *in silico* předpověď patogenicity

Tato bakalářská práce porovnává spolehlivost předpovědi patogenicity mutací v γ řetězci fibrinogenu pomocí různých předpovědních algoritmů. Práce je koncipována jako tzv. “slepá studie”, kde zkoumáme potenciální patogenicitu 70 missense mutací popsaných v literatuře. Pro analýzu spolehlivosti jednotlivých programů jsme použili statistické metriky, konkrétně citlivost, specificitu, přesnost a Matthewsův korelační koeficient. Jednotlivé testované programy, jmenovitě PANTHER-PSEP, PMut, SNPs&GO, PhD-SNP, SIFT, Mutation Taster, PolyPhen2 a Provean jsou představeny v úvodní, teoretické, části, která dále seznamuje čtenáře s fibrinogenem, jeho rolí při srážení krve a s chorobami souvisejícími s mutacemi ve fibrinogenu. Porovnání ukázalo, že se kvalita předpovědí pomocí jednotlivých programů velmi liší. Programy spolehlivěji předpoví patogenní než benigní mutaci.

Table of contents

1 Introduction

When you get a cut or injury, your body's ability to stop bleeding is crucial. That's where fibrinogen, a key protein in blood clotting, comes into play. Without this essential protein, you'd be at risk of excessive bleeding even from minor cuts. After injury of a blood vessels, fibrinogen, circulating in your blood, transforms into its activated form, fibrin, that forms a mesh-like structure, a fibrin clot, that participates in blood clot formation and stops the flow of blood from a wound. The blood clot formed is eventually dissolved by the fibrinolytic system, highlighting the essential role of fibrinogen not only in clot formation but also in the regulated process of clot resolution, ensuring that bleeding is effectively stopped while avoiding the risks of excessive clotting. Under pathophysiological conditions, a blood clot may form in an intact vessel, restraining a blood flow. This condition is known as thrombosis. Additionally, the clot or fragments of it may detach from the vessel wall and circulate within the bloodstream, a process known as embolism. Thromboembolic events may also result from impaired fibrinolysis.

Like all proteins, fibrinogen is a subject of mutagenesis, a process where a genetic information of an organism is altered by changes in DNA, mutations. These mutations can either be pathogenic, causing disease, or benign, having no effect on health. Understanding the impact of these mutations is vital, as doctors need to determine whether a mutation found in a patient is related to their health condition. One of the ways how to guess the potential impact of a mutation on one's health, especially some not characterized in literature, is to use computational predictors of variant's pathogenicity. There is plenty of *in silico* variant effect predictors reported in literature. These tools use different strategies to predict the mutation's pathogenicity and they differ in reliability of prediction. Although there are some benchmark studies reported in literature, none of them (to my best knowledge) deals with fibrinogen or its related proteins.

This thesis tests the reliability of prediction of 9 prediction tools on mutations in fibrinogen γ chain and reveals their variable reliability The results are useful for assessing whether a new mutation is pathogenic, especially when predictions about it conflict. Understanding these dynamics is essential for creating better diagnostic tools and treatment strategies, highlighting the importance of studying fibrinogen and its variants.

1.1 What is fibrinogen?

Fibrinogen is a glycoprotein present in vertebrate blood plasma, essential for hemostasis. It participates, inter alia, in wound healing, inflammation, tumorigenesis, and atherosclerosis, and its level changes significantly during pregnancy, reflecting its involvement in the increased coagulation activity essential for pregnancy maintenance (Hansen et al., 2011). In hemostasis, fibrinogen contributes to the aggregation of activated platelets and the formation of an insoluble blood clot by transforming into its active form fibrin. Fibrin clot is subsequently dissolved by the fibrinolytic system. Additionally, besides coagulation cascade fibrinogen is involved in various biochemical cascades such as platelet aggregation, and the fibrinolytic system.

1.2 Biosynthesis of fibrinogen

Fibrinogen is synthesized in the liver from three closely linked genes *FGA*, *FGB,* and *FGG* located on the long arm of human chromosome 4 (Harris, n.d.) (Platè et al., 2008), that are thought to have arisen through gene duplication (Thromb & Biol, 2017). The gene encoding the fibrinogen Aα chain (*FGA*) is 7.6 kb in size and consists of 6 exons. It translates into two proteins of 644 (major form) and 866 amino acids. The minor form of the Aα chain occurs only in 1 to 2 % of the fibrinogen molecules and it contains a fibrinogen related domain (FReD) at its C-terminus. Its molecular mass is 420 kDa, which is why this form is known as fibrinogen-420 (Fu & Grieninger, 1994). The gene for the Bβ chain (*FGB*) occupies an area of 8 kb, represents 8 exons, and its nascent form comprises 491 amino acids. The last γ chain (*FGG*) covers an area of 8.5 kb, contains 10 exons, and exists in two forms of 437 (major form) and 453 amino acids (nascent chain) (Tiscia & Margaglione, 2018). The minor form of the γ chain, reported as γ', is found in approx. 8 to 15% of the plasma fibrinogen molecules (Kattula et al., 2017). In this form, the four C-terminal amino acids of the major form are altered, and it is extended by additional 16 amino acids at the C-terminus.

Each gene is individually transcribed in the nucleus of the hepatocyte into mRNA and then translated to produce nascent protein including a signal peptide, which is removed from each chain as they move into the endoplasmic reticulum (ER) of the cell (Casini et al., 2021). The signal peptide of the Aα chain contains 19 amino acids, a signal peptide of the Bβ chain has 30

amino acids and the signal peptide of the γ chain has 26 amino acids (Haryadi et al., 2015). In the ER, the assembly starts with Aα-γ and Bβ-γ dimers formation. These intermediates further bind the missing chains and dimerize (Redman & Xia, 2001). It proceeds to Golgi, where it is post-translationally glycosylated and the 15 C-terminal amino acids are cleaved from the Aα chain (Suskiewicz, 2024). Next, it is directed to the extracellular pathway. Misfolded, misassembled, and excess proteins are retained in the ER, ultimately degraded by lysosomes and proteasomes (Xia & Redman, 1999). Aberrant molecules with the mutation in the last exon may escape this mechanism and be secreted into blood (B. D. Wang & Lee, 2018).

Mutations in fibrinogen may result in defects in protein synthesis, folding, assembly, and secretion, therefore the molecules having such mutation are not released into the blood. Most of such molecules are degraded, although some mutations in the C-terminal region of Aα chain and FReD domain of γ chain may be retained in the liver. Such conditions are referred to as hereditary fibrinogen alpha chain amyloidosis (AFib) and fibrinogen storage disease, respectively.

1.3 Structure of fibrinogen

The mature fibrinogen molecule is in its major form a 340-kDa (2964 amino acids) glycoprotein consisting of 2Aα, 2Bβ, and 2γ homologous polypeptide chains with molecular masses of 66,5 (610 amino acids), 52 (461 amino acids), and 46,5 (411 amino acids) kDa, respectively (Weisel & Litvinov, 2017)(Weisel, 2005). The disordered N-termini of the three chains are followed by α-helices, which, within a AαBβγ trimer, make a parallel triple coiled-coil domain. The Aα chain bends towards the N-terminus of the molecule, and following a short α-helix, the C-terminus of the molecule is mainly disordered(Kollman et al., 2009). The C-termini of the Bβ and γ chain, and the C-terminal extension of the minor of the Aα chain, make a fibrinogen-related domain (FReD). The FReD is dominated by a central anti-parallel β-sheet. It also contains short α-helices and a loop region, that participates in fibrin polymerization. The binding of Ca2+ is essential for the domain's stability (Doolittle et al., 2012). The C-terminal extension of the minor form of the γ chain is supposedly disordered. All six chains are connected by their N-termini to the central part of the molecule (E domain) (Weisel & Litvinov, 2017) and linked by 29 disulphide bridges (Thromb & Biol, 2017) see Figure 1. Fibrinogen structure has an elongated shape – 45 nm in length and ~ 2–5 nm in diameter (Harris, n.d.). The complexity of fibrinogen structure is further increased by the co-

and post-translational N-glycosylation to the Bβ, γ and the minor form of Aα chains, enhancing its total molecular mass. (Harris, n.d.).

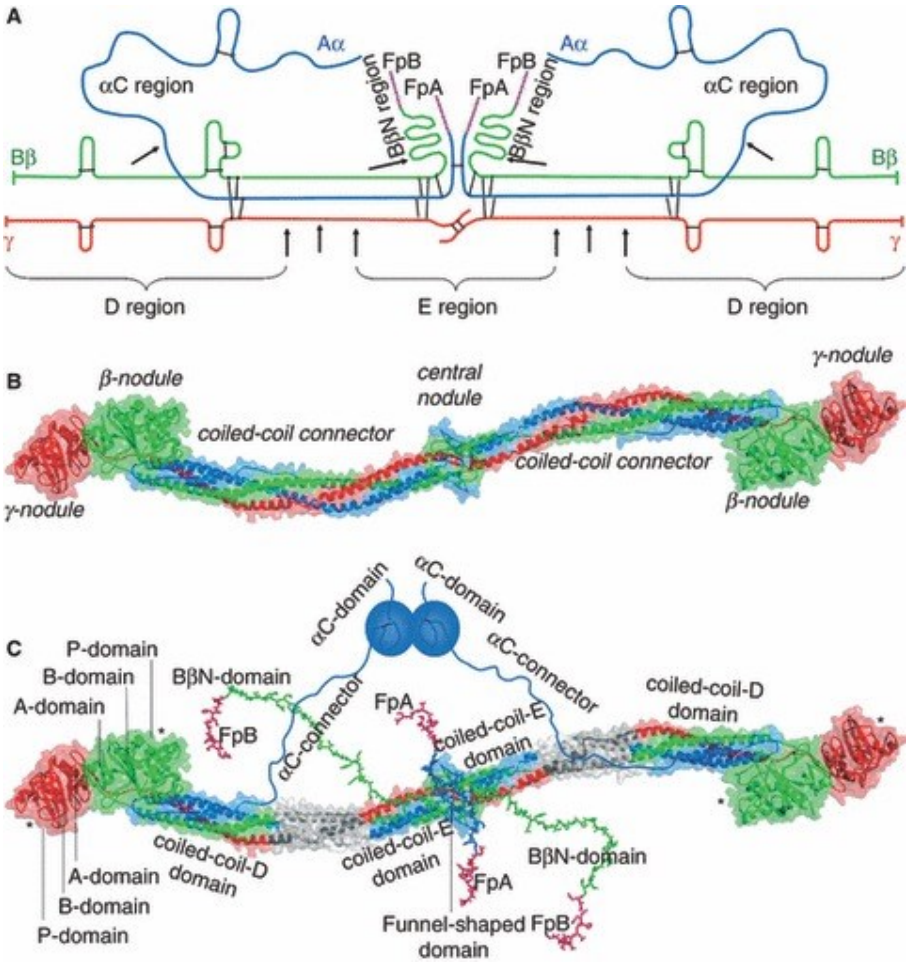

Figure 1 Structure of fibrinogen

(A) The composition of the polypeptide chain, with Aα shown in blue, Bβ in green, and γ in red. The FpA/B regions are illustrated in purple, black stripes represent disulphide bonds, and three arrows point to the site of proteolytic cleavage between the D and E regions. A single arrow indicates the cleavage that leads to the removal of the αC and BβN regions.

(B) The crystal structure of fibrinogen.

(C) The same molecule as in (B), but also showing interacting αC domains attached to α-connectors; N-terminal parts of Bβ chains forming functional BβN domains; the γN domain is located on the opposite side of the molecule, hence it is not visible here; "a" and "b" binding sites in the P domain of γ and β nodules are marked with asterisks; the cleavage site between D and E regions is shown in gray (Medved & Weisel, 2009).

1.4 Cell-based model of coagulation

The coagulation is a sequence of proteolytic events resulting in the formation of fibrin, an activated form of fibrinogen, that forms a fibrin clot. This process begins after an injury to the vessel wall and starts by aggregation and activation of platelets at the site of injury and release of von Willebrand factor (vWF) and factor V (FV), among others. Simultaneously, a protein called tissue factor (TF) is exposed to the bloodstream. TF forms a complex with factor VII (FVII), activating it to VIIa (FVIIa). The TF-FVIIa complex activates factors IX (FIX) and X (FX). Activated FX (FXa) activates and forms a complex with FV. This complex turns only a small amount of inactive prothrombin into activate thrombin. This is the end of the initiation phase, which occurs on the surface of TF-bearing cells, while the other two phases take place at the surface of activated platelets.

In the amplification phase, thrombin, which was produced in small amounts in the initial phase, activates additional platelets and increases the exposure of clotting factor receptors. This phase is marked by the activation of factor XI (FXI) on the platelet surface amplifying the clotting response. The generated thrombin also dissociates the factor VIII-vWF complex, leading to FVIII activation, and partially activates FV, helping more platelet adhesion and aggregation.(Green, n.d.)(Palta et al., 2014).

During the propagation phase, thrombin generation is maximized on the surface of activated platelets. The FVIIIa binds FIXa, that was activated by FXIa, and calcium ions, forming a "tenase" complex, that activates additional FX. Next, the "prothrombinase" complex comprising of FXa, FVa, and calcium ions is formed. Thrombin cleaves N-terminal fibrinopeptides from Aα and Bβ chains converting them into fibrin (Hoffman, 2003). Fibrinopeptide A is released after the cleavage of AαArg16-Gly17 and fibrinopeptide B is released when the BβArg14-Gly15 bond is cleaved. Mutations in any of these residues, especially the arginines do not let the fibrin form, which restrains fibrin clot formation and consequently results in dysfibrinogenemia (Hanss & Biot, n.d.).

1.4.1 From Fibrinogen to Fibrin clot

The conversion of fibrinogen to fibrin is the final stage in the complex process of blood coagulation (*Scheraga1957*, n.d.). It starts by thrombin-mediated cleavage of fibrinopeptides

A and B from the N-termini of fibrinogen chains Aα and Bβ. Fibrin first polymerizes into two-stranded, half-staggered protofibrils, those, after reaching a certain length, start laterally aggregate, forming fibrin fibres, and branches. This results in a sponge-like structure, and a fibrin clot (Duval et al., 2014).

Fibrin clot is stabilized by a cross-linking of fibrin α- and γ-chains by activated factor XIII (FXIIIa) (Duval et al., 2014). FXIIIa binds fibrin through an interaction with the C-terminal domain of the γ chain (Duval et al., 2014). Specifically, α-chain cross-linking leads to the thickening of fibrin fibres, enhancing clot stiffness, and reducing the rate of clot lysis. In contrast, γ-chain cross-linking determines the appearance time of fibrin fibres and their density within the clot (Duval et al., 2014). This process occurs at different rates, with γ-chains being more rapidly cross-linked in the early stages of clot formation, while α-chain cross-linking occurs at a slower pace (Duval et al., 2014). Cross-linking significantly influences the structural and functional properties of the clot, contributing to clot stiffness and influencing its resistance to fibrinolysis (Standeven et al., 2007). It is obvious, that intact binding sites of FXIIIa, as well as cross-linking sites, are necessary for proper fibrin polymerization and their mutations may be pathological.

