

Title

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

Authors

Kate Downes^{1,2,3,4*}, Karyn Megy^{1,2*}, Daniel Duarte^{1,2}, Minka Vries⁵, Johanna Gebhart⁶, Stefanie Hofer⁶, Olga Shamardina^{1,2}, Sri VV Deevi^{1,2}, Jonathan Stephens^{1,2}, Rutendo Mapeta^{1,2}, Salih Tuna^{1,2}, Namir Al Hasso⁴, Martin W Besser⁷, Nichola Cooper⁸, Louise Daugherty^{1,2}, Nick Gleadall^{1,3}, Daniel Greene^{1,3,16}, Matthias Haimel^{9,10}, Howard Martin⁴, Sofia Papadia^{1,2}, Shoshana Revel-Vilk¹¹, Suthesh Sivapalaratnam^{1,3,12} Emily Symington⁷, Will Thomas⁷, Chantal Thys¹³, Alexander Tolios¹⁴, Christopher J Penkett^{1,2}, NIHR BioResource², Willem H Ouwehand^{1,2,3,15}, Stephen Abbs⁴, Michael A Laffan⁸, Ernest Turro^{1,2,3,16}, Ilenia Simeoni^{1,2}, Andrew D Mumford¹⁷, Yvonne MC Henskens¹⁸, Ingrid Pabinger⁶, Keith Gomez^{19,20}, Kathleen Freson¹³

Author affiliations

1. Department of Haematology, University of Cambridge, Cambridge Biomedical Campus, Cambridge, United Kingdom
2. NIHR BioResource, Cambridge University Hospitals, Cambridge Biomedical Campus, Cambridge, United Kingdom
3. NHS Blood and Transplant, Cambridge Biomedical Campus, Cambridge, United Kingdom
4. East Midlands and East of England Genomic Laboratory Hub, Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge, United Kingdom
5. Department of Biochemistry, Maastricht University, Maastricht, The Netherlands
6. Clinical Division of Hematology and Hemostaseology, Department of Medicine I, Medical University of Vienna, Austria
7. Department of Haematology, Cambridge University Hospitals, Cambridge Biomedical Campus, Cambridge, United Kingdom
8. Centre for Haematology, Imperial College London, London, United Kingdom
9. Ludwig Boltzmann Institute for Rare and Undiagnosed Diseases, Vienna, Austria
10. CeMM Research Center for Molecular Medicine, Austrian Academy of Sciences, Vienna, Austria
11. Pediatric Hematology/Oncology Unit, Department of Pediatrics, Shaare Zedek Medical Center, Jerusalem, Israel
12. Blizard Institute, Queen Mary University London, London, United Kingdom.

13. Department of Cardiovascular Sciences, Center for Molecular and Vascular Biology, KU Leuven, Leuven, Belgium
14. Department of Blood Group Serology and Transfusion Medicine, Medical University of Vienna, Vienna, Austria
15. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, United Kingdom
16. MRC Biostatistics Unit, University of Cambridge, Cambridge Institute of Public Health, Cambridge Biomedical Campus, Cambridge, United Kingdom.
17. School of Cellular and Molecular Medicine, University of Bristol, Bristol, United Kingdom
18. Central Diagnostic Laboratory, Maastricht University Medical Centre, Maastricht, The Netherlands
19. Katharine Dormandy Haemophilia Centre and Thrombosis Unit, The Royal Free Hospital, United Kingdom
20. University College Medical School, London, United Kingdom

* Authors contributed equally to this work

Correspondence: Kate Downes, Department of Haematology, NHS Blood and Transplant, Long Road, Cambridge, CB20PT. email: kd286@cam.ac.uk

Recurrent abbreviations

ACMG	American College of Medical Genetics and Genomics
BTPD	Bleeding, Thrombotic and Platelet Disorders
MDT	Multi Disciplinary Team
HTS	High Throughput Sequencing
HPO	Human Phenotype Ontology
SNV	Single Nucleotide Variant
CNV	Copy Number Variant

