

RESEARCH ARTICLE

Whole exome sequencing in thrombophilic pedigrees to identify genetic risk factors for venous thromboembolism

Marisa L. R. Cunha^{1,2*}, Joost C. M. Meijers^{1,3}, Frits R. Rosendaal⁴, Astrid van Hylckama Vlieg^{4,5}, Pieter H. Reitsma^{5,6}, Saskia Middeldorp^{2*}

1 Department of Experimental Vascular Medicine, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands, **2** Department of Vascular Medicine, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands, **3** Department of Plasma Proteins, Sanquin, Amsterdam, the Netherlands, **4** Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands, **5** Department of Thrombosis and Hemostasis, Leiden University Medical Center, Leiden, the Netherlands, **6** Eindhoven Laboratory for Experimental Vascular Medicine, Leiden University Medical Center, Leiden, the Netherlands

* Current address: Personalised Healthcare and Biomarkers, AstraZeneca, Cambridge, United Kingdom
* s.middeldorp@amc.uva.nl



OPEN ACCESS

Citation: Cunha MLR, Meijers JCM, Rosendaal FR, Vlieg AvH, Reitsma PH, Middeldorp S (2017) Whole exome sequencing in thrombophilic pedigrees to identify genetic risk factors for venous thromboembolism. PLoS ONE 12(11): e0187699. <https://doi.org/10.1371/journal.pone.0187699>

Editor: Pablo Garcia de Frutos, Institut d'Investigacions Biomediques de Barcelona, SPAIN

Received: May 9, 2017

Accepted: October 24, 2017

Published: November 8, 2017

Copyright: © 2017 Cunha et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was funded by a grant (2010) from the Rembrandt Institute of Cardiovascular Science, Amsterdam/Leiden, The Netherlands. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: MLRC's current affiliation with AstraZeneca is not connected to the work

Abstract

Background

Family studies have shown a strong heritability component for venous thromboembolism (VTE), but established genetic risk factors are present in only half of VTE patients.

Aim

To identify genetic risk factors in two large families with unexplained hereditary VTE.

Methods

We performed whole exome sequencing in 10 affected relatives of two unrelated families with an unexplained tendency for VTE. We prioritized variants shared by all affected relatives from both families, and evaluated these in the remaining affected and unaffected individuals. We prioritized variants based on 3 different filter strategies: variants within candidate genes, rare variants across the exome, and SNPs present in patients with familial VTE and with low frequency in the general population. We used whole exome sequencing data available from 96 unrelated VTE cases with a positive family history of VTE from an affected sib study (the GIFT study) to identify additional carriers and compared the risk-allele frequencies with the general population. Variants found in only one individual were also retained for further analysis. Finally, we assessed the association of these variants with VTE in a population-based case-control study (the MEGA study) with 4,291 cases and 4,866 controls.

Results

Six variants remained as putative disease-risk candidates. These variants are located in 6 genes spread among 3 different loci: 2p21 (*PLEKHH2* NM_172069:c.3105T>C, *LRPPRC*

described here, and does not constitute a conflict of interest. The other authors have no conflicts of interest to declare. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

rs372371276, *SRBD1* rs34959371), 5q35.2 (*UNC5A* NM_133369.2:c.1869+23C>A), and 17q25.1 (*GPRC5C* rs142232982, *RAB37* rs556450784). In GIFT, additional carriers were identified only for the variants located in the 2p21 locus. In MEGA, additional carriers for several of these variants were identified in both cases and controls, without a difference in prevalence; no carrier of the *UNC5A* variant was present.

Conclusion

Despite sequencing of several individuals from two thrombophilic families resulting in 6 candidate variants, we were unable to confirm their relevance as novel thrombophilic defects.

Background

Venous thromboembolism (VTE), comprising deep vein thrombosis (DVT) and pulmonary embolism (PE), is a common disorder with a high mortality rate worldwide [1]. After a first episode of VTE patients have an elevated risk of a recurrent episode that is as high as 30% to 50% within 10 years in those with unprovoked VTE [2]. Major acquired risk factors for VTE include cancer, trauma, surgery, immobilization, pregnancy, use of oral contraceptives and the antiphospholipid syndrome. A family history of VTE increases the risk as well and indicates an important underlying genetic component [3–5].

Sequencing-based studies and, to a lesser extent, genome-wide association studies (GWAs), have contributed to the discovery of novel genetic risk factors for VTE [6–14]. Rare mutations identified by sequencing of candidate genes, i.e. antithrombin, protein C and protein S, are those that convey the highest risk [15]. Yet, the currently known genetic risk factors are present in approximately half of the patients with unexplained but familial VTE [16], which suggests that other genetic risk factors remain to be identified. The discovery of such genetic causes of VTE may unravel novel disease-causing mechanisms in or outside the coagulation cascade.

The aim of this study was to identify novel genetic risk factors for VTE by whole exome sequencing of 10 individuals from two unrelated Dutch families with an unexplained tendency for VTE, to estimate the prevalence of any candidate variants among patients with familial VTE, and to subsequently validate these in a large population-based case control study. We hypothesized that unexplained familial VTE is caused by the presence of genetic variants located in coding regions of genes with known or unknown roles in hemostasis.

Materials and methods

Subjects

Two large Dutch families, referred to as Family D and Family K, were selected from the GENES study, which has been described previously [17]. In both families, 5 or more individuals had had objectively confirmed VTE, in the absence of known hereditary thrombophilia (antithrombin-, protein C-, and protein S deficiency, factor V Leiden and prothrombin G20210A). All affected individuals from both families had experienced VTE at a relatively young age and some individuals had had DVT at unusual sites (abdomen or arms). The pedigrees are depicted in Fig 1.

In addition to the two families, two sets of patients with familial VTE, as well as a population-based case-control study of VTE were included in this study. The patients with familial VTE comprised of 36 unrelated cases with familial VTE from the GENES study [17], and 434

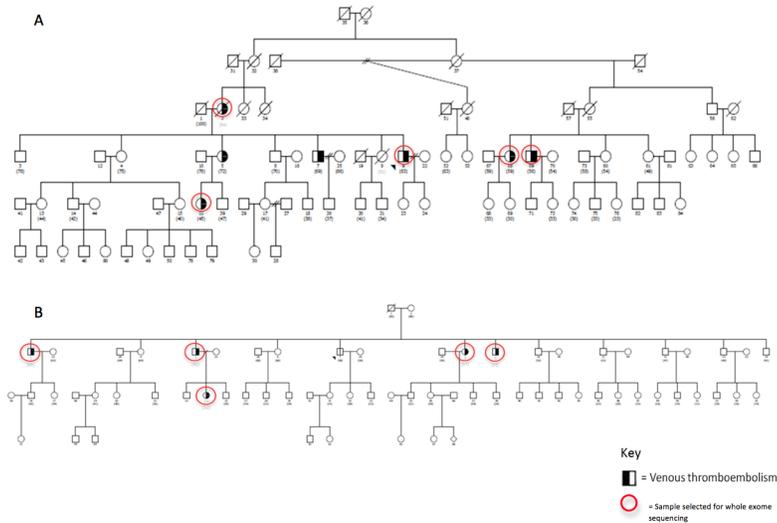


Fig 1. Pedigree of the families with inherited VTE. Half shaded symbols indicate affected individuals. In family D (A), 7 individuals had had objectively confirmed VTE while in family K (B), 5 individuals had had objectively confirmed VTE. Exome data was generated from all individuals surrounded by a red circle. Please notice that although the affected individual 5 from family D was not genetically tested, the genotype of her daughter (individual 11) who also has the same phenotype, is available. Therefore, it can be assumed that all candidate mutations validated in individual 11 are also present in the individual 5. Moreover, none of the individuals from the oldest generation of family K had had an objectively confirmed VTE event. The VTE events for each of the affected individuals from family D are the following: DVT in the leg at age of 46 for individual 59; DVT in the leg at age of 48 for individual 58; DVT in the leg and in the arm, and PE at age of 42, 50, and 52, respectively, for individual 9; DVT in the leg at age of 26 for individual 2; PE at age of 22 for individual 11; DVT in the leg at age of 30 for individual 5, and; DVT in the leg at age of 56 for individual 7. The VTE events for each of the affected individuals from family K are the following: PE at age of 44 for individual 9; DVT in the abdomen at age of 41, 42 and 45 for individual 8; DVT in the leg at age of 53 for individual 5; PE at the age of 43 and 45 for individual 3 and; PE at age of 23 for individual 24.

<https://doi.org/10.1371/journal.pone.0187699.g001>

cases with VTE from 201 families from the GIFT study, which included affected sibs with thrombosis before the age of 45, in the absence of known thrombophilia [18]. The population-based MEGA study consisted of 4,291 cases with a first VTE and 4,866 controls [19]. Identification of cases and controls has been described in detail previously [17–19]. The GENES study was approved by the Medical Ethics Committee of the Academic Medical Center, Amsterdam, the Netherlands. The GIFT and the MEGA studies were approved by the Medical Ethics Committee of the Leiden University Medical Center, Leiden, the Netherlands. Written informed consent was obtained from all participants.

Whole-exome sequencing

Families. Genomic DNA extracted from whole blood from 5 affected relatives from both families was sent to the Beijing Genomics Institute for whole exome sequencing Fig 1. The exomes were captured using the Agilent SureSelect Human all Exon (44M) kit and the enriched exome libraries were multiplexed and sequenced on the Illumina HiSeq2000 platform to generate 90-bp paired-end reads per individual with an average sequencing depth above 50x. The exome design covers 44 Mb of human genome corresponding to the exons and flanking intronic regions of ~18,000 genes in the National Center for Biotechnology Information Consensus CDS database (April 2011 release).

The sequencing reads were aligned to the human reference genome sequence NCBI Build 37 (hg19) using the Burrows-Wheeler Aligner (BWA) version (v) 0.6.2. Then, each SAM file

resulting from each sample was converted to a BAM file using SAMtools (v0.1.18). Picard tools (v1.78) and GATK tools (v2.2) were used to process the BAM files. Variant calling was performed with all samples simultaneously. Single nucleotide variants/polymorphisms (here globally called as SNPs), as well as insertions and deletions (INDELs), were called by the UnifiedGenotyper walker from GATK tools. After quality score recalibration and removal of low-confidence variants, all SNPs and INDELs were annotated using ANNOVAR (November 2012 release). All steps were performed according to GATK Best Practices recommendations [20].