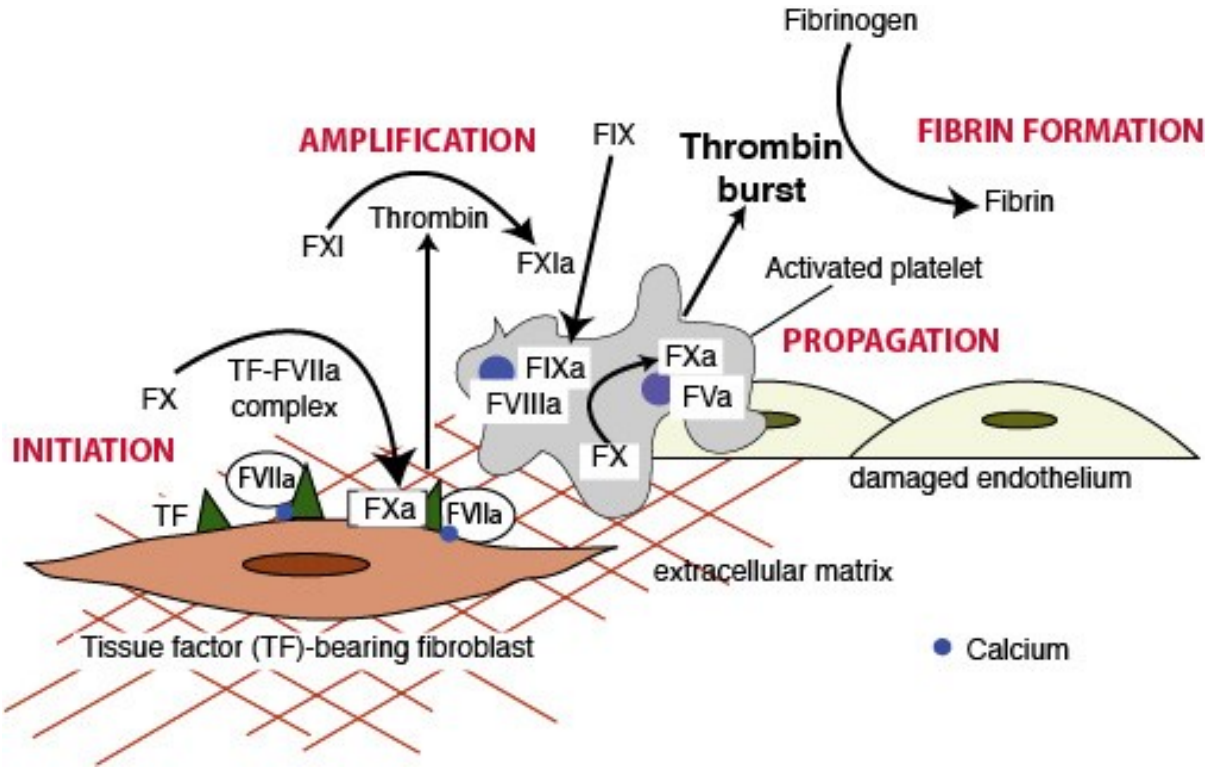

Figure 2 Cell-based model of coagulation image

The figure describes the coagulation cascade in three stages: initiation, where tissue factor (TF) and Factor VII activate Factor X, amplification, where thrombin enhances the activation of Factor IX and Factor XI and propagation, where there is a thrombin burst that converts fibrinogen into a fibrin clot. Calcium ions are present for the activation of the factors, and the process takes place on activated platelets and damaged endothelium. (eClinPath.com, Cornell University https://eclinpath.com/hemostasis/physiology/secondary-hemostasis/secondary-haemostasis/)

1.5 Fibrinolysis

Fibrin clot is dissolved into so-called fibrin degradation products in a process known as fibrinolysis. This process is enzymatically catalysed by plasmin, an active form of serine protease plasminogen. Activation of plasminogen requires catalysis by either tissue plasminogen activator (tPA), urokinase (uPA), kallikrein or factor XII. . Of these enzymes, tPA has the highest affinity to plasminogen and both plasminogen and tPA are incorporated into the fibrin clot (Chapin & Hajjar, 2015). Plasmin augments its own generation by creation of more reactive forms of tPA and uPA. The body controls protein breakdown by plasmin using inhibitors - plasminogen activator inhibitor-1 (PAI-1) and alpha-2-antiplasmin (A2AP). Thrombin-activated TAFI slows clot dissolution by removing lysine residues, further limiting plasminogen activation (Sillen & Declerck, 2021). Plasminogen cleaves fibrin after lysine and arginine residues. There are at least 34 plasmin cleavage sites recognized in fibrin, although they are cleaved with different affinity. Mutations at any of these sites, as well as mutations of binding sites of fibrinolytic enzymes, hamper fibrinolysis, what may result into thrombotic states in a patient. (Hudson, 2017)

1.6 Mutations

Mutations are alterations in the DNA sequence that can occur in various forms and have diverse effects on organisms. Mutations are either point mutations, which replace one nucleotide with another, or insertions and deletions (indels) of one or more nucleotides. Indels can change the protein sequence of the translated gene, because of a possible shift in the open reading frame (ORF). These changes are caused by errors during DNA replication, by exposure to mutagens like chemicals and radiation or can be inherited (Niroula & Vihinen, 2016). A specific type of point mutation, known as a missense mutation, involves a single nucleotide change that results in the substitution of one amino acid for another during protein translation (Stefl et al., 2013). These changes can affect protein stability, function and level of protein's expression leading to

a range of effects from minor to severe troubles in biological processes (Zhang et al., 2012). Mutations can occur on one (heterozygous mutations) or on both (homozygous mutations) alleles either in coding or non-coding regions of a gene. Studying mutations is crucial because they influence protein properties and interactions, which in return affect cellular processes and can lead to diseases and variations in traits among individuals (Stefl et al., 2013) (Zhang et al., 2012). This understanding helps us make better tests and treatments for genetic disorders.

1.7 Congenital fibrinogen disorders

Congenital fibrinogen disorders are classified into two types of plasma fibrinogen defects – quantitative and qualitative. In quantitative fibrinogen deficiency, there are low or absent plasma fibrinogen antigen levels. The antigen level of fibrinogen reflects the total amount of fibrinogen in the blood, no matter if it is functional or not. In qualitative there are normal or reduced antigen levels associated with low functional activity (Neerman-Arbez et al., 2016). There are two distinct quantitative disorders: afibrinogenemia which is characterized by the complete absence of fibrinogen in the blood and hypofibrinogenemia which is a condition with a proportional decrease in functional and antigenic fibrinogen levels. (Casini et al., 2018). There are also two quantitative disorders, dysfibrinogenemia when patients have decreased functional and normal antigenic fibrinogen levels, and hypodysfibrinogenemia which is characterized by a decrease in functional as well as antigenic fibrinogen levels (Casini et al., 2018). Apart from congenital fibrinogen disorders there can be mutations in the C-terminal part of the Aα chain of fibrinogen causing renal amyloidosis (Chapman & Dogan, 2019). Amyloidosis is a group of diseases characterized by the deposition of amyloid fibrils in tissues, leading to organ dysfunction and potentially death. Hereditary fibrinogen alpha chain amyloidosis is a rare autosomal dominant disorder caused by mutations in the *FGA* gene (Chapman & Dogan, 2019). Certain mutations in the FReD domain of the γ chain are associated with fibrinogen storage disease (FSD), where the aberrant fibrinogen is not degraded by proteolytic enzymes but is stored in the liver. FSD is associated with hypofibrinogenemia.

1.7.1 Clinical manifestations of fibrinogen disorders:

1.7.1.1 Afibrinogenemia

Congenital afibrinogenemia is an autosomal recessive disorder characterized by bleeding that varies from mild to severe and by complete absence or extremely low levels of plasma and antigen fibrinogen (Duga et al., n.d.). Afibrinogenemia mainly causes severe bleeding in various body parts, including muscles, and can be identified in newborns due to prolonged bleeding from the umbilical cord. Unlike haemophilia, joints bleeding (hemarthrosis) is less common and less severe, but it can still lead to joint diseases. The most dangerous symptom is spontaneous bleeding in the brain, which is the leading cause of death (Casini et al., 2016). Paradoxically, also thrombotic events are reported in some afibrinogenemic patients.

1.7.1.2 Hypofibrinogenemia

Hypofibrinogenemia generally leads to less severe symptoms compared to afibrinogenemia. It can result in bleeding, mainly after injuries or surgeries, with spontaneous bleeding being rare unless fibrinogen level drops very low. A notable issue with hypofibrinogenemia involves fibrinogen storage disease, which occurs in some genetic variants and causes liver inflammation and fibrosis due to abnormal fibrin build-up (Casini et al., 2016).

1.7.1.3 Dysfibrinogenemia

Dysfibrinogenemia is often discovered accidentally through blood tests. Symptoms range from none to mild bleeding, typically in the mucous membranes. Less frequently there is significant bleeding related to surgery, injury, or childbirth, with a notable risk by age 50. Additionally, it carries a risk of thrombosis, more so with certain genetic mutations. It is also linked to increased risks of chronic lung hypertension and kidney amyloidosis (Casini et al., 2016).

1.7.1.4 Hypodysfibrinogenemia

Hypodysfibrinogenemia, combines features of both quantitative and qualitative fibrinogen disorders, often leading to more symptoms compared to dysfibrinogenemia. Patients frequently experience spontaneous bleeding across various tissues, including the central nervous system, and face a high risk of both arterial and venous thrombosis (Casini et al., 2016; Mount, 2008).

1.8 Programs

Computational tools for predicting the clinical relevance of mutations on protein function, known as Variant Effect Predictors (VEPs) primarily focus on analysing missense mutations. VEPs assess whether genetic substitutions are benign or pathogenic, using parameters derived from evolutionary, physico-chemical, sequence homology, or structural and functional characteristics. Historically, these programs began with methodologies based on sequence alignment and probabilities, such as SIFT, which determines the probability of a mutation's impact after performing multiple sequence alignments (MSA), and PANTHER, which constructs phylogenetic trees. With advancements in technology, newer VEPs employed machine learning methods like artificial neural networks, decision trees, random forests, and support vector machines, then develop decision rules based on training (Livesey & Marsh, 2022a). However, there is a bias when it comes to comparisons among these predictors caused by evaluating their performance against the same data used for their training. Sequence conservation is a key element of every VEP. Several tools measure sequence conservation by considering different characteristics for model development and, in considerably less extent, these biases may origin in various training data.

Sequence conservation is a key element of almost every VEP. It is usually computed from MSA, that is obtained by BLAST (Altschul et al., 1997) or BLAT (Kent, 2002). Substitution matrices, like BLOSUM (Henikoff & Henikoff, 1993) and PAM reflect likelihood of interchange of wild-type amino acid to the mutated one. Structural and functional characteristics are adopted from databases, like UniProt (Bateman, 2019) or NCBI (Sayers et al., 2022).

Training data for ML-based predictors are taken from public databases. Only the widely used databases are mentioned. The dbSNP database (Sherry et al., 2001a) at the NCBI currently collects about 20 million validated human SNPs, although it usually not reports the clinical manifestations of the mutation. The manually curated UniProt database has approximately 61,000 missense SNPs. ClinVar (Landrum et al., 2014) reports over 125,000 clinically relevant mutations (López-Ferrando, Gazzo, De La Cruz, et al., 2017). Human Gene Mutation Database, HGMD, (Stenson et al., 2003) aims to gather all mutation available in literature, including their clinical manifestations. Its full version is a commercial product. Despite the extensive data

available, predicting the functional consequences of single amino acid mutations (SAVs) remains a significant challenge.

1.8.1 Machine learning

Machine learning (ML) in the field of variation interpretation primarily identifies patterns in features such as conservation, secondary structure, and amino acid properties to predict pathogenicity. A significant portion of recent methods in this area rely on ML, which typically offers binary outcomes—classifying genetic mutations as either benign or pathogenic (Thusberg et al., 2011). Among ML techniques, supervised learning is the most frequent, where models are trained using well-defined, quality-controlled examples to differentiate between multiple classes (Thusberg et al., 2011). Conversely, unsupervised learning operates without labelled training data, allowing the model to independently formulate methods for making predictions, thus minimizing bias from prior examples (Livesey & Marsh, 2022a). This highlights the importance of selecting the correct ML approach, considering the specific needs and data availability in mutations interpretation. Most used machine learning techniques: Gradient-boosted trees, Random Forest, Neural Networks, Naïve Bayes Classifiers, Support Vector Machines (SVMs), Variational Autoencoders (VAEs). Techniques used by VEP tested in this thesis are introduced below.

1.8.1.1 Random Forest

The Random Forest (RF) algorithm is a machine learning technique used in both classification and regression. It constructs multiple decision trees by process called bagging during the training phase. For classification tasks it gives us the mode of the classes predicted by the individual trees. For regression tasks, like estimating the effects of mutations, it provides the mean prediction from all trees. The key strength of RF in mutation prediction is its ability to improve accuracy and control over-fitting through its ensemble approach. Each tree in the forest is built from a random subset of the data and a random selection of features. This approach efficiently handles the complexity of relationships within the data while minimizing overfitting (Pellegrino et al., 2021).

1.8.1.2 Naïve Bayes

Naïve Bayes classifiers are supervised learning algorithms that use a feature vector for

classification based on Bayes' theorem (Livesey & Marsh, 2022a). They rely on the assumption that all input features are independent and that each feature equally influences the outcome. This simplifies the computations and makes the problems more tractable (https://www.ibm.com/topics/naive-bayes). Despite these simplifications, Naïve Bayes models are widely used because they are quick and easy to construct, require minimal space, and remain competitive in performance against more complex algorithms. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3128400/).

1.8.1.3 Support vector machine

Support Vector Machine (SVM) is a type of computer algorithm used to classify data into different groups. It finds the best line, or "hyperplane," that separates the data into classes as clearly as possible (Livesey & Marsh, 2022a). More classes can be separated using more hyperplanes. This line is drawn so that it has the most space possible from the nearest data points of each class, which are called support vectors (Huang et al., 2018).SVM can handle data that isn't naturally easy to separate by transforming it into a higher dimension where it can be more easily divided. To use SVM, you need to prepare your data, train the SVM model on a part of the data, and then test its accuracy on a separate part of the data. (https://academic.oup.com/bioinformatics/article/22/3/278/220718).