Abstract

A targeted high-throughput sequencing (HTS) panel test for clinical diagnostics requires careful consideration of the inclusion of appropriate diagnostic-grade genes, the ability to detect multiple types of genomic variation with high levels of analytic sensitivity and reproducibility, and variant interpretation by a multi-disciplinary team (MDT) in the context of the clinical phenotype. We have sequenced 2,396 index patients using the ThromboGenomics HTS panel test of diagnostic-grade genes known to harbour variants associated with rare bleeding, thrombotic or platelet disorders (BTPD). The molecular diagnostic rate was determined by the clinical phenotype, with an overall rate of 49.2% for all thrombotic, coagulation, platelet count and function disorder patients and a rate of 3.2% for patients with unexplained bleeding disorders characterized by normal hemostasis test results. The MDT classified 745 unique variants, including copy number and intronic variants, as Pathogenic, Likely Pathogenic or Variants of Uncertain Significance. Half (50.9%) of these variants are novel and 41 unique variants were identified in 7 genes recently found to be implicated in BTPD. Inspection of canonical hemostasis pathways identified 29 patients with evidence of oligogenic inheritance. A molecular diagnosis has been reported for 894 index patients providing evidence that introducing an HTS genetic test is a valuable addition to laboratory diagnostics in patients with a high likelihood of having an inherited BTPD.

Introduction

Inherited bleeding, thrombotic and platelet disorders (BTPD) are a heterogeneous group of rare disorders caused by DNA variants in a large number of loci. The most common bleeding disorders are von Willebrand Disease, affecting up to 0.01% of the population, and Hemophilia A and B, which together affect 0.01% of males.¹ There are no accurate estimates of the prevalence of the remaining rare inherited bleeding disorders, although registry data suggest the prevalence being <0.001%.² Venous thrombosis has an overall annual incidence of less than 1 in 1,000, but it is rare in the pediatric population, with rates of approximately 1 in 100,000, indicative of possible environmental and lifestyle effects in adult patients.³ To obtain a conclusive molecular diagnosis requires attendance at multiple outpatient consultations for a large portion of patients with an assumed diagnosis of a rare inherited BTPD.

The genetic architecture of inherited BTPDs is well determined, but new genes continue to be identified. To date there are close to 100 diagnostic-grade genes (hereafter TIER1 genes) associated with coagulation, thrombotic and platelet disorders. Since validation of the ThromboGenomics HTS test⁴, 33 TIER1 genes, including recently discovered BTPD genes, have been added to the HTS test, increasing the clinical utility. Others have reported on similar gene panel tests or used whole exome sequencing to identify DNA variants causing inherited BTPDs⁵⁻¹⁰. Most studies used patients with uniform clinical phenotypes, but all on relatively small numbers of patients (<160 index patients), preventing firm conclusions about the clinical utility of such tests. The diagnostic rates obtained in these studies cannot be compared as all focused on different sets of genes (with a subset of TIER1 genes, and also inclusion of 'research' genes), used different patient inclusion criteria and variant classification was not standardized as is now recommended.¹¹

Here we report on the results obtained with the ThromboGenomics HTS test for 2,396 index patients categorized into five classes of disorders based on the appended human phenotype ontology (HPO) terms¹² and referral information; thrombotic, platelet

count, platelet function, coagulation and unexplained bleeding. Variant interpretation, by a multi-disciplinary team (MDT), determined the contribution of variants in TIER1 genes to the observed phenotypes thereby providing insights in the clinical utility of HTS testing for different categories of patients. We also comment on the standardization of variant interpretation and how the reporting of a conclusive molecular diagnosis has immediately impacted on clinical management. Finally, due to the large number of patients tested, we are able to highlight the clinical importance of detecting copy number and deep intronic variants and possible oligogenic inheritance.

Methods

Patients

The 2,396 index patients were either referrals for the ThromboGenomics test or patients who joined the PANE and VIBB studies (Supplemental Table 1). Clinical and laboratory phenotypes were recorded using HPO terms as described.^{4,13} Further details of the study participants, including institutional review board or research ethics committee information are in the Supplemental Information.

ThromboGenomics referrals for diagnostic testing of inherited BTPDs

Samples with clinical and laboratory phenotype information from 1,608 index patients with a known or suspected diagnosis of inherited BTPD, according to criteria described (Supplemental Information),⁴ were recruited by clinicians in 72 UK and 46 non-UK hospitals (Supplemental Figure 1).

