haematologica

Journal of The Ferrata Storti Foundation

Linkage analysis combined with whole exome sequencing identifies a novel prothrombin (F2) gene mutation in a Dutch Caucasian family with unexplained thrombosis

by René Mulder, Ton Lisman, Joost C.M. Meijers, James A. Huntington, André B. Mulder, and Karina Meijer

Haematologica 2019 [Epub ahead of print]

Citation: René Mulder, Ton Lisman, Joost C.M. Meijers, James A. Huntington, André B. Mulder, and Karina Meijer. Linkage analysis combined with whole exome sequencing identifies a novel prothrombin (F2) gene mutation in a Dutch Caucasian family with unexplained thrombosis. Haematologica. 2019; 104:xxx
doi:10.3324/haematol.2019.232504

Linkage analysis combined with whole exome sequencing identifies a novel prothrombin (*F2*) gene mutation in a Dutch Caucasian family with unexplained thrombosis

René Mulder[1], Ton Lisman[2], Joost C.M. Meijers[3], James A. Huntington[4], André B. Mulder[1], and Karina Meijer[5]

[1]Department of Laboratory Medicine, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands

[2]Surgical Research Laboratory and Section of Hepatobiliary Surgery and Liver Transplantation, Department of Surgery, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

[3]Department of Molecular and Cellular Hemostasis, Sanquin Research, Amsterdam, the Netherlands; Amsterdam UMC, University of Amsterdam, Department of Experimental Vascular Medicine, Amsterdam Cardiovascular Sciences, Amsterdam, the Netherlands

[4]Department of Haematology, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK

[5]Department of Hematology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

Correspondence: Dr. René Mulder at the Department of Laboratory Medicine, University Medical Center Groningen, Hanzeplein 1, 9700 RB, Groningen, the Netherlands, r.mulder01@umcg.nl.

Running title: Novel F2 gene mutation in unexplained thrombosis

Total word count: 1478 (max 1500)

Total number of figures: 2 (max 3)

1 supplementary

Genetics play a major role in the development of venous thromboembolism (VTE). However, current thrombophilia testing identifies a known heritable defect in only a minority of VTE patients.[1] A part of this missing heritability for VTE may be caused by rare variants.[2] In the current study we combined genome-wide linkage analysis with whole-exome sequencing (WES) and identified a rare novel prothrombin (*F2*) gene mutation (c.1621C>T; p.Arg541Trp hereafter referred to as p.Arg173Trp according to chymotrypsin numbering) in a Dutch family with unexplained thrombosis.

The proband, from Dutch Caucasian ancestry, presented at age 31 with a first deep vein thrombosis (DVT) during pregnancy. Shortly after, her father had DVT after hospitalization at age 58. The family was aware that they were prone for thrombosis, and offered a family tree of the older generations. The proband's paternal uncle had had spontaneous PE at age 38 and age 42, a cousin had a spontaneous DVT at age 44 (data obtained from the subjects themselves, VTE diagnoses in generations I, II and III were obtained from family members). Routine thrombophilia testing in the proband was negative, excluding deficiencies of antithrombin, protein C, protein S, and factor V Leiden and prothrombin G20210A mutation. In the years after, both the proband and her father had a spontaneous DVT recurrence, and the proband's younger sister presented with estrogen-related PE. We hypothesized that this family carried a private, high risk mutation. After informed consent was obtained we included a total of 19 family members for testing.

First, genome-wide SNP genotyping was performed on 19 family members (5 affected and 14 unaffected; Figure 1A) (Online Supplementary) and identified twelve chromosomes with equally high significant linkage peaks (LOD > 1). These chromosomes were: 3, 4, 7, 9, 11, 13, 14, 18, 19, 20, 22, and X. Next, WES was performed on 5 family members (4 affected: 26101, 26103, 26105, 26206; 1 unaffected: 26312; Figure 1A) (Online Supplementary). In total, 138,672 variants were detected, of which, after rigorous filtering, 113 remained common among affected family members. When we combined these results with linkage data, e.g. filtering based on their distance (~ 3 Mb) to the linkage peak, 30 variants remained. In the final step, in silico prediction was performed with MetaSVM and