1.8.2 Mutation effect predictors tested in this thesis

1.8.2.1 PANTHER-PSEP

PANTHER-PSEP (Tang & Thomas, 2016) is a prediction tool based on sequence conservation. This method differs from other by introducing the concept of evolutionary preservation. Evolutionary preservation refers to the time (in millions of years) of conservation of the site in sequences of direct ancestors of the protein of interest (Marini et al., 2010). By analysing the frequency, we get the importance of the amino acids for the examined protein function. The longer the time of conservation for a site, the greater the likelihood that its mutation is associated with disease. To be more comprehensive, the preservation time is converted into the "Probability of Deleterious effect", PDel, based on results of a benchmark study (Capriotti et al., 2006). The mutation is referred as "probably damaging" if it preserved for more than 450 million of years (PDel > 0.4). It is "probably benign", if the preservation is shorter than 200 million of years (PDel < 0.2) and otherwise, it is classified as "possibly damaging".

PANTHER-PSEP speeds up predictions by using pre-calculated sequence alignments and phylogenetic trees from the PANTHER library (Thomas et al., 2003). PANTHER-PSEP is implemented as both a standalone tool and a web server.

1.8.2.2 PMut

PMut (López-Ferrando, Gazzo, de la Cruz, et al., 2017), uses a machine learning method called Random Forest (Rigatti, 2017), which looks at 12 specific features to decide if a mutation is likely to cause a disease. The twelve features include information about sequence conservation and predicted physico-chemical properties of wild type and mutated amino acids. The classifier is trained on SwissVar database (Mottaz et al., 2010). The prediction values range from zero to one, and scores higher than 0.5 are considered disease related. To make these predictions more reliable, PMut also looks at how confident it is in its prediction. It was found that predictions with very high or very low scores tend to be more accurate. This information helps researchers assess the reliability of the predictions made by PMut.

1.8.2.3 SNPs&GO and PhD-SNP

SNPs&GO is a support vector machine-based predictor, that, unlike other predictors, considers information derived from the Gene Ontology annotation (GO) (Aleksander et al., 2023) to predict if a mutation is pathogenic or not. GO information is supported by features obtained from PANTHER-PSEP and the PANTHER library (as mentioned earlier)(Thomas et al., 2003), as well as from the sequence profile, which includes details about the proximity of the mutated amino acid. The prediction scores range from zero to one, and values below 0.5 are considered disease related. SNPs&GO is trained on data derived from SwissProt database. SNPs&GO is an updated version of a prediction tool called PhD-SNP (Capriotti et al., 2006).

1.8.2.4 SIFT

SIFT (Sorting Intolerant From Tolerant) (Ng & Henikoff, 2003) is a program that uses sequence homology to predict how amino acids changes, might impact protein function, potentially altering the phenotype. If an amino acid substitution is found at a position highly conserved across species, it suggests that this change might be deleterious to the protein's function. In contrary, if the substitution occurs at a position where amino acids vary across species, it might be better tolerated without significant functional consequence. SIFT gives us

predictions for all 20 possible amino acid changes at each position in the protein. If the substitution scores less than 0.05 SIFT considers it as pathogenic. SIFT employs a multiple sequence alignment (MSA) based on the protein-of-interest, scanning each column to determine the frequency of substitutions and the probability of tolerance (Livesey & Marsh, 2022b). Users can select from SwissProt, SwissProt/TrEMBL (Bairoch & Apweiler, 2000), and NCBI's protein databases.

1.8.2.5 Mutation Taster

MutationTaster2 is a prediction tool not only for amino acid substitutions, but also for intronic and synonymous alterations, and indels mutations. MutationTaster2 maps the mutation to all suitable genes and transcripts, analyses the mutation in all of them and gives us a table summarizing the predictions. To generate those predictions, it uses a Bayes classifier. Mutations that are found pathogenic in ClinVar are automatically said to be pathogenic. Nowadays there is a new version MutationTaster2021 which instead of using Naive Bayes classifier, uses Random Forest models for obtaining better results. The output shows how many decision trees are suggestive of pathogenicity. If more than 50 trees reveal pathogenicity, the mutation is said to be pathogenic. If less, it is marked as benign. MutationTaster2021 also provides information on the disease the mutations cause (Lubeck et al., 2014; Steinhaus et al., 2021).

1.8.2.6 PolyPhen2

Polyphen-2 is a prediction tool that uses structural and comparative evolutionary considerations. It compares different versions of the same protein from lots of different animals to see how important certain parts of the protein are. Then, it looks at the structure of the protein to see how big impact it is going to have, if any. It also considers how often that part of the DNA has changed over a time. PolyPhen-2 uses the UniProtKB database as a reference source for all protein sequences and annotations. An MSA is employed to create independent count profiles specific to each position. When a sequence aligns with a known three-dimensional structure, further characteristics are integrated into the prediction. The original PolyPhen algorithm uses decision tree to calculate the score. PolyPhen2 uses a naive Bayes classifier to process the data obtained from sequence alignments and also uses protein structural properties, such as the accessible surface area of an amino acid residue, crystallographic beta-factor, and

others.The output is a probability score and three different labels - probably damaging, possibly damaging, benign. The lower the score is, the lower the probability of mutation being damaging is. (Adzhubei et al., 2013)

1.8.2.7 Provean

PROVEAN (Protein Variation Effect Analyzer) is an algorithm that predicts not only the functional impact for single amino acids substitutions but also insertions, deletions, and multiple substitutions. This alignment-based score measures how much a change in an amino acid affects the similarity between two protein sequences. Provean is trained on UniProtKB/ Swiss-Prot. It calculates a delta score which is not only determined by the amino acid position but also by the neighbourhood that surrounds the site of variation. Firstly, Provean gathers a collection of homologous and distantly related sequences. For each sequence within this supporting set, a delta score is computed using the BLOSUM62 substitution matrix, with gap penalties set at 10 for opening and 1 for extension. (Choi et al., 2012).

2 Methods

2.1 Selection of mutations

The work, designed as a blind study, was performed on mutations reported in the fibrinogen γ chain. We addressed mutations reported in the Human Fibrinogen Database (Sovova et al., 2022) and those are not reported in neither dbSNP (Sherry et al., 2001b) nor ClinVar (Landrum et al., 2014)databases. Further, we excluded mutations for which the description of phenotype is missing, or its report was unavailable for us. The clinical phenotype of the mutation was adapted from the original work describing the mutation. The mutation is considered benign only if no clinical manifestation but for congenial fibrinogen disorders is reported in any of its carriers. We checked whether the position of the mutation at both the protein and nucleotide levels (cDNA) matches the reference sequences NP_000500.2 and NM_000509.6, respectively. If necessary, missing information was supplied. We used the major transcript of the *FGG* gene because all tested mutations are in the region, where both transcripts are identical. The conversion of the c.DNA to g.DNA was performed by "position converter" utility by Mutalyzer3 (Wildeman et al., 2008). In total, we tested 70 mutations, 20 of which were benign and 50 were pathogenic. The tested mutations, including their clinical manifestations Table 5. For testing, we selected predictors, those that are either used at Institute of Hematology and Blood Transfusion or those that are recommended by American College of Medical Genetics and Genomics, ACMG (Richards et al., 2015). For SIFT, we consider two thresholds of detection. One is the original value by Ng and Hanikoff, that is referred as SIFT 0.05, the other is value by Pejaver et al. (Pejaver et al., 2022), referred as SIFT. We use the HGVS recommendations (den Dunnen et al., 2016) to describe the position of a mutation.

2.2 Prediction of impact of mutations

We accessed the predictors by the web interface using links see Table 1 and followed the instructions provided by the tool. Provean, that misses the web interface, was accessed by dbNSFP interface (Liu et al., 2020). Understanding the input requirements was essential before using any tool, including details such as the wild-type residue, the position of the mutation, and sequence data, often in FASTA format. Input involved mutations in the specified format together with uploading sequence data in FASTA format, making sure to check whether to exclude or include the header. After inputting mutations and sequence data, we submitted the

query through the provided interface, initiating the analysis process. The waiting time for results depended on the complexity of the analysis and server load, with some tools providing results instantly, while others took longer. Finally, saving or downloading the results for further analysis was an important step, with most tools offering options to download results in various formats, such as text files or spreadsheets, as well as selecting the option to send it by email.

PROGRAM	LINK
Panther	https://pantherdb.org/tools/csnpScoreForm.jsp
PANTHER-PSEP	https://pantherdb.org/tools/csnpScoreForm.jsp
PMut	http://mmb.irbbarcelona.org/PMut/analyses/new/
PhD-SNP	https://snps.biofold.org/phd-snp/phd-snp.html
SIFT	*https://sift.bii.a-star.edu.sg/index.html*
MutationTaster 2014	*https://www.mutationtaster.org/*
MutationTaster 2021	*https://www.genecascade.org/MutationTaster2021/ - transcript*
PolyPhen2	*http://genetics.bwh.harvard.edu/pph2/*
Provean	*http://provean.jcvi.org/protein_batch_submit.php?species=human*

Table 1 Links to bioinformatic tools

The table lists various bioinformatics tools with links for predicting the functional impact of genetic mutations, including Panther, PMut, PhD-SNP, SIFT, Mutation Taster (both 2014 and 2021 versions), PolyPhen2, and Provean. These tools assist in evaluating the potential pathogenicity of mutations by analyzing protein sequences and structures.

2.3 Data analysis and statistics

The outputs from the prediction programmes were uploaded into MS Excel. Evaluating the results involves identifying four basic categories: true positive, true negative, false positive, and false negative see Table 2. This was done manually. The predictions for every mutation in each programme are shown in an attachment in Table 6

True positive (TP):	Correct identification of positive cases
True negative (TN):	Correct identification of negative cases
False positive (FP):	Incorrect identification of positive cases
False negative (FN):	Incorrect identification of negative cases

Table 2 Definitions of outcomes of binary classifiers

The table categorizes the accuracy of analysis results into four groups: correctly identified positive and negative cases (TP and TN), and incorrectly identified cases where positive cases are labelled negative (FN) or negative cases are labelled positive (FP).

We used four statistical measurements to analyse the results - sensitivity, specificity, accuracy, and the Matthews correlation coefficient. These measurements were computed in MS Excel according to formulae listed see Table 3.

Statistical measurement	**Formula**
Sensitivity	TP/ (TP+ FN)
Specificity	TN/(TN+FP)
Accuracy	(TP + TN)/ (TP + FP + TN + FN)
Matthews correlation coefficient	(TP * TN – FP * FN) / √((TP + FP) * (TP + FN) * (TN + FP) * (TN + FN))

Table 3 Metrics for binary classification

The table presents formulas for various statistical measurements used in evaluating the accuracy of protein mutation analysis. These measurements include sensitivity, specificity, accuracy, and the Matthews correlation coefficient, each calculated based on the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

3 Results and Discussion

In our analysis of the performance of computational tools for predicting the effects of mutations on the γ chain of fibrinogen protein, we evaluated a series of metrics: sensitivity, specificity, accuracy, and the Matthews correlation coefficient (MCC) see Table 4. Each metric provided us with insights into the strengths and weaknesses of the tools. Then we calculated mean and median of all the metrics to help us analyse the results more accurately. When data is ordered from the lowest to the greatest and number of data is odd the central value is the median. If there is an even number of data, then it is the average of the middle two. The mean (average) t is calculated as a sum of all the values in a dataset divided by the total number of values (*Whitley2002*, n.d.).

	SENSITIVITY %	SPECIFICITY %	ACCURACY %	MCC
PANTHER-PSEP	98.0	0.0	70.0	-0.076
Panther	95.9	5.6	71.6	0.032
PhD_SNP	84.0	26.3	68.1	0.118
PMut	83.7	19.0	64.3	0.033
SIFT	61.3	50.0	59.5	0.085
SIFT 0.05	88.2	68.4	82.9	0.567
SNPs&GO	94.0	5.3	69.6	-0.014
PolyPhen2	96.2	22.2	77.1	0.287
MutTaster14	96.1	0.0	70.0	-0.105
MutTaster21	96.1	5.3	71.4	0.029
Provean	90.0	5.0	65.7	-0.081
Mean	89.4	18.8	70.0	0,080
Median	94.0	5.6	70.0	0,032

Table 4 Performance of bioinformatic tools

This table lists the performance of various prediction tools used to identify genetic mutations associated with diseases, measuring their sensitivity, specificity, accuracy, and Matthew's Correlation Coefficient (MCC). Tools like SIFT 0.05 and PolyPhen2 show a balance between sensitivity and specificity with relatively high accuracy and MCC values, indicating their effectiveness in predicting the impact of genetic mutations. The

table also contains mean and median for all four statistical metrics – sensitivity, specificity, accuracy and MCC.

3.1 Sensitivity

Sensitivity describes the true positive rate. It measures the proportion of actual positives that are correctly identified by a computational tool. It is a critical metric for determining a model's ability to detect true positive cases.

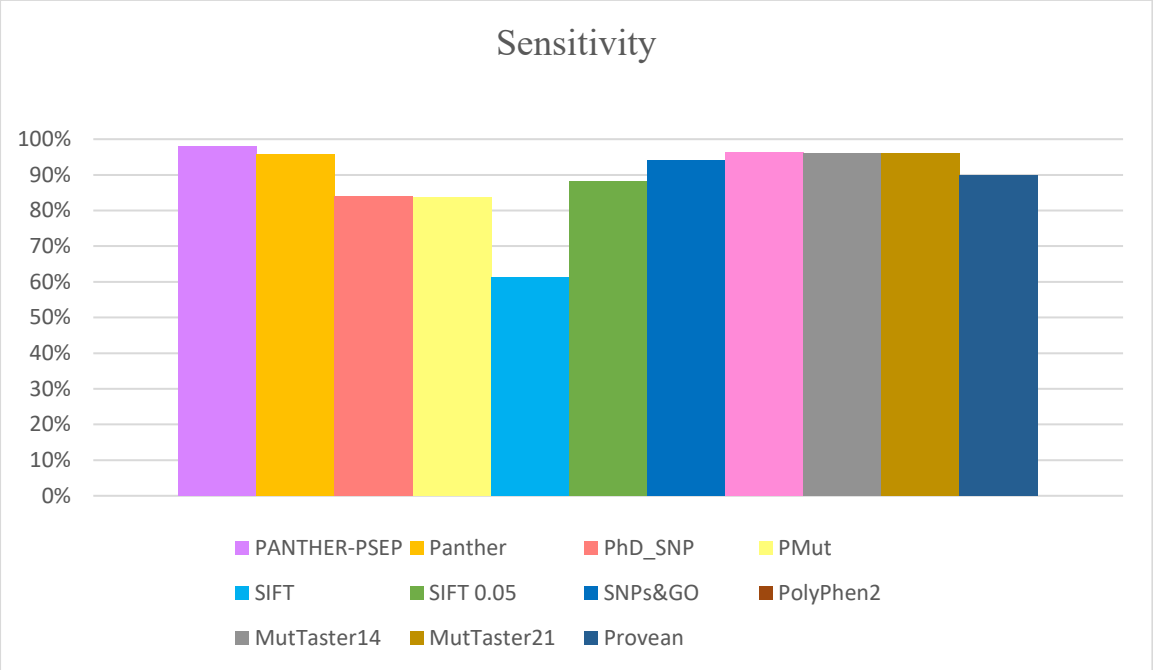

Figure 3 Sensitivity graph

This graph compares the sensitivity of all tested programs, with percentages on the y-axis and tool names on the x-axis. The best-performing tool is PANTHER-PSEP with just under 90% sensitivity, and the worst-performing tool is SIFT at around 40% sensitivity.