PANE: Preoperative screening for mild bleeding risk

A total of 212 patients, identified through preoperative assessment of bleeding risk at Maastricht University Medical Centre, were recruited and underwent a full hematological assessment including extensive laboratory testing for hemostasis parameters (Supplemental Table 2).^{14,15}

VIBB: Vienna Bleeding Biobank

A total of 599 patients referred to the Hematology and Hemostaseology specialist tertiary referral centre in Vienna for assessment of a mild to moderate bleeding disorder were recruited and subjected to a full hematological assessment, including extensive laboratory testing for hemostasis parameters (Supplemental Table 2).¹⁶

ThromboGenomics HTS test

The ThromboGenomics HTS test sample preparation and sequencing protocols are as described with minor modifications (Supplemental Information).⁴ The content of the test has been reversioned twice since first described to include additional (mostly recently discovered) TIER1 genes. ThromboGenomics version 2 (TG.V2) and version 3 (TG.V3) include 80 and 96 genes, respectively (Supplemental Table 3). TIER1 genes are curated and approved by the Scientific and Standardization Committee on Genomics in Thrombosis and Hemostasis (SSC-GinTH) of the International Society on Thrombosis and Haemostasis (ISTH) (https://www.isth.org/page/GinTH_GeneLists). For TG.V3, probes for 10,000 common single nucleotide variants (SNVs) were included to estimate relatedness and ancestry. Automated bioinformatics analysis pipeline methods, including variant calling, are described in the Supplemental Information.

Variant prioritisation and interpretation

Variants were annotated and prioritized for interpretation using the analytical process as reported⁴ based on the predicted effect in the curated transcript, presence in the Human Gene Mutation Database¹⁷ or in a curated set of known pathogenic variants (Supplemental Information) and the minor allele frequency (MAF) in the Exome Aggregation Consortium (ExAC) and Genome Aggregation (gnomAD) databases.¹⁸ On a patient-by-patient basis, DNA variants passing filtering were prioritized and interpreted by a MDT in the context of the appended HPO terms, clinical information and family history. Reported variants were characterized as Pathogenic, Likely Pathogenic and Variants of Uncertain Significance alongside a decision of the likely contribution of each variant to the patients phenotype. The MDT made use of Congenica's diagnostic decision support platform Sapienia™ (Cambridge, UK) to support the review process and record findings

in the form of research reports for return to referring clinicians. For all samples sequenced using TG.V2, variant interpretation was performed according to guidelines agreed by the members of the ThromboGenomics MDT (criteria in Supplemental Table 4 and details of MDT process in Supplemental Information). In 2017, the UK Association of Clinical Genomic Science published best practice guidelines for variant interpretation based on the earlier reported American College of Medical Genetics and Genomics (ACMG) guidelines.¹¹ Sapienta software implemented ACMG guidelines from release 1.7 (January 2018), allowing rapid variant interpretation following these guidelines. This updated system was applied for variant interpretation of all samples sequenced using TG.V3. Table 1 summarizes the main differences in panel content, methods, analysis and interpretation used for TG.V2 and TG.V3.

Results

Patient inclusion criteria and phenotypes

A total of 2,396 index patients and 156 samples from relatives and carriers were tested using the ThromboGenomics HTS test (Supplemental Table 1). The largest group of 1,608 index patients and 18 referrals for hemophilia carrier status were referred by specialist tertiary centres for diagnostic testing. To better appreciate the clinical utility of the HTS test we also included 193 and 595 index patients from the PANE and VIBB single-centre studies.

Based on the clinical and laboratory phenotypes of all patients, a total of 7,341 HPO terms were appended. HPO terms and clinical information were used to categorize all patients into five broad disease classes; thrombotic (n=284), platelet count (n=335), platelet function (n=430), coagulation (n=728) and unexplained bleeding (n=619) (Figure 1A). The ThromboGenomics referrals included patients of all five classes, while the majority of the PANE (80.3%) and VIBB (59.5%) patients are classed as unexplained bleeding.

Most of the patients categorized to the thrombotic class were referred because of reduced Protein C or Protein S levels as indicated by the HPO term 'Abnormality of the

protein C anticoagulation pathway' (Figure 1B). Patients with platelet count disorders were generally referred due to (macro)thrombocytopenia. Platelet function abnormalities were diverse, including defects in aggregation, reduced platelet membrane protein expression (particularly GPIb/IX/V and GPIIb/IIIa) and reduced granule secretion. Coagulation defects included patients with reduced VWF levels or abnormal coagulation parameters of the intrinsic, extrinsic and common pathways. HPO terms coding abnormal phenotypes outside the blood system were appended to 41 (1.7%) of patients. These terms related to eyes (ocular albinism), hearing (deafness), skeletal (abnormal radius, joint disorders) and kidney (functional insufficiency). Across the three study groups 619 patients with bleeding symptoms and normal hemostasis test results were classed as unexplained bleeding (Figure 1B). Over half (51.2%) of the patients in this class were characterized by the presence of both spontaneous and trauma-related bleeding symptoms, a similar distribution as found for the coagulation and platelet function disease class (Supplemental Figure 2). Chronic mild thrombocytopenia is generally not accompanied by spontaneous bleeding.¹⁹ However, 65.1% of the patients in the platelet count class presented with spontaneous bleeding, strongly suggesting an enrichment of 'thrombocytopenia with bleeding' patients.

