A Gene Variant-Based and Proteomic-Informed Contrastive Cross-Modal Ensemble Deep Attention Approach for Antithrombotic Candidate Selection and Rationalisation

Tim Dong *, Rhys Llewellyn, Melanie J. Hezzell, Gianni D. Angelini

Posted Date: 17 April 2026

doi: 10.20944/preprints202604.1287.v1

Keywords: deep learning; bioinformatics; drug discovery; transcriptomics; proteomics; thrombotic conditions; machine learning

 doi:10.20944/preprints202604.1287.v1

Article

A Gene Variant-Based and Proteomic-Informed Contrastive Cross-Modal Ensemble Deep Attention Approach for Antithrombotic Candidate Selection and Rationalisation

Tim Dong [1,*], Rhys Llewellyn [2], Melanie J. Hezzell [3] and Gianni D. Angelini [1]

[1] Bristol Heart Institute, Translational Health Sciences, University of Bristol, Bristol BS2 8HW, UK
[2] Pharmacy Department, Liverpool Heart and Chest Hospital, Thomas Dr, Liverpool L14 3PE, UK
[3] Bristol Veterinary School, University of Bristol, Langford House, Langford, Bristol BS40 5DU, UK
* Correspondence: qd18830@bristol.ac.uk

Abstract

Background: Genetic variations such as single nucleotide polymorphisms (SNPs) as part of pharmacogenomics play an important role in the metabolism of drug and hence their active concentrations in blood plasma. **Objectives:** The aim of this study is to select candidate compounds from a TCM dataset that may be repurposed for arterial and venous thromboses management. This shall be achieved through development and evaluation of an ensemble deep learning model that taking into account the genetic variations in protein sequences. **Methods:** BIOSNAP dataset was supplemented with 321,657 drug–target pairs consisting of SNP variants of wild-type proteins. The application dataset consisted of a TCM dataset containing 35,553 ingredients. The control group was set as the pathogenic group, whilst the treatment group was set as the non-pathogenic group. Contrastive and non-contrastive deep cross-modal attention ensemble modelling was developed, evaluated and applied. **Results:** Contrastive regularisation effect improved the performance of the Contrastive Learning (CL) Ensemble over the Non-CL Ensemble model as well as the Dong et al. (May 2025) CL model in the test set (AUPR 0.919 vs. 0.894 vs. 0.813). Safflower yellow A, Paeoniflorin and Notoginsenoside R6 were associated with existing TCM and highly ranked for interaction with Factor Xa genetic variants. Highly interacting protein targets were identified. **Conclusions:** Ensemble modelling with contrastive learning resulted in performance improvements and can be useful for selecting TCM compounds for antithrombotic management. This is a step towards personalised drug selection and can simultaneously facilitate interpretation of the biological rationales for risk vs benefit evaluations during decision making.

Keywords: deep learning; bioinformatics; drug discovery; transcriptomics; proteomics; thrombotic conditions; machine learning

1. Introduction

Artificial intelligence (AI) has the potential to advance phenotyping, procedural safety, and most notably individualised anticoagulation decisions, with early evidence suggesting potential reductions in overtreatment if validated prospectively.[1] Pharmacogenomics is at early stages of being used in healthcare systems but will play an important role in guiding therapy for thrombotic conditions. Genetic variations or polymorphisms such as single nucleotide polymorphisms (SNPs) can result in protein variations in terms of structure, binding, quantity, as well as interaction with other proteins or compounds.[2] Genetic polymorphisms as part of pharmacogenomics play an important role in the metabolism of drug and hence their active concentrations in blood plasma.[3] Specific SNPs that impact direct acting oral anticoagulants (DOACs) have been reported.[4] Hence,

 doi:10.20944/preprints202604.1287.v1

mutations should be taken into account when considering the appropriate dosage and drug to prescribe in order to optimise risk vs. benefit decisions.

Although bleeding risk scores have been developed such as the ABC, HAS-BLED and ORBIT bleeding risk scores,[5] studies on the management of thrombotic conditions have focused mainly on the decoupling thrombosis and haemostasis and novel anticoagulant and antiplatelet therapy development.[6] More importantly, there has been limited inclusion of genetic variation in models to account for interaction changes between proteins and medicinal compounds other than the commonly used DOACs and Vitamin K antagonists. The majority of existing studies considering SNPs have also focused on the use of bioinformatic approaches and Warfarin with limited studies on thrombotic conditions using deep learning with TCM dataset. Traditional Chinese Medicine compounds are often associated with low toxicity effects and their effects on thrombotic conditions have yet to be fully understood by researchers in the West I am. TCM serves a huge population in China and if this vast population in already using it then there would likely be benefit in their use for the Western population. Hence, research on how to translate this benefit should be considered.

1.1. Related Studies in Effect of SNP in Thrombotic Management

Human paraoxonase 1 (PON1) is a protein with key antioxidant effects in terms of preventing oxidation of cholesterol molecules LDL and HDL, as well as supporting the removal of unnecessary cholesterol from the body. It was found that rivaroxaban had lower inhibitory off-target effects on PON1, representing a better candidate for managing thrombotic conditions than dabigatran.[7]

One study that the majority of CYP2C9 missense mutations CYP2C9*1, CYP2C9*11, CYP2C9*2, CYP2C9*3, CYP2C9*5, CYP2C9*6 were predicted to be have an over-anticoagulation effect on warfarin.[8] Dos Santos et al. studied the three different types of the plasminogen activator inhibitors (PAIs), PAI-1, PAI-2, and PAI-3, which play a key role in thrombosis formation through their inhibitory effects on fibrinolysis.[9] They showed that mutations near to the active site resulted in increased rigidity, potential reducing function of the PAI proteins.

Another study looked into the relationship between genetic polymorphism in Thrombomodulin (TM) and effects on development of venous thromboembolism (VTE), for which atrial fibrillation (AF) is also a risk factor.[10] The individuals carrying any one of the qualifying mutations were compared to the non-qualifying group using a cox-regression, showing significant increase in Hazard Ratio of developing VTE.[10]

GGCX is an enzyme that carboxylates the coagulation factors and mutations can cause bleeding disorders due to dysregulation of these factors. One study looked into the effect of mutations in the *GGCX* gene which encodes Vitamin K-dependent γ-glutamyl carboxylase (VKGC).[11]

Another study characterised and identified new mutations in SERPINC1, which encodes antithrombin that is often tested as part of the hypercoagulable workup [12] (or Thrombophilia Screen).[13] Down-regulation of antithrombin is associated with higher risk of venous thrombosis. 5 novel mutations (p.Phe179Cys, p.Gln286Pro, p.Gln286Pro, c.763-1G>A, p.Val327Glyfs*16) were identified from 21 families using direct (targeting SERPINC1 region).[13] In addition, to common mutation variants, splice site mutation was analysed using Human Splicing Finder (HSF) V3.

1.2. Current Clinical Pharmacogenomic (PG) Practice in the NHS

Numerous hospitals in the United Kingdom including that in Bristol and Liverpool have started using CYP2C19 testing to guide Mavacamten drug dosing.[14] This has been supported by the European Medicines Agency, which have advised that all patients should undergo CYP2C19 genotyping before commencing Mavacamten therapy. In Australia, cost of testing for thiopurine methyltransferase (TPMT) gene variants (drugs: azathioprine, mercaptopurine, thioguanine) and HLA-B*57:01 (drug: abacavir) are covered by Australian government.[15] The limited PG testing in the healthcare systems have remained primarily reactive. That is, PG testing is conducted after an unexpected drug-related problem occurs at standard doses.[16]

 doi:10.20944/preprints202604.1287.v1

1.3. Aim

The aim of this study is to select candidate compounds from a Traditional Chinese Medicine (TCM) dataset that may be repurposed for thrombotic condition management. This shall be achieved through development and evaluation of a Contrastive Cross-Modal Ensemble Deep Attention framework that accounts for the genetic variations in protein sequences during target-compound screening.

2. Methods

2.1. Dataset and Materials

The primary training dataset is the MINER drug–target interaction (DTI) dataset from the BIOSNAP collection (Table 1).[17] The wild-type set includes 27,497 drug–target interacting and non-interacting pairs from DrugBank, based on 4510 drug compounds, and 2181 protein targets. This was supplemented with 321,657 drug–target pairs consisting of SNP variants of those wild-type proteins (see BIOSNAP dataset supplementation). The Database of Useful Decoys (DUD-E) is used to fine-tune the model using contrastive learning. The DUD-E dataset analysed consisted of 57 protein targets and active compounds that are known to interact with these targets as well as decoys that are known not to bind the targets but have very similar chemical structures.[18]

Table 1. Summary of the BIOSNAP and DUD-E dataset samples included. Training, validation, and test dataset values represent the total number of combinations (number of interaction pairs/number of non-interaction pairs); the DUD-E dataset is only used in the training process; hence, validation and test set boxes are left blank.

Dataset	Drugs	Targets	Training	Validation	Test
BIOSNAP Wild-type	4510	2181	19,248 (9676/9572)	2752 (1396/1352)	5497 (2770/2727)
BIOSNAP SNP Variants	4510	2181	225,161 (134363/90798)	32,165 (19194/12971)	64,331 (38389/25942)
DUD-E	852,292	57	415,204 (8996/406208)	—	—

2.2. BIOSNAP Dataset Supplementation

The humsavar dataset containing human SNP polymorphic variations metadata was linked to the corresponding protein sequences in UniProt database. Then based on the amino acid (AA) change column, a script was written to insert an updated sequence for each polymorphism variant at the AA sequence location specified using the SNP metadata, e.g. for p.Asn308Asp, the letter N (Asparagine) is replaced with D (Aspartic acid) at position 308 of the sequence. The wild-type proteins are assumed to be non-pathogenic in this study, in addition to any SNP variants specified as such in the metadata. Uncertain significance (US) was coded into the same group as likely benign or benign (LB/B) i.e. as 0 for pathogenicity. Likely pathogenic or pathogenic (LP/P) were coded as 1.