All the tested programs provided predictions with sensitivity above 80%, but for SIFT, whose sensitivity of 61.3% is considerably lower see Table 4 and Figure 3. This indicates that the threshold value proposed by Pejaver et al may not be optimally adjusted for all types of proteins, particularly fibrinogen γ chain. Using the originally intended threshold 0.05, the sensitivity of SIFT predictions rose to 88.2%. The Panther-PSEP algorithm showed superior sensitivity of 98%, therefore is very accurate in detection of true positives within the dataset. Panther-PSEP has an approach that uses both the physicochemical properties of amino acid changes and the evolutionary conservation of protein sequences. Such approach is beneficial for proteins like fibrinogen, where evolutionary constraints may play a significant role in

maintaining their structure and function. High sensitivity was also obtained by PolyPhen2 (96.2%), MutationTaster (96.1%), Panther (95.9%) and SNPs&GO (94.0%).

Note, the high median value of 94.0 %, which means, that all programmes are successful in the prediction of pathogenic mutations. The median and mean are quite close, with the median 94.0% and the mean 89.0 % Table 4. It implies a symmetrical distribution of data, suggesting a consistency in the sensitivity across programs. This means that there are not many outliers that significantly skew the data, which corresponds with the graph Figure 3. The mean is a bit lower than the median because of the SIFT's lower sensitivity, but it does not change the overall results much.

3.2 Specificity

Specificity describes the true negative rate. It measures a tool's ability to correctly identify negatives, meaning how it can accurately dismiss benign mutations. High specificity means that the method is good at avoiding detecting true negatives.

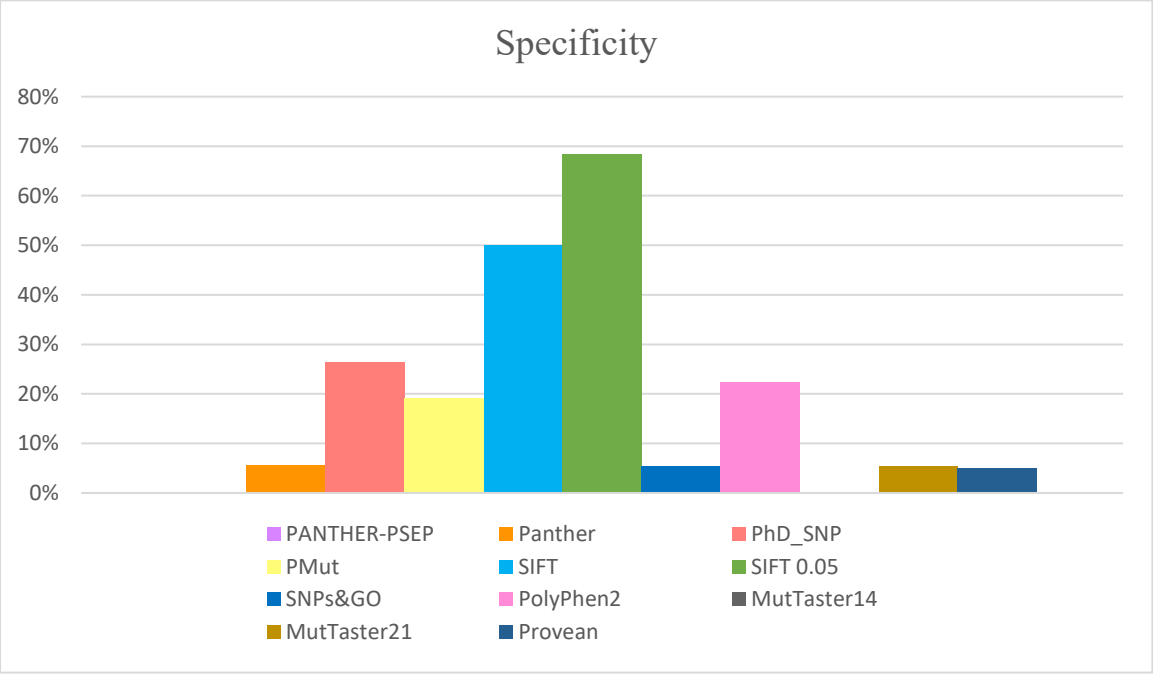

Figure 4 Specificity graph

This graph compares the specificity of all tested programs, with percentages on the y-axis and tool names on the x-axis. The best-performing tool is SIFT 0.05 with 68.4 % and the worst-performing tools are MutationTaster14 and PANTHER-PSEP with 0%.

Programs exhibited low performance in specificity (median of 5.6%) and significant disparity in the results. This diversity in performance illustrates a fundamental trade-off between sensitivity and specificity. Programs like Panther-PSEP and MutationTaster14, which demonstrated high sensitivity yet zero specificity, reflect this trade-off by preferring the detection of as many true positives as possible, even at the risk of a high false positive rate. While such a strategy may result in many benign mutations being marked as potentially harmful, it can be strategically advantageous in a clinical context where missing a pathogenic mutation could have critical consequences. SIFT with the VarSome detection threshold displayed a relatively good specificity (50.0%) and when using the originally intended threshold 0.05 the result was even better (68.4%) see Table 4 and Figure 4. Higher specificity was obtained by PhD_SNP (26.3%) and PolyPhen2 (22.2%) and PMut (19.0%). The other programmes (Panther (5.6%), SNPs&GO (5.3%), MutTaster21 (5.3%), Provean (5.0%)) have specificity below 6% and PANTHER-PSEP and MutationTaster14 despite their high sensitivity, showed the lowest specificity of 0%, indicating a tendency to classify most mutations as pathogenic.

The big difference between median (5.6%) and mean (18.8%) see Table 4 points to an asymmetrical distribution which can be see Figure 4. This means that many programs with low specificity significantly lowered the median. On the other hand, the mean is boosted by a minority of programs with high specificity rates, such as SIFT when using the original threshold (68.4%).

3.3 Accuracy

Accuracy is a proportion of true results (both true positives and true negatives) out of the total number of cases. It reflects the overall correctness of the tools' predictions.

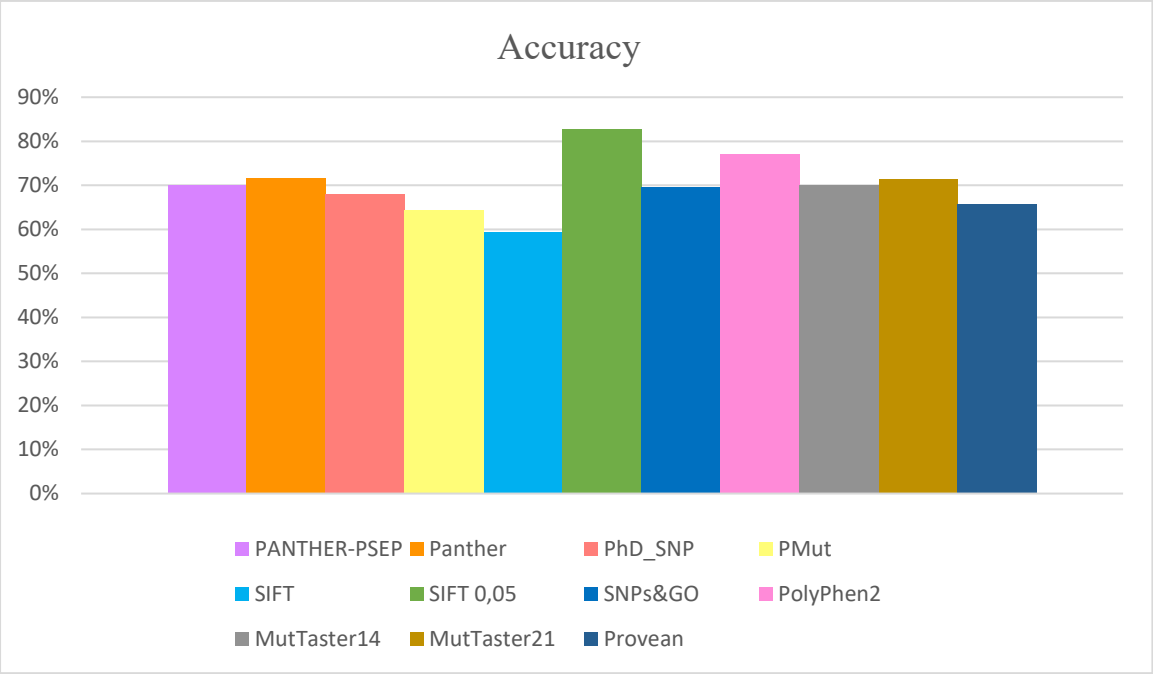

Figure 5 Accuracy graph

This graph compares the accuracy of all tested programs, with percentages on the y-axis and tool names on the x-axis. The best-performing tool is SIFT 0.05 with 82.9 % and the worst-performing tool is SIFT with 59.5%.

All tested programs provided predictions with accuracy above 60.0%, but for SIFT with the Pejaver's et al detection threshold, whose accuracy of 59.5 % is lower see Table 4 Figure 5. When the threshold was adjusted to 0.05 its accuracy to 82.9%, and no other programme reached higher accuracy. Shift of the threshold in the other programmes did not improve their performance. High accuracy was also obtained by PolyPhen2 (77.1%).

The accuracy shows the biggest consistency among all four metrics. It is likely affected by the dataset used for the evaluation, which has more pathogenic mutations than benign ones. This distribution reflects actual clinical situations since people without disease-causing mutations often don't seek medical help. This results in higher accuracy measurements that favour the detection of frequent pathogenic mutations, while the identification of benign ones may not be as reliable. Both the mean (70.0%) and median (70.0%) accuracy rates around 70.0%, indicating a balanced distribution without any extreme deviations see Table 4.

Although the accuracy is consistent, it may not be the best indicator of a tool's effectiveness, especially considering the varying levels of specificity and sensitivity revealed by other graphs see Figure 3 Figure 4. Accuracy could remain unaffected by the balance between false positives and negatives, as long as the number of correct predictions remains high. This situation suggests that while accuracy is an important factor, it should be considered alongside other performance metrics to fully judge the capabilities of these tools.

3.4 Matthews correlation coefficient

MCC is a very reliable metric for evaluating the performance of binary classifiers, such as the computational tools used for predicting the pathogenicity of mutations. Unlike accuracy, MCC considers the balance between all four categories: true positives, true negatives, false positives, and false negatives. This makes it valuable in situations where the datasets differ in size or in this case where our dataset has more pathogenic mutations than benign ones. A high MCC indicates that the tool is not only good at identifying the true positives (sensitivity) and true negatives (specificity) but also that it is effective at avoiding false positives and false negatives.

An MCC of +1 represents a perfect prediction, 0 is no better than random prediction, and -1 indicates total disagreement between prediction and observation. Therefore, a high MCC value close to +1 is the goal for a predictive tool.

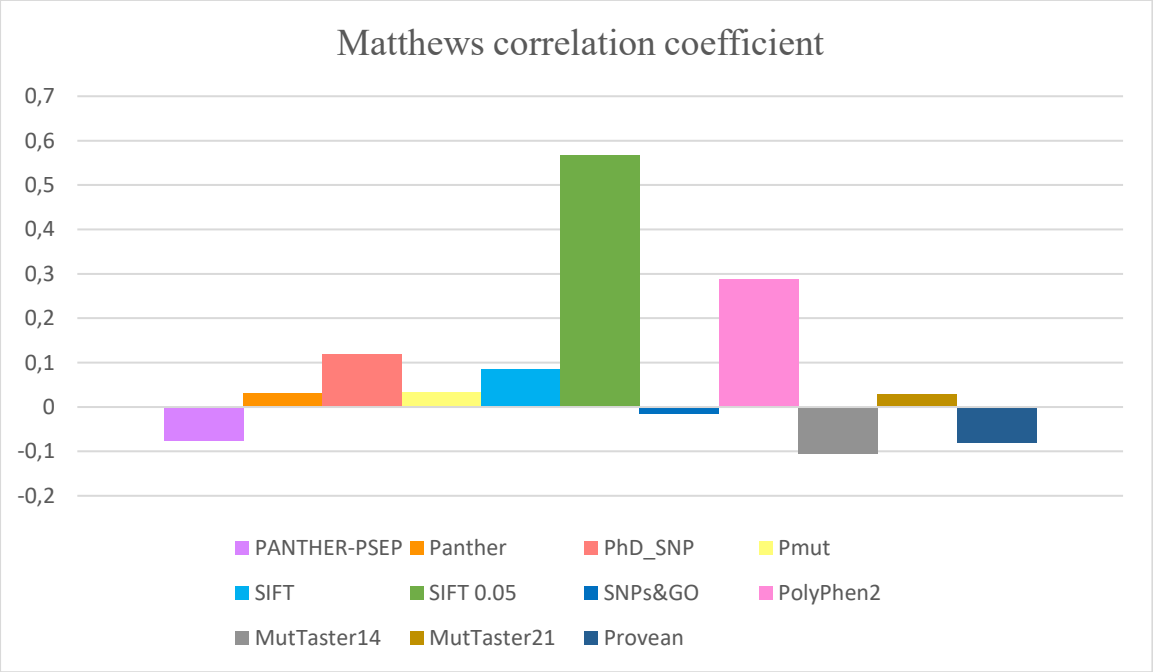

Figure 6 Matthews correlation coefficient

This graph compares the Matthews correlation coefficient of all tested programs, with percentages on the y-axis and tool names on the x-axis. The best-performing tool is SIFT 0.05 with 0.567 and the worst-performing tool is SIFT with -0.105

Figure 6 shows the variability in MCC among the tools which is high compared to other metrics – sensitivity, specificity, and accuracy. This inconsistency highlights the differences in the tools' abilities to balance the rate of false classifications against the correct predictions. The MCC for SIFT at a threshold of 0.05 is the highest (0.567), which means that it achieved the best balance between true and false positives and negatives. Polyphen2 (0.287) and PhD_SNP (0.118) have MCC above 0.1. In contrast, several tools such as PANTHER-PSEP (-0.076), SNPs&GO (-0.014), MutationTaster14 (-0.105) and Provean (-0.081) showed negative MCC values. The negative values for MCC means that on average it is worse than random guessing. This is concerning because they are likely to misclassify mutations, which could lead to incorrect clinical decisions.