Performance of the ThromboGenomics HTS test

The previously reported validation of the ThromboGenomics HTS test (TG.V1) used 296 samples from patients, with and without previously known disease causing variants, sequenced for 63 TIER1 genes.⁴ Here we sequenced 1,333 and 1,063 index patient samples with the TG.V2 and TG.V3 tests targeting 80 and 96 TIER1 genes, with a region of interest of 0.222 Mb and 0.275 Mb respectively (Table 1). Since validation, the analysis method for calling SNVs and short (<50 base pair (bp))–insertion/deletions (indels) has undergone minor modifications; however, the detection of CNVs has been substantially improved.²⁰ In short, sequencing read depth is computed over 500 bp elements, to improve the sensitivity for the detection of shorter CNVs, and an optimized reference set of data obtained from 10 samples has been generated, using genetically unrelated individuals (Supplemental Information). Despite the increase of the region of interest and increased multiplexing of samples, the read coverage has remained high,

with 99.99% and 99.98% of the region of interest with a read depth exceeding 30x for TG.V2 and TG.V3 tests, respectively. For each sample, an average of 146.7 and 190.9 SNVs, 9.4 and 11.2 indels and 0.20 and 0.21 CNVs were identified by the TG.V2 and TG.V3 tests, respectively. The proportions of variant types identified did not differ between test versions (Supplemental Figure 3). For all samples, an average 4.5 variants were prioritized for interpretation by the MDT (Supplemental Table 1).

Diagnostic rates and validation of recently discovered TIER1 genes

Prioritized variants were reviewed by the MDT in the context of the disease incidence, variant frequency, assigned HPO terms and family history. Variants were reported with pathogenicity and the contribution to the patients phenotype (full or partial). Variants for recessive BTPDs were reported if present in the homozygous or compound heterozygous states. Variants of Uncertain Significance were reported if the MDT predicted a future Likely Pathogenic status with additional evidence from cosegregation and functional studies. Screening of 2,396 index patients resulted in an overall molecular diagnostic rate of 37.3% by reporting a total of 1,031 variants in 894 index patients (Figure 2A, Supplemental Table 5). Most reported variants (81.9%) were rare (<0.01%) or absent in gnomAD (Supplemental Figure 4). There was a marked difference in diagnostic yield between the five classes; thrombotic, 48.9%; platelet count 47.8%; platelet function 26.1%; coagulation disorders 63.6% and unexplained bleeding 3.2% (Figure 2A). For patients with thrombotic and coagulation disorders, 68.7% and 70.6% of reported variants were known variants that had previously been associated with disease, while for the patients with platelet count and function disorders this proportion was lower; 44.4% and 28.4% respectively (Figure 2B). We reason that this difference reflects the fact that cataloging of pathogenic variants for coagulation disorders (especially von Willebrand disease and Hemophilia A and B) began over three decades ago, whilst the majority of TIER1 genes for platelet disorders have only been identified over the past decade. Of the 335 patients within the platelet count class, 29 were referred under a working diagnosis of 'immune thrombocytopenia refractory to treatment'. In 7 of these patients, variants were reported in genes known to be associated with thrombocytopenia (*ANKRD26*, *ETV6*, *ITGA2B*, *TUBB1*).

The implementation of the criteria of the ACMG guidelines, instead of our 'in-house' criteria (Supplemental Table 4), impacted on variant interpretation for patients tested using TG.V3. On comparing the ThromboGenomics cohort samples sequenced and analyzed using either TG.V2 or TG.V3 (PANE and VIBB samples were sequenced using only TG.V2 and TG.V3 respectively), there was minimal difference in the total number of reported variants for each of the five classes of patients (Supplemental Figure 5A). However, for all disease classes, an interpretation shift from Likely Pathogenic to Variant of Uncertain Significance was noted (Supplemental Figure 5B). This change in variant interpretation was mainly explained by novel missense variants (Supplemental Figure 6). For patients tested using TG.V2, variants that were not known as disease associated were often deemed as Likely Pathogenic by the MDT. On introduction of the ACMG guidelines, these novel missense variants did not reach the threshold for designation as Likely Pathogenic and thus were labeled as Variants of Uncertain Significance. Prior to submission, all variants have been reanalyzed according to ACMG guidelines, using the latest versions of variant databases, including the ThromboGenomics database, and results from co-segregation studies (Supplementary Table 5, column 'Reinterpretation 2019').