GIFT study. Genomic DNA extracted from whole blood from 96 unrelated cases with familial VTE from the GIFT study was sent to the The Post-Genomic Platform of the Pitié-Salpêtrière (P3S) for whole exome sequencing. The DNA libraries were performed using TruSeq DNA sample Prep kit (Illumina) and the exomes were captured using the TruSeq Exome Enrichment kit (Illumina). The enriched exome libraries were multiplexed and sequenced on the Illumina HiSeq2000 platform to generate 100-bp paired-end reads per individual with an average sequencing depth above 50x. The exome design covers 62 Mb of human genome corresponding to the exons and UTR regions of 20794 genes.

Illumina's CASAVA software v1.8.2 was used to demultiplex samples and convert BCL files to FASTQ format files. Quality controls of the raw data were performed with FastQC (v0.10.1). Reads with poor quality were removed (`fastq_illumina_filter-0.1`) and low quality bases of the reads were trimmed (`sickle v1.200`). After these preprocessing steps, sequenced data were aligned to the human genome reference sequence NCBI Build 37 (hg19) using BWA (v0.6.2). Reads in the resulting BAM files were then realigned around INDELs and the base quality score were recalibrated using GATK (v2.4–3). Duplicates and singletons were then removed along with non-well oriented reads using Picard (v1.85) and Samtools (v0.1.18). Variant calling was performed on bases with good quality ($Q\text{-score} > 20$) using the samtools `mpileup` command. Variants that were present in GIFT and in individuals from families D and K were identified.

Filter strategies

To identify the genetic variant responsible for the increased risk for VTE in members from the families D and K, we applied 3 different strategies, summarized in Fig 2. These strategies were based on: (i) variants within candidate genes (filter strategy 1), (ii) rare variants across the exome (filter strategy 2), (iii) SNPs present in patients with familial VTE and with low frequency (minor allele frequency $< 5\%$) in the general population (filter strategy 3). All strategies assume that familial VTE is an autosomal dominant disease, and that all affected family members from a certain family share the same genetic risk variant. In addition, variants located in the X chromosome were considered for analysis in filter strategies 1 and 2, as VTE affects both males and females.

For filter strategy 1, we created a list of genes known or possibly associated with VTE based on the available literature. This list comprises a total of 126 genes and is available in [S1 File](#). Following the selection of SNPs and INDELs located within these candidate genes, we enriched for variants more likely to affect protein function, by excluding non-coding variants without any potential splice site prediction or involving a low conserved base pair, as well as synonymous variants without any potential splice site prediction. Then, we tested the effect of these variants on hemostatic traits in cases from the GIFT study, although we realize that this approach is subject to collider bias [21,22]. We analysed 413 cases, for which previously investigated hemostatic traits and genotypes were available from a previous genome-wide association study [14]. For more information on the association analysis see further below in this section.

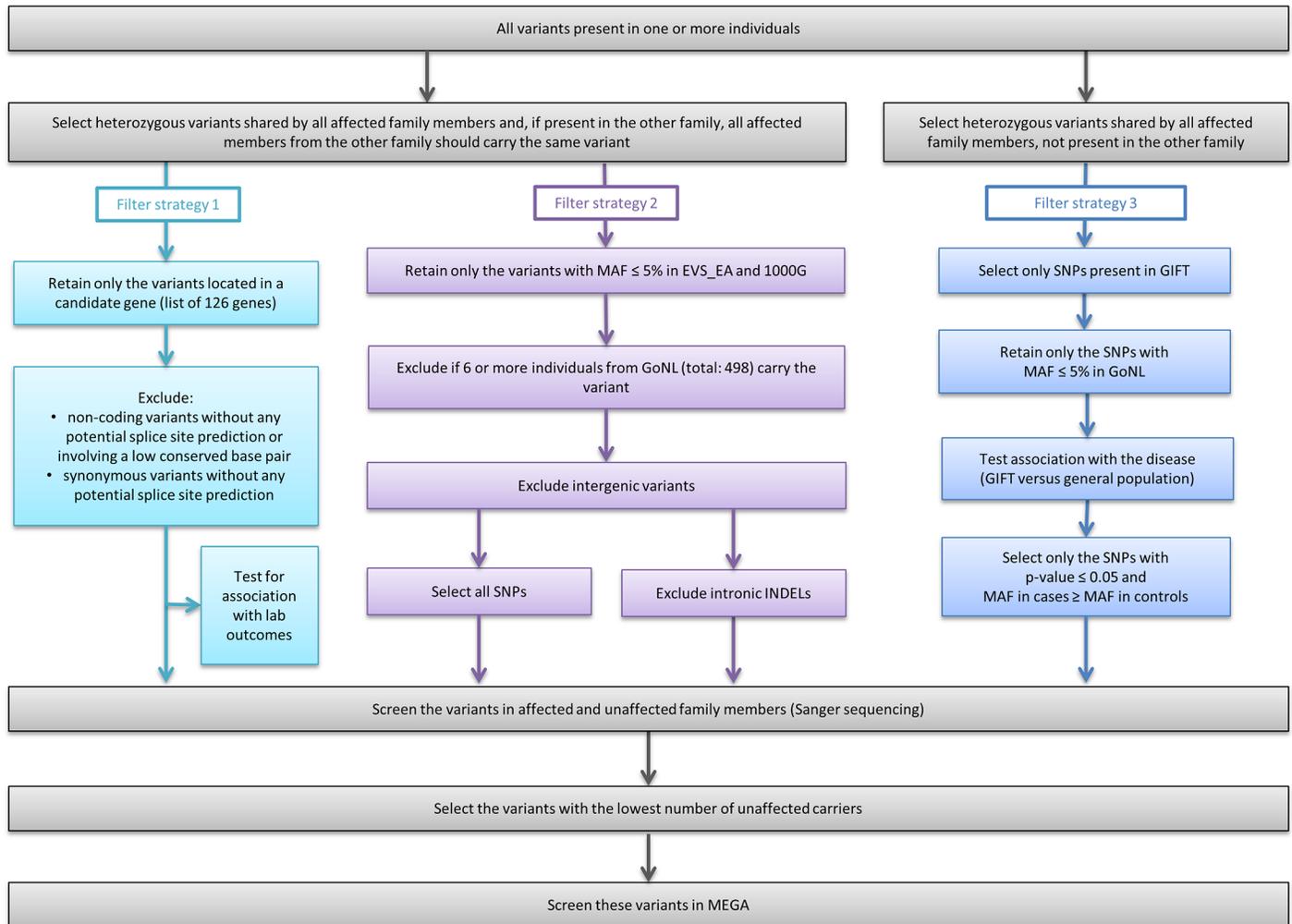


Fig 2. Overview of all filter strategies applied to the whole exome sequencing data derived from the DNA of 10 affected individuals from 2 Dutch families with inherited VTE (5 from each family). Three filter strategies were applied to the sequencing data. These strategies were based on: variants within candidate genes (filter strategy 1), rare variants across the exome (filter strategy 2) and, SNPs present in patients with familial VTE and rare in the general population, i.e., associated with VTE (filter strategy 3). Together, all these strategies might increase the chances of finding the genetic risk factor for VTE present in family D and family K. Abbreviations: SNPs, single nucleotide variants/polymorphisms; INDELS, insertions and deletions; MAF, minor allele frequency; GoNL, Genome of the Netherlands project database, 498 unrelated Dutch individuals; EVS_EA, NHLBI Exome Sequencing Project database, 4,300 European-American unrelated individuals; 1000G, 1000 Genomes Project, 2,500 individuals from about 25 populations around the world; GIFT, Genetics In Familial Thrombosis study, 96 unrelated VTE cases with positive family history of VTE; MEGA, Multiple Environmental and Genetic Assessment of risk factors for venous thrombosis study, up to 4,291 cases with VTE and 4,866 controls.

<https://doi.org/10.1371/journal.pone.0187699.g002>

For filter strategy 2, we identified all variants with minor allele frequency (MAF) equal or less than 5% in the NHLBI Exome Sequencing Project database, 4,300 European-American unrelated individuals (EVS_EA) [23] and 1000G database [24]. Then, we selected only variants identified in 5 or less ($\leq 1\%$) individuals from GoNL, as these individuals share the same ethnic origin and geographic region as families D and K. To minimize genotyping errors, all intergenic variants and INDELS located exclusively in intronic regions were excluded from further analysis.

For filter strategy 3, we identified SNPs that were present in one or more GIFT cases for whom whole exome sequencing data was available, and with MAF equal or below than 5% in GoNL. We tested these SNPs for association with VTE by comparing the risk allele frequencies

in GIFT ($n = 96$) with those present in GoNL ($n = 498$) or EVS_EA databases ($n = 4,300$). SNPs shared by members from both families were not considered in this filter strategy. Of note, SNPs available only in the GIFT database, i.e., only identified in cases with VTE, are retained for follow-up analysis in filter strategy 2.

After applying all the filter strategies, we evaluated each of the retained variants to minimize the presence of sequencing errors (e.g. those inherent to the technology used) [25,26]. Variants located in duplicated genome regions, in homopolymeric regions or with MAF above 10% in 1000G or EVS_EA databases (resulting from filter strategy 3) were excluded from further analysis. Furthermore, when 2 or more SNPs were located in the same gene, we used SNIpA [27] to evaluate whether these SNPs were in high linkage disequilibrium ($D' > 0.9$), and if so only one was taken for further analysis.

Finally, variants not present in dbSNP135 were classified as novel variants in this study. Yet, to facilitate the identification of all variants mentioned in this text, we have annotated these variants with a later version of dbSNP database, the dbSNP142.

Confirmation with Sanger sequencing or allele-specific PCR

DNA samples available from affected and unaffected members from both families were used for validation and segregation analysis of the candidate variants. This was performed by standard bidirectional Sanger sequencing. We ranked the variants in terms of number of asymptomatic carriers within a family, and we used samples from the GENES and the MEGA studies for follow-up validation of the top ranked variants (i.e., those with the lowest number of asymptomatic carriers). In total, 6 variants were selected for follow-up validation: 2 variants from family D and 4 variants from family K. These variants were validated with Sanger sequencing or allele-specific PCR on DNA available from 36 samples from the GENES study and 9,157 samples from the MEGA study, respectively (primers available on request).