The Gene column in the BIOSNAP (wild-type) dataset was then mapped using the UniProt ID mapping function tab to normalize[19] the gene names for consistency with UniProt-humsavar combined dataset. The BIOSNAP and UniProt-humsavar dataset were then vertically combined. In line with previous work, the 70%, 10% and 20% of the variant dataset (UniProt-humsavar), were randomly sampled without replacement and assigned for combination with the corresponding training, validation and test sets of the BIOSNAP wild-type dataset. Random stratified sampling without replacement was used to maintain the proportion of the pathogenicity and compound interaction status prior to sampling.

 doi:10.20944/preprints202604.1287.v1

This linkage resulted in 349,154 rows with one gene and drug to many protein mutation variants including wild-type (Table 2). The pathogenicity profile consisted of pathogenic (144,845) and non-pathogenic SNP protein variant (204,309). Following the combination of the BIOSNAP and UniProt-humsavar datasets, there were 349,154 drug-protein pairs in the entire dataset of which there were 205,792 interaction pairs including 73,541 that are interactions with pathogenic protein variants. So, for the 70% training data of the 349,154 drug-protein pairs, each base model of the 10 ensembles would have 349,154 * 0.7 / 10 = 24,440 rows of data.

The application dataset consisted of the TCMBank dataset as part of the case study. This dataset contained 61,962 ingredients from Traditional Chinese medicines, of which 26,409 compounds missing SMILES were removed, retaining 35,553 compounds to include for analysis.

Table 2. Summary of the BIOSNAP-UniProt-humsavar linked dataset stratified by interaction and non-interaction class labels and SNP variant pathogenicity status.

	Pathogenic	Non-Pathogenic	Overall
Interactions	73,541	132,251	205,792
Non-interactions	71,304	72,058	143,362
Total	144,845	204,309	349,154

2.3. Target Featurisation

The chemical compounds were transformed into features using the Morgan fingerprint method. The protein sequences were transformed into features using the Skipgram Neural Network embedding approach as described by ProtVec.[20] The rationale for using this embedding approach over that of ProtBERT as in [21] is that the latter requires CUDA (NVIDIA GPU) to be available and configured to the appropriate version of the torch library on the device, which is not always available to all users. In addition, ProtBERT requires enormous computational power and requires continuous rapid connectivity to platforms such as Hugging Face, which may also pose challenges.

2.4. Deep Ensemble Modelling Approach

The ensemble modelling consisted of partitioning of the entire dataset into 10 subsets of training, validation and test sets, one set for each of the 10 base models. The ensemble models are trained with 5 epochs for each ensemble in order to minimise risk of overfitting (pilot studies found increasing epochs did not improve performance). Due to the high computational cost of the algorithm with increasing data volume, bootstrapping approaches are excluded and this bagging approach was used instead. Stratified batch sampling using the interaction class labels was used to mitigate against class imbalance of during training (batch loading). Both contrastive and non-contrastive learning ensemble models shall be developed and evaluated, with the former used as the regularised applications model to use on TCMBank data for generalisability.

The interaction for the drug embeddings and the protein variant embeddings were separately visualised. Firstly, the Factor X wild-type and its SNP variants were inputted for screening against the relevant clinical anticoagulants together with the TCM compounds. Sequioflavone is also included as it was considered in a previous study.[19] Since mechanisms and metabolism for existing anticoagulants are already well established,[22] the primary focus of this study is on the TCM targets and associated metabolism. Using each of the three identified TCM compounds with existing evidence for association with thrombosis from the top 100 strongest interaction candidates, the PaCMAP plot was analysed with a single TCM compound at a time against the whole training set of protein variants. The metabolism perspective was also analysed by taking each compound and screening for which likely protein they are likely metabolised by.

The PaCMAP plot was aggregated across the 10 ensembles, with average PaCMAP component 1 and 2 coordinates calculated across the 10 models to derive pooled Euclidean distances for each pair of interactions. The control (decoys) group was set as the pathogenic group, whilst the treatment group (actives) was set as the non-pathogenic group. The strong interacting targets for each anti-

 doi:10.20944/preprints202604.1287.v1

thrombotic candidate shall also be considered for plausibility in terms of potential metabolism relationships. The pathogenicity status were used to color in the PaCMAP projections. The rationale for not building a separate pathogenicity model is that there already many existing models for predicting pathogenicity. The standard error of mean (SEM) across the PaCMAP coordinates across the X and Y directions of the variants were used to quantify the uncertainty in the projected data points. Further details of this representation approach are provided in Supplementary Materials, Representation Approach.

The base modelling methodology are provided in the Supplementary Materials, Base Modelling Approach. Details on model evaluation are also provided in Supplementary Materials, Model Evaluation. Due to the CPU and GPU memory requirements of the large scale dataset, the modelling was conducted using a 64 Core, 128-thread Threadripper workstation equipped with NVIDIA RTX 48GB A6000 GPU.

3. Results

The Dong et al. (May 2025) Non-CL achieved a satisfactory performance of 0.814 and 0.837 on the validation and test dataset, respectively (**Table 1**). The Non-CL Ensemble model demonstrated a higher magnitude of performance in both the validation (0.843 vs. 0.814) and test set (0.894 vs. 0.837).

Table 2. Evaluation of the performance of base models without contrastive learning; PR AUC: Precision–Recall Area Under the Curve.

	PR AUC	
Model	**Validation Set**	**Test Set**
Non-CL Ensemble	0.843	0.894
Dong et al. (May 2025) Non-CL	0.814	0.837

Both the Dong et al. (May 2025) and Ensemble models were supplemented with contrastive learning. Whilst the contrastive regularisation effect decreased the performance of the Dong et al. (May 2025) model slightly compared to its non-contrastive counterpart, it improved the performance of the CL Ensemble over the Non-CL Ensemble model as well as the Dong et al. (May 2025) CL model in both the validation set (0.901 vs. 0.843 vs. 0.812; **Table 2**) and test set (0.919 vs. 0.894 vs. 0.813). In addition, it also outperformed the Dong et al. (May 2025) model without contrastive learning in both the validation set (0.901 vs. 0.814; **Table 2**) and test set (0.919 vs. 0.837).

Table 3. Evaluation of performance of the best-performing model with contrastive learning; PR AUC: Precision–Recall Area Under the Curve.

	PR AUC	
Model	**Validation Set**	**Test Set**
CL Ensemble	0.901	0.919
Dong et al. (May 2025) CL	0.812	0.813

Apixaban and Rivaroxaban are shown to have the strongest interaction with Factor Xa variants on average out of the Direct Acting Oral Anticoagulants (DOACs) justifying their most common usages in human and animal patients, respectively (Figure 1). The candidate identified in our previous study, Sequoiaflavone is shown to have similar but slightly weaker interaction.[23] This is followed by Edoxaban and Enoxaparin.