The mean (0.078) and median (0.032) are relatively close. The closeness of the mean and median in the MCC despite the asymmetry means that there are extreme values on both sides of the scale that balance each other out see Figure 6.

3.5 The overall performance

In general, the tested programmes show higher sensitivity (median of 89.4 %) than specificity (median of 5.6%). This means, that their prediction of pathogenic mutations is more reliable than the prediction of benign mutations. The four tools that had negative Matthews correlation coefficient (MCC) values—Panther-PSEP (-0.076), SNPs&GO (-0.014), MutationTaster14 (-0.105), and Provean (-0.081)—are concerning because the predictions made by these tools could be less reliable than random guessing. SIFT, with the original 0.05 threshold, had the highest specificity (68.4%), accuracy (82.9%), and MCC (0.567), together with a good sensitivity rate of 88.0%. This makes it the most reliable of the tested programmes. However, SIFT's performance dropped with the Pejaver's threshold, exhibiting a sensitivity of 61.3%, specificity of 50.0%, accuracy of 59.5%, and MCC of 0.085. PhD_SNP, PMut, had moderate scores in sensitivity and specificity, with MCC values reflecting average predictive reliability. PolyPhen2 showed strong sensitivity (96.2%) and moderate specificity (22.2%), with a relatively high MCC (0.287), making it a reliable tool, even though it does not reach the

performance of SIFT 0.05. Panther and MutationTaster21 showed moderate specificity and positive but low MCC, indicating some balance but also the need for improvement. The least reliable tools are PANTHER-PSEP and MutationTaster14, primarily due to their complete lack of specificity (0.0%) and negative MCC value.

4 Conclusion

This thesis compares performance of 9 programmes for prediction of potential pathogenicity of mutations on 70 mutations in the fibrinogen γ chain, those are absent in dbSNP and ClinVar databases, e.i. databases, those are used for training of AI-based programmes. We used sensitivity, specificity, accuracy and MCC to evaluate the performance of the programmes. In general, the programs have considerably higher sensitivity than specificity, which means, that they are more likely to identify pathogenic mutation than the benign one.

We consider SIFT with threshold 0.05 as the best program. It has the highest values of specificity, accuracy and MCC. Its sensitivity (88.0%) is not the highest, but it is high enough to support the reliability of the tool. SIFT has the best balance between all the studied metrics and can provide the most accurate reflection of reality, minimizing both false positives and false negatives. This is particularly important in the context of fibrinogen due to its significant role in blood clotting. On the other hand, Panther-PSEP, SNPs&GO, MutationTaster14 and Provean provide negative values of MCC, which means, that their reliability is inferior to random guessing. Thus, using these programmes, one must be very cautious.

5 Table of figures and tables

6 References

Aleksander, S. A., Balhoff, J., Carbon, S., Cherry, J. M., Drabkin, H. J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N. L., Hill, D. P., Lee, R., Mi, H., Moxon, S., Mungall, C. J., Muruganugan, A., Mushayahama, T., Sternberg, P. W., Thomas, P. D., Van Auken, K., … Westerfield, M. (2023). The Gene Ontology knowledgebase in 2023. *Genetics*, *224*(1). https://doi.org/10.1093/genetics/iyad031

Asselta, R., Platè, M., Robusto, M., Borhany, M., Guella, I., Soldà, G., Afrasiabi, A., Menegatti, M., Tahir, S., & Peyvandi, F. (2015). Clinical and molecular characterisation of 21 patients affected by quantitative fibrinogen deficiency. *Thrombosis and Haemostasis*, *113*(03), 567–576.

Asselta, R., Robusto, M., Platé, M., Santoro, C., Peyvandi, F., & Duga, S. (2015). Molecular characterization of 7 patients affected by dys-or hypo-dysfibrinogenemia: identification of a novel mutation in the fibrinogen Bbeta chain causing a gain of glycosylation. *Thrombosis Research*, *136*(1), 168–174.

Bairoch, A., & Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. In *Nucleic Acids Research* (Vol. 28, Issue 1). http://www.expasy.

Bateman, A. (2019). UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research*, *47*(D1), D506–D515. https://doi.org/10.1093/nar/gky1049

Bentolila, S., Samama, M. M., Conard, J., Horellou, M. H., & Ffrench, P. (1995). Association of dysfibrinogenemia and thrombosis. Apropos of a family (Fibrinogen Melun) and review of the literature. *Annales de Medecine Interne*, *146*(8), 575–580.

Brennan, S. O., Davis, R. L., & Chitlur, M. (2010a). New fibrinogen substitution (γSer313Arg) causes diminished γ chain expression and hypodysfibrinogenaemia. *Thrombosis and Haemostasis*, *103*(02), 478–479.

Brennan, S. O., Davis, R. L., & Chitlur, M. (2010b). New fibrinogen substitution (γSer313Arg) causes diminished γ chain expression and hypodysfibrinogenaemia. *Thrombosis and Haemostasis*, *103*(02), 478–479.

Brennan, S. O., & Laurie, A. (2014). Functionally compromised FGG variant (γ320Asp→ Glu) expressed at low level in plasma fibrinogen. *Thrombosis Research*, *134*(3), 744–746.

Brennan, S. O., Laurie, A., & Smith, M. (2015). Novel FGG variant (γ339C→ S) confirms importance of the γ326–339 disulphide bond for plasma expression of newly synthesised fibrinogen. *Thrombosis and Haemostasis*, *113*(04), 903–905.

Brennan, S. O., Sheen, C. R., & Peter, M. (2005). Novel γ230 Asn→ Asp substitution in fibrinogen Middlemore associated hypofibrinogenaemia. *Thrombosis and Haemostasis*, *93*(06), 1196–1197.

Brennan, S. O., Wyatt, J. M., Ockelford, P., & George, P. M. (2000). Defective fibrinogen polymerization associated with a novel gamma279Ala--> Asp mutation. *British Journal of Haematology*, *108*(2), 236–240.

Brennan, S. O., Wyatt, J., Medicina, D., Callea, F., & George, P. M. (2000). Fibrinogen Brescia: hepatic endoplasmic reticulum storage and hypofibrinogenemia because of a γ284 Gly→ Arg mutation. *The American Journal of Pathology*, *157*(1), 189–196.

Callea, F., Giovannoni, I., Sari, S., Aksu, A. U., Esendagly, G., Dalgic, B., Boldrini, R., Akyol, G., Francalanci, P., & Bellacchio, E. (2017). A novel fibrinogen gamma chain mutation (c. 1096C> G; p. His340Asp), fibrinogen Ankara, causing hypofibrinogenaemia and hepatic storage. *Pathology*, *49*(5), 534–537.

Cao, Z., Dong, Y., Zeng, J., Zhu, H., Xie, X., Liu, J., Zhai, Y., & Li, L. (2019). Whole-exome sequencing identified novel mutations in FGA and FGG genes in the patients with decreased fibrinogen. *Thrombosis Research*, *177*, 79–82.

Capriotti, E., Calabrese, R., & Casadio, R. (2006). Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics*, *22*(22), 2729–2734. https://doi.org/10.1093/bioinformatics/btl423

Casini, A., Blondon, M., Lebreton, A., Koegel, J., Tintillier, V., De Maistre, E., Gautier, P., Biron, C., Neerman-Arbez, M., & De Moerloose, P. (2015). Natural history of patients with congenital dysfibrinogenemia. *Blood, The Journal of the American Society of Hematology*, *125*(3), 553–561.

Casini, A., De Moerloose, P., & Neerman-Arbez, M. (2016). Clinical Features and Management of Congenital Fibrinogen Deficiencies. *Seminars in Thrombosis and Hemostasis*, *42*(4), 366–374. https://doi.org/10.1055/s-0036-1571339

Casini, A., Neerman-Arbez, M., & de Moerloose, P. (2021). Heterogeneity of congenital afibrinogenemia, from epidemiology to clinical consequences and management. In *Blood Reviews* (Vol. 48). Churchill Livingstone. https://doi.org/10.1016/j.blre.2020.100793

Casini, A., Undas, A., Palla, R., Thachil, J., & de Moerloose, P. (2018). Diagnosis and classification of congenital fibrinogen disorders: communication from the SSC of the ISTH. *Journal of Thrombosis and Haemostasis*, *16*(9), 1887–1890. https://doi.org/10.1111/jth.14216

Castaman, G., Giacomelli, S. H., Biasoli, C., Contino, L., & Radossi, P. (2019). Risk of bleeding and thrombosis in inherited qualitative fibrinogen disorders. *European Journal of Haematology*, *103*(4), 379–384.

Chapin, J. C., & Hajjar, K. A. (2015). Fibrinolysis and the control of blood coagulation. *Blood Reviews*, *29*(1), 17–24. https://doi.org/10.1016/j.blre.2014.09.003

Chapman, J., & Dogan, A. (2019). Fibrinogen alpha amyloidosis: insights from proteomics. In *Expert Review of Proteomics* (Vol. 16, Issue 9, pp. 783–793). Taylor and Francis Ltd. https://doi.org/10.1080/14789450.2019.1659137

Chinni, E., Tiscia, G., Favuzzi, G., Cappucci, F., Malcangi, G., Bagna, R., Izzi, C., Rizzi, D., De Stefano, V., & Grandone, E. (2019a). Identification of novel mutations in patients with fibrinogen disorders and genotype/phenotype correlations. *Blood Transfusion*, *17*(3), 247.

Chinni, E., Tiscia, G., Favuzzi, G., Cappucci, F., Malcangi, G., Bagna, R., Izzi, C., Rizzi, D., De Stefano, V., & Grandone, E. (2019b). Identification of novel mutations in patients with fibrinogen disorders and genotype/phenotype correlations. *Blood Transfusion*, *17*(3), 247.

Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., & Chan, A. P. (2012). Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS ONE*, *7*(10). https://doi.org/10.1371/journal.pone.0046688

de Raucourt, E., de Mazancourt, P., Maghzal, G. J., Brennan, S. O., & Mosesson, M. W. (2005). Fibrinogen Saint-Germain II: Hypofibrinogenemia due to heterozygous γ N345S mutation. *Thrombosis and Haemostasis*, *94*(11), 965–968.

Dear, A., Brennan, S. O., & George, P. M. (2005). Familial hypodysfibrinogenaemia associated with second occurrence of γ326 Cys→ Tyr mutation. *Thrombosis and Haemostasis*, *93*(03), 612–613.

Dear, A., Dempfle, C. E., Brennan, S. O., Kirschstein, W., & George, P. M. (2004). Fibrinogen Mannheim II: a novel γ307 His→ Tyr substitution in the γD domain causes hypofibrinogenemia. *Journal of Thrombosis and Haemostasis*, *2*(12), 2194–2199.

den Dunnen, J. T., Dalgleish, R., Maglott, D. R., Hart, R. K., Greenblatt, M. S., Mcgowan-Jordan, J., Roux, A. F., Smith, T., Antonarakis, S. E., & Taschner, P. E. M. (2016). HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Human Mutation*, *37*(6), 564–569. https://doi.org/10.1002/humu.22981

Doolittle, R. F., McNamara, K., & Lin, K. (2012). Correlating structure and function during the evolution of fibrinogen-related domains. In *Protein Science* (Vol. 21, Issue 12, pp. 1808–1823). https://doi.org/10.1002/pro.2177

Duga, S., Asselta, R., Santagostino, E., Zeinali, S., Simonic, T., Malcovati, M., Mannuccio Mannucci, P., & Luisa Tenchini, M. (n.d.). *Missense mutations in the human fibrinogen gene cause congenital afibrinogenemia by impairing fibrinogen secretion*. http://ashpublications.org/blood/article-pdf/95/4/1336/1661069/1336.pdf

Duval, C., Allan, P., Connell, S. D. A., Ridger, V. C., Philippou, H., & Ariëns, R. A. S. (2014). Roles of fibrin α- and γ-chain specific cross-linking by FXIIIa in fibrin structure and function. *Thrombosis and Haemostasis*, *111*(5), 842–850. https://doi.org/10.1160/TH13-10-0855

Fu, Y., & Grieninger, G. (1994). *Fib420: A normal human variant of fibrinogen with two extended a chains (cotting factor/blood plasma/fibrinogen a chain)* (Vol. 91). https://www.pnas.org

Galanakis, D. K., Neerman-Arbez, M., Brennan, S., Rafailovich, M., Hyder, L., Travlou, O., Papadakis, E., Manco-Johnson, M. J., Henschen, A., & Scharrer, I. (2014). Thromboelastographic phenotypes of fibrinogen and its variants: clinical and non-clinical implications. *Thrombosis Research*, *133*(6), 1115–1123.

Gindele, R., Kerényi, A., Kállai, J., Pfliegler, G., Schlammadinger, Á., Szegedi, I., Major, T., Szabó, Z., Bagoly, Z., & Kiss, C. (2021). Resolving differential diagnostic problems in von willebrand disease, in fibrinogen disorders, in prekallikrein deficiency and in hereditary hemorrhagic telangiectasia by next-generation sequencing. *Life*, *11*(3), 202.

Green, D. (n.d.). *Coagulation cascade*.

Guglielmone, H. A., Sanchez, M. C., Daga, D. A., & Bocco, J. L. (2004). A new heterozygous mutation in gamma fibrinogen gene leading to 326 Cys→ Ser substitution in fibrinogen Córdoba is associated with defective polymerization and familial hypodysfibrinogenemia. *Journal of Thrombosis and Haemostasis*, *2*(2), 346–379.

Hamano, A., Mimuro, J., Aoshima, M., Itoh, T., Kitamura, N., Nishinarita, S., Takano, K., Ishiwata, A., Kashiwakura, Y., & Niwa, K. (2004). Thrombophilic dysfibrinogen Tokyo V with the amino acid substitution of γ Ala327Thr: formation of fragile but fibrinolysis-resistant fibrin clots and its relevance to arterial thromboembolism. *Blood*, *103*(8), 3045–3050.

Hansen, A. T., Andreasen, B. H., Salvig, J. D., & Hvas, A. M. (2011). Changes in fibrin D-dimer, fibrinogen, and protein S during pregnancy. *Scandinavian Journal of Clinical and Laboratory Investigation*, *71*(2), 173–176. https://doi.org/10.3109/00365513.2010.545432

Hanss, M., & Biot, F. (n.d.). *A Database for Human Fibrinogen Variants*. http://www.med.unc.edu/isth

Hanss, M., Ffrench, P., Vinciguerra, C., Bertrand, M. A., & De Mazancourt, P. (2005). LETTERS TO THE EDITOR: Four cases of hypofibrinogenemia associated with four novel mutations. *Journal of Thrombosis and Haemostasis*, *3*(10), 2347–2349.