Statistical analysis

Testing for association between a SNP and VTE. For each SNP resulting from filter strategy 3, the allele frequencies between cases (GIFT) and controls (GoNL or EVS_EA) were compared using the Pearson's chi-square test with 1 degree of freedom. Associations with $p < 0.05$ were defined as putative associations. After applying Bonferroni correction, associations with $p < 0.0003$ ($0.05/135$) were considered to be statistically significant. SNPs located in the X chromosome were excluded from this analysis, since sex information from the individuals enrolled in GoNL was not available. Instead, we decided to select for further analysis all coding SNPs or novel non-coding SNPs located in the X chromosome, which were absent in GoNL database. The association between each SNP genotyped in MEGA and VTE was investigated using Fisher's exact test. The threshold for significance was set as $p < 0.05$. The statistical analysis was performed using the software R (version 2.15.2).

Association of SNPs from candidate genes on the variability of hemostatic traits. We performed linear regression analyses adjusted for age and sex to estimate the effect of each SNP resulted from filter strategy 1 on hemostatic traits known to associate with VTE. Investigated hemostatic traits were endogenous thrombin generation, plasma antigen or activity levels of fibrinogen, coagulation factors II, VII, VIII, IX, X, and XII, von Willebrand factor, antithrombin, protein C, protein S (total and free), protein Z, C4b binding protein, and anti- β 2-glycoprotein I. The threshold for significance was set as $p < 0.05$ and when we corrected for multiple testing, $p < 0.0004$ was considered to be statistically significant. We repeated this analysis using some SNPs resulted from filter strategy 2 but solely for 3 hemostatic traits:

fibrinogen, protein S total and protein S free levels. The statistical analysis was performed using the software SPSS (version 22).

Sample size considerations

Using standard power calculation methods for case-control studies, the MEGA study would be sufficient to detect relative risks of 4 for a dominant trait with a minor allele frequency of 0.1%, with a power of 80% and 5% two-sided type I-error[28].

Results

Whole exome sequencing

Families. The sequencing yielded an average sequencing depth of ~53x per sample, with at least 86% of the targeted region covered more than 10x. Each individual had ~64,000 SNPs and ~12,000 INDELS, making in total 121,347 SNPs and 23,364 INDELS to undergo our filtering strategies (Fig 2).

GIFT study. Sequence data was successfully generated for all 96 subjects. In the combined 96 samples, the average read depth was 63X and the median was 62X. This value ranged from 38X to 80X according to individuals. In 91 samples, the average is above 50X, in the remaining 5 samples, the average read depth was around 40X. For all samples, the coverage was over 10x at greater than 88% of the targeted region. The median [range] number of SNPs and INDELS with read depth above 10x detected per sample was 87,461 [78,956–90,627] and 2,768 [2,249–2,997], respectively.

Overall, 13,286 genes carried at least one new or rare (<0.5%) coding variant (based on reported 1000g2012apr_eur frequency), and the average number of such variants per individual was 1,456.

Variants shared by all affected relatives

We identified all single nucleotide polymorphisms and insertions and deletions shared by all affected relatives from the families D and K. Tables 1 and 2 list the variants identified in autosomes and X chromosome, respectively.

Autosomal variants. We identified 276 SNPs and 22 INDELS exclusively shared by all affected individuals from family D, 733 SNPs and 40 INDELS exclusively shared by all affected individuals from family K, and 207 SNPs and 101 INDELS shared by all affected individuals from both families. All these variants were present in a heterozygous form. Only few SNPs were novel: 3 SNPs exclusive to family D, 36 SNPs exclusive to family K and 9 SNPs shared by both families.

Variants located in the X chromosome. We identified 2 and 20 SNPs, but no INDELS, shared by all affected individuals from family D and family K, respectively. Only 1 SNP was novel. The 10 individuals with these 22 SNPs shared no SNPs.

Selection of the candidate variants

Tables 1 and 2 display the number of variants retained after each filtering step. Concerning the variants exclusive to family D, 2 SNPs, 10 SNPs and 3 INDELS, and 5 SNPs remained after applying all the filtering steps from the filter strategies 1, 2, and 3. One SNP, *MAP3K6* rs55841735, was common between 2 of the 3 filter strategies. Concerning the variants exclusive to family K, 3 SNPs, 55 SNPs and 3 INDELS, and 10 SNPs remained after applying all the filtering steps from the filter strategies 1, 2, and 3. Four SNPs, *LRPPRC* rs372371276, *SRBD1*

Table 1. Number of variants located in autosomes retained after each filter step. Abbreviations: SNPs, single nucleotide variants/polymorphisms; INDELs, insertions and deletions; MAF, minor allele frequency; GoNL, Genome of the Netherlands project database, 498 unrelated Dutch individuals; EA, NHLBI Exome Sequencing Project database, 4,300 European-American unrelated individuals; 1000G, 1000 Genomes Project, 2,500 individuals from about 25 populations around the world; VTE, venous thromboembolism; GIFT, Genetics In Familial Thrombosis study, 96 unrelated VTE cases with positive family history of VTE.

Filter strategy	Filter step	SNPs			INDELs		
		Exclusive to family D	Exclusive to family K	Shared by both families	Exclusive to family D	Exclusive to family K	Shared by both families
1,2,3	1. Present in all affected individuals	995	1467	21029	80	89	2073
	2. Only heterozygous	276	733	207	22	40	101
1	3. Within candidate genes	4	7	0	0	0	0
	4. Retain only with functional effect	2	3	0	0	0	0
	Selected for Sanger sequencing	2	3	0	0	0	0
2	3. MAF_EA <= 5%	111	391	140	14	28	92
	4. MAF_1000G <= 5%	36	201	90	12	20	86
	5. Less than 6 carriers in GoNL (total of 498 unrelated controls)	10	56	10	12	20	83
	6. Exclude intergenic	10	54	2	10	20	69
	7. INDELs: Exclude intronic	10	54	2	3	3	4
	Selected for Sanger sequencing	10	48	0	2	2	1
3	3. Present in GIFT (total of 96 index cases)	265	663	-	-	-	-
	4. MAF_GoNL <= 5%	28	167	-	-	-	-
	5. Retain putative associations with VTE (p-value <= 0.05)	6	13	-	-	-	-
	6. Remove if MAF in controls > MAF in cases	5	10	-	-	-	-
	Selected for Sanger sequencing	3	6	-	-	-	-
1,2,3	Total variants for Sanger sequencing	14	53	0	2	2	1

<https://doi.org/10.1371/journal.pone.0187699.t001>

rs34959371, *FSCN3* rs34941808, and *ZNF816* rs369724877 were common between 2 of the 3 filter strategies. Regarding the variants shared by all 10 affected individuals, no variants remained for analysis, except for filter strategy 2. Using this filter strategy, 2 SNPs and 4 INDELs remained for analysis. Of note, 1 of the variants that resulted from filter strategy 2 was identified only among cases of VTE (family K and GIFT). This was an intronic SNP in the *NEBL* gene (rs61849814).

After all the variants have been individually checked, we excluded 12 SNPs and 5 INDELs from further validation with Sanger sequencing. Five of these variants were intronic SNPs from family K, which we considered not relevant for further analysis (details can be found in [S2 File](#)).

The 73 candidate variants retained for validation (68 SNPs and 5 INDELs) are listed in the [Table 3](#) along with the predicted impact on protein function, and MAF in various databases.

Validation of candidate variants

We designed primers to validate the 73 candidate variants resulting from all filter strategies.

Table 2. Number of variants located in the X chromosome retained after each filter step. Abbreviations: SNPs, single nucleotide variants/polymorphisms; INDELS, insertions and deletions; MAF, minor allele frequency; GoNL, Genome of the Netherlands project database, 498 unrelated Dutch individuals; EA, NHLBI Exome Sequencing Project database, 4,300 European-American unrelated individuals; 1000G, 1000 Genomes Project, 2,500 individuals from about 25 populations around the world.

Filter strategy	Filter step	SNPs			INDELS		
		Exclusive to family D	Exclusive to family K	Shared by both families	Exclusive to family D	Exclusive to family K	Shared by both families
1,2	1. Present in all affected individuals	8	30	339	0	0	45
	2. Only heterozygous females and homozygous males	2	20	0	0	0	0
1	3. Within candidate genes	0	0	0	0	0	0
	4. Retain only with functional effect	0	0	0	0	0	0
	Selected for Sanger sequencing	0	0	0	0	0	0
2	3. MAF_EA <= 5%	1	15	0	0	0	0
	4. MAF_1000G <= 5%	0	6	0	0	0	0
	5. Less than 6 carriers in GoNL	0	1	0	0	0	0
	6. Exclude intergenic	0	1	0	0	0	0
	7. INDELS: Exclude intronic	0	1	0	0	0	0
	Selected for Sanger sequencing	0	1	0	0	0	0
1,2	Total variants for Sanger sequencing	0	1	0	0	0	0

<https://doi.org/10.1371/journal.pone.0187699.t002>

The presence of 5 variants, 4 SNPs (*IFT57* rs199895727, *IRF3* NM_001197122.1:c.1115-8C>T, *MUC4* rs199718845 and *TRIM41* rs376725370) and 1 INDEL (*TRAK1* NM_001265609:c.1844_1845insGGA) could not be validated due to PCR problems. The presence of the remaining variants was evaluated by Sanger sequencing in affected and unaffected family members (including 1 additional affected member from family D for whom no whole exome sequencing data were available) (S3 File). We identified the variants with the lowest number of unaffected carriers based on the families' genotyping results. We selected 6 variants. These variants were located in 3 different loci: 2p21 (*PLEKHH2*, *LRPPRC* and *SRBD1*) and 5q35.2 (*UNC5A*) based on family K, and 17q25.1 (*GPRC5C* and *RAB37*) based on family D. Three of these variants, all from the 2p21 locus, were also present in the GIFT study. One individual carried the 3 variants and one individual carried the *SRBD1* variant only.