 doi:10.20944/preprints202604.1287.v1

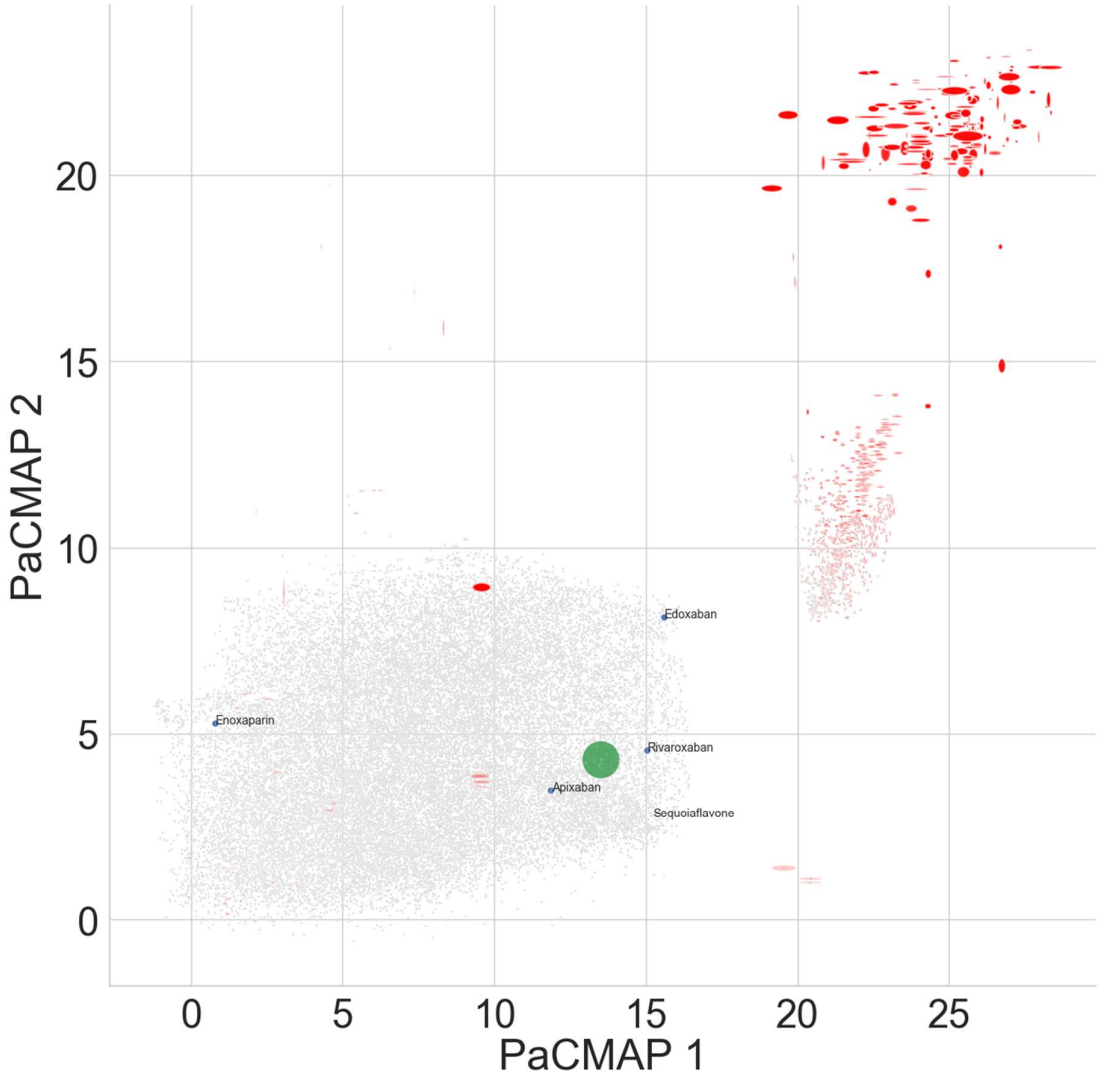

Figure 1. Using PaCMAP, the learned latent space for the average over Factor Xa variants (green), commonly used clinical anticoagulants (blue), and Traditional Chinese Medicine (TCM) compounds (grey) are shown. Red: shows the variation (standard error of mean) in interaction across Factor Xa variants.

Five notable compounds in the top 100 ranked interaction scores for Factor Xa variants were found to be associated with existing TCM used. These were specifically, Safflower yellow A, Paeoniflorin, Ginsenoside Rd2, cucurbitacin j 2-o-β-glucopyranoside and Notoginsenoside R6. Following consultation with Bristol Chinese Medicine Centre (see acknowledgement section), Ginsenoside Rd2 and cucurbitacin j 2-o-β-glucopyranoside were excluded from follow-on analysis due to these being less commonly used in TCM prescriptions. For additional reasons for exclusion, see discussion section.

Regions of high variation in interaction effect due to gene type variation are shown in red. The smaller red bands are likely due to the variation corresponding to interaction effect variants in fewer factor Xa variant, whereas the thicker red band correspond to higher proportion of genetic variants having a varying interaction effect with a specific compound.

The reason why the two clusters further away from Factor Xa have larger extent of variation can be explained using an analogy. If an enzyme is not specific, it would be more open to variation, i.e. it would sometimes not fit and other times fit a substrate. That is, it goes from having a large response to having no response at all; whereas if it is strongly binding, the effect of a change might not make much difference. The latter scenario is likely reflected in the main cluster.

 doi:10.20944/preprints202604.1287.v1

Figure 2 shows that DOACs interact significantly more strongly with Factor Xa variants than TCMs on average. However, it can be seen that the lower end of the whisker for the TCM compounds extend further downwards toward low values of Euclidean distance score, suggesting presence of a group of TCM compounds that likely interact more strongly than DOACs.

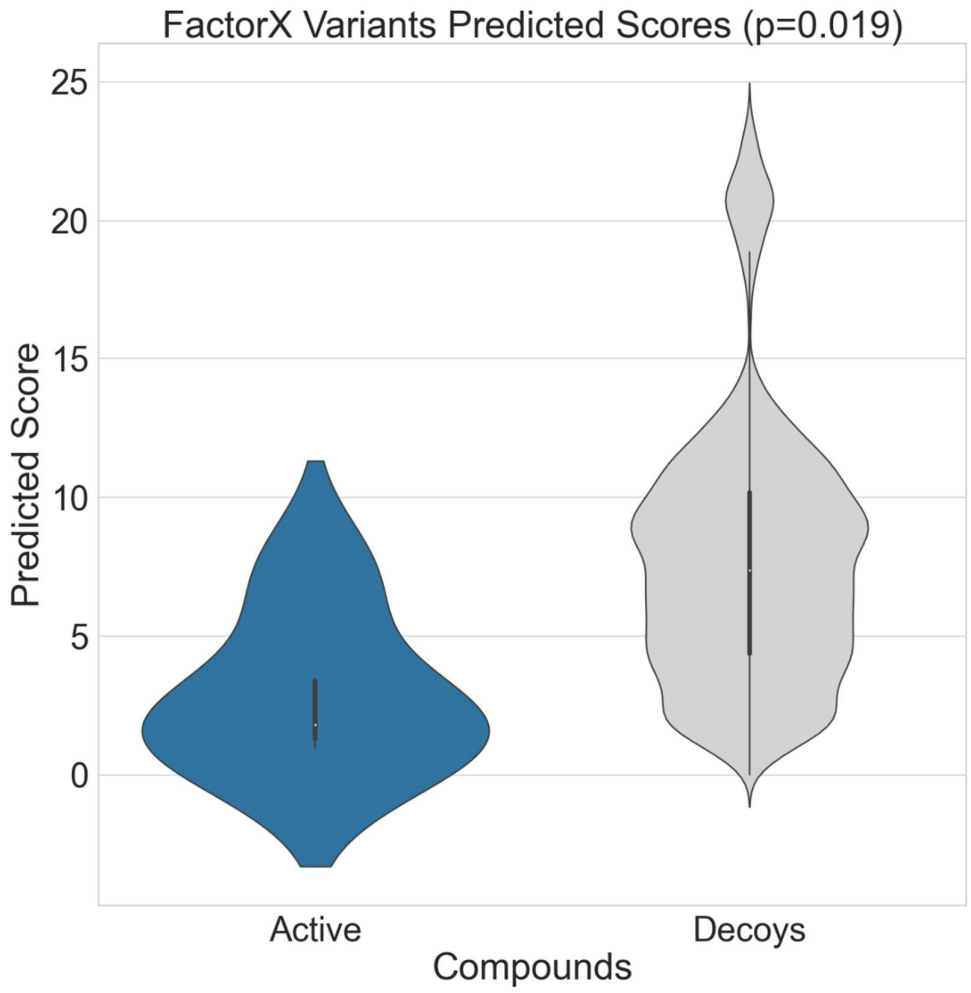

Figure 2. Using a violin plot, the distribution of interaction scores for average Factor Xa SNP variations against the clinical anticoagulants (blue) and TCM compounds (grey) are shown; p-value shows results from one-sided t test.

Safflower yellow A is shown in the projection plot in terms of its interaction with pathogenic (blue) and non-pathogenic (grey) SNP genetic variations (red ellipses) of protein targets (Figure 3). A limited number of very strongly interacting candidates are shown in the vicinity of Safflower yellow A. Apart from a few pathogenic proteins, interaction in close vicinity are generally heavily influenced by variations in SNP changes to protein targets sequences. In addition, pathogenic SNP proteins tended on average to interact more strongly with Safflower yellow A than their non-pathogenic counterpart (Supplementary Figure 1).

 doi:10.20944/preprints202604.1287.v1

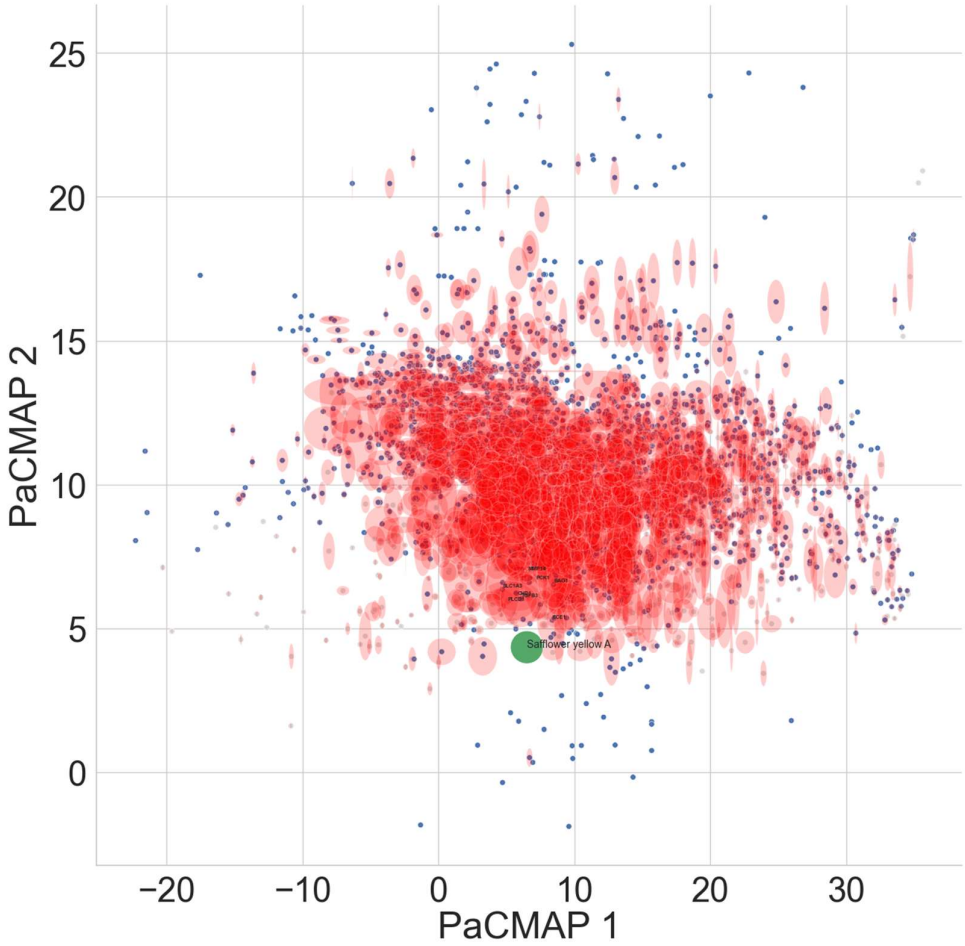

Figure 3. Using PaCMAP, the learned latent space for the Safflower yellow A (green), average of pathogenic SNP variations of protein (blue), and average of non-pathogenic SNP variations of protein (grey) are shown. Red: shows the variation (standard error of mean) in interaction across SNP variations of protein.

