Supplemental Information
Diagnostic high-throughput sequencing of 2,396 patients with bleeding, thrombotic and platelet disorders

Patient cohorts and inclusion criteria

ThromboGenomics cohort

International recruitment of 1,608 index patients with mostly known or suspected inherited bleeding, thrombotic or platelet disorders following previously described criteria used for the validation of the ThromboGenomics HTS test (Supplemental Table 1). [1] Patients entered this study through two routes.

1) Patients recruited to the bleeding and platelet disorder arm of the NIHR-BioResource Rare Diseases study with a suspected disease etiology (as below) were sequenced using the HTS test as a pre screen prior to whole genome sequencing. Patients provided written consent according to the study ethical approval (East of England Cambridge South national research ethics committee (REC) reference 13/EE/0325). Obtaining consent from overseas patients was the responsibility of the respective principal investigators at the enrolling hospitals. Material and transfer agreements were applied to regulate the exchange of samples and data between the donor institutions and the University of Cambridge.

2) Patients were referred for diagnostic testing and consented by the referring clinician, according to local clinical practice. For UK patients, a UKHCDO approved patient information leaflet and consent form was provided for use. Consent for genetic testing was also obtained from a small number of relatives and suspected carriers (Supplemental Table 1).

Referral guidelines with inclusion criteria were provided on the ThromboGenomics website for referring clinicians. For patients with bleeding and platelet disorders, testing was recommend in patients with a suspected disease etiology where there was a high likelihood of the condition being genetic. Exclusion criteria included the use of prescription or over-the-counter drugs known to be associated with bleeding or abnormal platelet phenotypes or a high likelihood of a medical condition associated with abnormal platelet phenotypes. For the thrombotic disorders, testing was recommended in patients with a suspected etiology if there was a high likelihood of the condition being genetic due to the patient suffering from a thrombotic event before 40 years of age and a positive family history. Exclusion criteria included late onset thrombosis, thrombosis caused by trauma or surgical challenge or if an acquired thrombotic disorder was suspected.

Patients referred for diagnostic testing with severe hemophilia A were tested for the pathogenic intron 1 and intron 22 inversions in the F8 gene using inverse shifting-polymerase chain reaction prior to testing using the ThromboGenomics test.[2]

All likely pathogenic and pathogenic variants in patients referred for diagnostic testing have been confirmed using Sanger sequencing or Multiplex ligation-dependent probe amplification (MLPA) copy number variation (CNV) analysis in the Cambridge University Hospitals Genetic Laboratories.

Preoperative screening for mild bleeding disorders cohort (PANE)

Individuals were identified using responses to a pre operative anesthesiology bleeding questionnaire when admitted for elective surgery at Maastricht University Medical Centre (MUMC). Patients were invited to join the PANE study if they reported one or more bleeding symptoms and were over the age of 18, had no anaemia and were not using drugs that could interfere with hemostasis.[3,4] Ethical approval was obtained from the Medical Ethics Committee of MUMC (NL38767.068.11/METC11-2-096). Written informed consent was obtained from all patients. During the study visit, bleeding symptoms of the participants were evaluated by an experienced hematologist. Blood was drawn and subjected to a panel of 30 laboratory tests (Supplemental Table 2). Patients with excluded from the study with platelet counts $<100x10^9$/l. Using local reference values, HPO codes were appended to patients with abnormal test parameters and clinical bleeding symptoms. HTS was performed for 212 unrelated study subjects that included 193 index patients and 19 subjects that did not have any bleeding symptoms or clinical test results that indicated a BTPD-related HPO code (Supplemental Table 1).

Vienna Bleeding Biobank cohort (VIBB)

The VIBB study was established in collaboration with the MedUni Wien Biobank (Department of Laboratory Medicine, Medical University of Vienna, Austria, www.biobank.at) as a single-centre study.[5] Patients who were referred to the hemostasis outpatient department for investigation of a mild to moderate bleeding disorder were recruited into the The Vienna Bleeding Biobank (VIBB) study. The Ethics Committee of the Medical University of Vienna approved the project (EK No 603/2009 and 039/2006). All patients gave written informed consent before inclusion in the VIBB study. On recruitment to the study, blood was drawn and subjected to a panel of 42 laboratory tests (Supplemental Table 2). Patients with excluded from the study with platelet counts $<100x10^9$/l. Using local reference values, HPO codes were appended to patients with abnormal test parameters and clinical bleeding symptoms. HTS was performed for 595 index patients (Supplemental Table 1).

ThromboGenomics HTS BTPD test design

ROCHE NimbleGen SeqCap capture baits (ROCHE NimbleGen, Inc. Madison, WI USA) were designed to target all consensus coding sequences (CCDS) of the ThromboGenomics genes, the first and last 100 bp of introns, 5' and 3' UTRs, regions 1,000bp upstream of the transcription start site and the position of known pathogenic variants (see below). Regions of four allosome genes (SRY, TSPY1, AMELY, AMELX) were included to assign genomics sex. For TG.V3, capture baits were also included for a panel of 10,000 common SNVs to calculate ethnicity and relatedness estimates.

Known variants

A curated list of known disease associated variants were included in the HTS test design and for variant prioritisation.

At the time of design, all known Human Gene Mutation Database (HGMD) variants associated with the BTPD genes were included in the design.[6] For TG.V2, 9,280 variants were included (HGMDPro 2015.2) and for TG.V3, 11,442 variants were included (HGMDPro 2016.4).