Variants genotyped in individuals from GENES and MEGA

We examined the prevalence of the 6 rare candidate SNPs, *PLEKHH2* NM_172069: c.3105T>C, *LRPPRC* rs372371276, *SRBD1* rs34959371, *UNC5A* NM_133369.2:c.1869+23C>A, *GPRC5C* rs142232982 and, *RAB37* rs556450784, in 36 affected individuals from GENES and, 4,291 affected and 4,866 unaffected individuals from MEGA. We found additional carriers in MEGA, except for the *UNC5A* SNP. We initially found 2 carriers of the *PLEKHH2* variant (1 case and 1 control, and surprisingly, the case was homozygous for the minor allele) but after validation with Sanger sequencing, the case was found to be wild type. We also found 2 carriers of the *LRPPRC* variant (1 case and 1 control), 38 carriers of the *SRBD1* variant (14 cases and 24

Table 3. List of the 73 variants selected for validation. Abbreviations: Chr, Chromosome; Ref, reference allele; Obs, observed allele; HOM, homozygous; HET, heterozygous; Func, variant function; SNPs, single nucleotide variants/polymorphisms; MAF, minor allele frequency; GoNL, Genome of the Netherlands project database, 498 unrelated Dutch individuals; EVS_EA, NHLBI Exome Sequencing Project database, 4,300 European-American unrelated individuals; 1000G, 1000 Genomes Project, 2,500 individuals from about 25 populations around the world; GIFT, Genetics In Familial Thrombosis study, 96 unrelated VTE cases with positive family history of VTE; GoNL, Genome of the Netherlands project database, 498 unrelated Dutch individuals.

Chr	Start	End	Ref	Obs	Func	Gene	Exonic Func	dbSNP135	MAF EA_6500_%	MAF GIFT_% (assuming_allCalled)	MAF 1000G 2012apr_ALL_%	GoNL number of genotypes (HOM_Ref/HET/HOM_Obs)	Filter strategy
1	1905479	1905479	C	T	exonic	KIAA1751	nonsynonymous SNP		NA	0.00	0	NA	2
1	6211042	6211042	C	G	intronic	CHD5			0.05	0.00	0	NA	2
1	27688743	27688743	T	C	splicing	MAP3K6		rs55841735	0.41	1.56	0.41	494/4/0	2.3
1	117122285	117122285	-	TCC	exonic	IGSF3	nonframeshift insertion		NA	21.77	0	NA	2
1	156354348	156354348	C	-	splicing	RHBG		rs11303415	NA	0.00	0	NA	2
1	231377078	231377078	C	A	UTR5	GNPAT			0.01	0.00	0	493/1/0	2
1	234536943	234536943	T	G	exonic	TARBP1	nonsynonymous SNP		NA	0.00	0	NA	2
2	43965641	43965641	T	C	exonic	PLEKHH2	synonymous SNP		NA	0.52	0	NA	2
2	44184604	44184604	T	C	intronic	LRRPFC			0.02	0.52	0	NA	2.3
2	45640302	45640302	G	C	exonic	SRBD1	nonsynonymous SNP	rs34959371	0.43	1.04	0.27	497/1/0	2.3
2	207175476	207175476	G	T	exonic	ZDBF2	nonsynonymous SNP	rs36095066	0.87	0.00	0.46	493/5/0	2
3	42251580	42251580	-	GGA	exonic	TRAK1	nonframeshift insertion		NA	0.00	0	NA	2
3	107932728	107932728	G	A	intronic	IFT57			0.40	0.52	0.09	496/2/0	2
3	111667982	111667982	C	T	intronic	PHLDB2		rs116705395	NA	0.00	0.18	497/1/0	2
3	121097497	121097497	C	T	intronic	STXBP5L		rs182808568	NA	0.00	0.14	494/4/0	2
3	121260373	121260373	G	A	intronic	POLQ		rs183249693	0.08	0.00	0.09	494/4/0	2
3	124829057	124829057	A	G	exonic	SLC12A8	synonymous SNP	rs11714448	3.09	6.25	1	470/28/0	3
3	129308276	129308276	G	A	exonic	PLXND1	nonsynonymous SNP	rs146072632	0.21	0.00	0	481/1/0	2
3	195510838	195510838	G	A	exonic	MUC4	nonsynonymous SNP		NA	0.00	0	NA	2
3	195591020	195591020	T	G	UTR3	TNK2		rs112362099	1.00	2.60	1	480/18/0	3
4	144346515	144346515	C	G	intronic	GAB1		rs187533266	NA	0.00	0.23	497/1/0	2
5	176305151	176305151	C	A	intronic	UNC5A			0.01	0.00	0	NA	2
5	180661310	180661310	G	T	exonic	TRIM41	synonymous SNP		0.03	0.00	0	NA	2
7	100016781	100016781	T	C	exonic	ZCWPW1	nonsynonymous SNP	rs141450215	0.50	0.00	0.27	496/2/0	2
7	103061356	103061356	T	A	intronic	SLC26A5			NA	0.00	0	497/1/0	2
7	110763503	110763503	A	G	exonic	LRRN3	synonymous SNP	rs35229264	0.47	0.00	0.09	495/3/0	2
7	111936315	111936315	C	T	exonic	ZNF277	synonymous SNP	rs142208106	0.00	0.00	0	496/2/0	2

(Continued)

Table 3. (Continued)

Chr	Start	End	Ref	Obs	Func	Gene	Exonic Func	dbSNP135	MAF EA_6500_%	MAF GIFT_% (assuming_allCalled)	MAF 1000G 2012apr_ALL_%	GoNL number of genotypes (HOM_Ref/HET/HOM_Obs)	Filter strategy
7	127229265	127229265	C	T	intronic	ARF5			0.56	1.56	0.23	487/4/0	2
7	127239585	127239585	C	T	exonic	FSCN3	nonsynonymous SNP	rs34941808	0.60	1.56	0.14	495/3/0	2.3
9	2039776	2039776	-	CAG	exonic	SMARCA2	nonframeshift insertion		NA	11.54	0	NA	2
9	2648286	2648286	G	A	exonic	VLDLR	nonsynonymous SNP	rs35339834	0.13	0.00	0.05	497/1/0	2
9	5919929	5919929	T	C	exonic	KIAA2026	nonsynonymous SNP		0.07	0.00	0	NA	2
9	8486132	8486132	G	C	exonic	PTPRD	synonymous SNP	rs14411555	0.10	0.00	0.27	496/2/0	2
9	136917482	136917482	G	A	exonic	BRD3	synonymous SNP	rs61731642	0.00	0.00	0.09	NA	2
9	140145895	140145895	C	G	ncRNA_intronic	LOC100129722			0.17	0.52	0	NA	2
10	21318672	21318672	A	C	intronic	NEBL		rs61849814	NA	1.04	0	NA	2
11	62601915	62601915	G	A	exonic	WDR74	nonsynonymous SNP		NA	0.00	0	NA	2
11	63883847	63883847	C	T	exonic	FLRT1	synonymous SNP		NA	0.00	0	NA	2
11	65293664	65293664	G	A	exonic	SCYL1	nonsynonymous SNP		0.01	0.00	0	NA	2
11	66472221	66472221	G	A	exonic	SPTBN2	synonymous SNP	rs144939155	0.25	0.52	0.18	479/3/0	2
11	66638474	66638474	C	A	intronic	PC			NA	0.00	0	NA	2
11	68177405	68177405	C	T	exonic	LRP5	synonymous SNP	rs145456776	0.03	0.00	0	NA	2
12	130919182	130919182	G	A	intronic	RIMBP2			NA	0.00	0	NA	2
12	131311749	131311749	A	C	exonic	STX2	nonsynonymous SNP	rs137928907	2.55	2.08	1	463/32/1	1
12	133306808	133306808	C	T	exonic	ANKLE2	nonsynonymous SNP	rs149597645	1.63	3.65	1	477/21/0	3
13	113824899	113824899	T	C	intronic	PROZ		rs3024787	NA	0.00	0	493/5/0	2
14	105360291	105360291	C	G	intronic	KIAA0284		rs185909273	NA	0.00	0.18	496/2/0	2
14	105418571	105418571	C	T	exonic	AHNAK2	nonsynonymous SNP		0.00	0.00	0	NA	2
16	602505	602505	C	T	exonic	SOLH	synonymous SNP		0.00	0.00	0	NA	2
16	4940366	4940366	A	G	intronic	PPL			0.14	0.52	0.18	497/1/0	2
16	5083391	5083391	C	G	exonic	NAGPA	nonsynonymous SNP		NA	0.00	0	NA	2
17	45360730	45360730	T	C	exonic	ITGB3	nonsynonymous SNP	rs5918	15.17	14.58	9	341/145/12	1
17	64210757	64210757	C	A	exonic	APOH	nonsynonymous SNP	rs4581	23.76	21.88	47	311/166/21	1
17	65713943	65713943	-	A	upstream	NOL11			NA	0.00	0	NA	2

(Continued)

Table 3. (Continued)

Chr	Start	End	Ref	Obs	Func	Gene	Exonic Func	dbSNP135	MAF EA_6500_%	MAF GIFT_% (assuming_allCalled)	MAF 1000G 2012apr_ALL_%	GoNL number of genotypes (HOM_Ref/HET/HOM_Obs)	Filter strategy
17	66274384	66274384	C	T	exonic	SLC16A6	synonymous SNP	rs149995088	0.29	0.52	0.09	494/4/0	2
17	67310584	67310584	A	G	UTR5	ABCA5			NA	0.00	0	NA	2
17	72443156	72443156	G	A	exonic	GPRC5C	nonsynonymous SNP	rs142232982	0.05	0.00	0.09	NA	2
17	72740355	72740355	G	A	intronic	RAB37			NA	0.00	0	NA	2
18	43496563	43496563	G	A	intronic	EPG5			0.12	0.00	0	497/1/0	2
19	15569378	15569378	G	C	exonic	RASAL3	nonsynonymous SNP	rs58123634	4.84	6.77	3	457/25/0	3
19	47207685	47207685	G	C	intronic	PRKD2			0.08	0.00	0	494/1/0	2
19	49254538	49254538	T	G	exonic	FUT1	nonsynonymous SNP		NA	0.00	0	NA	2
19	49713534	49713534	G	A	exonic	TRPM4	nonsynonymous SNP		NA	0.00	0	NA	2
19	49878275	49878275	G	A	exonic	DKK1	nonsynonymous SNP	rs35389403	0.33	0.00	1	NA	2
19	49910139	49910139	C	G	intronic	CCDC155		rs112074780	0.39	0.00	1	NA	2
19	50163098	50163098	G	A	intronic	IRF3			NA	0.00	0	NA	2
19	51503285	51503285	C	T	exonic	KLK8	nonsynonymous SNP	rs16988799	5.15	2.60	5	441/56/1	1
19	51528041	51528041	C	T	exonic	KLK11	nonsynonymous SNP	rs9745539	7.53	7.29	6	421/7/5	1
19	53454862	53454862	T	C	intronic	ZNF816			0.08	0.52	0	497/1/0	2.3
20	744270	744270	G	A	exonic	SLC52A3	synonymous SNP	rs139430185	0.05	0.00	0	497/1/0	2
22	38890624	38890624	A	C	intronic	DDX17			NA	0.00	0	497/1/0	2
22	41736075	41736075	C	A	exonic	ZC3H7B	synonymous SNP		NA	0.00	0	494/2/0	2
X	117702143	117702143	C	G	intronic	DOCK11			0.07		0	NA	2

<https://doi.org/10.1371/journal.pone.0187699.t003>

Table 4. Candidate variants genotyped in MEGA.