MetaLR thereby identifying 1 potentially pathogenic novel prothrombin (*F2*) gene mutation (p.Arg173Trp) (Figure 1B) located on chromosome 11. Using Sanger sequence analysis we confirmed the heterozygous presence of a novel prothrombin (*F2*) gene mutation (Figure 1C). Next, with a custom made TaqMan qPCR assay (Applied Biosystems) we confirmed the presence of the prothrombin (*F2*) gene mutation p.Arg173Trp in 4 affected family members (26101, 26103, 26105, 26206), whereas 1 unaffected family member (216209) also carried this mutation (Figure 1D). Furthermore, one family member who did not carry the mutation (216418) had a history of deep vein thrombosis.

Functional testing was complicated due to the use of vitamin K antagonists in the carriers of the prothrombin mutation with available plasma. We therefore performed functional assays with recombinant prothrombin molecules (U-Protein Express BV Utrecht, the Netherlands). A hydroxyapatite column was used to obtain fully γ-carboxylated wild-type and mutant prothrombin. Prothrombin levels were measured a by one-stage clotting assay on an automated coagulation analyser (ACL 300 TOP) with reagents and protocols from the manufacturer (Instrumentation Laboratory, Breda, The Netherlands). At a final concentration of 90 µg/ml in prothrombin deficient plasma, the recombinant wild-type and mutant prothrombins had prothrombin activity of 82% and 67%, respectively. Furthermore, thrombin generation was performed using calibrated automated thrombography and Thrombinoscope software according to the manufacturer's instruction (Thrombinoscope BV, Maastricht, The Netherlands). Recombinant wild-type or p.Arg173Trp prothrombin was added to prothrombin deficient plasma (Siemens) at a final concentration of 90 µg/ml in the absence or presence of 0.1 U/ml unfractionated heparin (LEO Pharma BV) and in vitro thrombin generation was assessed. Figure 1E depicts representative thrombin generation curves. In the absence of unfractionated heparin, plasma containing p.Arg173Trp prothrombin had a substantially higher endogenous thrombin potential as compared to plasma to which wild-type prothrombin was added (2400 ± 113 vs 797 ± 42 nM IIa*min, $p<0.0001$, n=3). Addition of unfractionated heparin reduced thrombin generation in plasma containing the p.Arg173Trp

prothrombin by only 7.3 ± 3.9% compared to a 31.4 ± 9.3% reduction in thrombin generation in plasma containing wild-type prothrombin (p=0.01, n=3).

Finally, we performed molecular dynamics in order to simulate the effect of the p.Arg173Trp mutation (Online Supplementary). The prothrombin (F2) gene mutation is located in the 170s loop, near the non-primed side of the active site (Figure 2A). The results of molecular dynamics are consistent with a conformational alteration in the non-prime side of the active site due to the repositioning of the Trp173 side chain to minimize its solvent accessibility by packing up against Trp215 (Figure 2B).

Venous thrombosis is a disorder in which genetic predisposition plays an important role. However, in a large portion of patients with VTE no genetic defect can be detected. Several aspects contribute to the missing heritability for VTE, including rare variants. To identify such variants we combined linkage analysis with WES and found a novel prothrombin (*F2*) gene mutation p.Arg173Trp.

This mutation is assumed to be rare as it was absent in 2 control groups, e.g. in 24 probands of families with unexplained thrombosis and 100 healthy volunteers without VTE (selection from GoNL; http://www.nlgenome.nl/) (data not shown). Furthermore, this mutation was not present in dbSNP (https://www.ncbi.nlm.nih.gov/snp) or HGMD (http://www.hgmd.cf.ac.uk/ac/index.php).