The genes with reported variants in index patients are ranked in Figure 3. For the thrombotic, platelet count and coagulation disease classes, a quarter of patients had variants reported in just one gene, *PROS1*, *MYH9* and *VWF*, respectively. The fourth quarter of patients for each disease class had reported variants in at least 6 genes. All reported variants per patient are summarized in Supplemental Table 5 and have been submitted to the ClinVar database.²¹

Since TG.V1, 17 TIER1 genes associated with BTPD have been included in TG.V2 and a further 16 more TIER1 genes in TG.V3. In addition to introducing known and recently discovered genes, those with new modes of inheritance were added (Supplemental Table 3). Diagnostic reports for 41 patients have been issued with variants in one of the 19 recently discovered TIER1 genes or by applying a new mode of inheritance.

Copy number variation, deep intronic variants and oligogenic findings.

The analytical pipeline for the identification of CNVs has been modified with improved quality scoring and visualization tools. Overall, CNVs were reported in 40 patients, predicted to affect single exons (n=11), multiple exons (n=15) or whole genes (n=14) (Supplemental Table 5). Manual inspection of reads revealed the presence of complex CNVs, including an inversion with breakpoints in introns 26 and 27 of *DIAPH1* resulting in an in-frame deletion of exon 26, and an inversion flanked by two deletions within *F8* causing severe hemophilia A (Supplemental Figure 7).²²

Aside from the core dinucleotide splice sites at the 5' and 3' of introns, the lack of reliable prediction tools makes it difficult to determine the likely functional consequences of potential splicing altering variants. Outside the SnpEff annotated splice regions (8 bp), intronic variants were only prioritized if previously associated with disease. A deep intronic homozygous *ITGA2B* variant identified in one index patient with Glanzmann Thrombasthenia was validated using platelet RNA expression studies confirming alternative splicing and the absence of normal *ITGA2B* transcript (Supplemental Figure 8). These functional data, together with co-segregation analysis in the pedigree of the index patient, resulted in a reclassification of this *ITGA2B* variant from Variant of Uncertain Significance to Likely Pathogenic.

Of the 894 index patients where variants were reported, 772 had a single reported variant while 122 had at least two reported variants. Most patients with two variants have a recessive disease, but for 29 patients the reported variants were in two or more genes. For most of these examples, the variants identified were within first or second order interactors in the known canonical hemostasis pathways (Supplemental Table 6). For the thrombotic and coagulation classes, we identified 11 (3.9%) and 13 (1.8%) patients with oligogenic variants, respectively (Figure 4).

Incidental findings

By sequencing a large number of patients for the TIER1 genes underlying known BTPDs, we expected to observe incidental secondary findings. In four females, not

referred for hemophilia carrier testing or known to have reduced factor levels, we identified carriership of Likely Pathogenic variants in *F8* or *F9*. In addition, in two patients we identified a heterozygous deletion of the *RBM8A* gene. A heterozygous *RBM8A* loss-of function variant (generally a deletion), if accompanied by a low frequency non-coding regulatory variant on the alternate allele, results in thrombocytopenia with absent radius syndrome.²³ These incidental secondary findings were reported to the referring clinician as they are actionable with respect to family planning. In contrast, sex chromosome aneuploidy, identified in three patients, was not reported in line with local best practice guidelines.