A set of known variants was used to prioritise the variants identified in patient samples (see below). All variants in HGMD_PRO_2017.2 were used to perform the prioritisation analysis of all TG.V2 samples. HGMD_PRO_2017.4 variants were used for the prioritisation of TG.V3 samples alongside a curated list of variants from disease specific databases. These included; 2,498 variants, with a gnomAD minor allele frequency (MAF) <0.25, from the F7, F8, F9 and VWF gene EAHAD databases (http://www.eahad-db.org/, accessed February 2017); 438 MYH9, GP1BA, GP1BB, GP9, and WAS variants with a MAF <0.01 from the LOVD databases[7] (accessed May 2017); 203 ITGA2B and ITGB3 variants from the Glanzmann Thrombasthenia Database (https://glanzmann.mcw.edu/, accessed May 2017). In addition, all variants previously reported by ThromboGenomics were used for prioritisation.

Library preparation, enrichment and sequencing

DNA samples were processed as previously described with minor modifications.[1] In short, samples were processed in batches with 500ng of each sample fragmented using a Covaris E220 (Covaris Inc., Woburn, MA, USA). Samples were processed using the ROCHE KAPA HTP Library Preparation kit (Roche Diagnostics Ltd., Burgess Hill, UK). DNA libraries were captured using ROCHE NimbleGen SeqCap ThromboGenomics capture baits (ROCHE NimbleGen, Inc. Madison, WI USA). Final libraries were quantified,

samples pooled and sequenced using an Illumina Hiseq 4000 sequencer, 150 base pair (bp) paired-end (PE) run.

Initially for TG.V2, 48 samples were multiplexed per sequencing reaction. In an effort to increase sample throughput and reduce costs, methods were adjusted to multiplex 96 samples. At the same time, modifications were made to increase the DNA library fragment size with the aim of increasing read coverage in poorly performing regions. DNA was fragmented to obtain an average insert size of 220 bp for TG.V2 and 350 bp for TG.V3.

Variant calling

Single nucleotide variants (SNVs) and short insertions or deletions (INDELs) were called using GATK 3.3 using GRCh37.[8] HaplotypeCaller in a single sample mode and filtered using the following VariantFiltration expressions "MQ< 40.0 || QD < 2.0 || FS > 100.0" for SNVs and "FS > 200.0 || QD < 2.0 || ReadPosRankSum < -20.0" for INDELs. Variants were merged into multi-sample VCF files. SNVs and INDELs were annotated with their predicted impact against Ensembl 75, presence in the human gene using SnpEff 4.0.[9]

Relatedness and ancestry estimation

A panel of 10,000 SNVs were incorporated into the design of TG.V3 to estimate the degree of relatedness between individuals and to categorise an individual's ancestry into European, African, East Asian, South Asian or Other. These variants were selected from a larger panel of SNVs recommended by ROCHE for this purpose (personal communication, Todd Richmond, Roche Sequencing Solutions). Principal component analysis of samples from the 1000 Genomes Project with known ethnicity were used to generate a reference data set. Patient samples were then compared to this reference to estimate ancestry.

Relatedness was estimated using the 10,000 common SNVs using the PC-Relate function from GENESIS R package. For each sample pair, a relatedness score was calculated, ranging from 0 to 1.

Copy Number Variation

CNVs were called using a custom pipeline based on the ExomeDepth R-package (version 1.1.10).[10] ExomeDepth makes a copy number gain or loss call in a specified genomic interval by comparing the read depths in a sample and an optimised reference set of other samples. Our customisation reduces false negative and positive calls by specifically defining a set of ten unrelated reference samples. To detect small CNVs within large exons, genomic intervals of no more than 500bp were used to calculate read depth. Modifications were also made to the ExomeDepth read counting method to avoid inflation

caused by reads overlapping two adjacent genomic intervals. CNVs observed in more than 10% of samples within a batch were filtered out as technical artefacts or common CNVs. Raw sequencing reads of all potentially pathogenic CNVs were visually inspected by the MDT.

Region of interest

The region of interest (ROI) for variant prioritisation was defined as:

- all coding regions for each curated gene transcript
- +/- 15 bps into the introns
- 5' and 3' UTRs sequences
- the position of all known variants at the time of the panel design

Variant prioritisation

For each patient sample, variants identified were prioritised to provide a list of potentially pathogenic variants for interpretation by the multi-disciplinary team (MDT).

Variants were prioritised if:

- predicted to have a moderate or a high impact effect according to SnpEff
- within the snRNA gene RNU4ATAC
- located at the same nucleotide position as a known variant with a gnomAD MAF <0.025
  or
- novel with a gnomAD MAF <0.001

Variants were not prioritised if they had >3 alternate alleles (to guard against sequencing errors in repetitive regions) or if observed in the HTS samples with a frequency >=10% (to remove systematic artifacts).

Problematic regions

A number of regions are not well covered by aligned sequence generated using the HTS test and bioinformatic pipeline. A region in the UTRs of the ORAI1 gene has poor sequencing coverage, although only 6bp have a read depth less than 20x (GRCh37:Chr12:122064774-122064779).

Exon 26 of the VWF gene has poor aligned sequence coverage due to a homologous region within VWFP1, an unprocessed pseudogene on chromosome 22 (<20x aligned reads at GRCh37:Chr12:6131938-6132010). Raw sequencing reads that align to this region are manually inspected in patients with an indication of von Willebrand Disease to

identify any possible variants within this exon. Any potential pathogenic variants are confirmed using PCR primers specific to the VWF gene sequence before issuing a report.

Oligogenic Findings

The oligogenic visualisation plot was made using Circos software[11] available to download (http://circos.ca/software/). Amino acid conservation scores were calculated using ConSurf (http://consurf.tau.ac.il/2016).