Prothrombin deficiency has long been considered to only be associated with bleeding tendency until Miyawaki et al. reported in 2012 on a family with dysprothrombinemia and venous thrombosis.[3] This dysprothrombinemia was caused by a substitution of arginine to leucine at position 221a in *F2* gene (prothrombin Yukuhashi).[3] In the following years, 2 additional mutations at position 221a were reported, e.g. prothrombin Belgrade[4] and prothrombin Padua 2.[5] Apart from prothrombin Belgrade, these mutations were all associated with a strong risk of VTE at young age.[3, 5] The association with VTE was explained by antithrombin resistance caused by mutations at position 221a.[3-5]

All carriers of the prothrombin (*F2*) gene mutation p.Arg173Trp, with the exception of 216209 have had venous thrombosis. One family member who did not carry the mutation (216418) also had a history of deep vein thrombosis. We cannot exclude that additional unknown thrombophilic factors are present in the family. We have also observed in studies of protein C, S and antithrombin deficiency that non-affected family members have a higher incidence of VTE than the normal population.[6] However, venous thromboembolism is a common disorder and this one case in a family member without the mutation might also have been due to chance.

Our functional studies showed a clear heparin resistance that is suggestive for antithrombin resistance. In line with this, Tamura et al. performed an in vitro exploratory study using prothrombin mutants and showed that a recombinant p.Arg173Trp mutation exhibited a mild antithrombin resistance.[7]

The side chain of p.Arg173 is not fully modeled in the structure of the antithrombin-thrombin Michaelis complex, suggesting that it is not making a single distinct contact with antithrombin. It is in a position where it could exert favorable long-range electrostatic influence on docking, since it is in the vicinity of glutamates 232, 377 and 378 on antithrombin. However, such an influence would be overwhelmed by the magnitude of the effect of heparin on the association rate. The heparin resistance of the p.Arg173Trp thrombin variant therefore suggests something more profound than the loss of a favorable interaction or electrostatic influence. What is more likely is a structural rearrangement of the active site caused by the introduction of a large hydrophobic residue to a solvent-exposed loop (Figure 2A). The 170s loop has been implicated in the zymogen to protease conformational change and allosteric activation. The molecular dynamics results suggest a distinct conformational change in the non-prime side of the active site due to the Trp173 burying itself as much as possible thereby interfering with antithrombin inhibition (Figure 2B). However, more evidence is still needed to conclude that the prothrombin mutation p.Arg173Trp leads to antithrombin resistance.

Finally, at a final concentration of 90 µg/ml in deficient plasma, the recombinant wild-type and mutant showed borderline low prothrombin activity levels, which is in contrast with the increased thrombin potential observed with the thrombin generation assay. Such a discrepancy has also been observed for 4 antithrombin-resistant mutations in position 221a,[7] . This discrepancy could be explained by the fact that clotting assays measure only the clotting time corresponding to the initiation phase of the coagulation process, whereas thrombin generation assays also provide information about the amplification and propagation phases of the hemostatic system.

In conclusion, in this study we identified a rare novel prothrombin (*F2*) gene mutation p.Arg173Trp that causes heparin resistance which is suggestive for antithrombin resistance. Evidently further investigations are needed to fully disentangle the mechanism responsible for the potential association between the prothrombin (*F2*) gene mutation p.Arg173Trp and venous thrombosis.

References

1. Morange PE, Tregouet DA. Deciphering the molecular basis of venous thromboembolism: where are we and where should we go? Br J Haematol. 2010;148(4):495-506.

2. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet. 2008;9(5):356-369.

3. Miyawaki Y, Suzuki A, Fujita J, et al. Thrombosis from a prothrombin mutation conveying antithrombin resistance. N Engl J Med. 2012;366(25):2390-2396.

4. Djordjevic V, Kovac M, Miljic P, et al. A novel prothrombin mutation in two families with prominent thrombophilia--the first cases of antithrombin resistance in a Caucasian population. J Thromb Haemost. 2013;11(10):1936-1939.

5. Bulato C, Radu CM, Campello E, et al. New Prothrombin Mutation (Arg596Trp, Prothrombin Padua 2) Associated With Venous Thromboembolism. Arterioscler Thromb Vasc Biol. 2016;36(5):1022-1029.