Table 4. The interactions between Safflower yellow A (SYA) and protein targets likely involved in thrombosis stratified across pathogenicity. * represents target identified with a high level of variation in interaction within vicinity of SYA.

Compound_name	pathogenicity	mean_Euclid_distance	Rank
Safflower yellow A	NA	0.00	
MTHFD2	1	0.51	1
QARS1	1	0.73	2
TUBB4B	1	0.85	3
SLC6A11	1	1.01	4
MARS2	1	1.03	5
SCN4A	0	1.07	6
EPHB4	0	1.34	7
PLA2G4A	0	1.63	11
TGFB3*	1	1.74	12
SOD1	1	1.81	16
ECE1*	1	2.09	21
FTH1	1	2.27	29
ENPP1	0	2.32	30
MMP14*	0	2.63	41

 doi:10.20944/preprints202604.1287.v1

Table 3 shows both targets that have known associations (bold) with thrombosis and those without known existing associations with thrombosis. The top 5 ranked targets are associated with inflammation, which can increase the propensity for thrombosis by creating a prothrombotic state. These were all strongly interacting with Safflower yellow A, the TCM ingredient in safflower (*Carthamus tinctorius*), following pathogenic SNP mutations. SCN4A is a sodium channel that is strongly interacting in the non-pathogenic state. Abnormalities in sodium concentration, specifically hypernatremia (high sodium), are associated with a significant increase in the risk of coagulation and thrombosis. There are an equal balance of pathogenic and non-pathogenic targets with known evidence for association with thrombosis. Using an approach akin to anomaly detection, three of these target TGFB3, ECE1 and MMP14 were found to display high level of variation in interaction across their SNP genetic variants within the vicinity of Safflower yellow A. For the purpose of this study, the discussion later focus shall be on the targets with existing evidence for thrombosis.

Table 4 shows the strongly interacting targets from the top 100 interactors that are likely to be associated with Safflower yellow A metabolism based on existing evidence.

Table 5. Strongly interacting protein targets likely associated with metabolism of Safflower yellow A.

Compound_name	pathogenicity	mean_Euclid_distance	Rank
Safflower yellow A	NA	0.00	
SOD1	1	1.81	16
ACOX1	0	2.49	37
ECI1	1	2.64	42
POR	0	3.02	64

Notoginsenoside R6 is shown in the Supplementary Figure 2 in terms of its interaction with pathogenic (blue) and non-pathogenic (grey) SNP genetic variations (red ellipses) of protein targets. On average, the interaction with pathogenic groups were not significantly different to that of the non-pathogenic group (**Supplementary Figure 3**). The interacting targets with known associations to thrombosis are provided in ***Supplementary Table 1***.

The strongly interacting targets likely involved in Notoginsenoside R6 metabolism are provided in **Supplementary Table 2**. The POR gene encodes the cytochrome P450 Reductase, which is the body's main protein for metabolising xenobiotics. The UGT1A6 gene is expressed instead as one isoform of UDP-glucuronosyltransferases family member 1 and promotes drug metabolism in vivo through glucuronidation.[24]

Paeoniflorin is shown to embed to the periphery of the central cluster (**Supplementary Figure 4**), surrounded by both pathogenic and non-pathogenic targets. Strong variations in interaction effect is seen for the main cluster across genetic variants of each target, with targets beyond the cluster showing little variations.

Table 5 shows the main targets with existing thrombotic evidence, which interacted strongly with Paeoniflorin. Of the targets detected using anomaly detection with high interaction fluctuations based on genetic variation, TGFB3 (rank 97) and RHOA (rank 25) are known to be associated with thrombosis, with the latter also identified one of the top 50 candidates.

 doi:10.20944/preprints202604.1287.v1

Table 6. The interactions between Paeoniflorin and protein targets likely involved in thrombosis stratified across pathogenicity. * represents target identified with a high level of variation in interaction within vicinity of Paeoniflorin.

Compound_name	pathogenicity	mean_Euclid_distance	Rank
Paeoniflorin	NA	0	
CHRND	0	0.40	1
ITK	0	0.47	2
ADORA2B	1	0.78	8
MMP21	0	0.90	12
RHOA*	0	1.24	25
MPO	0	1.36	30
KDR	0	1.41	31

On average, there were no significant difference between pathogenic and non-pathogenic groups of targets in terms of interaction with Paeoniflorin (**Supplementary Figure 5**).

Table 7. Strongly interacting protein targets likely associated with metabolism of Paeoniflorin. .

Compound_name	pathogenicity	mean_Euclid_distance	Rank
Paeoniflorin	NA	0	
FMO3	0	1.12	21
CYP27B1	0	1.83	74

Table 6 shows Flavin-containing monooxygenase 3 (FMO3) and the cytochrome 1-alpha-hydroxylase (CYP27B1) are potential candidates involved in metabolising Paeoniflorin. FMO3 is known to metabolise a wide range of therapeutic sulphur-, nitrogen- and phosphate-based compounds.[25,26] CYP27B1's involvement in metabolism is predominantly through hydroxylation and this has been shown to be a significant part of Paeoniflorin's metabolism. [27].

4. Discussion

Existing work in DTI prediction has focused on areas of antithrombosis management but there has been limited consideration of the use of single nucleotide polymorphism's effect of Traditional Chinese Medicine (TCM) compounds on the proteome for this specific purpose. Although TCM practice is increasingly gaining momentum in West, many clinicians in the West still have limited knowledge of the effects of TCM compounds and their mechanisms, requiring continued paradigm shift to further enable integration of both practices.[28] In addition, previous work on DTI prediction have primarily focused on the use of single models without consideration of the benefits of regularised ensemble modelling.[19,21,23] Here, it was shown that contrastive regularisation improved over the performance of the models without regularisation, suggesting a benefit in preventing issues of overfitting when an ensemble approach is taken. By accounting for interindividual variations in genetic profiles, specifically both pathogenic and non-pathogenic single nucleotide polymorphisms, our approach provides the stepping stone towards personalised drug selection and at the same time provides meaningful interpretations of corresponding variations in affected protein pathways.

Since the three TCM compounds bind many pathogenic targets, these are expected to be potentially useful for discovery of secondary/alternative uses of that drug i.e. for future repurposing of the drug, as the compound may be tailored to treating such a disease through allosteric regulation[29] (reversibly activate or inhibit) or degradation.[30] If a compound binds many non-pathogenic targets, then it is likely to be disruptive of normal healthy functioning of the body system i.e. contributing negatively. In such case, future adverse effect screening would be required.[19] Additionally, compound metabolism is important as exemplified in the report that polymorphisms in one gene superfamily, the cytochrome P450s (CYP) are associated with up to 60% of drug-induced toxicity.[31] The cytochromes in the liver are the main proteins for metabolising drug compounds,

 doi:10.20944/preprints202604.1287.v1

either accelerating or deaccelerating their processing. To add to the complexity, proteins involved in metabolism can also be induced or inhibited to result in faster or slower body clearance of drugs by other proteins or drugs (e.g. antimicrobials, antiarrhythmics, antihyperlipidemic),[22,32] depending a variety of factors from effects on transcription (transcriptomics),[33] and post-transcriptional mechanisms (e.g. translational efficiency, protein stabilization, and reduced protein degradation).[34] Evidence have shown that the gut microbiome and antibiotics play a significant role due to the former's rich source of metabolites and vitamin K,[22] which are required for coagulation Factor IX and the downstream cascade. However, it should be noted that excessively high or low levels of the normal (wild-type) protein due to dysfunction in autophagy is also main driver of disease[30] and in such cases inhibition or agnostic activation effects of the compound could also be beneficial depending on the dosage used.

4.1. Paeoniflorin

This study identified Paeoniflorin as a potential antithrombotic candidate for targeting Factor Xa given these considerations. Paeoniflorin is the active ingredient of Paeonia lactiflora Pall. (*P. lactiflora*), also known as peony and Shaoyao, a Chinese traditional herbal medicine with an extensive set of properties including anti-inflammation and anti-coagulation.[35] Paeoniflorin has been found to have anti-thrombotic, anti-inflammatory and anticoagulation effects.[36] Its anti-thrombotic effects have been attributed to the recovery towards the prethrombotic state and recanalisation of thrombosis by increasing the expression of urokinase-type plasminogen activator (uPA), which may be mediated via regulation of the p38 and JNK MAPK signalling pathways.[37] However, a review found that both in-vitro and in-vivo studies on its mechanism on anticoagulation is lacking.[38]