Variant interpretation by the multi-disciplinary team (MDT)

Two hour MDT teleconference meetings were organised weekly with participation of one MDT chair (ML and KG mostly for coagulation genes, AM and KF mostly for platelet genes), scientific lead and coordinator (KD), Clinical Scientist (NAH), bioinformatics expert (ET, OS and KM), pertinent findings specialist (KM) and a responsible scientist for laboratory issues and preparation of the MDT meeting (DD). Clinicians were often invited or alternatively additional patient information was requested via email. The role of the MDT is to discuss clinical, laboratory and genetic findings and to interpret and classify variants.

Supplemental References:

1. Simeoni I, Stephens JC, Hu F, et al. A high-throughput sequencing test for diagnosing inherited bleeding, thrombotic, and platelet disorders. *Blood*. 2016;127(23):2791-2803.
2. Rossetti LC, Radic CP, Larripa IB, De Brasi CD. Developing a new generation of tests for genotyping hemophilia-causative rearrangements involving int22h and int1h hotspots in the factor VIII gene. *J Thromb Haemost*. 2008;6(5):830-836.
3. Vries MJ, van der Meijden PE, Kuiper GJ, et al. Preoperative screening for bleeding disorders: A comprehensive laboratory assessment of clinical practice. *Res Pract Thromb Haemost*. 2018;2(4):767-777.
4. Moenen F, Vries MJA, Nelemans PJ, et al. Screening for platelet function disorders with Multiplate and platelet function analyzer. *Platelets*. 2017:1-7.
5. Gebhart J, Hofer S, Panzer S, et al. High proportion of patients with bleeding of unknown cause in persons with a mild-to-moderate bleeding tendency: Results from the Vienna Bleeding Biobank (VIBB). *Haemophilia*. 2018;24(3):405-413.
6. Stenson PD, Mort M, Ball EV, et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet*. 2017;136(6):665-677.
7. Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT. LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat*. 2011;32(5):557-563.
8. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297-1303.
9. Cingolani P, Platts A, Wang le L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6(2):80-92.
10. Plagnol V, Curtis J, Epstein M, et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics*. 2012;28(21):2747-2754.
11. Krzywinski M, Schein J, Birol I, et al. Circos: an information aesthetic for comparative genomics. *Genome Res*. 2009;19(9):1639-1645.

Supplemental Tables and Figures

Supplemental Table 3

Genes and disorders included in the ThromboGenomics HTS test.

TIER1 genes sequenced using the ThromboGenomics HTS test. Table includes, HGNC-approved gene symbol, gene name, disease category, disorder, mode of inheritance, reporting transcript, LRG accession and version of test.
AD, autosomal dominant; AR, autosomal recessive; XLR, X-linked recessive.
*Gene without sufficient evidence for TIER1 status as of July 2018.

Supplemental Table 5

Variants reported in index patients using the ThromboGenomics HTS test.
Table includes all index patients with reported variants. Single nucleotide variants and small insertion-deletions are in a separate tab to the CNVs. Information includes patient disease class, version of ThromboGenomics used for testing, variant information reported pathogenicity and contribution to phenotype, known or novel status at time of MDT, frequency in gnomAD and reinterpretation of variant pathogenicity (2019).

Supplemental Table 1

Summary of the subjects from the ThromboGenomics, PANE and VIBB cohorts sequenced using the HTS test

	Cohort size					2,396 Index patients (prioritised variants)		Diagnostic yield
	Index patients	Hemophilia carrier query[$]	Affected relatives	Non affected relatives[&]	Non affected individuals[#]	TG.V2	TG.V3	% patients
TG	1,608	18	98*	33*	-	1,140 patients (4.75 variants)	468 patients (4.70 variants)	51.3%
PANE	193	-	-	-	19	193 patients (3.28 variants)	-	4.7%
VIBB	595	-	7 (mother child pairs)	-	4	-	595 patients (4.21 variants)	10.3%

TG: ThromboGenomics

TG.V2: ThromboGenomics version 2

TG.V3: ThromboGenomics version 3

* 225 subjects are related, in 97 families.

[$] Hemophilia A and B carrier query referred with no index case

[&] Non-affected relatives - includes four hemophilia carrier queries

[#] Non-affected subjects - participants in the study without bleeding symptoms of abnormal laboratory test results.

Supplemental Table 2
Key laboratory tests performed in the PANE and VIBB study subjects.

	PANE	**VIBB**
Coagulation	aPTT, PT, TT, Fibrinogen, Factors II, V, VII, VIII, IX, X, XI, XII and XIII activity	aPTT, PT, TT, Fibrinogen, Factors VIII, IX and XIII activity
vWF	VWF:Ag, VWF:Rco	VWF:Ag, VWF:Rco, VWF:CB
Platelet function	PFA, LTA with AA, TRAP, Collagen, Epinephrine, Ristocetin, ADP	PFA, LTA with AA, TRAP, Collagen, Epinephrine, Ristocetin, ADP. Platelet receptor phenotyping (GPIIb, GPIIIa, GPIX, GPIbalpha and GPIV)
Fibrinolysis	tPA, PAI, α2-antiplasmin activity, Plasminogen activity	tPA, PAI, Tpa-PAI complex, α2-antiplasmin activity

Supplemental Table 4
Variant interpretation guidelines used by ThromboGenomics multi-disciplinary team for study participants sequenced using TG.V2.