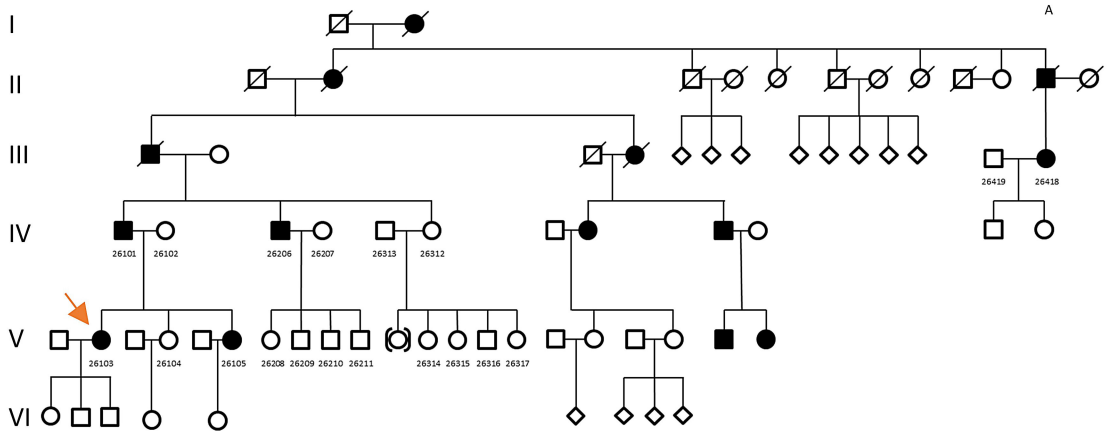
6. Mahmoodi BK, Brouwer JL, Ten Kate MK, et al. A prospective cohort study on the absolute risks of venous thromboembolism and predictive value of screening asymptomatic relatives of patients with hereditary deficiencies of protein S, protein C or antithrombin. J Thromb Haemost. 2010;8(6):1193-1200.

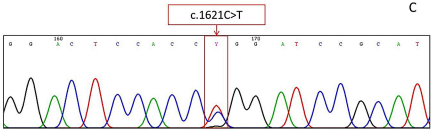
7. Tamura S, Murata-Kawakami M, Takagi Y, et al. In vitro exploration of latent prothrombin mutants conveying antithrombin resistance. Thromb Res. 2017;159:33-38.

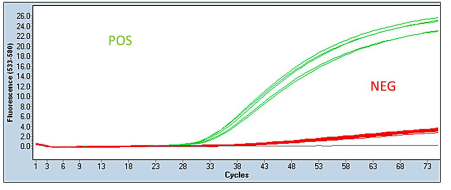
11. Takagi Y, Murata M, Kozuka T, et al. Missense mutations in the gene encoding prothrombin corresponding to Arg596 cause antithrombin resistance and thrombomodulin resistance. Thromb Haemost. 2016;116(6):1022-1031.

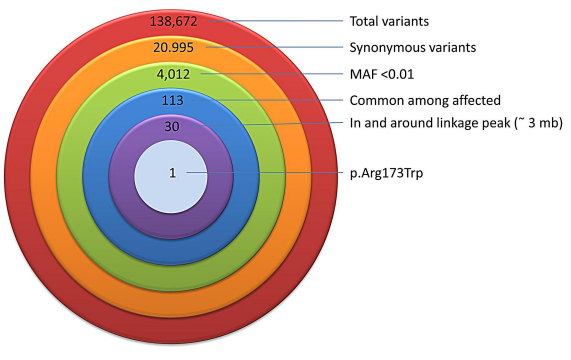
Figure 1. Identification and functional characterization of novel prothrombin (F2) gene mutation in a Dutch family with unexplained thrombosis. A. The pedigree. The arrow indicates the proband. Roman numbers denote the generation in the pedigree; study number of included family members is depicted below each symbol; square symbols, male sex; round symbols, female sex; solid symbols represent VTE-affected family members, open symbols represent unaffected family members. B. WES combined with linkage analysis. The diagram depicts effects of filtering WES data based on specific aspects including among other things such as type of mutation and location to linkage peak. C. Direct sequencing results of exon 12 of *F2* gene. The prothrombin (*F2*) gene mutation (c.1621C>T; p.Arg173Trp) is highlighted. D. This figure shows the genotyping results of a custom made Taqman qPCR assay for 19 included family members. The green lines represent the carriers of the prothrombin (*F2*) gene mutation (26101, 26103, 26105, 26206, and 26209), whereas red lines represent the non-carriers of the mutation (26102, 26104, 26207, 26208, 26210, 26211, 26312, 26313, 26314, 26315, 26316, 26317, 26418, and 26419). E. Thrombin generation in prothrombin depleted plasma supplemented with wild-type (WT) and mutant protein in the presence or absence of 0.1 U/ml unfractionated heparin.