Harris, J. R. (n.d.). *Subcellular Biochemistry Volume 82*. http://www.springer.com/series/6515

Haryadi, R., Ho, S., Kok, Y. J., Pu, H. X., Zheng, L., Pereira, N. A., Li, B., Bi, X., Goh, L. T., Yang, Y., & Song, Z. (2015). Optimization of heavy chain and light chain signal peptides for high level expression of therapeutic antibodies in CHO cells. *PLoS ONE*, *10*(2). https://doi.org/10.1371/journal.pone.0116878

Henikoff, S., & Henikoff, J. G. (1993). Performance Evaluation of Amino Acid Substitution Matrices. In *PROTEINS: Structure, Function, and Genetics* (Vol. 17).

Hoffman, M. (2003). Remodeling the Blood Coagulation Cascade. In *Journal of Thrombosis and Thrombolysis* (Vol. 16, Issue 2). Kluwer Academic Publishers.

Huang, S., Nianguang, C. A. I., Penzuti Pacheco, P., Narandes, S., Wang, Y., & Wayne, X. U. (2018). Applications of support vector machine (SVM) learning in cancer genomics. In *Cancer Genomics and Proteomics* (Vol. 15, Issue 1, pp. 41–51). International Institute of Anticancer Research. https://doi.org/10.21873/cgp.20063

Hudson, N. E. (2017). Biophysical Mechanisms Mediating Fibrin Fiber Lysis. In *BioMed Research International* (Vol. 2017). Hindawi Limited. https://doi.org/10.1155/2017/2748340

Ikeda, M., Kobayashi, T., Arai, S., Mukai, S., Takezawa, Y., Terasawa, F., & Okumura, N. (2014). Recombinant γT305A fibrinogen indicates severely impaired fibrin polymerization due to the aberrant function of hole 'a'and calcium binding sites. *Thrombosis Research*, *134*(2), 518–525.

Kagami, K., Yamazaki, R., Minami, T., Okumura, N., Morishita, E., & Fujiwara, H. (2016). Familial discrepancy of clinical outcomes associated with fibrinogen Dorfen: A case of huge genital hematoma after episiotomy. *Journal of Obstetrics and Gynaecology Research*, *42*(6), 722–725.

Kamijo, T., Kaido, T., Yoda, M., Arai, S., Yamauchi, K., & Okumura, N. (2021). Recombinant γY278H fibrinogen showed normal secretion from CHO cells, but a corresponding heterozygous patient showed hypofibrinogenemia. *International Journal of Molecular Sciences*, *22*(10), 5218.

Kattula, S., Byrnes, J. R., & Wolberg, A. S. (2017). Fibrinogen and Fibrin in Hemostasis and Thrombosis. In *Arteriosclerosis, Thrombosis, and Vascular Biology* (Vol. 37, Issue 3,

pp. e13–e21). Lippincott Williams and Wilkins. https://doi.org/10.1161/ATVBAHA.117.308564

Kent, W. J. (2002). BLAT —The BLAST -Like Alignment Tool . *Genome Research*, *12*(4), 656–664. https://doi.org/10.1101/gr.229202

Kollman, J. M., Pandi, L., Sawaya, M. R., Riley, M., & Doolittle, R. F. (2009). Crystal Structure of Human Fibrinogen. *Biochemistry*, *48*(18), 3877–3886. https://doi.org/10.1021/bi802205g

Kotlín, R., Pastva, O., Štikarová, J., Hlaváčková, A., Suttnar, J., Chrastinová, L., Riedel, T., Salaj, P., & Dyr, J. E. (2014). Two novel mutations in the fibrinogen γ nodule. *Thrombosis Research*, *134*(4), 901–908.

Kotlín, R., Reicheltová, Z., Malý, M., Suttnar, J., Sobotková, A., Salaj, P., Hirmerová, J., Riedel, T., & Dyr, J. E. (2009). Two cases of congenital dysfibrinogenemia associated with thrombosis–Fibrinogen Praha III and Fibrinogen Plzeň. *Thrombosis and Haemostasis*, *102*(09), 479–486.

Kotlín, R., Sobotková, A., Suttnar, J., Salaj, P., Walterová, L., Riedel, T., Reicheltová, Z., & Dyr, J. E. (2008). A novel fibrinogen variant–Liberec: dysfibrinogenaemia associated with γ Tyr262Cys substitution. *European Journal of Haematology*, *81*(2), 123–129.

Kumar, R., Dawson, J., Varga, E., Canini, J. T., Monda, K. L., & Dunn, A. L. (2019). Fibrinogen Columbus II: A novel c. 1075G> T mutation in the FGG gene causing hypodysfibrinogenemia and thrombosis in an adolescent male. *Pediatric Blood & Cancer*, *66*(9), e27832.

Landrum, M. J., Lee, J. M., Riley, G. R., Jang, W., Rubinstein, W. S., Church, D. M., & Maglott, D. R. (2014). ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research*, *42*(D1). https://doi.org/10.1093/nar/gkt1113

Liao, Z., Tang, H., Xie, Y., Duan, X., Xu, S., Liu, C., Cheng, Y., Chen, Y., Tan, Y., & Wang, D. (2014). Fibrinogen Hangzhou: congenital dysfibrinogenemia caused by the novel missense mutation in FGG (γ308Asn→ Thr). *Clinica Chimica Acta; International Journal of Clinical Chemistry*, *428*, 106–109.

Liu, X., Li, C., Mou, C., Dong, Y., & Tu, Y. (2020). dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Genome Medicine*, *12*(1). https://doi.org/10.1186/s13073-020-00803-9

Livesey, B. J., & Marsh, J. A. (2022a). Interpreting protein variant effects with computational predictors and deep mutational scanning. In *DMM Disease Models and Mechanisms* (Vol. 15, Issue 6). Company of Biologists Ltd. https://doi.org/10.1242/DMM.049510

Livesey, B. J., & Marsh, J. A. (2022b). Interpreting protein variant effects with computational predictors and deep mutational scanning. In *DMM Disease Models and Mechanisms* (Vol. 15, Issue 6). Company of Biologists Ltd. https://doi.org/10.1242/DMM.049510

López-Ferrando, V., Gazzo, A., De La Cruz, X., Orozco, M., & Gelpí, J. L. (2017). PMut: A web-based tool for the annotation of pathological variants on proteins, 2017 update. *Nucleic Acids Research*, *45*(W1), W222–W228. https://doi.org/10.1093/nar/gkx313

López-Ferrando, V., Gazzo, A., de la Cruz, X., Orozco, M., & Gelpí, J. L. (2017). PMut: a web-based tool for the annotation of pathological variants on proteins, 2017 update. *Nucleic Acids Research*, *45*(W1), W222–W228. https://doi.org/10.1093/nar/gkx313

Lounes, K. C., Soria, C., Valognes, A., Turchini, M. F., Soria, J., & Koopman, J. (1999). Fibrinogen Bastia (γ 318 Asp→ Tyr) a novel abnormal fibrinogen characterized by defective fibrin polymerization. *Thrombosis and Haemostasis*, *82*(12), 1639–1643.

Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M., & Cai, L. (2014). Single-cell in situ RNA profiling by sequential hybridization. In *Nature Methods* (Vol. 11, Issue 4, pp. 360–361). Nature Publishing Group. https://doi.org/10.1038/nmeth.2892

Luo, S., Xu, Q., Xie, Y., Li, X., Jin, Y., Yang, L., Liu, S., & Wang, M. (2020). A novel heterozygous mutation (γIIe367Thr) causes congenital dysfibrinogenemia in a Chinese family. *Blood Coagulation & Fibrinolysis*, *31*(8), 569–574.

Marini, N. J., Thomas, P. D., & Rine, J. (2010). The use of orthologous sequences to predict the impact of amino acid substitutions on protein function. *PLoS Genetics*, *6*(5), 3. https://doi.org/10.1371/journal.pgen.1000968

Medved, L., & Weisel, J. W. (2009). Recommendations for nomenclature on fibrinogen and fibrin. *Journal of Thrombosis and Haemostasis*, *7*(2), 355–359. https://doi.org/10.1111/j.1538-7836.2008.03242.x

Meyer, M., Bergmann, F., & Brennan, S. O. (2006). Novel fibrinogen mutation (γ 313 Ser→ Asn) associated with hypofibrinogenemia in two unrelated families. *Blood Coagulation & Fibrinolysis*, *17*(1), 63–67.

Meyer, M., Franke, K., Richter, W., Steiniger, F., Seyfert, U. T., Schenk, J., Treuner, J., Haberbosch, W., Eisert, R., & Barthels, M. (2003a). New molecular defects in the γ subdomain of fibrinogen D-domain in four cases of (hypo) dysfibrinogenemia: fibrinogen variants Hannover VI, Homburg VII, Stuttgart and Suhl. *Thrombosis and Haemostasis*, *89*(04), 637–646.

Meyer, M., Franke, K., Richter, W., Steiniger, F., Seyfert, U. T., Schenk, J., Treuner, J., Haberbosch, W., Eisert, R., & Barthels, M. (2003b). New molecular defects in the γ subdomain of fibrinogen D-domain in four cases of (hypo) dysfibrinogenemia: fibrinogen variants Hannover VI, Homburg VII, Stuttgart and Suhl. *Thrombosis and Haemostasis*, *89*(04), 637–646.

Miesbach, W., Scharrer, I., Henschen, A., Neerman-Arbez, M., Spitzer, S., & Galanakis, D. (2010). Inherited dysfibrinogenemia: clinical phenotypes associated with five different fibrinogen structure defects. *Blood Coagulation & Fibrinolysis*, *21*(1), 35–40.

Mimuro, J., Kawata, Y., Niwa, K., Muramatsu, S., Madoiwa, S., Takano, H., Sugo, T., Sakata, Y., Sugimoto, T., & Nose, K. (1999). A new type of Ser substitution for γ Arg-275 in fibrinogen Kamogawa I characterized by impaired fibrin assembly. *Thrombosis and Haemostasis*, *81*(06), 940–944.

Mottaz, A., David, F. P. A., Veuthey, A. L., & Yip, Y. L. (2010). Easy retrieval of single amino-acid polymorphisms and phenotype information using SwissVar. *Bioinformatics*, *26*(6), 851–852. https://doi.org/10.1093/bioinformatics/btq028

Mount, D. W. (2008). Using BLOSUM in sequence alignments. *Cold Spring Harbor Protocols*, *3*(6). https://doi.org/10.1101/pdb.top39

Mukaddam, A., Kulkarni, B., Jadli, A., Ghosh, K., & Shetty, S. (2015). Spectrum of mutations in Indian patients with fibrinogen disorders and its application in genetic diagnosis of the affected families. *Haemophilia*, *21*(6), e519–e523.

Mukai, S., Ikeda, M., Takezawa, Y., Sugano, M., Honda, T., & Okumura, N. (2015). Differences in the function and secretion of congenital aberrant fibrinogenemia between heterozygous γD320G (Okayama II) and γΔN319-ΔD320 (Otsu I). *Thrombosis Research*, *136*(6), 1318–1324.

Mullin, J. L., Brennan, S. O., Ganly, P. S., & George, P. M. (2002). Fibrinogen Hillsborough: a novel γGly309Asp dysfibrinogen with impaired clotting. *Blood, The Journal of the American Society of Hematology*, *99*(10), 3597–3601.

Neerman-Arbez, M., De Moerloose, P., & Casini, A. (2016). Laboratory and Genetic Investigation of Mutations Accounting for Congenital Fibrinogen Disorders. *Seminars in Thrombosis and Hemostasis*, *42*(4), 356–365. https://doi.org/10.1055/s-0036-1571340

Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research*, *31*(13), 3812–3814. https://doi.org/10.1093/nar/gkg509

Niroula, A., & Vihinen, M. (2016). Variation Interpretation Predictors: Principles, Types, Performance, and Choice. In *Human Mutation* (Vol. 37, Issue 6, pp. 579–597). John Wiley and Sons Inc. https://doi.org/10.1002/humu.22987

Niwa, K., Takebe, M., Sugo, T., Kawata, Y., Mimuro, J., Asakura, S., Sakata, Y., Mizushima, J., Maeda, A., & Endo, H. (1996). *A gamma Gly-268 to Glu substitution is responsible for impaired fibrin assembly in a homozygous dysfibrinogen Kurashiki I*.

Palta, S., Saroa, R., & Palta, A. (2014). Overview of the coagulation system. In *Indian Journal of Anaesthesia* (Vol. 58, Issue 5, pp. 515–523). Indian Society of Anaesthetists. https://doi.org/10.4103/0019-5049.144643

Pejaver, V., Byrne, A. B., Feng, B. J., Pagel, K. A., Mooney, S. D., Karchin, R., O'Donnell-Luria, A., Harrison, S. M., Tavtigian, S. V., Greenblatt, M. S., Biesecker, L. G., Radivojac, P., Brenner, S. E., Tayoun, A. A., Berg, J. S., Cutting, G. R., Ellard, S., Kang, P., Karbassi, I., … Topper, S. (2022). Calibration of computational tools for missense variant pathogenicity classification and ClinGen recommendations for PP3/BP4 criteria. *American Journal of Human Genetics*, *109*(12), 2163–2177. https://doi.org/10.1016/j.ajhg.2022.10.013

Pellegrino, E., Jacques, C., Beaufils, N., Nanni, I., Carlioz, A., Metellus, P., & Ouafik, L. H. (2021). Machine learning random forest for predicting oncosomatic variant NGS analysis. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-01253-y

Platè, M., Asselta, R., Spena, S., Spreafico, M., Fagoonee, S., Peyvandi, F., Tenchini, M. L., & Duga, S. (2008). Congenital hypofibrinogenemia: Characterization of two missense mutations affecting fibrinogen assembly and secretion. *Blood Cells, Molecules, and Diseases*, *41*(3), 292–297. https://doi.org/10.1016/j.bcmd.2008.06.004

Puls, F., Goldschmidt, I., Bantel, H., Agne, C., Bröcker, V., Dämmrich, M., Lehmann, U., Berrang, J., Pfister, E.-D., & Kreipe, H. H. (2013). Autophagy-enhancing drug carbamazepine diminishes hepatocellular death in fibrinogen storage disease. *Journal of Hepatology*, *59*(3), 626–630.

Redman, C. M., & Xia, H. (2001). Fibrinogen biosynthesis: Assembly, intracellular degradation, and association with lipid synthesis and secretion. *Annals of the New York Academy of Sciences*, *936*, 480–495. https://doi.org/10.1111/j.1749-6632.2001.tb03535.x

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., & Rehm, H. L. (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

Ridgway, H. J., Brennan, S. O., Loreth, R. M., & George, P. M. (1997). Fibrinogen Kaiserslautern (γ 380 Lys to Asn): A new glycosylated fibrinogen variant with delayed polymerization. *British Journal of Haematology*, *99*(3), 563–569.

Rigatti, S. J. (2017). Random forest. *Journal of Insurance Medicine*, *47*(1), 31–39.