Variant classification	**Criteria**
Pathogenic	Variant that has been described in at least 3 unrelated index patients that have been reported in: - HGMD as disease mutation (DM) - F7/F8/F9/VWF EAHAD variant database (http://eahad-db.org) - ThromboGenomics BPD variant dataset
Likely Pathogenic	Variant that has been described in <3 unrelated index patients but has been reported in: - HGMD as DM - F7/F8/F9/VWF EAHAD variant database - ThromboGenomics BPD variant dataset OR Variant that is absent from control datasets (1000G, ExAC and UK10K sequencing data) and present in a gene that strongly matches the clinical and laboratory phenotype for the patient. OR Variant that is absent from control datasets (1000G, ExAC and UK10K sequencing data) and predicted to cause a loss of function
Variant of uncertain clinical significance (VUS)	Variant that has a low minor allele frequency (MAF) or is absent from control datasets (1000G, gnomAD and UK10K sequencing data) and present in a gene that matches the clinical and laboratory phenotype for the patient

Supplemental Table 6

Details of index patients with oligogenic findings. VUS – Variants of Uncertain Significance

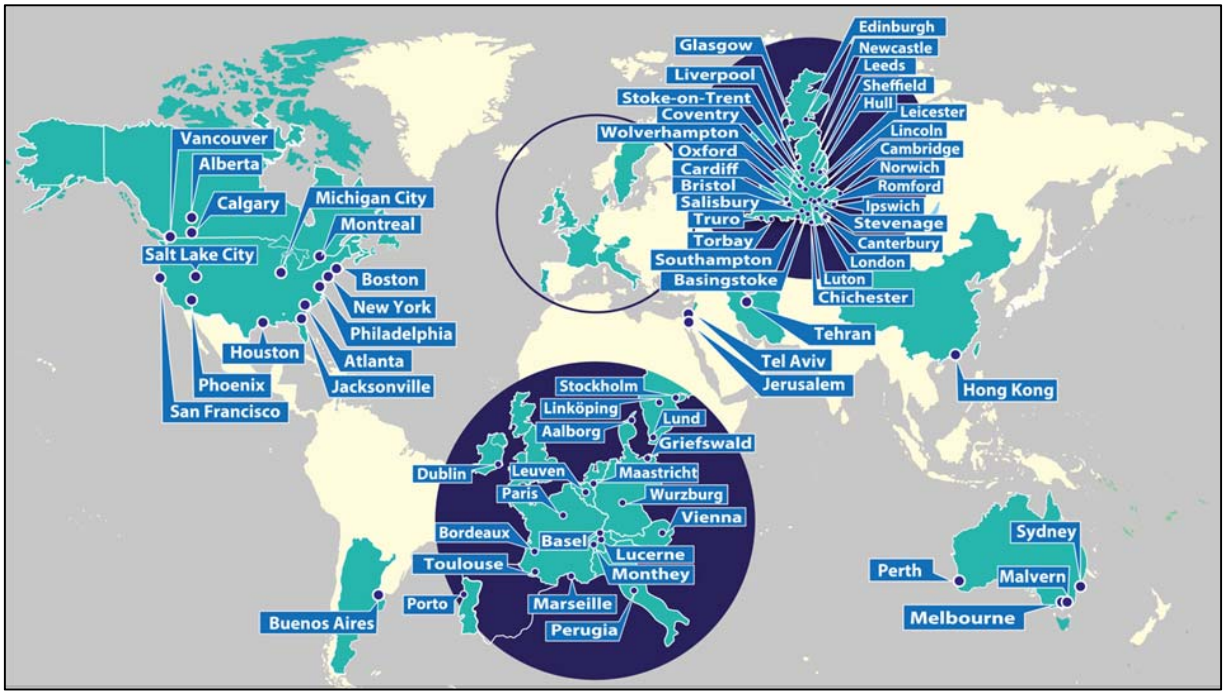
Disease class	Index patient	Gene 1	Gene 2	Gene 3
		Variant (heterozygous/homozygous) – Variant classification		
Thrombotic	1	F5 (het)- Pathogenic	PROS1 (het)- Likely Pathogenic	
	2	F5 (het)- Pathogenic	PROS1 (het)- VUS	
	3	SERPIND1 (het)- VUS	FGB (het)- VUS	F5 (het)- Pathogenic
	4	SERPINC1 (het)- Likely Pathogenic	F5 (het)- Pathogenic	
	5	F2 (het)- Likely Pathogenic	PROS1 (het)- Likely Pathogenic	
	6	SERPINC1 (het)- Pathogenic	F5 (het)- Pathogenic	
	7	PROS1 (het)- Likely Pathogenic	PROC (het)- Likely Pathogenic	F2 (het)- Pathogenic
	8	F2 (het)- Pathogenic	PROS1 (het:variant 1)- Likely Pathogenic PROS1 (het:variant 2)- Likely Pathogenic	
	9	F2 (het)- Pathogenic	PROS1 (het)- Pathogenic	
	10	PROC (het)- Likely Pathogenic	PROS1 (het)- Pathogenic	
	11	PROC (het)- Likely Pathogenic PROC (het)- Likely Pathogenic	SERPINC1 (het)- VUS	
Coagulation	1	F8 (hom) - Pathogenic	F9 (hom)- VUS	
	2	F11 (het:variant 1)- Likely Pathogenic F11 (het:variant 2) - Likely Pathogenic	F8 (het)- VUS	

Coagulation	**3**	F2 (het)- Likely Pathogenic	FGB (het)- Likely Pathogenic	F9 (het)- Pathogenic
	4	VWF (het)- Likely Pathogenic	F11 (het)- Likely Pathogenic	
	5	F2 (het)- Likely Pathogenic	F7 (het)- Likely Pathogenic	
	6	VWF (het:variant1)- VUS VWF (het:variant2)- VUS	F8 (het)- VUS	
	7	VWF (het)- VUS	F8 (het)- Pathogenic	
	8	F11 (het)- VLP Likely Pathogenic	F10 (het)- VUS	F7 (het)- VUS
	9	F11 (het)- Pathogenic	VWF (het)- Likely Pathogenic	
	10	F8 (hom)- Likely Pathogenic	VWF (het)- VUS	
	11	F5 (het)- Likely Pathogenic	F10 (het)- VUS	
	12	F8 (hom)- VUS	F11 (het)- Likely Pathogenic	
	13	F13A1 (het:variant 1) - Likely Pathogenic F13A1 (het:variant 2) - Likely Pathogenic	F7 (het)- VUS	
Coagulation and Platelet function	**1**	VWF (het)- Likely Pathogenic	P2RY12 (het)- Likely Pathogenic	
	2	VWF (het)- Likely Pathogenic	GP1BB (het)- VUS	
Coagulation and Platelet count	**1**	VWF (het)- Pathogenic	GATA1 (het)- VUS	
	2	F11 (het)- VUS	TUBB1 (het)- VUS	MYH9 (het)- VUS
Unexplained Bleeding	**1**	SERPINF2 (het)- VUS	THBD (het)- VUS	