Gene	Variant	Reference/ alternate alleles	Number of MEGA cases			Number of MEGA controls			P- value	OR	95% CI
			Ref/ Ref	Ref/ Alt	Alt/ Alt	Ref/ Ref	Ref/ Alt	Alt/ Alt			
<i>PLEKHH2</i>	c.3105T>C	T/C	4031	0	0	4691	1	0	NA	NA	NA
<i>LRPPRC</i>	rs372371276	A/G	4044	1	0	4662	1	0	1	1.15	0.01– 90.44
<i>SRBD1</i>	rs34959371	G/C	3829	14	0	4400	24	0	0.26	0.67	0.32– 1.35
<i>UNC5A</i>	c.1869 +23C>A	C/A	3981	0	0	4613	0	0	NA	NA	NA
<i>GPRC5C</i>	rs142232982	G/A	4021	7	0	4637	12	0	0.49	0.67	0.22– 1.86
<i>RAB37</i>	rs556450784	G/A	4036	6	0	4598	13	0	0.25	0.53	0.16– 1.48

<https://doi.org/10.1371/journal.pone.0187699.t004>

controls), 19 carriers of the *GPRC5C* variant (7 cases and 12 controls), and 19 carriers of the *RAB37* variant (6 cases and 13 controls). Some individuals carried 2 of these disease candidate variants simultaneously: 1 control carried both *LRPPRC* and *SRBD1* variants, and 4 controls carried both *GPRC5C* and *RAB37* variants. We compared the distribution of the disease risk alleles between affected and unaffected individuals and found no statistically significant differences or clear signals of an association (Table 4).

Association of SNPs from candidate genes with variability of hemostatic traits

We investigated whether any of the 5 candidate SNPs resulting from filter strategy 1 (*STX2* rs137928907, *ITGB3* rs5918, *APOH* rs4581, *KLK8* rs16988799, and *KLK11* rs3745539) were associated with one or more hemostatic traits known to associate with VTE. We investigated 413 individuals from GIFT, for whom both genotype and hemostatic traits data were available. No genotype was available or could be imputed for the *STX2* rs137928907. We found associations for the variants located in chromosome 17, *APOH* rs4581 and *ITGB3* rs5918. After correcting for multiple testing only one association remained significant. The *ITGB3* rs5918 remained associated with FX levels. All putative associations are shown in Table 5.

We also investigated whether the 3 candidate variants that were identified from filter strategy 2, *RAB37* rs556450784, *GPRC5C* rs142232982 and *SRBD1* rs34959371, were associated with fibrinogen, total protein S and free protein S levels in MEGA (S4 File).

Discussion

Our approach enabled us to narrow down the list of genetic variants to 3 candidate loci, but we were unable to identify the causative variant underlying the risk for disease in 2 families with unexplained familial VTE. Either variants in known hemostatic genes as well as rare variants across the exome are unlikely to explain the VTE tendency in these families.

A major strength of this study was the availability of families with multiple affected individuals that included cases with an early age of onset. Although family members are more likely to share environmental risk factors than unrelated individuals, young individuals without acquired causes for VTE are more likely to carry variants with large effect sizes [29]. Assuming that affected individuals from a family would all share the same susceptibility allele, we could reduce the list of candidate SNPs from 121,347 to 1,238, and candidate INDELS from 23,364 to

Table 5. Results of SNPs with P-values less than 0.05 for association with a haemostatic trait (samples from GIFT study). Putative associations were found for 2 of the candidate SNPs resulting from filter strategy 1. No genotype data was available for the *STX2* rs137928907. Abbreviations: NA, not available; -, no association; Ref, reference allele; Alt, alternate allele; N, number; CI, confidence interval.

Gene variant	Lab outcome	N individuals by genotype			Mean <lab outcome> by genotype			Multiple regression analysis				Single regression analysis
		Ref/Ref	Ref/Alt	Alt/Alt	Ref/Ref	Ref/Alt	Alt/Alt	Variance explained (%)	Per-Alt allele effect size ± Std. Error	P-value	95% CI	P-value
<i>STX2</i> rs137928907	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
<i>ITGB3</i> rs5918	Factor X levels	197	70	4	99 ± u/dL	107 ± 23 u/dL	114 ± 3 u/dL	6.7	9 ± 2 u/dL	3 × 10 ⁻⁴	3.91–13.19	0.001
	Factor IX levels	197	70	4	101 ± u/dL	106 ± u/dL	106 ± u/dL	5.6	5 ± 2 u/dL	0.013	1.14–9.68	0.029
	Protein S levels	197	70	4	93 ± u/dL	98 ± u/dL	101 ± u/dL	10.5	6 ± 2 u/dL	0.002	2.35–10.57	0.013
	Protein S free levels	197	70	4	88 ± u/dL	97 ± u/dL	88 ± u/dL	11.4	13 ± 3 u/dL	0.001	6.25–15.63	0.006
	Thrombin generation lagtime (Tissue factor = 1pM)	197	70	4	8.7 ± 1 min	9.7 ± 0 min	8.4 ± 4 min	8.2	0.9 ± 0.4 min	0.029	0.09–1.71	0.062
<i>APOH</i> rs4581	Protein S free levels	182	78	11	88 ± u/dL	91 ± u/dL	106 ± u/dL	8.3	3 ± 3 u/dL	0.108	0.00–9.38	0.035
	Factor V ag	274	117	19	112 ± u/dL	104 ± u/dL	98 ± u/dL	2.6	-7 ± 3 u/dL	0.027	-12.49–0.77	0.02
	Factor V act	274	116	18	103 ± u/dL	100 ± u/dL	95 ± u/dL	2.3	-3 ± 2 u/dL	0.037	-0.66–0.21	0.032
	Thrombin generation lagtime (Tissue factor = 1pM)	182	78	11	8.7 ± 0 min	9.6 ± 0 min	10.0 ± 6 min	8	0.7 ± 0.4 min	0.04	0.03–1.44	0.032
	Thrombin generation lagtime (Tissue factor = 10pM)	182	78	11	1.9 ± 0.3 min	2.0 ± 0.4 min	2.0 ± 0.2 min	4.6	0.1 ± 0.0 min	0.018	0.02–0.17	0.014
<i>KLK8</i> rs16988799	-	-	-	-	-	-	-	-	-	-	-	-
<i>KLK11</i> rs3745539	-	-	-	-	-	-	-	-	-	-	-	-

<https://doi.org/10.1371/journal.pone.0187699.t005>

163 before applying more stringent filter strategies. It is worth mentioning that an analysis-by-exclusion approach with exomes from unaffected relatives could have reduced the number of shared variants to a minimum. We chose to only sequence exomes from affected relatives because an analysis-by-exclusion also has its limitations in VTE family studies, as we could have missed disease risk variants because unaffected individuals from the same family can still develop the disease at older age.

Our choice to pursue a high-throughput exome sequencing approach enabled us to identify both rare and common coding variants. Moreover, this approach allowed us to focus not only on variants located in genes previously reported to be associated with VTE, but also in novel genes.

Many patients and families included in gene identification studies still remain without a molecular diagnosis [30]. One of the main reasons is because ranking variants is often a

challenge in these studies [31]. We reported our results from applying three different filter strategies, as reporting the results from different approaches is essential to improve variant ranking strategies. One common approach consists of prioritizing non-synonymous variants located in or near genes known to be associated with the disease of interest or a related phenotype. These variants are much more likely to be involved in the disease than variants elsewhere in the genome. This is the foundation of our filter strategy 1. Another approach consists of prioritizing rare and low frequency variants. Rare and low frequency variants are often at the basis of inherited risk for VTE, like in the case of deficiencies in one of the natural anticoagulants and the factor V Leiden and prothrombin mutations [30]. The use of rare and low frequency variants is the foundation of the remaining two filter strategies. Compared with the filter strategy 2, the filter strategy 3 has a more relaxed MAF threshold since we cannot exclude that young controls included in our study might develop VTE associated with inherited thrombophilia at a later time. Data from a set of cases with family history of VTE (GIFT) was key to rank these variants although we cannot ignore the possibility of generating false negative results due to the small sample size (only 96 unrelated cases). Of note, since there is growing evidence that supports a functional role for synonymous rare codons [32,33], we retained synonymous variants for analysis in both filter strategies 2 and 3.

Several genome-wide studies, and more recently also whole exome sequencing studies have been conducted to identify genetic risk factors for VTE [9–14,34–36]. However, nearly all variants that have been robustly found to be associated with the risk for this disease are located within known susceptibility genes. In our study, the candidate variants resulting from filter strategy 1 are the least likely disease-causing variants, as there were a high number of unaffected carriers, including homozygotes. Furthermore, because the affected individuals from these families do not share abnormalities in hemostatic traits known to affect the risk of VTE, variants with a large effect size are unlikely to be present among the known VTE susceptibility genes.

All variants tested for segregation with the disease showed incomplete penetrance, i.e. unaffected individuals carrying the putative pathogenic variant were found among the families. Yet, an autosomal dominant incomplete penetrance is likely. Despite the inheritance pattern being consistent with an autosomal dominant mode for both families, family K oldest generation (proband's parents) did not experience VTE. Furthermore, given the different clinical manifestations of VTE within the affected individuals, it is plausible that other factors, such as environmental or other genetic modifiers, are also involved in the pathogenesis of the disease in these families. Although some studies suggest that the proportion of the variance attributable to shared familial environment factors is small [37,38], it is difficult to predict the likelihood of finding VTE-causing variants based solely on the severity of the disease within a single family.