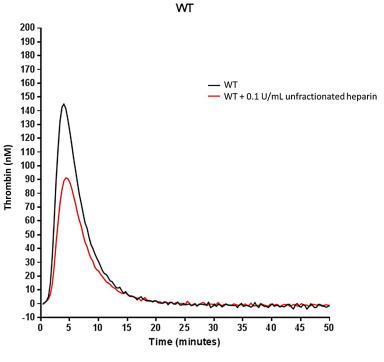
Figure 2. Structural characterization of novel prothrombin (F2) gene mutation. A. A ribbon diagram of the heavy chain of thrombin in the standard orientation showing the position of the mutation (side chain of Arg173 as sticks) and the Na+ (purple ball) binding site (colored from N-to-C terminus, blue to red). B. Surface representation of the resulting structures after molecular dynamics simulation of wild-type (left) and the p.Arg173Trp mutant variant (right) thrombin coloured according to electrostatics. Oval indicates the non-primed side of the active site cleft, and arrow indicates the position of the mutation.

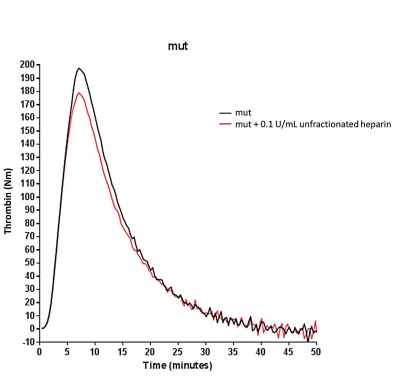

A
I
II
III
IV
V
VI
26101
26102
26206
26207
26313
26312
26419
26418
26103
26104
26105
26208
26209
26210
26211
26314
26315
26316
26317