Supplemental Figures

Supplemental Figure 1

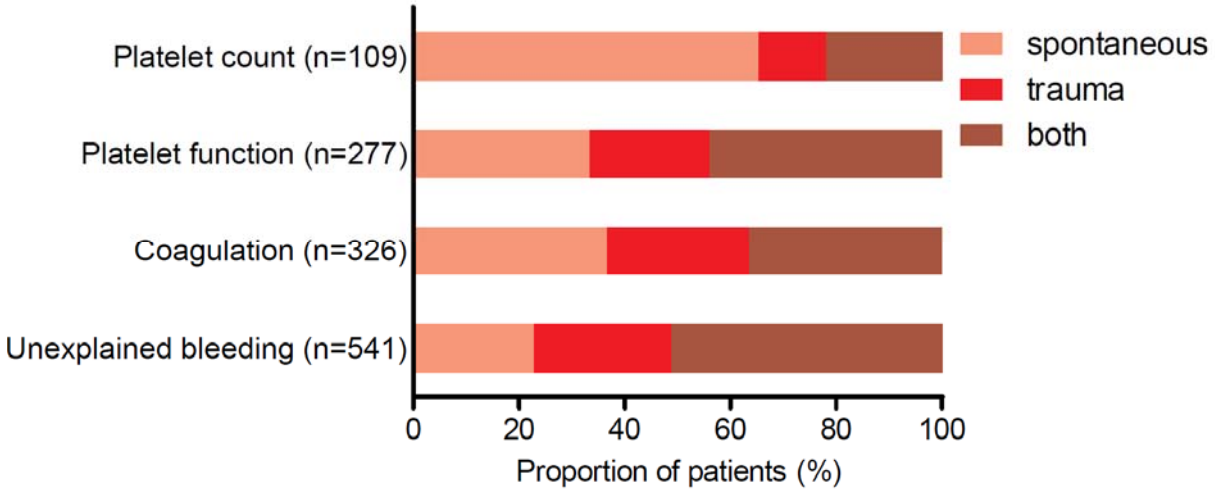
A global map with the location of the 72 UK and 46 non-UK hospitals which recruited patients used in this study.



Supplemental Figure 2

Distribution of bleeding symptoms present in the different disease classes.

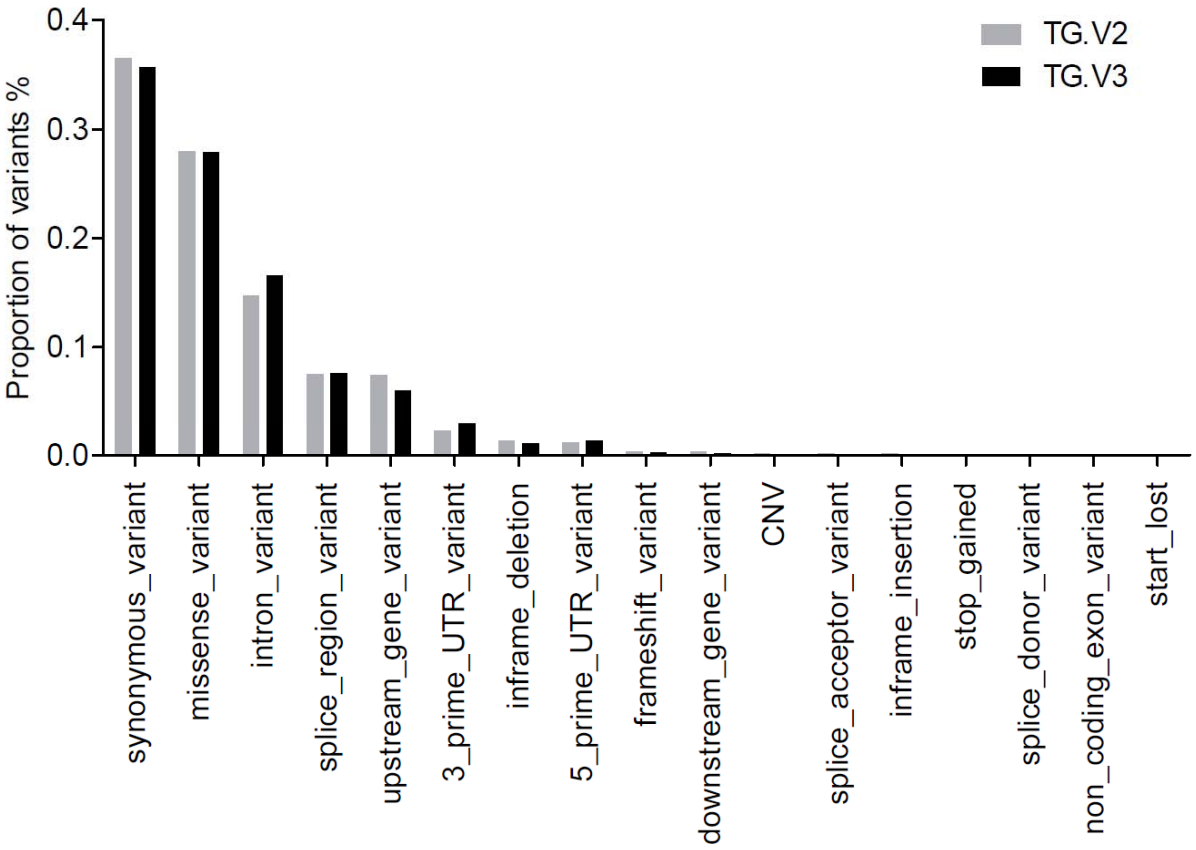
Figure representing bleeding symptoms of 1,253 patients with assigned HPO codes for 'clinically significant bleeding events' that were present in at least 10 patients. The proportion of patients per type of bleeding is on the horizontal axis. The HPO term for menorrhagia was excluded when selecting patients for this analysis. Between brackets are the number of patients for each disease class with bleeding phenotypes used in this analysis.