With regard to the 6 variants genotyped in MEGA, we found additional carriers among cases and controls, except for the *UNC5A* variant, for which no carrier was found. The 2 putative pathogenic variants identified for family D, the *GPRC5C* and *RAB37* SNPs, were found to be more prevalent in the Netherlands than in the worldwide population. This was unexpected because none of these variants was present in GoNL database. An explanation could be that these genomic regions were not covered in GoNL database. With regard to the 4 putative pathogenic variants identified in family K, the rarity of the variants identified in *UNC5A*, *LRPPRC* and *PLEKHH2* genes suggests that these variants are deleterious. However, whether these variants predispose to VTE remains unknown. The *UNC5A* variant seems to be a family-specific variant, as we have not identified additional carriers nor it is reported in any of the available database. The *PLEKHH2* variant was found in 1 unaffected individual while the *LRPPRC* variant was found 1 affected and 1 unaffected individual, both with negative family history for

VTE. Noteworthy that despite being so rare, both *LRPPRC* and *PLEKHH2* variants were found in 1 affected individual with positive family history for VTE in GIFT. Co-segregation with the phenotype in other affected family members may shed further light on the causality of this locus. This is particularly important because these variants are predicted to have no functional consequence and biological evidence might be difficult to address.

According to the GWAS catalog database, variants in *SRBD1*, *LRPPRC* and *RAB37* genes have been reported to be associated with other phenotypes: *SRBD1* rs3213787 has been associated with normal tension glaucoma[39], *LRPPRC* rs13387221 has been associated with cognitive ability (intelligence) in childhood[40] and *RAB37* rs10512597 has been associated with fibrinogen levels[41,42]. This last association is intriguing because elevated total fibrinogen levels and reduced fibrinogen gamma' levels are associated with increased risk for VTE[43]. Common SNPs at the locus 17q25.1, which includes the *RAB37* rs10512597-C, have been associated with reduced fibrinogen and C-reactive protein levels[41,42], which suggests that regulators of fibrinogen and C-reactive protein levels might be present at this locus. However, in the GIFT study, the *RAB37* rs10512597-C was not associated with fibrinogen levels ($\beta = 0.07$, 95% CI = -0.05–0.20, $p = 0.251$) but instead, it was associated with protein S levels ($\beta = 4.37$, 95% CI = 1.00–7.74, $p = 0.011$) but not free protein S levels ($\beta = 0.010$, 95% CI = -0.006–0.026, $p = 0.236$) (MLRC, PHR, unpublished observation August 2015). Both candidate variants from family D, *RAB37* rs556450784 and *GPRC5C* rs142232982, were not associated with fibrinogen or protein S levels in MEGA.

Variants located within the nearest upstream or the nearest downstream genes of the following candidate genes, *PLEKHH2*, *LRPPRC*, and *SRBD2*, all located in the 2p21 locus, have also been reported as associated with various phenotypes according to the GWAS catalog database. Variants in *THADA* have been associated with age-related hearing impairment[44], Crohn's disease[45], DNA methylation variation[46], hair morphology[47], inflammatory bowel disease[48], mitochondrial DNA levels[49], orofacial clefts[50,51], platelet counts[52], polycystic ovary syndrome[53,54], prostate cancer[55], response to amphetamines[56], and type 2 diabetes[57]. Variants in *ABCG8* have been associated with LDL cholesterol[58–63], total cholesterol[58–60], campesterol levels[64], and gallstones[65]. Finally, variants in *PRKCE* have been associated with various red blood cell traits (red blood cell counts[66], hematocrit [66–69], and hemoglobin[66,67]), as well as pulmonary function decline[70], QT interval[71], metabolite levels (X-11787)[72], and suicide risk[72]. While some of these phenotypes might be linked with the risk of developing VTE, the association of the 2p21 locus with VTE risk has not been reported before.

Our study has some limitations. First, although different types of genetic variation exist, we have investigated only SNPs and INDELS. Second, we had to employ different filtering strategies to select only potentially clinically relevant variants, as the number of variants shared by the affected family members was very high. We cannot exclude the possibility of missing the causative variant after applying these strategies. Nevertheless, because there are no established filtering strategies for identification of VTE causing genes, we have validated with Sanger sequencing additional variants that were not retained after applying the different filter strategies (details in [S2 File](#)). Yet, none of these additional variants is likely to explain the increased risk for VTE in families D and K. As a side note, researchers can reutilize these data and apply filter strategies other than the ones described in our manuscript to identify putative disease-causing variants (the information about all variants shared by the relatives can be found in [S2 File](#)). For example, researchers can prioritize the variants based on other candidate genes [73] or on scores that estimate the variant effect [74] and compare these with their relevant datasets. Third, given that all 10 samples from the 2 families have undergone the same experimental protocol and bioinformatics analysis, systematic errors are likely

to be present in this dataset. Some of these errors might lead to erroneous variant calls. To minimize this type of errors, for each variant identified in our analysis, we have obtained the genotype information for all samples. This allowed us to distinguish samples with homozygous reference calls from samples with no genotype calls. We used this information and selected only SNPs for which the genotype information was available in all 10 individuals. Therefore, although discerning rare variants from sequencing errors remains a big challenge in next generation sequencing data analysis, we think that rare variants exclusively shared by one family are less likely to be sequencing errors in our study. Yet, we cannot exclude the possibility of incomplete coverage of some genes, a limitation inherent to the next generation sequencing technique. Fourth, although all variants analyzed are based on families D and K results, the GIFT, EVS_EA and GoNL data were generated with different library preparation, sequencing and bioinformatics methodology. Cross-dataset comparisons (comparing GIFT cases with EVS_EA or GoNL) are in general prone to technical bias. Fifth, our sample size in MEGA was powered to detect genotype relative risks greater than 4 at the significance level of 0.05 for risk allele frequencies greater than 0.10%, and the power to detect a relative risk of 1.5 was low. Sixth, although our results argue that coding variants in known genes or rare coding variants across the exome are unlikely to explain the VTE tendency in these families, these results are not generalizable to other families with unexplained VTE. Finally, our whole exome sequencing data was generated by the end of 2011 and a significant improvement on sequencing technology and data analysis has been made over the last years. The feasibility of whole exome sequencing for identifying new VTE genetic variants is only as good as the quality of the data, so it could be that better coverage might lead to identification of more potential variants of interest.

In conclusion, despite extensive investigation, we did not find a definitive genetic cause for the increased risk of VTE in the 2 evaluated families. Our study suggests that rare variants within 3 candidate loci, 2p21, 5q35.2 and 17q25.1, have an impact on the risk of VTE in these families, but we could not find any evidence for the 6 rare variants genotyped in a large case-control association study.

Supporting information

S1 File. List of 126 genes. An excel workbook with information concerning the genes names and genomic location.

(XLSX)

S2 File. Variants shared by all affected members. An excel workbook that includes the following information: all variants retained for manual inspection after applying the 3 filter strategies, all variants shared by family members with Filter = PASS, and (iii) all variants shared by family members with Filter = NO PASS. All variants validated with Sanger sequencing are discernible. Please note that some of these variants did not follow our filter strategies described in the manuscript.

(XLSX)

S3 File. Results of Sanger sequencing in affected and unaffected family members for all candidate variants.

(XLSX)

S4 File. Association analysis of fibrinogen, total and free protein S levels in healthy individuals from MEGA.

(DOCX)

Acknowledgments

We thank C. Koch and K. Los for collection of blood samples of family members. We thank J. Peter, A. Hoenderdos, P. Noordijk, and L. Mahic for technical support. We thank D. Trégouët, and M. Germain for performing the whole exome sequencing analysis in GIFT. We thank individuals and families for their participation in this study.

Author Contributions

Conceptualization: Marisa L. R. Cunha, Joost C. M. Meijers, Saskia Middeldorp.

Data curation: Marisa L. R. Cunha.

Formal analysis: Marisa L. R. Cunha, Astrid van Hylckama Vlieg.

Funding acquisition: Frits R. Rosendaal, Saskia Middeldorp.

Investigation: Marisa L. R. Cunha, Pieter H. Reitsma.

Methodology: Marisa L. R. Cunha, Joost C. M. Meijers, Saskia Middeldorp.

Project administration: Marisa L. R. Cunha.

Resources: Joost C. M. Meijers, Frits R. Rosendaal, Astrid van Hylckama Vlieg, Pieter H. Reitsma, Saskia Middeldorp.

Software: Marisa L. R. Cunha.

Supervision: Joost C. M. Meijers, Saskia Middeldorp.

Validation: Marisa L. R. Cunha.

Visualization: Marisa L. R. Cunha.

Writing – original draft: Marisa L. R. Cunha.

Writing – review & editing: Marisa L. R. Cunha, Joost C. M. Meijers, Frits R. Rosendaal, Astrid van Hylckama Vlieg, Pieter H. Reitsma, Saskia Middeldorp.

References

1. Konstantinides S V, McCumber M, Ozaki Y, Wendelboe a, Weitz JI. Thrombosis: A Major Contributor to Global Disease Burden Raskob G.E., Angchaisuksiri P., Blanco A.N., Buller H., Gallus A., Hunt B.J., Hylek E.M., Kakkar A., Konstantinides S.V., McCumber M., Ozaki Y., Wendelboe A. and Weitz J.I. *Arterioscler Thromb Vasc*. 2014; 2363–2372.
2. Prandoni P, Barbar S, Milan M, Vedovetto V, Pesavento R. The risk of recurrent thromboembolic disorders in patients with unprovoked venous thromboembolism: New scenarios and opportunities. *Eur J Intern Med*. 2014; 25: 25–30. <https://doi.org/10.1016/j.ejim.2013.09.005> PMID: 24120221
3. Bezemer ID, van der Meer FJM, Eikenboom JCJ, Rosendaal FR, Doggen CJM. The value of family history as a risk indicator for venous thrombosis. *Arch Intern Med*. 2009; 169: 610–615. <https://doi.org/10.1001/archinternmed.2008.589> PMID: 19307525
4. Reitsma PH, Versteeg HH, Middeldorp S. Mechanistic view of risk factors for venous thromboembolism. *Arterioscler Thromb Vasc Biol*. 2012; 32: 563–568. <https://doi.org/10.1161/ATVBAHA.111.242818> PMID: 22345594
5. Zöller B, Li X, Sundquist J, Sundquist K. Familial transmission of venous thromboembolism a cohort study of 80 214 swedish adoptees linked to their biological and adoptive parents. *Circ Cardiovasc Genet*. 2014; 7: 296–303. <https://doi.org/10.1161/CIRCGENETICS.113.000341> PMID: 24795348
6. Bertina RM, Koeleman BP, Koster T, Rosendaal FR, Dirven RJ, de Ronde H, et al. Mutation in blood coagulation factor V associated with resistance to activated protein C. *Nature*. 1994; 369: 64–7. <https://doi.org/10.1038/369064a0> PMID: 8164741