Here, we tackle this lack of mechanistic understanding by identification of several anticoagulation associated proteins that interact strongly with Paeoniflorin. Acetylcholine receptor subunit delta is a protein that in humans is encoded by the CHRND gene, with research data suggesting that drugs targeting acetylcholine receptor subunits might inhibit thrombosis.[39] Myeloperoxidase (MPO) has been shown to mediate thrombosis formation and AF associated stroke risk through directly damaging blood vessel walls, provoking the formation of Neutrophil Extracellular Traps (NETs) and exerting pro-oxidative and pro-inflammatory effects.[40] Research have shown that functional ADORA2B is associated with mediation of the inhibitory effects on NETs formation.[41] Here, *P. lactiflora* was found to interact strongly with the pathogenic variant of ADORA2B, suggesting potential beneficial compensatory potential through allosteric interactions. Matrix Metallopeptidase 21 (MMP21) is a member of the Matrix metalloproteinase family, which are associated with thrombus regulation, vascular tissue remodelling and wound repair.[42] Previously, MMP19 have been shown to regulate thrombus stability,[43] with MMP-1 and MMP-2 shown to stimulate platelet activation and thrombus formation under arterial flow, whereas MMP-9 and MMP-14 (identified as strong interactor for Safflower yellow A) acted as inhibitors of these processes.[44] Evidence on MMP21 is lacking but it has been associated with tissue remodelling.[45]

Flavin-containing monooxygenase 3 (FMO3) was identified as a metabolism protein strongly interacting with Paeoniflorin. Research have shown that polymorphisms at the FMO3 gene locus play a significant role in the interethnic variation of the metabolism of certain amines including drugs, dietary agents and other xenobiotics, which may contribute to their therapeutic efficacy.[46] FMO3 has been shown to catalyse the synthesis of gut microbiota-derived metabolite trimethylamine N-oxide (TMAO), a major risk factor in cardiovascular thrombotic events.[47] Interestingly, Paeoniflorin exerts an inhibitory effects on FMO3 activity,[48] reducing platelet reactivity and thrombus formation in vivo, suggesting here that instead of FMO3 metabolising Paeoniflorin, Paeoniflorin is modulating the activity of the metaboliser.

 doi:10.20944/preprints202604.1287.v1

4.2. Safflower Yellow A

Safflower yellow A is the primary bioactive compound extracted from safflower (Carthamus tinctorius L.), a TCM compound known as Honghua.[49] Research has shown that Safflower yellow A extracts from safflower have anticoagulation activity and can prolong clotting (prothrombin time) time.[50] It also has simultaneous antiplatelet effects through suppression of adenosine diphosphate (ADP) and platelet activating factor (PAF).[51] One translational study found that Safflower Yellow A injection in rats can synergistically enhance the anticoagulant effects of warfarin as well as reducing its metabolism.[52] SNP variations (−173C>A, −208C>T, and rs2868177) in the strongly interacting Safflower yellow A metabolising protein Cytochrome P450 oxidoreductase (POR) has previously been shown to result in interindividual variability in warfarin maintenance dose (which is determined partly by rates of warfarin metabolism).[53]

Research has shown that normal functioning Endothelin-converting enzyme-1 (ECE1) is associated with lower levels of LDL,[54] and hence circulating level of coagulation factors.[55] One study in the Chinese Han population found that the C-338A polymorphism of the ECE-1b gene was associated with an increased risk of ischemic stroke.[56] Our finding that Safflower yellow A strongly interacts with the pathogenic ECE1 suggests the compensatory potential of Safflower yellow A on the SNP induced ECE1 dysfunction.

Here, TGFβ3 was identified as one of the strongly interacting proteins highly affected by SNP variations. TGFβ1 is responsible for inducing in endothelial-to-mesenchymal transition (EndMT) leading to fibrotic processes in heart,[56] as well as venous thrombofibrosis.[57,58] However, translational evidence in mice suggested that TGFβ1, TGFβ2, and TGFβ3 have a protective role in preventing growth of existing plaque. In addition, this protective effect seems to depend on the potent deactivating effects of TGF-β on macrophages and T lymphocytes.[59]

4.3. Notoginsenoside R6

The triterpenoid saponin,[60] Notoginsenoside R6 is one of the main components of *Panax notoginseng*, which is a traditionally used Chinese medicine for anticoagulation and improving blood circulation.[61] However, there is a lack of knowledge of the specific mechanisms of its ingredients with the suggestion that these could have different.[61] Whilst Notoginsenoside R1 (NG-R1) has been more widely studied, the mechanism of other ingredients including R6 are less well understood.[62]

Here, the top interacting protein was found to be 3-hydroxy-3-methylglutaryl-CoA reductase (HMGCR), which is also the key target responsible for statin's pleiotropic effects (cholesterol reduction & anticoagulation). This is a transmembrane glycoprotein that is the rate-limiting enzyme in cholesterol biosynthesis as well as in the biosynthesis of nonsterol isoprenoids that are essential for normal cell function including ubiquinone and geranylgeranyl proteins.[63] Research has shown that inhibition of HMGCR limits exposure of phosphatidylserine at the cell surface by restricting the cellular pool of mevalonate-derived isoprenoids.[64] This membrane alteration restricts the activity of proteolytic enzyme (factor VIIa, tissue factor and Prothrombinase) complexes that propagate the coagulation cascade.[65]

CPB1 encodes for the protein pancreatic procarboxypeptidase B, more commonly known as Carboxypeptidase B1. Whilst CPB1 is secreted by the pancreas into the small intestine, there has been limited evidence on its precise role. However, the closely related protein CPB2 (also called TAFIa) which is primarily localised in blood plasma is known to be a fibrinolysis inhibitor that acts by cleaving lysine and arginine residues from partially degraded fibrin, thereby lowering its binding capacity for plasminogen.[66] This leads to the inhibition of tissue plasminogen activator-mediated fibrin degradation and increases the risk of thrombosis.[66] Given emerging evidence that has examined the interactions between molecular markers across the heart and the gut, showing that following myocardial infarction, the gut lining becomes more permeable,[67] one could question whether the CPB1 could similarly have prothrombotic effects. Specifically, research have shown that CPB1 is involved in severe inflammatory responses and coagulation dysfunction through the caspase-11 mediating carboxypeptidase B1(CpB1)-C3-C3aR pathway.[68,69]

 doi:10.20944/preprints202604.1287.v1

Membrane metallo-endopeptidase (MME) also called Neprilysin exerts an inhibitory effect on coagulation by causing a reduction in the rate of conversion of fibrinogen to fibrin mediated proteolysis of fibrinogen by MME.[70] This highlights the importance of this protein in the final steps of the coagulation cascade involving the conversion of fibrinogen to fibrin monomers, which polymerises to form fibrin polymer mesh and the resulting cross-linked fibrin clot.[71]

4.4. Cardiovascular Medicine and Surgery Relevance

Antiplatelets primarily are primarily effective for platelet induced arterial thrombosis and given in heart attacks prior to coronary bypass surgery.[72,73] Anticoagulants on the other hand are effective for venous and atrial appendage formed thrombosis which primarily related to fibrin induced clots.[74] However, most of these are only prescribed after thrombus detection has been confirmed i.e. typically after cardiac surgery. This is the case even for patients at high risk of AF since the anticoagulants are licenced on the condition of AF rather than pre-emptively, except enoxaparin (see below). This poses a problem given the high prevalence of new-onset AF following surgery and the concomitant high risk in embolism and stroke. As such, improved prophylactic agents would be highly desirable. Although Dual Antiplatelet therapies (DAPTs) have been considered, these are generally found to be less effective compared to anticoagulants in elderly patients.[75] However, more recent trial evidence are beginning to favour the use of antiplatelet (aspirin) monotherapy over anticoagulants plus antiplatelet combined therapy.[76] Where combined therapy are used, short term (one-month) administrations are preferred compared to long term (one-year) administrations. The second antiplatelet is typically started sometime after surgery e.g. one week or months after. As such, the compounds and novel targets identified in this study may present additional options for clinical trial and repurposing investigations to shift the management of AF stroke risk from reactive to proactive, the consideration of targets beyond Factor Xa, as well as the possibility of more potent prophylactic agents than those currently considered for DAPTs.

Enoxaparin is expected to have lower interaction than the other anticoagulants as it is less direct for Factor Xa and licensed for both management of heart attack and AF. It is injected preoperatively if the patient has signs of AF or arrythmia because it can be more easily reversed than some other anticoagulants. In some hospitals, such as Liverpool Heart and Chest Hospital, dalteparin is routinely given as a low dose prophylactic to most patients and can be escalated to treatment with enoxaparin instead based on body weight, renal function and other real time characteristics of the patients. Following surgery, if AF is presented, the patients are prescribed oral anticoagulants alone for further management with antiplatelets discontinued, although an antiplatelet is kept if they have recently received stents.[22] For further details on anticoagulation guidelines, see British Society for Haematology (BSH) publication.[77]

4.5. Future Work and Limitations

Future work should aim to incorporate the information from SNP genetic variations during the adverse effects screening process when considering drug selection decisions. Future work should also further categorise the SNP mutations as active site affecting or not although this information may be partly implied by pathogenicity categorisation already. Transcription and translation factors also play a significant role in regulating protein activities and drug metabolism.[29] Further work should aim to dedicate focus to these topics. It is also currently challenging to simultaneously perform screening for multiple compounds vs multiple targets simultaneously. However, such tasks may benefit from further research in topological theories and find applicability in drug-drug or compound-compound interaction studies. Since, only a limited DUD-E targets were considered in studies so far, it would be potentially helpful to develop and apply synthetic data generation methods to augment the contrastive samples for each of the targets in BIOSNAP dataset using approaches such as conditional GAN or drug MPNN, with the aim to provide more representative decoys.[78] As effect of SNPs may be subtle, future research should investigate a dataset updated with other types of structural Variant mutations such as deletions, insertions inversions, and translocations etc.

 doi:10.20944/preprints202604.1287.v1

Future research could consider including recombinant factor X variants that could support haemostasis when injected into patients taking anticoagulants. It would also be useful to curate a SNP protein variants based dataset for screening adverse effects.[19]

5. Conclusions

Ensemble modelling with contrastive learning resulted in performance improvements and can be useful for selecting Traditional Chinese Medicine compounds for antithrombotic management. This is a step towards personalised drug selection and can simultaneously facilitate interpretation of the biological rationales for risk vs benefit evaluations during decision making.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