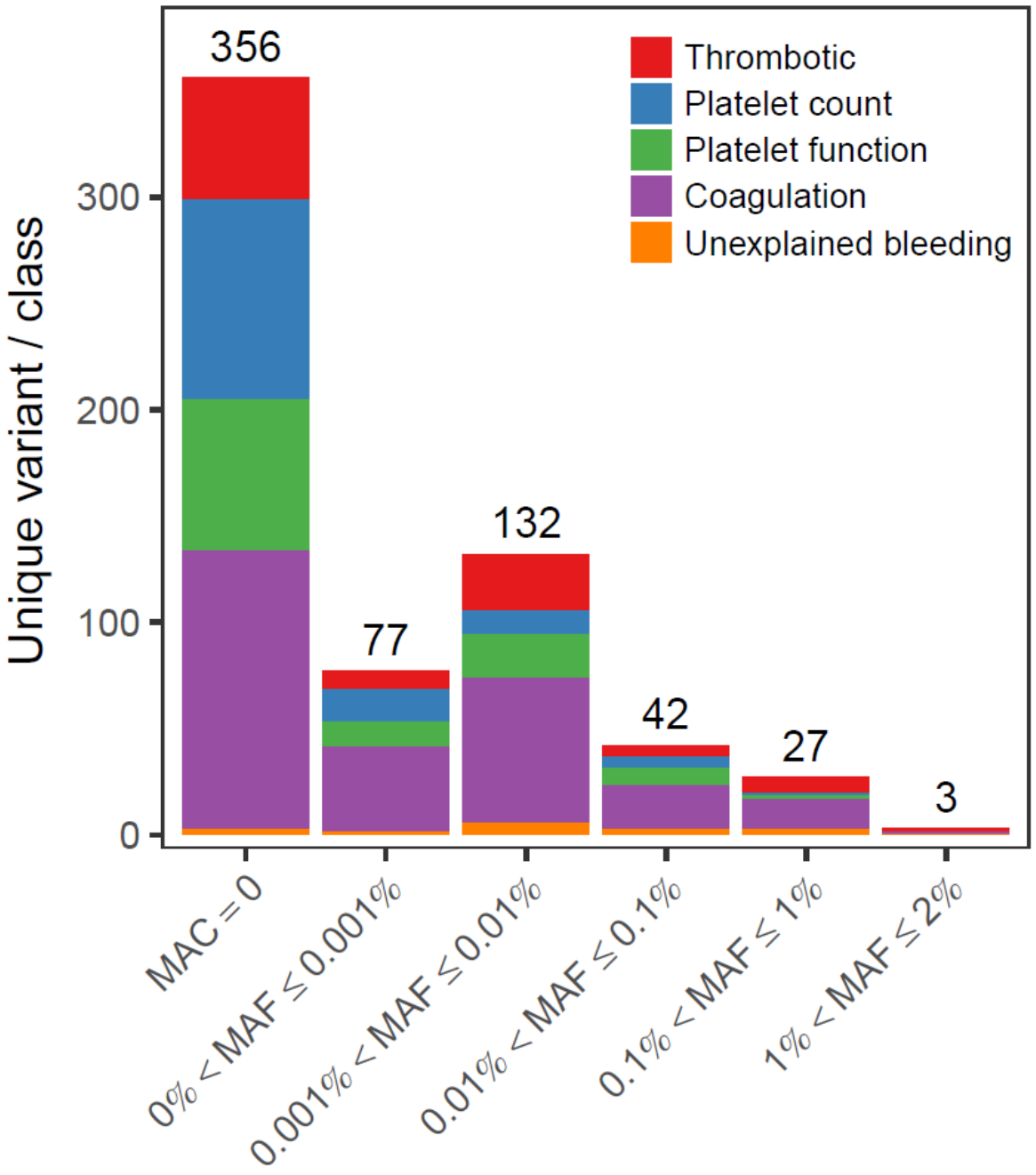
Supplemental Figure 3

The average proportion of variants and predicted effects identified in patients sequenced using TG.V2 or TG.V3.

Only variants within the region of interest were included. The average number of variants within the region of interest for each patient was 156.3 and 202.4 for TG.V2 and TG.V3 respectively.



Supplemental Figure 4
Minor allele frequency or allele count (gnomAD) for the reported autosomal SNVs and indels for each disease class.
MAC: Minor Allele Count. MAF: Minor Allele Frequency. The first bin in the plot (MAC=0) corresponds to variants not observed in gnomAD.