7. Poort SR, Rosendaal FR, Reitsma PH, Bertina RM. A common genetic variation in the 3'-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. *Blood*. 1996; 88: 3698–3703. PMID: [8916933](https://pubmed.ncbi.nlm.nih.gov/8916933/)
8. S NL, H LA, H SR, et al. ASsociation of genetic variations with nonfatal venous thrombosis in postmenopausal women. *JAMA*. 2007; 297: 489–498. <https://doi.org/10.1001/jama.297.5.489> PMID: [17284699](https://pubmed.ncbi.nlm.nih.gov/17284699/)
9. Trégouët DA, Heath S, Saut N, Biron-Andreani C, Schved JF, Pernod G, et al. Common susceptibility alleles are unlikely to contribute as strongly as the FV and ABO loci to VTE risk: Results from aGWAS approach. *Blood*. 2009; 113: 5298–5303. <https://doi.org/10.1182/blood-2008-11-190389> PMID: [19278955](https://pubmed.ncbi.nlm.nih.gov/19278955/)
10. Germain M, Saut N, Greliche N, Dina C, Lambert JC, Perret C, et al. Genetics of Venous thrombosis: Insights from a new genome wide association study. *PLoS One*. 2011; 6. <https://doi.org/10.1371/journal.pone.0025581> PMID: [21980494](https://pubmed.ncbi.nlm.nih.gov/21980494/)
11. Heit J a., Armasu SM, Asmann YW, Cunningham JM, Matsumoto ME, Petterson TM, et al. A genome-wide association study of venous thromboembolism identifies risk variants in chromosomes 1q24.2 and 9q. *J Thromb Haemost*. 2012; 10: 1521–1531. <https://doi.org/10.1111/j.1538-7836.2012.04810.x> PMID: [22672568](https://pubmed.ncbi.nlm.nih.gov/22672568/)
12. Greliche N, Germain M, Lambert J-C, Cohen W, Bertrand M, Dupuis A-M, et al. A genome-wide search for common SNP x SNP interactions on the risk of venous thrombosis. *BMC Med Genet*. 2013; 14: 36. <https://doi.org/10.1186/1471-2350-14-36> PMID: [23509962](https://pubmed.ncbi.nlm.nih.gov/23509962/)
13. Tang W, Teichert M, Chasman DI, Heit J a., Morange PE, Li G, et al. A Genome-Wide Association Study for Venous Thromboembolism: The Extended Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium. *Genet Epidemiol*. 2013; 37: 512–521. <https://doi.org/10.1002/gepi.21731> PMID: [23650146](https://pubmed.ncbi.nlm.nih.gov/23650146/)
14. Germain M, Chasman DI, de Haan H, Tang W, Lindström S, Weng L-C, et al. Meta-analysis of 65,734 Individuals Identifies TSPAN15 and SLC44A2 as Two Susceptibility Loci for Venous Thromboembolism. *Am J Hum Genet*. 2015; 532–542. <https://doi.org/10.1016/j.ajhg.2015.01.019> PMID: [25772935](https://pubmed.ncbi.nlm.nih.gov/25772935/)
15. Reitsma PH, Rosendaal FR. Past and future of genetic research in thrombosis. *J Thromb Haemost*. 2007; 5 Suppl 1: 264–269. <https://doi.org/10.1111/j.1538-7836.2007.02502.x> PMID: [17635735](https://pubmed.ncbi.nlm.nih.gov/17635735/)
16. Middeldorp S, Coppens M. Evolution of thrombophilia testing. *Hema*. 2013; 7: 375–382.
17. Wichers IM, Tanck MWT, Meijers JCM, Lisman T, Reitsma PH, Rosendaal FR, et al. Assessment of coagulation and fibrinolysis in families with unexplained thrombophilia. *Thromb Haemost*. 2009; 101: 465–470. <https://doi.org/10.1160/TH08-06-0405> PMID: [19277406](https://pubmed.ncbi.nlm.nih.gov/19277406/)
18. de Visser MCH, van Minkelen R, van Marion V, den Heijer M, Eikenboom J, Vos HL, et al. Genome-wide linkage scan in affected sibling pairs identifies novel susceptibility region for venous thromboembolism: Genetics in familial thrombosis study. *J Thromb Haemost*. 2013; 11: 1474–1484. <https://doi.org/10.1111/jth.12313> PMID: [23742623](https://pubmed.ncbi.nlm.nih.gov/23742623/)
19. Blom JW, Doggen CJM, Osanto S, Rosendaal FR. Malignancies, prothrombotic mutations, and the risk of venous thrombosis. *JAMA*. 2005; 293: 715–722. <https://doi.org/10.1001/jama.293.6.715> PMID: [15701913](https://pubmed.ncbi.nlm.nih.gov/15701913/)
20. DePristo MA, Banks E, Poplin RE, Garimella K V, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011; 43: 491–498. <https://doi.org/10.1038/ng.806> PMID: [21478889](https://pubmed.ncbi.nlm.nih.gov/21478889/)
21. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*. 2003; 14: 300–306. <https://doi.org/10.1097/01.EDE.0000042804.12056.6C> PMID: [12859030](https://pubmed.ncbi.nlm.nih.gov/12859030/)
22. Wirth KE, Tchetgen Tchetgen EJ. Accounting for selection bias in association studies with complex survey data. *Epidemiology*. 2014; 25: 444–53. <https://doi.org/10.1097/EDE.0000000000000037> PMID: [24598413](https://pubmed.ncbi.nlm.nih.gov/24598413/)
23. EVS. Exome Variant Server. NHLBI GO Exome Seq Proj. 2014;
24. 1000 Genomes Project Consortium T 1000 GP, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; 491: 56–65. <https://doi.org/10.1038/nature11632> PMID: [23128226](https://pubmed.ncbi.nlm.nih.gov/23128226/)
25. Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, et al. Characterizing and measuring bias in sequence data. *Genome Biol*. BioMed Central Ltd; 2013; 14: R51. <https://doi.org/10.1186/gb-2013-14-5-r51> PMID: [23718773](https://pubmed.ncbi.nlm.nih.gov/23718773/)
26. Cunha MLR, Meijers JCM, Middeldorp S. Introduction to the analysis of next generation sequencing data and its application to venous thromboembolism. *Thromb Haemost*. Schattauer GmbH; 2015; 114: 920–932.

27. Arnold M, Raffler J, Pfeufer A, Suhre K, Kastenmüller G. SNIIPA: an interactive, genetic variant-centered annotation browser. *Bioinforma*. 2015; 31: 1334–1336. <https://doi.org/10.1093/bioinformatics/btu779> PMID: 25431330
28. Fleiss J, Levin B, Cho Paik M. *Statistical Methods for Rates and Proportions*. John Wiley & Sons. 2003. <https://doi.org/10.1198/tech.2004.s812>
29. Zöller B, Li X, Sundquist J, Sundquist K. Age- and Gender-Specific Familial Risks for Venous Thromboembolism: A Nationwide Epidemiological Study Based on Hospitalizations in Sweden. *Circ*. 2011; 124: 1012–1020. <https://doi.org/10.1161/CIRCULATIONAHA.110.965020> PMID: 21824919
30. Morange P-E, Suchon P, Tregouet D-A. Genetics of Venous Thrombosis: update in 2015. *Thromb Haemost*. Schattauer Publishers; 2016; <https://doi.org/10.1160/TH15-05-0410> PMID: 26354877
31. Alkuraya FS. Discovery of mutations for Mendelian disorders. *Human Genetics*. 2016. pp. 615–623. <https://doi.org/10.1007/s00439-016-1664-8> PMID: 27068822
32. Chaney JL, Steele A, Carmichael R, Rodriguez A, Specht AT, Ngo K, et al. Widespread position-specific conservation of synonymous rare codons within coding sequences. *PLoS Comput Biol*. 2017; 13. <https://doi.org/10.1371/journal.pcbi.1005531> PMID: 28475588
33. Fernandez-Cadenas I, Penalba A, Boada C, MsC CC, Bueno SR, Quiroga A, et al. Exome Sequencing and Clot Lysis Experiments Demonstrate the R458C Mutation of the Alpha Chain of Fibrinogen to be Associated with Impaired Fibrinolysis in a Family with Thrombophilia. *J Atheroscler Thromb*. 2015; <https://doi.org/10.5551/jat.30676> PMID: 26581183
34. Su J, Shu L, Zhang Z, Cai L, Zhang X, Zhai Y, et al. A small deletion in SERPINC1 causes type I anti-thrombin deficiency by promoting endoplasmic reticulum stress. *Oncotarget*. Impact Journals LLC; 2016; 7: 76882–76890. <https://doi.org/10.18632/oncotarget.12349> PMID: 27708219
35. Lee E-J, Dykas DJ, Leavitt AD, Camire RM, Ebberink E, García de Frutos P, et al. Whole-exome sequencing in evaluation of patients with venous thromboembolism. *Blood Adv*. 2017; 1: 1224 LP–1237.
36. Larsen TB, Sørensen HT, Skytthe A, Johnsen SPSP, Vaupel JW, Christensen K, et al. Major genetic susceptibility for venous thromboembolism in men: a study of Danish twins. *Epidemiology*. 2003; 14: 328–32. <https://doi.org/10.1097/01.EDE.0000060457.51194.BC> PMID: 12859034
37. Zöller B, Ohlsson H, Sundquist J, Sundquist K. A sibling based design to quantify genetic and shared environmental effects of venous thromboembolism in Sweden. *Thromb Res*. 2017; 149: 82–87. <https://doi.org/10.1016/j.thromres.2016.10.014> PMID: 27793415
38. Society = Writing Committee for the Normal Tension Glaucoma Genetic Study Group of Japan Glaucoma, Meguro A, Inoko H, Ota M, Mizuki N, Bahram S. Genome-wide association study of normal tension glaucoma: common variants in SRBD1 and ELOVL5 contribute to disease susceptibility. *Ophthalmology*. Department of Ophthalmology and Visual Science, Yokohama City University Graduate School of Medicine, Yokohama, Kanagawa, Japan.; 2010; 117: 1331–8.e5. <https://doi.org/10.1016/j.ophtha.2009.12.001> PMID: 20363506
39. Benyamin B, Pourcain B, Davis OS, Davies G, Hansell NK, Brion M-JA, et al. Childhood intelligence is heritable, highly polygenic and associated with FBNP1L. *Mol Psychiatry*. 1] The University of Queensland, Queensland Brain Institute, St Lucia, Queensland, Australia [2] Queensland Institute of Medical Research, Brisbane, Queensland, Australia.; 2014; 19: 253–258. <https://doi.org/10.1038/mp.2012.184> PMID: 23358156
40. Sabater-Lleal M, Huang J, Chasman D, Naitza S, Dehghan A, Johnson AD, et al. Multiethnic meta-analysis of genome-wide association studies in >100 000 subjects identifies 23 fibrinogen-associated Loci but no strong evidence of a causal association between circulating fibrinogen and cardiovascular disease. *Circulation*. 2013; 128: 1310–1324. <https://doi.org/10.1161/CIRCULATIONAHA.113.002251> PMID: 23969696
41. Danik JS, Pare G, Chasman DI, Zee RYL, Kwiatkowski DJ, Parker A, et al. Multiple Novel Loci, Including Those Related to Crohn's Disease, Psoriasis and Inflammation, Identified in a Genome-Wide Association Study of Fibrinogen in 17,686 Women: the Women's Genome Health Study. *Circ Cardiovasc Genet*. 2009; 2: 134–141.
42. Uitte de Willige S, de Visser MCH, Houwing-Duistermaat JJ, Rosendaal FR, Vos HL, Bertina RM. Genetic variation in the fibrinogen gamma gene increases the risk for deep venous thrombosis by reducing plasma fibrinogen γ' levels. *Blood*. 2005; 106: 4176–4183. <https://doi.org/10.1182/blood-2005-05-2180> PMID: 16144795
43. Fransen E, Bonneux S, Corneveaux JJ, Schrauwen I, Di Berardino F, White CH, et al. Genome-wide association analysis demonstrates the highly polygenic character of age-related hearing impairment. *Eur J Hum Genet*. 2014; 1–6. <https://doi.org/10.1038/ejhg.2014.56> PMID: 24939585
44. Franke A, McGovern DPB, Barrett JC, Wang K, Radford-Smith GL, Ahmad T, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet*. 2010; 42: 1118–25. <https://doi.org/10.1038/ng.717> PMID: 21102463