Author Contributions: Conceptualization, Tim Dong; Methodology, Tim Dong; Software, Tim Dong; Validation, Tim Dong, Rhys Llewellyn and Melanie J Hezzell; Formal analysis, Tim Dong; Investigation, Tim Dong, Rhys Llewellyn, Melanie J Hezzell and Gianni D. Angelini; Resources, Tim Dong; Data curation, Tim Dong; Writing – original draft, Tim Dong; Writing – review & editing, Tim Dong, Rhys Llewellyn, Melanie J Hezzell and Gianni D. Angelini; Visualization, Tim Dong; Supervision, Melanie J Hezzell and Gianni D. Angelini; Project administration, Tim Dong and Gianni D. Angelini. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: We are grateful to Sarah Searle from the South West Genomics Laboratory Hub for an initial meeting and subsequent NHS genetic testing workflow guidance on the project.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lampert, J.; Kauffman, J.; Bhatt, D.; Vleck, T.V.; Vaid, A.; Halperin, J.L.; Koruth, J.S.; Turagam, M.K.; Musikantow, D.R.; Miller, M.A.; et al. EN-499638-001 GRAPH NEURAL NETWORK AUTOMATION OF ANTICOAGULATION DECISION MAKING OUTPERFORMS CHA2DS2-VASC AND HAS-BLED. *Heart Rhythm* **2025**, *22*, S13. https://doi.org/10.1016/j.hrthm.2025.03.028.
2. Paliwal, A.; Jain, S.; Kumar, S.; Wal, P.; Khandai, M.; Khandige, P.S.; Sadananda, V.; Anwer, M.K.; Gulati, M.; Behl, T.; et al. Predictive Modelling in Pharmacokinetics: From in-Silico Simulations to Personalized Medicine. *Expert Opin Drug Metab Toxicol* **2024**, *20*, 181–195. https://doi.org/10.1080/17425255.2024.2330666.
3. Cullell, N.; Carrera, C.; Muiño, E.; Torres, N.; Krupinski, J.; Fernandez-Cadenas, I. Pharmacogenetic Studies with Oral Anticoagulants. Genome-Wide Association Studies in Vitamin K Antagonist and Direct Oral Anticoagulants. *Oncotarget* **2018**, *9*, 29238–29258. https://doi.org/10.18632/oncotarget.25579.
4. Chin, P.K.L. Which Patients May Benefit from Dose Adjustment of Non-Vitamin K Antagonist Oral Anticoagulants? *Semin Thromb Hemost* **2015**, *41*, 195–207. https://doi.org/10.1055/s-0035-1546465.
5. Hijazi, Z.; Oldgren, J.; Lindbäck, J.; Alexander, J.H.; Connolly, S.J.; Eikelboom, J.W.; Ezekowitz, M.D.; Held, C.; Hylek, E.M.; Lopes, R.D.; et al. The Novel Biomarker-Based ABC (Age, Biomarkers, Clinical History)-Bleeding Risk Score for Patients with Atrial Fibrillation: A Derivation and Validation Study. *The Lancet* **2016**, *387*, 2302–2311. https://doi.org/10.1016/S0140-6736(16)00741-8.
6. Barnes, G.D. New Targets for Antithrombotic Medications: Seeking to Decouple Thrombosis from Hemostasis. *Journal of Thrombosis and Haemostasis* **2025**, *23*, 1146–1159. https://doi.org/10.1016/j.jtha.2024.12.003.
7. Sudhan, M.; Janakiraman, V.; Ahmad, S.F.; Attia, S.M.; Subramanian, R.; Devi, D.; Ahmed, S.S.S.J. A Comprehensive Insight from Molecular Docking and Dynamics with Clinical Investigation on the Impact of Direct Oral Anticoagulants on Atheroprotective Protein in Atrial Fibrillation. *BMC Pharmacol Toxicol* **2024**, *25*, 56. https://doi.org/10.1186/s40360-024-00785-z.

 doi:10.20944/preprints202604.1287.v1

8. Avsar, O. Identification of the Effects of Pathogenic Genetic Variations of Human CYP2C9 and CYP2D6: An in Silico Approach. *Nucleosides, Nucleotides & Nucleic Acids* **2024**, *43*, 356–376. https://doi.org/10.1080/15257770.2023.2262519.
9. dos Santos, R.V.; Barrionuevo, M.V.F.; Vieira, M.R.F.; Mazoni, I.; Tasic, L. Plasminogen Activator Inhibitors in Thrombosis: Structural Analysis and Potential Natural Inhibitors. *ACS Omega* **2025**, *10*, 27348–27362. https://doi.org/10.1021/acsomega.5c02926.
10. Manderstedt, E.; Halldén, C.; Lind-Halldén, C.; Elf, J.; Svensson, P.J.; Engström, G.; Melander, O.; Baras, A.; Lotta, L.A.; Zöller, B. Thrombomodulin (*THBD*) Gene Variants and Thrombotic Risk in a Population-based Cohort Study. *Journal of Thrombosis and Haemostasis* **2022**, *20*, 929–935. https://doi.org/10.1111/jth.15630.
11. Hassan, M.O.; Gassim, D.A.; Albakrye, A.M.; Elnasri, H.A.; Khaier, M.A.M. In Silico Analysis of Likely Pathogenic Variants in Human *GGCX* Gene. *Informatics in Medicine Unlocked* **2020**, *19*, 100337. https://doi.org/10.1016/j.imu.2020.100337.
12. Shen, Y.-M.; Nagalla, S. Hypercoagulable Workup in Thrombotic Cardiovascular Diseases. *Circulation* **2018**, *138*, 229–231. https://doi.org/10.1161/CIRCULATIONAHA.117.031699.
13. Mulder, R.; Croles, F.N.; Mulder, A.B.; Huntington, J.A.; Meijer, K.; Lukens, M.V. SERPINC1 Gene Mutations in Antithrombin Deficiency. *British Journal of Haematology* **2017**, *178*, 279–285. https://doi.org/10.1111/bjh.14658.
14. Burford, E.; Kasolo, Y.; Draper, J.; Sheikh, N.; Carr-White, G.; Marvao, A. de; Dungu, J.N.; Bastiaenen, R.; Cooper, R.M. Tale of Three Services: Early UK Experience with Mavacamten Treatment for Hypertrophic Cardiomyopathy with Left Ventricular Outflow Tract Obstruction. *Open Heart* **2025**, *12*. https://doi.org/10.1136/openhrt-2025-003218.
15. Stocker, S.L.; Polasek, T.M. Using Pharmacogenomics to Personalise Drug Therapy: Which Drugs, When and How. *Aust Prescr 48*, 82–86. https://doi.org/10.18773/austprescr.2025.021.
16. Polasek, T.M. Calculation of the Pharmacogenomics Benefit Score for Patients with Medication-Related Problems. *Front Genet* **2023**, *14*, 1152585. https://doi.org/10.3389/fgene.2023.1152585.
17. Zitnik, M.; Sosič, R.; Maheshwari, J.L. Stanford Biomedical Network Dataset Collection Available online: http://snap.stanford.edu/biodata/ (accessed on 20 February 2025).
18. Mysinger, M.M.; Carchia, M.; Irwin, John.J.; Shoichet, B.K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55*, 6582–6594. https://doi.org/10.1021/jm300687e.
19. Dong, T.; Llewellyn, R.; Hezzell, M.; Angelini, G.D. Rapid Screening of Anticoagulation Compounds for Biological Target-Associated Adverse Effects Using a Deep-Learning Framework in the Management of Atrial Fibrillation. *Bioengineering* **2025**, *12*, 972. https://doi.org/10.3390/bioengineering12090972.
20. Asgari, E.; Mofrad, M.R.K. Continuous Distributed Representation of Biological Sequences for Deep Proteomics and Genomics. *PLOS ONE* **2015**, *10*, e0141287. https://doi.org/10.1371/journal.pone.0141287.
21. Singh, R.; Sledzieski, S.; Bryson, B.; Cowen, L.; Berger, B. Contrastive Learning in Protein Language Space Predicts Interactions between Drugs and Protein Targets. *Proceedings of the National Academy of Sciences* **2023**, *120*, e2220778120. https://doi.org/10.1073/pnas.2220778120.
22. Mar, P.L.; Gopinathannair, R.; Gengler, B.E.; Chung, M.K.; Perez, A.; Dukes, J.; Ezekowitz, M.D.; Lakkireddy, D.; Lip, G.Y.H.; Miletello, M.; et al. Drug Interactions Affecting Oral Anticoagulant Use. *Circulation: Arrhythmia and Electrophysiology* **2022**, *15*, e007956. https://doi.org/10.1161/CIRCEP.121.007956.
23. Dong, T.; Llewellyn, R.D.; Hezzell, M.; Angelini, G.D. A Deep Learning Methodology for Screening New Natural Therapeutic Candidates for Pharmacological Cardioversion and Anticoagulation in the Treatment and Management of Atrial Fibrillation. *Biomedicines* **2025**, *13*, 1323. https://doi.org/10.3390/biomedicines13061323.
24. Chan, J.; Narayan, P.; Law, J.J.; Dong, T.; Angelini, G.D. Socioeconomic Deprivation Is Associated with Worse In-Hospital Survival after Isolated Coronary Artery Bypass Grafting in the UK. *Interdiscip Cardiovasc Thorac Surg* **2025**, *40*, ivaf119. https://doi.org/10.1093/icvts/ivaf119.
25. Krueger, S.K.; Williams, D.E. Mammalian Flavin-Containing Monooxygenases: Structure/Function, Genetic Polymorphisms and Role in Drug Metabolism. *Pharmacol Ther* **2005**, *106*, 357–387. https://doi.org/10.1016/j.pharmthera.2005.01.001.