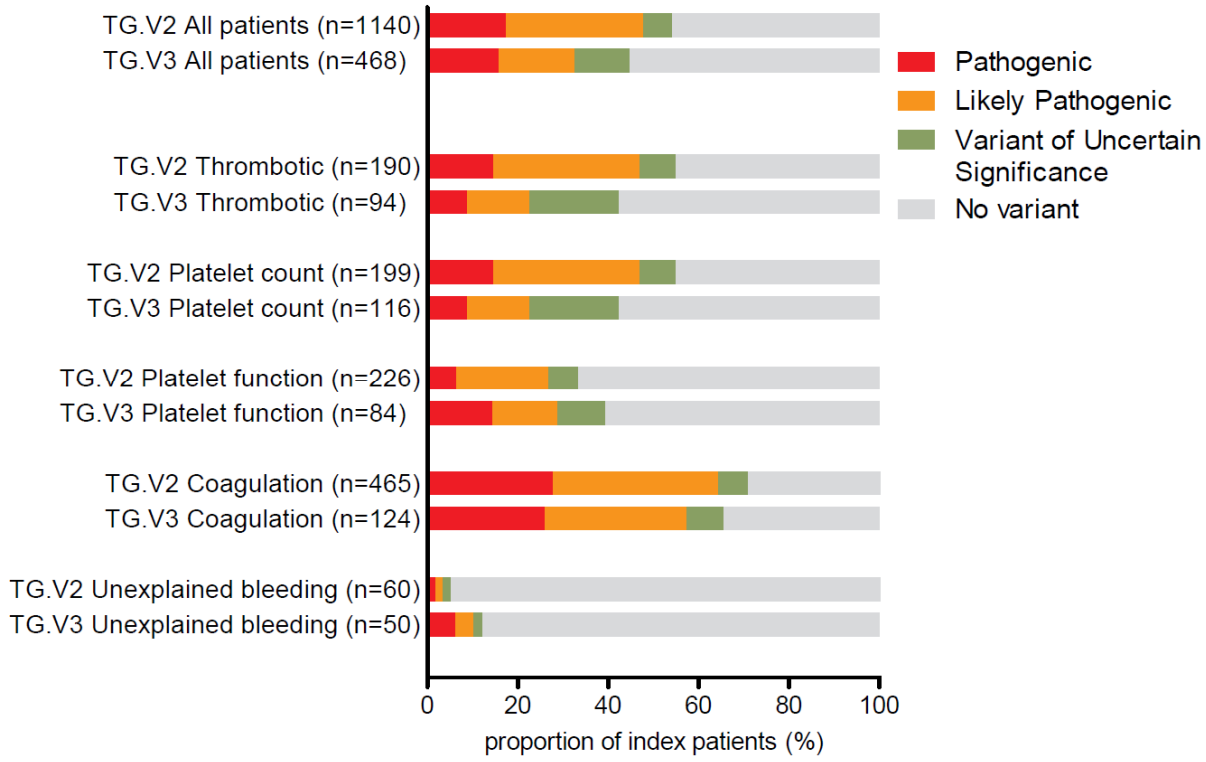


Supplemental Figure 5

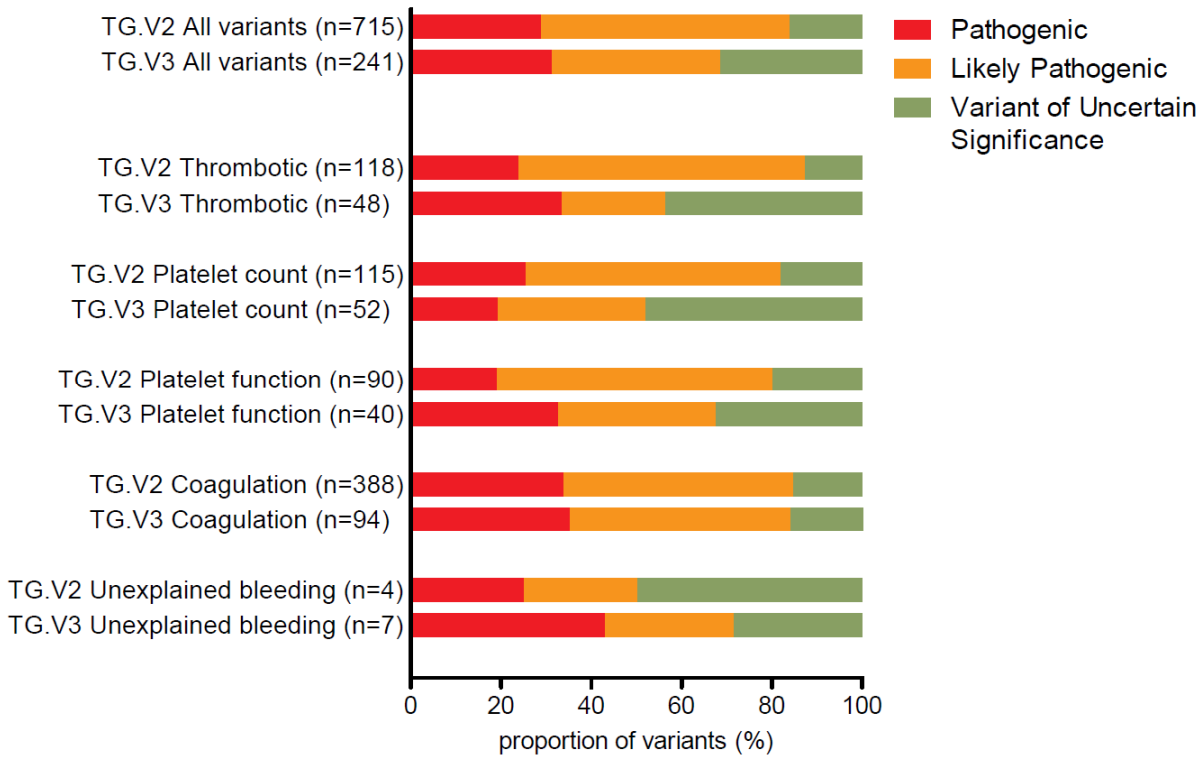
(A) Diagnostic yield in patients and (B) pathogenicity of reported variants using the TG.V2 and TG.V3 HTS test.