Discussion

We evaluated the performance of a targeted HTS panel test for TIER1 genes in over 2,500 subjects drawn from three distinct groups; (1) patients with a high likelihood of having an inherited BTPD, (2) patients undergoing an extended preoperative assessment for bleeding risk and (3) patients with a bleeding disorder of unknown etiology referred to a tertiary referral centre. Using HPO coding of clinical and laboratory phenotypes alongside clinical information, patients were assigned to one of five diseases classes; thrombotic, platelet count, platelet function, coagulation or unexplained bleeding. DNA samples were sequenced with the ThromboGenomics HTS test and prioritized variants reviewed and classified by an MDT and reported to referring clinicians. The resulting data were used to assess the effectiveness of the HTS test, analytical pipeline and MDT variant interpretation in generating a conclusive molecular diagnosis. For 1,777 patients of the thrombotic, coagulation and platelet count or function disease class, variants were reported for half (49.2%). In contrast, variants were reported for 3.2% of the 619 index unexplained bleeding patients with normal hemostasis test results. Overall, 20.1% of reported variants were Variants of Uncertain Significance that require additional evidence including estimation of variant odds ratios using the results from large genotyped cohort studies, functional testing and cosegregation analysis. These data illustrate the diagnostic yield obtained when applying the ThromboGenomics test for patients with a high likelihood of having an inherited BTPD.

Screening patients with the TG.V2 and TG.V3 HTS test presented several improvements compared to the TG.V1 test. Adding more TIER1 genes on TG.V2 and TG.V3 supported the molecular diagnosis of 41 patients and important genotype-phenotype association were observed for recently discovered BTPD genes such as *DIAPH1*²², *ETV6*²⁴, *GFI1B*²⁵, *GNE*^{26,27} and *RASGRP2*²⁸ or for the alternative mode of inheritance recently reported for *GP1BB*.²⁹ In comparison to TG.V1, the detection of CNVs was optimized, resulting in the detection of CNVs in 40 patients, including previously unobserved deletions in 13 genes, indicating CNVs as an important variant class for all categories of BTPDs. In addition, 2 novel duplications, 1 novel inversion and a complex CNV were also reported. Nevertheless, the optimized ExomeDepth method is sensitive to variable read depths, has a minimum resolution and cannot detect inversions or predict the location of duplicated regions. The introduction of a split-read or read-pair CNV calling method, alongside the ExomeDepth analysis method, may further improve CNV detection in the future.

Previously, Sanger sequencing of most BTPD genes was performed with primer sets flanking intron/exon boundaries and therefore, the frequency of non-coding or silent variants, aside from those disrupting the immediate splice site (<+/- 8 bp from the exon boundary), that are associated with BTPDs is unknown. With the use of HTS, deep intronic variants have been identified and apparent silent variants in BTPD genes have been shown to alter splicing.³⁰⁻³³ Variants disrupting transcription regulatory motifs located in gene promoters and enhancer regions have also been associated with BTPDs, and with the recent mapping of endothelial and blood cell specific enhancers, it is likely that more variants located in these regions will be identified as associated with disease in the near future.³⁴⁻³⁶ Nevertheless, due to the challenges in interpretation of non-coding variants we have only targeted and prioritized intronic and regulatory variants if previously associated with disease. Therefore, a research analysis of novel deep intronic, silent and likely regulatory variants detected is required, and such studies are best performed using whole genome sequencing data from well characterized patient cohorts.

A diagnostic HTS platform requires careful selection of diagnostic-grade TIER1 genes. The decision to include a gene to the BTPD TIER1 list is made by the GinTH SSC of the ISTH. The designation of genes for diagnostic reporting involved the review of associated literature to evaluate if a gene is associated with disease in more than three independent pedigrees with convincing cosegregation data or in less than three pedigrees, but with strong functional evidence (mouse models or cell/protein studies) in addition to cosegregation data. In the future, this task will also be coordinated by the NIH supported Hemostasis/Thrombosis Clinical Domain Working Group in close partnership with the ISTH and ASH expert working groups.³⁷

Our results indicate that when screening large number of patients with a HTS test, half (50.1%) of unique reported variants are novel and following ACMG guidelines for TG.V3, the majority (76.1%) of novel unique missense variants were reported as Variants of Uncertain Significance (Supplemental Figure 6). International initiatives for sharing of sequence data generated for BTPD patients and the sequencing of the genomes or exomes of large prospective population cohorts, like the UK Biobank, the Million Veteran Program and the 100,000 Genomes Project will lead to statistically robust approaches for the functional labelling of DNA variants, including the Variants of Uncertain Significance reported in this study.³⁸⁻⁴⁰