Robert-Ebadi, H., De Moerloose, P., El Khorassani, M., El Khattab, M., & Neerman-Arbez, M. (2009). A novel frameshift mutation in FGA accounting for congenital afibrinogenemia predicted to encode an aberrant peptide terminating 158 amino acids downstream. *Blood Coagulation & Fibrinolysis*, *20*(5), 385–387.

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S.,

Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., … Sherry, S. T. (2022). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, *50*(D1), D20–D26. https://doi.org/10.1093/nar/gkab1112

scheraga1957. (n.d.).

Shapiro, S. E., Phillips, E., Manning, R. A., Morse, C. V, Murden, S. L., Laffan, M. A., & Mumford, A. D. (2013). Clinical phenotype, laboratory features and genotype of 35 patients with heritable dysfibrinogenaemia. *British Journal of Haematology*, *160*(2), 220–227.

Sheen, C. R., Low, J., Joseph, J., Kotlyar, E., George, P. M., & Brennan, S. O. (2006). Fibrinogen Darlinghurst: hypofibrinogenaemia caused by a W253G mutation in the gamma chain in a patient with both bleeding and thrombotic complications. *Thrombosis and Haemostasis*, *96*(11), 685–687.

Sherry, S. T., Ward, M.-H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001a). dbSNP: the NCBI database of genetic variation. In *Nucleic Acids Research* (Vol. 29, Issue 1). http://www.ncbi.nlm.nih.gov/SNP.

Sherry, S. T., Ward, M.-H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001b). dbSNP: the NCBI database of genetic variation. In *Nucleic Acids Research* (Vol. 29, Issue 1). http://www.ncbi.nlm.nih.gov/SNP.

Sillen, M., & Declerck, P. J. (2021). Thrombin activatable fibrinolysis inhibitor (Tafi): An updated narrative review. In *International Journal of Molecular Sciences* (Vol. 22, Issue 7). MDPI. https://doi.org/10.3390/ijms22073670

Smith, N., Bornikova, L., Noetzli, L., Guglielmone, H., Minoldo, S., Backos, D. S., Jacobson, L., Thornburg, C. D., Escobar, M., & White-Adams, T. C. (2018). Identification and characterization of novel mutations implicated in congenital fibrinogen disorders. *Research and Practice in Thrombosis and Haemostasis*, *2*(4), 800–811.

Song, K. S., Park, N. J., Choi, J. R., Doh, H. J., & Chung, K. H. (2006). Fibrinogen Seoul (FGG Ala341Asp): a novel mutation associated with hypodysfibrinogenemia. *Clinical and Applied Thrombosis/Hemostasis*, *12*(3), 338–343.

Sovova, Z., Pecankova, K., Majek, P., & Suttnar, J. (2022). Extension of the human fibrinogen database with detailed clinical information—The αc-connector segment. *International Journal of Molecular Sciences*, *23*(1). https://doi.org/10.3390/ijms23010132

Standeven, K. F., Carter, A. M., Grant, P. J., Weisel, J. W., Chernysh, I., Masova, L., Lord, S. T., & Ariëns, R. A. S. (2007). Functional analysis of fibrin γ-chain cross-linking by activated factor XIII: Determination of a cross-linking pattern that maximizes clot stiffness. *Blood*, *110*(3), 902–907. https://doi.org/10.1182/blood-2007-01-066837

Stefl, S., Nishi, H., Petukh, M., Panchenko, A. R., & Alexov, E. (2013). Molecular mechanisms of disease-causing missense mutations. In *Journal of Molecular Biology* (Vol. 425, Issue 21, pp. 3919–3936). Academic Press. https://doi.org/10.1016/j.jmb.2013.07.014

Steinhaus, R., Proft, S., Schuelke, M., Cooper, D. N., Schwarz, J. M., & Seelow, D. (2021). MutationTaster2021. *Nucleic Acids Research*, *49*(W1), W446–W451. https://doi.org/10.1093/nar/gkab266

Steinmann, C., Bogli, C., Jungo, M., Lammle, B., Heinemann, G., Wermuth, B., Redaelli, R., Baudo, F., & Furlan, M. (1994). *A new substitution, gamma 358 Ser--> Cys, in fibrinogen Milano VII causes defective fibrin polymerization*.

Stenson, P. D., Ball, E. V., Mort, M., Phillips, A. D., Shiel, J. A., Thomas, N. S. T., Abeysinghe, S., Krawczak, M., & Cooper, D. N. (2003). Human Gene Mutation Database (HGMD®): 2003 Update. In *Human Mutation* (Vol. 21, Issue 6, pp. 577–581). https://doi.org/10.1002/humu.10212

Suskiewicz, M. J. (2024). The logic of protein post-translational modifications (PTMs): Chemistry, mechanisms and evolution of protein regulation through covalent attachments. In *BioEssays* (Vol. 46, Issue 3). John Wiley and Sons Inc. https://doi.org/10.1002/bies.202300178

Tang, H., & Thomas, P. D. (2016). PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation. *Bioinformatics*, *32*(14), 2230–2232. https://doi.org/10.1093/bioinformatics/btw222

Terasawa, F., Okumura, N., Kitano, K., Hayashida, N., Shimosaka, M., Okazaki, M., & Lord, S. T. (1999). Hypofibrinogenemia associated with a heterozygous missense mutation γ153Cys to Arg (Matsumoto IV): in vitro expression demonstrates defective secretion of the variant fibrinogen. *Blood, The Journal of the American Society of Hematology*, *94*(12), 4122–4131.

Thomas, P. D., Campbell, M. J., Kejariwal, A., & Mi, H. (2003). PANTHER: a library of protein families and subfamilies indexed by function. *Genome …*, *13*.

Thromb, A., & Biol, V. (2017). *Endogenous Mediators of Thrombin Generation and Fibrin Downloaded from*. http://atvb.ahajournals.org/

Thusberg, J., Olatubosun, A., & Vihinen, M. (2011). Performance of mutation pathogenicity prediction methods on missense variants. *Human Mutation*, *32*(4), 358–368. https://doi.org/10.1002/humu.21445

Tiscia, G. L., & Margaglione, M. (2018). Human fibrinogen: Molecular and genetic aspects of congenital disorders. In *International Journal of Molecular Sciences* (Vol. 19, Issue 6). MDPI AG. https://doi.org/10.3390/ijms19061597

Treliński, J., Witkowski, M., Chojnowski, K., Neerman-Arbez, M., Wypasek, E., & Undas, A. (2019). Fibrinogen Łódź: a new cause of dysfibrinogenemia associated with recurrent thromboembolic arterial events. *Polskie Archiwum Medycyny Wewnętrznej= Polish Archives of Internal Medicine*, *129*(12).

Undas, A., Zdziarska, J., Iwaniec, T., Stepien, E., Skotnicki, A. B., de Moerloose, P., & Neerman-Arbez, M. (2009). Fibrinogen Krakow: a novel hypo/dysfibrinogenemia mutation in fibrinogen gamma chain (Asn325Ile) affecting fibrin clot structure and function. *Thrombosis and Haemostasis*, *101*(05), 975–976.

Ushijima, A., Komai, T., Masukawa, A., Oikawa, K., Morita, N., Asai, S., Mukai, S., Okumura, N., Kobayashi, Y., & Miyachi, H. (2017). Hypodysfibrinogenemia with a heterozygous mutation of γCys326Ser by the novel transversion of TGT to TCT in a patient with pulmonary thromboembolism and right ventricular thrombus. *Cardiology*, *137*(3), 167–172.

van der Vorm, L., Remijn, J., de Laat, B., & Huskens, D. (2018). Effects of Plasmin on von Willebrand Factor and Platelets: A Narrative Review. *TH Open*, *02*(02), e218–e228. https://doi.org/10.1055/s-0038-1660505

Vu, D., De Moerloose, P., Batorova, A., Lazur, J., Palumbo, L., & Neerman-Arbez, M. (2005). Hypofibrinogenaemia caused by a novel FGG missense mutation (W253C) in the γ chain globular domain impairing fibrinogen secretion. *Journal of Medical Genetics*, *42*(9), e57–e57.

Wang, B. D., & Lee, N. H. (2018). Aberrant RNA splicing in cancer and drug resistance. In *Cancers* (Vol. 10, Issue 11). MDPI AG. https://doi.org/10.3390/cancers10110458

Wang, Y., Chen, W., Ma, P., Zhu, L., & Wang, M. (2018). Clinical and molecular characterization of nine Chinese patients affected by hypofibrinogenemia or dysfibrinogenemia. *Blood Coagulation & Fibrinolysis*, *29*(4), 404–409.

Wang, Y., Zhu, L., Hao, X., Xie, Y., Jin, Y., & Wang, M. (2014). Unique de-novo mutation of fibrinogen gene in a Chinese girl with hypofibrinogenemia. *Blood Coagulation & Fibrinolysis*, *25*(7), 780–782.

Wei, A., Wu, Y., Xiang, L., Yan, J., Cheng, P., Deng, D., & Lin, F. (2021). Congenital dysfibrinogenemia caused by γAla327Val mutation: Structural abnormality of D region. *Hematology*, *26*(1), 305–311.

Weisel, J. W. (2005). *FIBRINOGEN AND FIBRIN*. https://doi.org/10.1016/S0065-3233(04)70008-X

Weisel, J. W., & Litvinov, R. I. (2017). Fibrin formation, structure and properties. *Sub-Cellular Biochemistry*, *82*, 405–456. https://doi.org/10.1007/978-3-319-49674-0_13

whitley2002. (n.d.).

Wildeman, M., Van Ophuizen, E., Den Dunnen, J. T., & Taschner, P. E. M. (2008). Improving sequence variant descriptions in mutation databases and literature using the mutalyzer sequence variation nomenclature checker. *Human Mutation*, *29*(1), 6–13. https://doi.org/10.1002/humu.20654

Xia, H., & Redman, C. (1999). The Degradation of Nascent Fibrinogen Chains Is Mediated by the Ubiquitin Proteasome Pathway. *Biochemical and Biophysical Research Communications*, *261*(3), 590–597. https://doi.org/https://doi.org/10.1006/bbrc.1999.1081

Zhang, Z., Miteva, M. A., Wang, L., & Alexov, E. (2012). Analyzing effects of naturally occurring missense mutations. In *Computational and Mathematical Methods in Medicine* (Vol. 2012). Hindawi Limited. https://doi.org/10.1155/2012/805827

Zhou, J., Ding, Q., Chen, Y., Ouyang, Q., Jiang, L., Dai, J., Lu, Y., Wu, X., Liang, Q., & Wang, H. (2015). Clinical features and molecular basis of 102 Chinese patients with congenital dysfibrinogenemia. *Blood Cells, Molecules, and Diseases*, *55*(4), 308–315.

Zhou, P., Yu, M., Peng, Y., Ma, P., & Wan, L. (2021). Identification and characterization of novel mutations in Chinese patients with congenital fibrinogen disorders. *Blood Cells, Molecules, and Diseases*, *86*, 102489.

Zhu, L., Wang, M., Xie, H., Jin, Y., Yang, L., & Xu, P. (2013). A novel fibrinogen mutation (γ Thr277Arg) causes hereditary hypofibrinogenemia in a Chinese family. *Blood Coagulation & Fibrinolysis*, *24*(6), 642–644.

Zhu, L., Wang, Y., Zhao, M., Hao, X., Xie, H., Xie, Y., Wang, M., & Ding, H. (2014). Novel mutations (γTrp208Leu and γLys232Thr) leading to congenital hypofibrinogenemia in two unrelated Chinese families. *Blood Coagulation & Fibrinolysis*, *25*(8), 894–897.

7 Attachments

7.1 Clinical manifestation of the mutations

Program	c.DNA	Ensembl	Pac1	Pac2	Pac3	Pac4	Pac5
A305D	c.914C>A	4:154606920	meno+U1:Y70rrhagia	post-traumatic bleeding			
A315D	c.944C>A	4:154606890	stroke, spontaneous + induced bleeding	DVT, mild spontaneous bleeding			
A315V	c.944C>T	4:154606890	bl during pregnancy and post-partum, after tooth extraction	asym	post-partum bleeding		
A353T	c.1057G>A	4:154606777	cerebral infarction, PE, thrombosis, hypodys, hetero				
A353V	c.1058C>T	4:154606776	asym				
A367D	c.1100C>A	4:154606734	arterial thrombosis	asym	asym		
A367P	c.1099G>C	4:154606735	bleeding				
A367V	c.1100C>T	4:154606734	thrombotic				
A383T	c.1147G>A	4:154605049	bruising, post-operative bl, menorrhagia	bruising, gum bl	asym	asym	
C179R	c.535T>C	4:154609761	asym				
C352F	c.1055G>T	4:154606779	renal and splenic infarcts	asym			
C352S	c.1054T>A	4:154606780	post-operative bleeding, bruising, intracranial bl	asym	asym	PE	
C352S	c.1055G>C	4:154606779	PE, DVT	asym	PE		
C352Y	c.1055G>A	4:154606779	PE	DVT			
C365S	c.1094G>C	4:154606740	spontaneous bl, post-partum bl	bleeding			
D342G	c.1025A>G	4:154606809	menorrhagia	menorrhagia	menorrhagia		
D342H	c.1024G>C	4:154606810	asym				
D342N	c.1024G>A	4:154606810	FSD				
D344V	c.1031A>T	4:154606803	post-operative venous thrombosis				
D344Y	c.1030G>T	4:154606804	asym				
D346A	c.1037A>C	4:154606797	asym				
D346E	c.1038T>A	4:154606796	post-traumatic bl (venipuncture) since pregnency				
D346G	c.1037A>G	4:154606797	asym	asym			
D390N	c.1168G>A	4:154605028	asym				
D390V	c.1169A>T	4:154605027	DVT, arterial thrombosis				
E239A	c.716A>C	4:154608601	prolonged bleeding from cuts, bruising				
F307L	c.921C>G	4:154606913	asym				
G191V	c.572G>T	4:154609724	asym				
G294E	c.881G>A	4:154606953	asym				homo
G310R	c.928G>C	4:154606906	FSD	epistaxis, FSD			