Index patients from the ThromboGenomics cohort were used to compare the TG.V2 and TG.V3 tests (TG.V2; 1,140 index patients and 715 reported variants and TG.V3; 468 index patients and 241 reported variants). The most pathogenic variant reported for each index patient was used to calculate diagnostic yield.

(A)

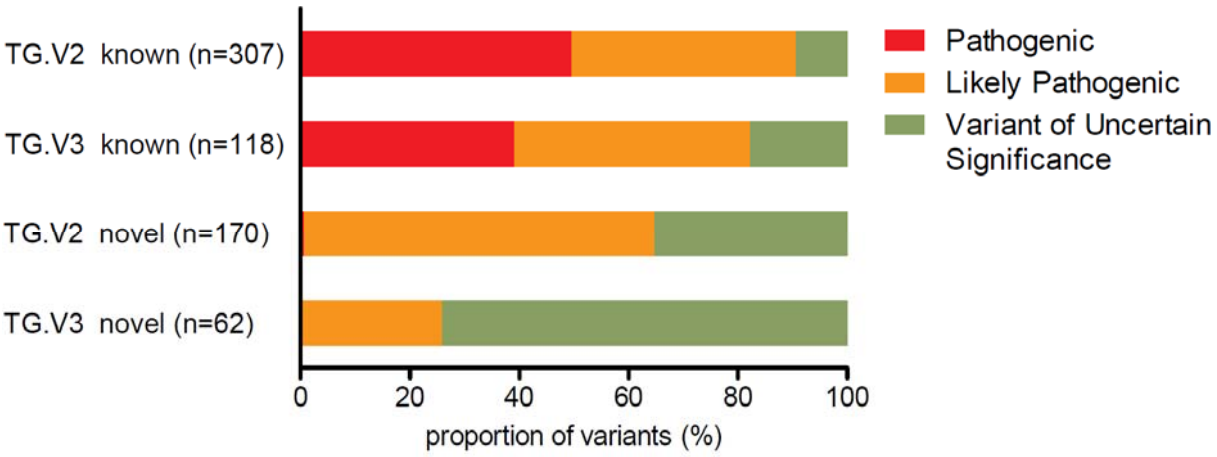


(B)


TG.V2 All variants (n=715)
TG.V3 All variants (n=241)
TG.V2 Thrombotic (n=118)
TG.V3 Thrombotic (n=48)
TG.V2 Platelet count (n=115)
TG.V3 Platelet count (n=52)
TG.V2 Platelet function (n=90)
TG.V3 Platelet function (n=40)
TG.V2 Coagulation (n=388)
TG.V3 Coagulation (n=94)
TG.V2 Unexplained bleeding (n=4)
TG.V3 Unexplained bleeding (n=7)
0
20
40
60
80
100
proportion of variants (%)
Pathogenic
Likely Pathogenic
Variant of Uncertain Significance

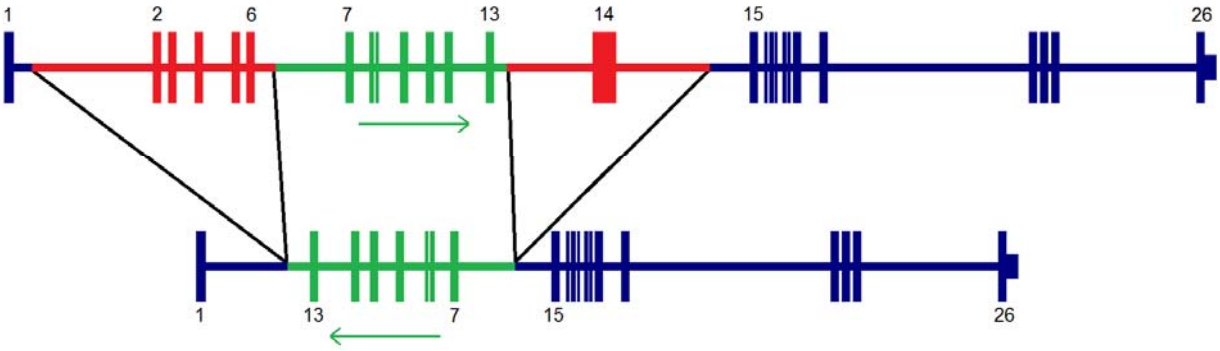
Supplemental Figure 6
The pathogenicity of novel and known reported missense variants for TG.V2 and TG.V3 in patients from the ThromboGenomics cohort.

ACMG guidelines were followed for TG.V3 patients. Analysis included 657 missense variants reported in index patients using TG.V2 (477 variants) and TG.V3 (180 variants).



Supplemental Figure 7

Hemizygous complex CNV associated with severe hemophilia A identified using the ThromboGenomics HTS test. Deletion of intron 1 to intron 6 (red) and intron 13 to intron 14 (red), flanking an inversion of intron 6 to intron 13 (green).



Supplemental Figure 8
Pedigree of proband with Glanzmann Thrombasthenia caused by a homozygous deep intronic variant.

Electrophoresis gel results of PCR amplicons of platelet cDNA reveals that aberrant splicing of the *ITGA2B* mRNA is associated with the deep intronic variant (c.1210+105A>G). Sanger sequencing of the aberrant splice products revealed two mRNAs, one with a deletion of all of exon 13 coding sequence and the second with a 123 bp deletion in the exon 13 coding sequence.