These data provide evidence of how a molecular diagnosis influences the clinical management and counselling of patients and their close relatives. In 30 patients, variants were reported in the ANKRD26, ETV6 and RUNX1 genes leading to counseling and follow up due to the increased risk of hematological malignancies. These findings are associated with an important ethical debate regarding predictive testing in these genes⁴¹. It is therefore advisable to inform clinicians and patients about the presence of genes associated with malignancies in a panel and to provide the possibility of an opt out choice. As for whole genome or exome sequencing clinical testing strategies, virtual subpanels can be used to independently analyze genes only associated with bleeding, thrombotic or platelet disorders (with and without inclusion of leukemic risk genes). In 7 patients with treatment refractory immune thrombocytopenia evidence was obtained of rare germline

variants likely causing their condition. In six of these cases, the identification of a Variant of Uncertain Significance has prompted follow-up studies in the probands and their relatives to obtain additional evidence for pathogenicity. Finally, in 24 patients with thrombotic or coagulation disorders, we reported variants in two or more TIER1 genes. We postulate that defects in hemostasis are due to the disruption of two interacting proteins in the known canonical pathways. These results of a large single study, further extend the reports of possible oligogenic architecture for thrombotic and coagulation. Together these justify further functional and genetic follow-up studies to provide patients with a molecular diagnosis and better estimates of risk to offspring.

We report the results of the largest gene panel sequencing study of patients with suspected inherited BTPDs. We included 619 index patients with an unexplained bleeding disorder characterized by normal hemostasis test results, and as per our hypothesis, we observed a low diagnostic rate for this group. It is possible that unknown genes are responsible or that the propensity of bleeding in these patients is the result of the aggregation of a large number of small effects emanating from common variants at hundreds of loci which are modifying the overall effectiveness of the hemostasis system. Estimation of polygenic risk scores has recently been reported for several common diseases such as coronary artery disease, type 2 diabetes and breast cancer⁴²⁻⁴⁵.

In patients with a known or suspected disease etiology, a molecular diagnosis was not discovered in half (50.8%) of patients. Future versions of the ThromboGenomics panel will include additional genes as they are identified and designated TIER1 status. However, this is unlikely to significantly increase the diagnostic rate as hundreds of genes are predicted to be currently unknown. Others have shown that reanalysis of negative clinical exome sequencing data at 1-3 years increased the diagnostic rate by 10% due to new evidence for gene causality and candidate variants not available during the initial analysis.⁴⁶ Such evidences have emerged due to the curation of disease variants, by ClinVar and the Human Gene Mutation Database, and the availability of large exome and genome sequencing datasets (ExAC and gnomAD). To further increase diagnostic rates in the future, the sequencing of many more individuals, through large consortium projects,

will provide more accurate estimates of the rareness of potential pathogenic variants,⁴⁷ alongside the development of novel genotype-phenotype algorithms,⁴⁸ studies of gene-environment interactions and the building of polygenic risk scores.

In conclusion the ThromboGenomics test is a valuable addition to the diagnostic algorithm for patients with a high likelihood of having an inherited BTPD. The results provide clinicians with a molecular diagnosis for approximately half of patients allowing for more precise prognostication and management of disease and, with cascade testing, better informed counselling of patients and their close relatives.

Data availability

Variants and pathogenicity (Reinterpretation 2019) have been deposited in ClinVar.

Acknowledgments

This study makes use of data generated by the NIHR BioResource. A full list of investigators who contributed to the generation of the data is available from <https://bioresource.nihr.ac.uk/researchers/researchers/acknowledgement/>. We gratefully acknowledge the participation of all NIHR BioResource volunteers and thank the NIHR BioResource centre and staff for their contribution. Funding for the project was provided by the National Institute for Health Research (NIHR) under grant number RG65966. The authors would like to thank the Locus Reference Genomic team (LRG) for their assistance in curating gene transcripts. Research in the Ouwehand laboratory receives funding from the British Heart Foundation, European Commission (TrainMALTA), International Society on Thrombosis and Haemostasis, Medical Research Council, NHS Blood and Transplant and the Rosetrees Trust. The Vienna Bleeding Biobank was supported by an unrestricted grant of CSL Behring. K.D. is supported as a NHS HSST trainee by Health Education England. K.F. is supported by the Research Council of the University of Leuven (BOF KU Leuven, Belgium; OT/14/098) and by an unrestricted grant of SOBI.

Authorship

Contributions: D.D., J.S. and C.T. performed experiments; K.M., M.V., O.S., S.V.V.D., R.M., S.T., L.D., N.G., D.G., M.H., A.T., C.J.P., E.T. analysed data; J.G., S.H., M.W.B., N.C., S.P., S.R-V., S.S., E.S., W.T., I.S., Y.M.C.H. and I.P., collected samples and provided clinical support; N.A.H., H.M., and S.A. provided clinical support; K.D., W.H.O., M.A.L., A.D.M., K.G. and K.F. lead the ThromboGenomics MDT and designed the study and wrote the manuscript, which was reviewed by all remaining authors.