G335C	c.1003G>T	4:154606831	asym				
G335D	c.1004G>A	4:154606830	asym				
G359C	c.1075G>T	4:154606759	DVT	associated DVT			
G392S	c.1174G>A	4:154605022	FSD				
H333R	c.998A>G	4:154606836	post-partum bleeding	thromboembolism, miscarriages			
H333Y	c.997C>T	4:154606837	post-partum bleeding	bleeding	thrombosis	thromb	stroke
H366N	c.1096C>G	4:154606738	FSD				
I105T	c.314T>C	4:154611892	bl				
I393T	c.1178T>C	4:154605018	menorrhagia				
K258T	c.773A>C	4:154608544	menorrhagia	asym			
K406N	c.1218G>T	4:154604978	post-operative thrombosis	asym	asym	asym	asym
L370F	c.1108C>T	4:154606726	post-operative bl				
M362K	c.1085T>A	4:154606749	stroke				
N256D	c.766A>G	4:154608551	miscarriage followed by prolonged bleeding				
N256H	c.766A>C	4:154608551	asym				
N256K	c.768T>G	4:154608549	severe nose bleeding, stillbirth with fetal thrombotic vasculopathy				
N334T	c.1001A>C	4:154606833	asym	asym			
N351I	c.1052A>T	4:154606782	post-operative DVT				
N371D	c.1111A>G	4:154606723	asym				
N371S	c.1112A>G	4:154606722	DVT, PE	post-operative bl, hematuria (after extreme sport activity)			
N391K	c.1173T>A	4:154605023	minor bl				
R301S	c.901C>A	4:154606933	asym				
R401L	c.1202G>T	4:154604994	asym				
S339G	c.1015A>G	4:154606819	thrombosis, bruising				
S339N	c.1016G>A	4:154606818	asym	asym	asym	asym	
S339R	c.1015A>C	4:154606819	epistaxis, bruising	bruising, vaginal bleeding during			
S358C	c.1073C>G	4:154606761	thrombosis, minor bleeding				
S384C	c.1151C>G	4:154605045	miscarriage	asym	asym	asym	asym
T303R	c.908C>G	4:154606926	asym				
T331A	c.991A>G	4:154606843	asym				
T340P	c.1018A>C	4:154606816	FSD				
W234L	c.701G>T	4:154608616	asym	easy bruising			
W253C	c.759G>T	4:154608558	spontaneous and induced bleeding, miscarriage				
W279C	c.837G>C	4:154608480	thrombotic				
W279G	c.835T>G	4:154608482	bleeding, thrombosis, miscarriages				homo
W395L	c.1184G>T	4:154605012	bleeding				
Y288C	c.863A>G	4:154606971	bruising				
Y304H	c.910T>C	4:154606924	post-partum bleeding, miscarriage				
Y380C	c.1139A>G	4:154605057	DVT				
Y389N	c.1165T>A	4:154605031	DVT, PE				

Table 5 Clinical manifestaions of the mutations

7.2 Performance of the programs

Substitution	PANTHER - PSEP	Panther	PMut	SnP&GO	PhD_SNP	SIFT 0.05	SIFT	Mutation Taster21	Mutaion Taster2014	PolyPhen2	Provean
A305D	0,19	0,692	0.1988	0,957	0,588	0,029	0,029	95	0,999991546	possibly damaging	-1.402
A315D	0,5	0,692	0.6911	0,951	0,79	0,005	0,005	96	0,999999999920006	probably damaging	-4.056
A315V	0,5	0,554	0.6871	0,892	0,657	0,002	0,002	97	0,999999999655284	probably damaging	-2.233
A353T	0,57	0,49	0.6501	0,896	0,802	0,037	0,037	77	0,999999999992831	probably damaging	-2.465
A353V	0,57	0,545	0.6501	0,831	0,725	0,003	0,003	94	0,999999999994019	probably damaging	-3.346
A367D	0,57	0,684	0.5010	0,892	0,492	0	0	99	0,99999999999709	probably damaging	-4.175
A367P	0,57	0,685	0.4922	0,915	0,663	0,001	0,001	97	0,999999999995209	probably damaging	-3.693
A367V	0,57	0,545	0.4922	0,77	0,332	0,001	0,001	97	0,999999999987462	probably damaging	-3.208
A383T	0,57	0,188	0.0557	0,295	0,223	0,682	0,682	13	0,9947216564497	benign	0.371
C179R	0,57	0,924	0.9107	0,953	0,948	0	0	97	1,0	probably damaging	-9.548
C352F	0,57	0,931	0.8999	0,932	0,918	0	0	100	0,999999999999928	probably damaging	-9.949
C352S	0,57	0,821	0.9055	0,9	0,864	0	0	100	0,999999999997709	probably damaging	-8.878
C352S	0,57		0.9055			0	0	100	0,999999999999579	probably damaging	-8.878
C352Y	0,57	0,944	0.8999	0,928	0,92	0	0	100	0,999999999999805	probably damaging	-9.883
C365S	0,57	0,821	0.9055	0,925	0,904	0,002	0,002	99	0,999999999999078	probably damaging	-8.441
D342G	0,57	0,801	0.8830	0,928	0,836	0,003	0,003	100	0,999999999950274	probably damaging	-6.216
D342H	0,57	0,9	0.9107	0,91	0,878	0	0	100	0,999999999993704	probably damaging	-6.116
D342N	0,57	0,762	0.7570	0,932	0,855	0,001	0,001	99	0,999999999967447	probably damaging	-4.397
D344V	0,57	0,913	0.8999	0,95	0,928	0,002	0,002	98	0,999999999999862	probably damaging	-8.185
D344Y	0,57	0,937	0.8570	0,961	0,954	0,001	0,001	97	0,999999999037483	probably damaging	-8.185
D346A	0,57	0,811	0.8570	0,914	0,852	0,008	0,008	94	0,999999999930367	probably damaging	-7.193
D346E	0,57	0,658	0.8184	0,895	0,788	0,095	0,095	98	0,999996835789626	probably damaging	-3.633
D346G	0,57	0,801	0.8184	0,925	0,854	0,005	0,005	100	0,999999999950858	probably damaging	-6.306
D390N	0,57	0,763	0.6738	0,862	0,393	0,017	0,017	51	0,999999999802725	probably damaging	-4.148
D390V	0,57	0,913	0.7299	0,881	0,492	0,008	0,008	99	0,999999999999969	probably damaging	-7.097
E239A	0,57	0,57	0.4810	0,641	0,173	0,003	0,003	97	0,999999997308363	probably damaging	-3.247
F307L	0,57	0,416	0.5519	0,799	0,559	0,001	0,001	50	0,998177442272407	possibly damaging	-3.606
G191V	0,57	0,785	0.8715	0,922	0,913	0,055	0,055	100	0,999999999999999	probably damaging	-4.716
G294E	0,57	0,539	0.4385	0,771	0,5	0,052	0,052	92	0,999998094047148	possibly damaging	-2.357
G310R	0,57	0,923	0.8962	0,9	0,875	0,003	0,003	100	0,999999999916227	probably damaging	-3.431
G335C	0,57	0,87	0.8550	0,92	0,849	0,056	0,056	99	0,999999999990757	probably damaging	-7.587
G335D	0,57	0,754	0.4405	0,881	0,681	0,059	0,059	89	0,999999999985375	probably damaging	-5.816
G359C	0,57	0,944	0.9107	0,95	0,935	0,002	0,002	100	0,999999999803975	probably damaging	-7.733
G392S	0,57	0,829	0.7652	0,873	0,781	0	0	100	0,999999999994319	probably damaging	-4.990
H333R	0,57	0,842	0.8999	0,929	0,779	0,002	0,002	99	0,999999998582905	probably damaging	-6.701
H333Y	0,78	0,826	0.8520	0,911	0,746	0,003	0,003	97	0,999999999573312	probably damaging	-5.057
H366N	0,78	0,875	0.7652	0,951	0,836	0,03	0,03	100	0,999999999873907	probably damaging	-5.623
I105T	0,78	0,377	0.3513	0,474	0,252	0,095	0,095	9	0,999998301895412	probably damaging	-1.152

I393T	0,78	0,584	0.7823	0,819	0,704	0,001	0,001	85	0,999999999562387	probably damaging	-3.517
K258T	0,78	0,625	0.5131	0,778	0,502	0,023	0,023	100	0,999999999382463	probably damaging	-3.796
K406N	0,85	0,599	0.7581	0,727	0,786	0,002	0,002	100	0,999999998783263	probably damaging	-3.630
L370F	0,85	0,541	0.6620	0,914	0,875	0,005	0,005	86	0,999999993097948	probably damaging	-2.601
M362K	0,85	0,651	0.7083	0,935	0,682	0	0	92	0,999999999813922	probably damaging	-3.205
N256D	0,85	0,719	0.8715	0,901	0,663	0,013	0,013	99	0,999999997823785	probably damaging	-4.232
N256H	0,85	0,9	0.8715	0,889	0,676	0,469	0,469	99	0,999999999007694	probably damaging	-4.232
N256K	0,85	0,795	0.8715	0,926	0,738	0,004	0,004	100	0,999999945689623	probably damaging	-5.078
N334T	0,85	0,812	0.6888	0,862	0,484	0,076	0,076	97	0,999957135987378	possibly damaging	-4.396
N351I	0,85	0,589	0.8494	0,8	0,712	0,001	0,001	83	0,999998340179717	probably damaging	-2.601
N371D	0,85	0,716	0.8999	0,923	0,841	0,052	0,052	98	0,999999986587345	probably damaging	-4.247
N371S	0,85	0,753	0.9055	0,928	0,853	0	0	100	0,999999908666767	probably damaging	-4.244
N391K	0,85	0,391	0.5664	0,611	0,261	0,007	0,007	96	0,99997134060394	possibly damaging	-3.588
R301S	0,85	0,643	0.5752	0,852	0,692	0,006	0,006	100	0,999999980122279	probably damaging	-2.752
R401L	0,85	0,672	0.4692	0,846	0,524	0,055	0,055	96	0,999999999471364	probably damaging	-2.833
S339G	0,85	0,7	0.8570	0,809	0,598	0,001	0,001	98	0,99999887629952	probably damaging	-3.604
S339N	0,85	0,799	0.8570	0,875	0,76	0,003	0,003	94	0,999999325368493	probably damaging	-2.662
S339R	0,85	0,844	0.8570	0,922	0,848	0,003	0,003	81	0,999999780250065	probably damaging	-4.447
S358C	0,85	0,664	0.7916	0,841	0,726	0,026	0,026	99	0,999999999957631	probably damaging	-3.538
S384C	0,85	Unclassifie	0.4537	0,34	0,513	0,083	0,083	87	0,996270221860272	probably damaging	-1.402
T303R	0,85	0,664	0.5194	0,863	0,331	0,386	0,386	94	0,999999913220628	possibly damaging	-3.089
T331A	0,85	Unclassifie	0.7694	0,133	0,114	0	0	95	0,999999999426608	probably damaging	-3.372
T340P	0,85	0,848	0.9055	0,836	0,677	0	0	99	0,999999999270544	probably damaging	-5.457
W234L	0,85	0,857	0.9070	0,903	0,885	0	0	96	0,999999999996614	probably damaging	-10.891
W253C	0,85	0,961	0.7178	0,911	0,905	0	0	99	1,0	probably damaging	-10.388
W279C	0,85	0,961	0.8056	0,94	0,784	0,001	0,001	99	1,0	probably damaging	-10.230
W279G	0,85	0,893	0.9055	0,935	0,689	0,001	0,001	100	0,999999999968805	probably damaging	-10.329
W395L	0,85	0,999	0.9107	0,831	0,842	0	0	100	0,999999999995178	probably damaging	-10.561
Y288C	0,85	0,947	0.8962	0,915	0,878	0	0	100	0,999999999541062	probably damaging	-7.775
Y304H	0,85	0,674	0.4597	0,827	0,35	0,061	0,061	76	0,999999978193757	probably damaging	-2.977
Y380C	0,85	0,946	0.6865	0,927	0,763	0,029	0,029	97	0,99999999784611	probably damaging	-6.680
Y389N	0,85	0,724	0.6865	0,902	0,71	0,038	0,038	93	0,998875736119652	probably damaging	-2.931

Table 6 Overall perfomance of the programs

7.3 References of the mutations

Substitution	Citation
A305D	(Castaman et al., 2019)
A315D	(Terasawa et al., 1999)
A315V	(Chinni et al., 2019a)
A353T	(Zhu et al., 2014)
A353V	(Mukaddam et al., 2015)
A367D	(Vu et al., 2005)
A367P	(Hanss et al., 2005)
A367V	((Brennan et al., 2005)
A383T	(Asselta, Robusto, et al., 2015)
C179R	(Y. Wang et al., 2018; Zhu et al., 2014)
C352F	(Casini et al., 2015)
C352S	(Sheen et al., 2006)
C352S	(Kotlín et al., 2008)
C352Y	(Niwa et al., 1996)
C365S	(Mimuro et al., 1999)
D342G	(Zhu et al., 2013)
D342H	(Kamijo et al., 2021)
D342N	(Brennan, Wyatt, Ockelford, et al., 2000; Castaman et al., 2019)
D344V	Zhou et al., 2021a)
D344Y	(Brennan, Wyatt, Medicina, et al., 2000; Puls et al., 2013)
D346A	(Smith et al., 2018)
D346E	(Dear et al., 2004; Kagami et al., 2016)
D346G	(Ikeda et al., 2014)
D390N	(Dear et al., 2004)
D390V	(Chinni et al., 2019b; Treliński et al., 2019)
E239A	(Liao et al., 2014; J. Zhou et al., 2015)
F307L	(Mullin et al., 2002)
G191V	(J. Zhou et al., 2015)
G294E	(Meyer et al., 2006)
G310R	(Brennan et al., 2010a)
G335C	(Kotlín et al., 2014)
G335D	(Brennan et al., 2010b)
G359C	(Y. Wang et al., 2014)
G392S	(Asselta et al., 2015b)
H333R	(Shapiro et al., 2013)
H333Y	(Lounes et al., 1999)
H366D	(Robert-Ebadi et al., 2009)

I105T	(Brennan & Laurie, 2014)
I393T	(Castaman et al., 2019; Mukai et al., 2015)
K258T	(Galanakis et al., 2014)
K406N	(Undas et al., 2009)
L370F	(Guglielmone et al., 2004)
M362K	(Ushijima et al., 2017)
N256D	(Dear et al., 2005; Meyer et al., 2003a)
N256H	(Hanss et al., 2005)
N256K	(Hamano et al., 2004)
N334T	(Wei et al., 2021)
N351I	(Casini et al., 2015)
N371D	(Kumar et al., 2019)
N371S	(Gindele et al., 2021)
N391K	(Brennan et al., 2015)Castaman et al., 2019)
R301S	(Callea et al., 2017)
R401L	(Song et al., 2006)
S339G	(J. Zhou et al., 2015)
S339N	(Galanakis et al., 2014)
S339R	(Castaman et al., 2019)
S358C	(Meyer et al., 2003b)
S384C	(de Raucourt et al., 2005; van der Vorm et al., 2018)
T303R	(Meyer et al., 2003a)
T331A	(Miesbach et al., 2010)
T340P	(Castaman et al., 2019; Steinmann et al., 1994)
W234L	(Kotlín et al., 2009)
W253C	(Bentolila et al., 1995)
W279C	(P. Zhou et al., 2021)
W279G	(Casini et al., 2015)
W395L	(Asselta, Platè, et al., 2015)
Y288C	(Luo et al., 2020)
Y304H	(Zhou et al., 2015)
Y380C	(Cao et al., 2019; Ridgway et al., 1997)
Y389N	