 doi:10.20944/preprints202604.1287.v1

26. Cashman, J.R. Human Flavin-Containing Monooxygenase Substrate Specificity and Role in Drug Metabolism. *Current Drug Metabolism 1*, 181–191. https://doi.org/10.2174/1389200003339135.
27. Chen, Z.-W.; Tong, L.; Li, S.-M.; Li, D.-X.; Zhang, Y.; Zhou, S.-P.; Zhu, Y.-H.; Sun, H. Identification of Metabolites of *Radix Paeoniae Alba* Extract in Rat Bile, Plasma and Urine by Ultra-Performance Liquid Chromatography–Quadrupole Time-of-Flight Mass Spectrometry. *Journal of Pharmaceutical Analysis* **2014**, *4*, 14–25. https://doi.org/10.1016/j.jpha.2013.06.004.
28. Yagüe, E.; Sun, H.; Hu, Y. East Wind, West Wind: Toward the Modernization of Traditional Chinese Medicine. *Front. Neurosci.* **2022**, *16*. https://doi.org/10.3389/fnins.2022.1057817.
29. Cooper, G.M. Regulation of Protein Function. In *The Cell: A Molecular Approach. 2nd edition*; Sinauer Associates, 2000.
30. Bhattacharya, A.; Chatterji, U. Exosomal Misfolded Proteins Released by Cancer Stem Cells: Dual Functions in Balancing Protein Homeostasis and Orchestrating Tumor Progression. *Discov Oncol* **2024**, *15*, 392. https://doi.org/10.1007/s12672-024-01262-z.
31. ADRs, ADEs and SEDs: A Bird's Eye View. In *Side Effects of Drugs Annual*; Elsevier, 2018; Vol. 40, pp. xxvii–xliv.
32. Modulating Cardiac-Gut Microbiome Interaction Post-Myocardial Infarction with Engineered Bacteria | Request PDF Available online: https://www.researchgate.net/publication/393702234_Modulating_Cardiac-Gut_Microbiome_Interaction_Post-Myocardial_Infarction_with_Engineered_Bacteria (accessed on 24 November 2025).
33. Belfiori, M.; Lazzari, L.; Hezzell, M.; Angelini, G.D.; Dong, T. Transcriptomics, Proteomics and Bioinformatics in Atrial Fibrillation: A Descriptive Review. *Bioengineering* **2025**, *12*, 149. https://doi.org/10.3390/bioengineering12020149.
34. Enzyme Induction - an Overview | ScienceDirect Topics Available online: https://www.sciencedirect.com/topics/neuroscience/enzyme-induction (accessed on 24 November 2025).
35. Ma, W.; Ren, H.; Meng, X.; Liu, S.; Du, K.; Fang, S.; Chang, Y. A Review of the Ethnopharmacology, Phytochemistry, Pharmacology, Pharmacokinetics and Quality Control of *Paeonia Lactiflora* Pall. *Journal of Ethnopharmacology* **2024**, *335*, 118616. https://doi.org/10.1016/j.jep.2024.118616.
36. Fang, Y.; Wu, L.-C.; Ma, K.; Pan, G.; Yang, S.; Zheng, Y.; Li, Y. Paeoniflorin Alleviates Lipopolysaccharide-Induced Disseminated Intravascular Coagulation by Inhibiting Inflammation and Coagulation Activation. *Drug Development Research* **2020**, *81*, 517–525. https://doi.org/10.1002/ddr.21647.
37. Ye, S.; Mao, B.; Yang, L.; Fu, W.; Hou, J. Thrombosis Recanalization by Paeoniflorin through the Upregulation of Urokinase-type Plasminogen Activator via the MAPK Signaling Pathway. *Molecular Medicine Reports* **2016**, *13*, 4593–4598. https://doi.org/10.3892/mmr.2016.5146.
38. Li, X.; Sun, C.; Zhang, J.; Hu, L.; Yu, Z.; Zhang, X.; Wang, Z.; Chen, J.; Wu, M.; Liu, L. Protective Effects of Paeoniflorin on Cardiovascular Diseases: A Pharmacological and Mechanistic Overview. *Front. Pharmacol.* **2023**, *14*. https://doi.org/10.3389/fphar.2023.1122969.
39. Bennett, J.A.; Ture, S.K.; Schmidt, R.A.; Mastrangelo, M.A.; Cameron, S.J.; Terry, L.E.; Yule, D.I.; Morrell, C.N.; Lowenstein, C.J. Acetylcholine Inhibits Platelet Activation. *J Pharmacol Exp Ther* **2019**, *369*, 182–187. https://doi.org/10.1124/jpet.118.253583.
40. Liu, W.; Wang, Y.-R.; Wu, H.; Cui, W.; Xu, X. The Role of Myeloperoxidase in the Pathogenesis of Stroke. *Brain Research* **2025**, *1861*, 149705. https://doi.org/10.1016/j.brainres.2025.149705.
41. Ngamsri, K.-C.; Putri, R.A.; Jans, C.; Schindler, K.; Fuhr, A.; Zhang, Y.; Gamper-Tsigaras, J.; Ehnert, S.; Konrad, F.M. CXCR4 and CXCR7 Inhibition Ameliorates the Formation of Platelet–Neutrophil Complexes and Neutrophil Extracellular Traps through Adora2b Signaling. *International Journal of Molecular Sciences* **2021**, *22*, 13576. https://doi.org/10.3390/ijms222413576.
42. Wang, X.; Khalil, R.A. Matrix Metalloproteinases, Vascular Remodeling, and Vascular Disease. *Adv Pharmacol* **2018**, *81*, 241–330. https://doi.org/10.1016/bs.apha.2017.08.002.
43. Causal Relationship between Matrix Metalloproteinase and Pulmonary Embolism: A Bidirectional Two-Sample Mendelian Randomization Study | Scientific Reports Available online: https://www.nature.com/articles/s41598-024-83735-3 (accessed on 19 November 2025).

 doi:10.20944/preprints202604.1287.v1

44. Mastenbroek, T.G.; Feijge, M.A.H.; Kremers, R.M.W.; van den Bosch, M.T.J.; Swieringa, F.; De Groef, L.; Moons, L.; Bennett, C.; Ghevaert, C.; Johnson, J.L.; et al. Platelet-Associated Matrix Metalloproteinases Regulate Thrombus Formation and Exert Local Collagenolytic Activity. *Arteriosclerosis, Thrombosis, and Vascular Biology* **2015**, *35*, 2554–2561. https://doi.org/10.1161/ATVBAHA.115.306153.
45. Amin, M.; Pushpakumar, S.; Muradashvili, N.; Kundu, S.; Tyagi, S.C.; Sen, U. Regulation and Involvement of Matrix Metalloproteinases in Vascular Diseases. *Frontiers in bioscience (Landmark edition)* **2016**, *21*, 89. https://doi.org/10.2741/4378.
46. Cashman, J.R. Human Flavin-Containing Monooxygenase (Form 3): Polymorphisms and Variations In Chemical Metabolism. *Pharmacogenomics* **2002**, *3*, 325–339. https://doi.org/10.1517/14622416.3.3.325.
47. Yu, H.; Chai, X.; Geng, W.-C.; Zhang, L.; Ding, F.; Guo, D.-S.; Wang, Y. Facile and Label-Free Fluorescence Strategy for Evaluating the Influence of Bioactive Ingredients on FMO3 Activity via Supramolecular Host-Guest Reporter Pair. *Biosensors and Bioelectronics* **2021**, *192*, 113488. https://doi.org/10.1016/j.bios.2021.113488.
48. Wang, Y.-R.; Gao, Y.-Q.; Liu, Z.; Liu, W.; Cui, W.-Q.; Wu, H.-Y.; Xu, X.-Q. Paeoniflorin as a Potential Therapeutic Agent for Cerebrovascular Diseases: A Comprehensive Review. *Pharmacological Research* **2025**, *221*, 107965. https://doi.org/10.1016/j.phrs.2025.107965.
49. Wang, Y.; An, J.; Zhou, J.; Chang, L.; Zhang, Q.; Peng, F. Hydroxysafflor Yellow A: A Natural Pigment with Potential Anticancer Therapeutic Effect. *Front. Pharmacol.* **2025**, *15*. https://doi.org/10.3389/fphar.2024.1495393.
50. Wang, K.-H.; Li, S.-F.; Zhao, Y.; Li, H.-X.; Zhang, L.-W. In Vitro Anticoagulant Activity and Active Components of Safflower Injection. *Molecules* **2018**, *23*, 170. https://doi.org/10.3390/molecules23010170.
51. Ao, H.; Feng, W.; Peng, C. Hydroxysafflor Yellow A: A Promising Therapeutic Agent for a Broad Spectrum of Diseases. *Evid Based Complement Alternat Med* **2018**, *2018*, 8259280. https://doi.org/10.1155/2018/8259280.
52. Liu, Y.; Sun, J.; Yang, C.; Qin, M.; Xu, S.; Zhao, Y.; Liu, G. Safflower Yellow for Injection Enhances Anti-Coagulation of Warfarin in Rats: Implications in Pharmacodynamics and Pharmacokinetics. *Xenobiotica* **2024**, *54*, 75–82. https://doi.org/10.1080/00498254.2024.2326987.
53. Zhang, X.; Li, L.; Ding, X.; Kaminsky, L.S. Identification of Cytochrome P450 Oxidoreductase Gene Variants That Are Significantly Associated with the Interindividual Variations in Warfarin Maintenance Dose. *Drug Metabolism and Disposition* **2011**, *39*, 1433–1439. https://doi.org/10.1124/dmd.111.038836.
54. Ruschitzka, F.; Moehrlen, U.; Quaschning, T.; Lachat, M.; Noll, G.; Shaw, S.; Yang, Z.; Teupser, D.; Subkowski, T.; Turina, M.I.; et al. Tissue Endothelin-Converting Enzyme Activity Correlates With Cardiovascular Risk Factors in Coronary Artery Disease. *Circulation* **2000**, *102*, 1086–1092. https://doi.org/10.1161/01.CIR.102.10.1086.
55. Zhang, Z.; Rodriguez, M.; Zheng, Z. Clot or Not? Reviewing the Reciprocal Regulation Between Lipids and Blood Clotting. *Arteriosclerosis, Thrombosis, and Vascular Biology* **2024**, *44*, 533–544. https://doi.org/10.1161/ATVBAHA.123.318286.
56. Li, R.; Cui, M.; Zhao, J.; Yu, M.; Ying, Z.; Zhou, S.; Zhou, H. Association of Endothelin-Converting Enzyme-1b C-338A Polymorphism with Increased Risk of Ischemic Stroke in Chinese Han Population. *J Mol Neurosci* **2013**, *51*, 485–492. https://doi.org/10.1007/s12031-013-0100-y.
57. Bochenek, M.L.; Saar, K.; Nazari-Jahantigh, M.; Gogiraju, R.; Wiedenroth, C.B.; Münzel, T.; Mayer, E.; Fink, L.; Schober, A.; Hübner, N.; et al. Endothelial Overexpression of TGF-β-Induced Protein Impairs Venous Thrombus Resolution. *JACC Basic Transl Sci* **2023**, *9*, 100–116. https://doi.org/10.1016/j.jacbts.2023.08.005.
58. Davis, J.; Maranto, M.; Kennedy, J.; Wang, X.; Azhar, M.; Jain, A.; Evans, C.E. Transforming Growth Factors in Venous Thrombus Formation and Resolution. *Arteriosclerosis, Thrombosis, and Vascular Biology* **2025**, *45*, 643–653. https://doi.org/10.1161/ATVBAHA.124.322395.
59. Mallat, Z.; Gojova, A.; Marchiol-Fournigault, C.; Esposito, B.; Kamaté, C.; Merval, R.; Fradelizi, D.; Tedgui, A. Inhibition of Transforming Growth Factor-β Signaling Accelerates Atherosclerosis and Induces an Unstable Plaque Phenotype in Mice. *Circulation Research* **2001**, *89*, 930–934. https://doi.org/10.1161/hh2201.099415.