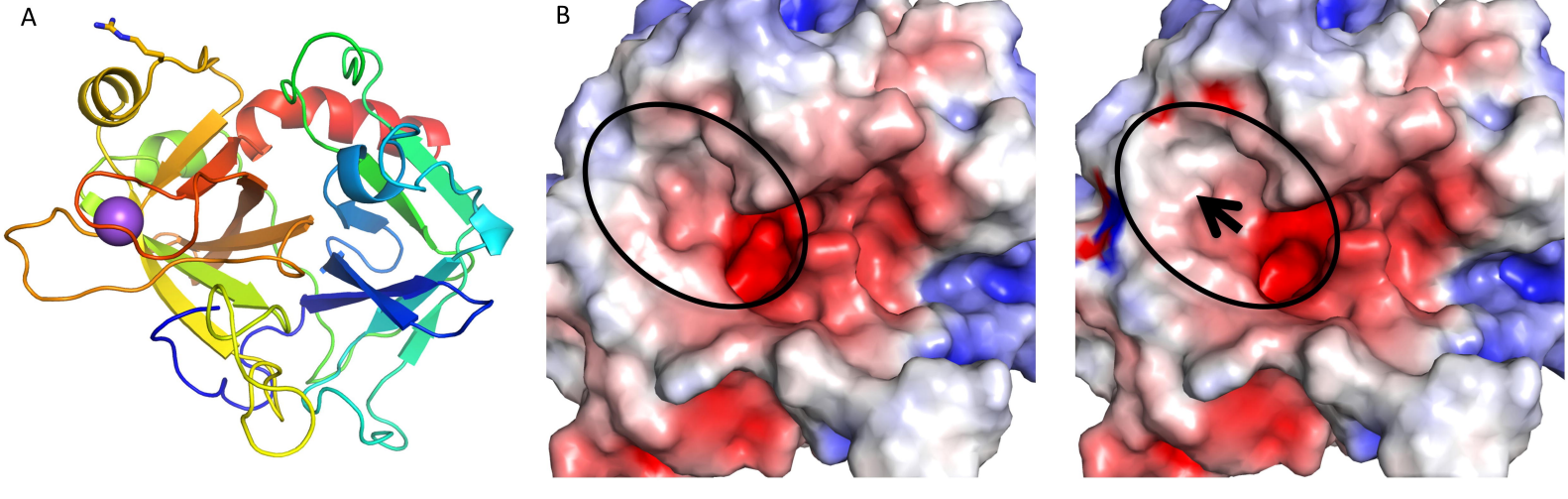

C
c.1621C>T


D
POS
NEG
Fluorescence (533-580)
Cycles


B
138,672
Total variants
20.995
Synonymous variants
4,012
MAF <0.01
113
Common among affected
30
In and around linkage peak (~ 3 mb)
1
p.Arg173Trp


WT
WT
WT + 0.1 U/mL unfractionated heparin
Thrombin (nM)
Time (minutes)


E
mut
mut
mut + 0.1 U/mL unfractionated heparin
Thrombin (Nm)
Time (minutes)


A
B

Supplementary Appendix

Linkage analysis combined with whole exome sequencing identifies a novel prothrombin (*F2*) gene mutation in a Dutch Caucasian family with unexplained thrombosis

René Mulder[1], Ton Lisman[2], Joost C.M. Meijers[3], James A. Huntington[4], André B. Mulder[1], and Karina Meijer[5]

[1]Department of Laboratory Medicine, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands; [2]Surgical Research Laboratory and Section of Hepatobiliary Surgery and Liver Transplantation, Department of Surgery, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; [3]Department of Molecular and Cellular Hemostasis, Sanquin Research, Amsterdam, the Netherlands; Amsterdam UMC, University of Amsterdam, Department of Experimental Vascular Medicine, Amsterdam Cardiovascular Sciences, Amsterdam, the Netherlands; [4]Department of Haematology, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK; [5]Department of Hematology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands;

Correspondence: Dr. René Mulder at the Department of Laboratory Medicine, University Medical Center Groningen, Hanzeplein 1, 9700 RB, Groningen, the Netherlands, r.mulder01@umcg.nl.

Linkage analysis combined with whole exome sequencing identifies a novel prothrombin (*F2*) gene mutation in a Dutch Caucasian family with unexplained thrombosis

René Mulder[1], Ton Lisman[2], Joost C.M. Meijers[3], James A. Huntington[4], André B. Mulder[1], and Karina Meijer[5]

[1]Department of Laboratory Medicine, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands

[2]Surgical Research Laboratory and Section of Hepatobiliary Surgery and Liver Transplantation, Department of Surgery, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

[3]Department of Molecular and Cellular Hemostasis, Sanquin Research, Amsterdam, the Netherlands; Amsterdam UMC, University of Amsterdam, Department of Experimental Vascular Medicine, Amsterdam Cardiovascular Sciences, Amsterdam, the Netherlands

[4]Department of Haematology, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK

[5]Department of Hematology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

Correspondence: Dr. René Mulder at the Department of Laboratory Medicine, University Medical Center Groningen, Hanzeplein 1, 9700 RB, Groningen, the Netherlands, r.mulder01@umcg.nl.

Genomic DNA

Genomic DNA was obtained from citrated and EDTA samples using the Qiacube system.

Linkage analysis

Genome-wide SNP genotyping was performed on 19 family members (5 affected and 14 unaffected; Figure 1A) using Illumina Infinium Core Exome-24v1.1 bead chip. The bead chip included 551,839 SNP markers distributed evenly across the genome. The mean and median intervals between markers were 5.26 kb and 1.82 kb, respectively.

For linkage analysis, we first excluded SNPs with genotype call rate less than 95%, MAF less than 0.001 and HWE test p-value less than 0.0001. Next, frequency and genotyping pruning was performed. From the subset, SNPs were checked for strong LD, since high LD within dense SNP regions can result in false positive results. We applied prune SNPs within strong LD using –indep 50 5 10 option in PLINK v1.07. Strong LD pruning can be defined as removing SNPs within a 50 SNP window, with 5 SNPs to shift the window at each step that had r^2>0.9 (corresponding to a variance inflation factor, VIF, greater than 10) with all other SNPs in the window. We then checked the SNPs with Mendelian inheritance error using Pedstats program[1] and masked the genotypes for the SNPs with error using Pedwipe program. For pruned and masked SNP data, non-parametric linkage analyses will be performed using the Merlin 1.1.2.[2] We performed non-parametric analyses to get LOD scores and statistics. A LOD score of more than 0 would be suggestive for evidence of linkage.