45. Rentería ME, Coolen MW, Statham AL, Seong R, Choi M, Qu W, et al. GWAS of DNA Methylation Variation Within Imprinting Control Regions Suggests Parent-of-Origin Association. *Twin Res Hum Genet.* 2013; 16: 767–781. <https://doi.org/10.1017/thg.2013.30> PMID: 23725790
46. Medland SE, Nyholt DR, Painter JN, McEvoy BP, McRae AF, Zhu G, et al. Common Variants in the Trichohyalin Gene Are Associated with Straight Hair in Europeans. *Am J Hum Genet.* 2009; 85: 750–755. <https://doi.org/10.1016/j.ajhg.2009.10.009> PMID: 19896111
47. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature.* 2012; 491: 119–24. <https://doi.org/10.1038/nature11582> PMID: 23128233
48. López S, Buil A, Souto JC, Casademont J, Martinez-Perez A, Almasy L, et al. A genome-wide association study in the genetic analysis of idiopathic thrombophilia project suggests sex-specific regulation of mitochondrial DNA levels. *Mitochondrion.* 2014; 18: 34–40. <https://doi.org/10.1016/j.mito.2014.09.004> PMID: 25240745
49. Mangold E, Ludwig KU, Birnbaum S, Baluardo C, Ferrian M, Herms S, et al. Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. *Nat Genet.* 2010; 42: 24–6. <https://doi.org/10.1038/ng.506> PMID: 20023658
50. Ludwig KU, Mangold E, Herms S, Nowak S, Reutter H, Paul A, et al. Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. *Nat Genet.* 2012; 44: 968–71. <https://doi.org/10.1038/ng.2360> PMID: 22863734
51. Gieger C, Kühnel B, Radhakrishnan a, Cvejic a, Serbanovic-Canic J, Meacham S, et al. New gene functions in megakaryopoiesis and platelet formation. *Nature.* 2011; 480: 201–208. <https://doi.org/10.1038/nature10659> PMID: 22139419
52. Chen Z-J, Zhao H, He L, Shi Y, Qin Y, Shi Y, et al. Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. *Nat Genet.* 2011; 43: 55–59. <https://doi.org/10.1038/ng.732> PMID: 21151128
53. Shi Y, Zhao H, Shi Y, Cao Y, Yang D, Li Z, et al. Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. *Nat Genet.* 2012; 44: 1020–1025. <https://doi.org/10.1038/ng.2384> PMID: 22885925
54. Eeles R a, Kote-Jarai Z, Al Olama AA, Giles GG, Guy M, Severi G, et al. Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nat Genet.* 2009; 41: 1116–1121. <https://doi.org/10.1038/ng.450> PMID: 19767753
55. Hart AB, Engelhardt BE, Wardle MC, Sokoloff G, Stephens M, de Wit H, et al. Genome-wide association study of d-amphetamine response in healthy volunteers identifies putative associations, including cadherin 13 (CDH13). *PLoS One.* 2012; 7: e42646. <https://doi.org/10.1371/journal.pone.0042646> PMID: 22952603
56. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet.* 2008; 40: 638–45. <https://doi.org/10.1038/ng.120> PMID: 18372903
57. Aulchenko YS, Ripatti S, Lindqvist I, Boomsma D, Iris M. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet.* 2009; 41: 47–55. <https://doi.org/10.1038/ng.269> PMID: 19060911
58. Teslovich TM, Musunuru K, Smith A V, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature.* 2010; 466: 707–713. <https://doi.org/10.1038/nature09270> PMID: 20686565
59. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet.* 2013; 45: 1274–83. <https://doi.org/10.1038/ng.2797> PMID: 24097068
60. Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, Schadt EE, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet.* 2009; 41: 56–65. <https://doi.org/10.1038/ng.291> PMID: 19060906
61. Coram MA, Duan Q, Hoffmann TJ, Thornton T, Knowles JW, Johnson NA, et al. Genome-wide characterization of shared and distinct genetic components that influence blood lipid levels in ethnically diverse human populations. *Am J Hum Genet.* 2013; 92: 904–16. <https://doi.org/10.1016/j.ajhg.2013.04.025> PMID: 23726366
62. Chasman DI, Paré G, Mora S, Hopewell JC, Peloso G, Clarke R, et al. Forty-three loci associated with plasma lipoprotein size, concentration, and cholesterol content in genome-wide analysis. *PLoS Genet.* 2009; 5: e1000730. <https://doi.org/10.1371/journal.pgen.1000730> PMID: 19936222
63. Teupser D, Baber R, Ceglarek U, Scholz M, Illig T, Gieger C, et al. Genetic regulation of serum phytosterol levels and risk of coronary artery disease. *Circ Cardiovasc Genet.* 2010; 3: 331–339. <https://doi.org/10.1161/CIRCGENETICS.109.907873> PMID: 20529992

64. Buch S, Schafmayer C, Volzke H, Becker C, Franke A, von Eller-Eberstein H, et al. A genome-wide association scan identifies the hepatic cholesterol transporter ABCG8 as a susceptibility factor for human gallstone disease. *Nat Genet.* 2007; 39: 995–999. <https://doi.org/10.1038/ng2101> PMID: 17632509
65. Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, Daigo Y, et al. Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat Genet.* 2010; 42: 210–215. <https://doi.org/10.1038/ng.531> PMID: 20139978
66. Ganesh SK, Zakai N a, van Rooij FJ a, Soranzo N, Smith A V, Nalls M a, et al. Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat Genet.* 2009; 41: 1191–1198. <https://doi.org/10.1038/ng.466> PMID: 19862010
67. van der Harst P, Zhang W, Mateo Leach I, Rendon A, Verweij N, Sehmi J, et al. Seventy-five genetic loci influencing the human red blood cell. *Nature.* 2012; 492: 369–75. <https://doi.org/10.1038/nature11677> PMID: 23222517
68. Chen Z, Tang H, Qayyum R, Schick UM, Nalls M a., Handsaker R, et al. Genome-wide association analysis of red blood cell traits in African Americans: The cogent network. *Hum Mol Genet.* 2013; 22: 2529–2538. <https://doi.org/10.1093/hmg/ddt087> PMID: 23446634
69. Imboden M, Bouzigon E, Curjuric I, Ramasamy A, Kumar A, Hancock DB, et al. Genome-wide association study of lung function decline in adults with and without asthma. *J Allergy Clin Immunol.* 2012; 129: 1–11. <https://doi.org/10.1016/j.jaci.2012.01.074> PMID: 22424883
70. Smith JG, Avery CL, Evans DS, Nalls M a., Meng Y a., Smith EN, et al. Impact of ancestry and common genetic variants on QT interval in African Americans. *Circ Cardiovasc Genet.* 2012; 5: 647–655. <https://doi.org/10.1161/CIRCGENETICS.112.962787> PMID: 23166209
71. Yu B, Zheng Y, Alexander D, Manolio TA, Alonso A, Nettleton JA, et al. Genome-wide association study of a heart failure related metabolomic profile among African Americans in the Atherosclerosis Risk in Communities (ARIC) study. *Genet Epidemiol.* 2013; 37: 840–845. <https://doi.org/10.1002/gepi.21752> PMID: 23934736
72. Perlis RH, Huang J, Purcell S, Fava M, Rush a J, Sullivan PF, et al. Genome-wide association study of suicide attempts in mood disorder patients. *Am J Psychiatry.* 2010; 167: 1499–1507. <https://doi.org/10.1176/appi.ajp.2010.10040541> PMID: 21041247
73. Simeoni I, Stephens JC, Hu F, Deevi SVV, Megy K, Bariana TK, et al. A high-throughput sequencing test for diagnosing inherited bleeding, thrombotic, and platelet disorders. *Blood.* 2016; 127: 2791–2803. <https://doi.org/10.1182/blood-2015-12-688267> PMID: 27084890
74. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 2014; 46: 310–5. <https://doi.org/10.1038/ng.2892> PMID: 24487276