 doi:10.20944/preprints202604.1287.v1

60. Olas, B.; Urbańska, K.; Bryś, M. Saponins as Modulators of the Blood Coagulation System and Perspectives Regarding Their Use in the Prevention of Venous Thromboembolic Incidents. *Molecules* **2020**, *25*, 5171. https://doi.org/10.3390/molecules25215171.
61. Olas, B.; Urbańska, K.; Bryś, M. Saponins as Modulators of the Blood Coagulation System and Perspectives Regarding Their Use in the Prevention of Venous Thromboembolic Incidents. *Molecules* **2020**, *25*, 5171. https://doi.org/10.3390/molecules25215171.
62. Latest Evidence and Perspectives of Panax Notoginseng Extracts and Preparations for the Treatment of Cardiovascular Diseases Available online: https://oce-ovid-com.bris.idm.oclc.org/article/00005344-202504000-00003/HTML (accessed on 20 November 2025).
63. De Giorgi, M.; Jarrett, K.E.; Burton, J.C.; Doerfler, A.M.; Hurley, A.; Li, A.; Hsu, R.H.; Furgurson, M.; Patel, K.R.; Han, J.; et al. Depletion of Essential Isoprenoids and ER Stress Induction Following Acute Liver-Specific Deletion of HMG-CoA Reductase. *Journal of Lipid Research* **2020**, *61*, 1675–1686. https://doi.org/10.1194/jlr.RA120001006.
64. Liao, J.K. Isoprenoids as Mediators of the Biological Effects of Statins. *J Clin Invest* **2002**, *110*, 285–288. https://doi.org/10.1172/JCI16421.
65. Dietzen, D.J.; Page, K.L.; Tetzloff, T.A.; Bohrer, A.; Turk, J. Inhibition of 3-Hydroxy-3-Methylglutaryl Coenzyme A (HMG CoA) Reductase Blunts Factor VIIa/Tissue Factor and Prothrombinase Activities via Effects on Membrane Phosphatidylserine. *Arteriosclerosis, Thrombosis, and Vascular Biology* **2007**, *27*, 690–696. https://doi.org/10.1161/01.ATV.0000255949.51053.ce.
66. Barbosa Pereira, P.J.; Segura-Martín, S.; Oliva, B.; Ferrer-Orta, C.; Avilés, F.X.; Coll, M.; Gomis-Rüth, F.X.; Vendrell, J. Human Procarboxypeptidase B: Three-Dimensional Structure and Implications for Thrombin-Activatable Fibrinolysis Inhibitor (TAFI). *Journal of Molecular Biology* **2002**, *321*, 537–547. https://doi.org/10.1016/S0022-2836(02)00648-4.
67. Madaan, T.; Nieman, M.L.; Rakheja, T.; Siddiqui, N.; Gherardini, D.; Sertorio, M.; Kamble, N.S.; Thomas, S.C.; Hartman, A.; Koch, S.E.; et al. Modulating Cardiac-Gut Microbiome Interaction Post-Myocardial Infarction with Engineered Bacteria 2025, 2025.07.10.664025.
68. Agnew, A.; Nulty, C.; Creagh, E.M. Regulation, Activation and Function of Caspase-11 during Health and Disease. *International Journal of Molecular Sciences* **2021**, *22*, 1506. https://doi.org/10.3390/ijms22041506.
69. Yu, N.; Liu, X.; Shi, D.; Bai, L.; Niu, T.; Liu, Y. CD63 and C3AR1: The Potential Molecular Targets in the Progression of Septic Shock. *Int J Gen Med* **2022**, *15*, 711–728. https://doi.org/10.2147/IJGM.S338486.
70. Burrell, M.; Henderson, S.J.; Ravnefjord, A.; Schweikart, F.; Fowler, S.B.; Witt, S.; Hansson, K.M.; Webster, C.I. Neprilysin Inhibits Coagulation through Proteolytic Inactivation of Fibrinogen. *PLOS ONE* **2016**, *11*, e0158114. https://doi.org/10.1371/journal.pone.0158114.
71. LaPelusa, A.; Dave, H.D. Physiology, Hemostasis. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2025.
72. Arterial and Venous Thrombosis | Mackman Lab Available online: https://www.med.unc.edu/mackmanlab/nigel-mackman-laboratory/cross-talk-between-coagulation-and-inflammation/ (accessed on 22 November 2025).
73. Xu, R.-G.; Ariëns, R.A.S. Insights into the Composition of Stroke Thrombi: Heterogeneity and Distinct Clot Areas Impact Treatment. *Haematologica* **2020**, *105*, 257–259. https://doi.org/10.3324/haematol.2019.238816.
74. Nesheiwat, Z.; Goyal, A.; Jagtap, M. Atrial Fibrillation. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2025.
75. Rajan', 'Rajesh; Jarallah', 'Mohammed Al; Jayakumar', 'Rohini B.; Loricchio', 'Maria Luisa; Dashti', 'Raja; Hemeda', 'Asem Abdallah Treating Atrial Fibrillation with Antiplatelet Drugs in the Elderly: Pro and Contra Arguments Available online: https://www.escardio.org/Journals/E-Journal-of-Cardiology-Practice/Volume-17/treating-atrial-fibrillation-with-antiplatelet-drugs-in-the-elderly-pro-and-contra-arguments (accessed on 22 November 2025).
76. New Trials Redefine Antithrombotic and Stroke Prevention Strategies in AFib Available online: https://www.acc.org/latest-in-cardiology/articles/2025/11/03/16/19/http%3a%2f%2fwww.acc.org%2flatest-in-cardiology%2farticles%2f2025%2f11%2f03%2f16%2f19%2fsat-315pm-afib-aha-2025 (accessed on 22 November 2025).

 doi:10.20944/preprints202604.1287.v1

77. Keeling, D.; Tait, R.C.; Watson, H.; Haematology, the B.C. of S. for Peri-Operative Management of Anticoagulation and Antiplatelet Therapy. *British Journal of Haematology* **2016**, *175*, 602–613. https://doi.org/10.1111/bjh.14344.
78. Krishnan, A.; Anahtar, M.N.; Valeri, J.A.; Jin, W.; Donghia, N.M.; Sieben, L.; Luttens, A.; Zhang, Y.; Modaresi, S.M.; Hennes, A.; et al. A Generative Deep Learning Approach to de Novo Antibiotic Design. *Cell* **2025**, *188*, 5962-5979.e22. https://doi.org/10.1016/j.cell.2025.07.033.
79. Huang, H.; Wang, Y.; Rudin, C.; Browne, E.P. Towards a Comprehensive Evaluation of Dimension Reduction Methods for Transcriptomic Data Visualization. *Commun Biol* **2022**, *5*, 1–11. https://doi.org/10.1038/s42003-022-03628-x.
80. Lee, S.B.; Gui, X.; Manquen, M.; Hamilton, E.R. Use of Training, Validation, and Test Sets for Developing Automated Classifiers in Quantitative Ethnography. In Proceedings of the Advances in Quantitative Ethnography; Eagan, B., Misfeldt, M., Siebert-Evenstone, A., Eds.; Springer International Publishing: Cham, 2019; pp. 117–127.