Whole exome sequencing (WES)

The SureSelect Human All Exon V5 was used for WES on 5 family members (4 affected: 26101, 26103, 26105, 26206; 1 unaffected: 26312; Figure 1A).

Read mapping

Paired-end sequences produced by HiSeq Instrument were first mapped to the human genome, with the reference sequence as UCSC assembly hg19 (original GRCh37 from NCBI, Feb. 2009), and without unordered sequences or alternate haplotypes. 'BWA' (version 0.7.12) was used as the mapping tool, which generates the mapping result file in BAM format using 'BWA-MEM'. Then, programs within Picard-tools (ver.1.130) were used in order to remove PCR duplicates, reducing those reads to identically match to a position at start into a single one, using MarkDuplicates.jar, which requires reads to be sorted. The local realignment process were performed to consume BAM files and to locally realign reads such that the number of mismatching bases is minimized across all reads. Base quality score recalibration (BQSR) and local realignment around indels were performed using Genome Analysis Toolkit to locally realign reads such that the number of mismatching bases was minimized.

Variant calling and annotation

Based on the BAM file previously generated, variant genotyping for each sample was performed with Haplotype Caller of GATK (v3.4.0). In this stage SNPs and short indels candidates were detected at the nucleotide resolution. These variants were annotated by a program called SnpEff v4.1g and converted to vcf file format. They were filtered with dbSNP version 142, and SNPs from the 1000

genome project. Then, in-house program and SnpEff were applied to filter additional databases, including ESP6500, ClinVar, dbNSFP2.9.

Family analysis

To identify the genetic variants underlying rare diseases, variants in all family members were joint-called with GATK (v3.4.0) and filtered with an in-house developed script. Firstly, exon-, exonic-splicing, and splicing regions are selected to merge variant data. Synonymous single nucleotide variants (SNVs) within the remaining variants were then removed. Furthermore, dbSNP135 common (freq>=1%) and 1000 genome freq >0.01 variants were eliminated. Next, variants uncommon among affected family members were filtered out. In addition, only variants at chromosomes with linkage peaks were further filtered based on their distance (~ 3 Mb) to the linkage peak. Remaining variants with variant ids were then filtered out. Finally, variants in genes known to be associated with coagulation were selected and further filtered based on the in silico prediction where only variants that scored deleterious with MetaSVM and MetaLR remained.[3]

Molecular dynamics

The starting model for the molecular dynamics simulation was thrombin from 1JOU (chains A and B), including surrounding water molecules and sodium ion, modified by mutation of Ala195 back to Ser and Arg173 to Trp. The simulation was conducted using the program YASARA within a hydrated box surrounding the thrombin molecule.[4, 5] The YAMBER forcefield was used with simulated annealing at 335K. The parent molecule (p.Ala195Ser) was run for 18 ps to obtain the minimized starting structure which did not result in significant changes. Arg173 was then mutated to Trp and the simulation was restarted. In order to bury the side chain of Trp173 it was selected and pulled toward the S4 pocket, followed by 26 ps of energy minimization with simulated annealing.

References

1. Wigginton JE, Abecasis GR. PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. Bioinformatics. 2005;21(16):3445-3447.

2. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet. 2002;30(1):97-101.

3. Dong C, Wei P, Jian X, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum Mol Genet. 2015;24(8):2125-2137.

4. Krieger E, Vriend G. New ways to boost molecular dynamics simulations. J Comput Chem. 2015;36(13):996-1007.

5. Krieger E, Darden T, Nabuurs SB, Finkelstein A, Vriend G. Making optimal use of empirical energy functions: force-field parameterization in crystal space. Proteins. 2004;57(4):678-683.