Conflict of Interest Disclosure: The authors have no conflict of interests.

References

1. Castaman G, Linari S. Diagnosis and Treatment of von Willebrand Disease and Rare Bleeding Disorders. *J Clin Med.* 2017;6(4).
2. UKHCDO. UKHCDO Annual Report. 2017.
3. Khan S, Dickerman JD. Hereditary thrombophilia. *Thromb J.* 2006;4:15.
4. 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.
5. Johnson B, Lowe GC, Futterer J, et al. Whole exome sequencing identifies genetic variants in inherited thrombocytopenia with secondary qualitative function defects. *Haematologica.* 2016;101(10):1170-1179.
6. Bastida JM, Lozano ML, Benito R, et al. Introducing high-throughput sequencing into mainstream genetic diagnosis practice in inherited platelet disorders. *Haematologica.* 2018;103(1):148-162.
7. Leinoe E, Zetterberg E, Kinalis S, et al. Application of whole-exome sequencing to direct the specific functional testing and diagnosis of rare inherited bleeding disorders in patients from the Oresund Region, Scandinavia. *Br J Haematol.* 2017;179(2):308-322.
8. Lee EJ, Dykas DJ, Leavitt AD, et al. Whole-exome sequencing in evaluation of patients with venous thromboembolism. *Blood Adv.* 2017;1(16):1224-1237.
9. Johnson B, Doak R, Allsup D, et al. A comprehensive targeted next-generation sequencing panel for genetic diagnosis of patients with suspected inherited thrombocytopenia. *Res Pract Thromb Haemost.* 2018;2(4):640-652.
10. Andres O, König E-M, Althaus K, et al. Use of Targeted High-Throughput Sequencing for Genetic Classification of Patients with Bleeding Diathesis and Suspected Platelet Disorder. Vol. 02; 2018.
11. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17(5):405-424.
12. Kohler S, Vasilevsky NA, Engelstad M, et al. The Human Phenotype Ontology in 2017. *Nucleic Acids Res.* 2017;45(D1):D865-D876.
13. Westbury SK, Turro E, Greene D, et al. Human phenotype ontology annotation and cluster analysis to unravel genetic defects in 707 cases with unexplained bleeding and platelet disorders. *Genome Med.* 2015;7(1):36.

14. 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.
15. Moenen F, Vries MJA, Nelemans PJ, et al. Screening for platelet function disorders with Multiplate and platelet function analyzer. *Platelets.* 2017;1-7.
16. 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.
17. 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.
18. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 2016;536(7616):285-291.
19. Noris P, Pecci A. Hereditary thrombocytopenias: a growing list of disorders. *Hematology Am Soc Hematol Educ Program.* 2017;2017(1):385-399.
20. Simeoni I, Shamardina O, Deevi SV, et al. GRID - Genomics of Rare Immune Disorders: a highly sensitive and specific diagnostic gene panel for patients with primary immunodeficiencies. *bioRxiv.* 2018.
21. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* 2018;46(D1):D1062-D1067.
22. Westbury SK, Downes K, Burney C, et al. Phenotype description and response to thrombopoietin receptor agonist in DIAPH1-related disorder. *Blood Adv.* 2018;2(18):2341-2346.
23. Albers CA, Paul DS, Schulze H, et al. Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. *Nat Genet.* 2012;44(4):435-439, S431-432.
24. Poggi M, Canault M, Favier M, et al. Germline variants in ETV6 underlie reduced platelet formation, platelet dysfunction and increased levels of circulating CD34+ progenitors. *Haematologica.* 2017;102(2):282-294.
25. Chen L, Kostadima M, Martens JHA, et al. Transcriptional diversity during lineage commitment of human blood progenitors. *Science.* 2014;345(6204):1251033.
26. Revel-Vilk S, Shai E, Turro E, et al. GNE variants causing autosomal recessive macrothrombocytopenia without associated muscle wasting. *Blood.* 2018;132(17):1851-1854.
27. Futterer J, Dalby A, Lowe GC, et al. Mutation in GNE is associated with severe congenital thrombocytopenia. *Blood.* 2018;132(17):1855-1858.
28. Westbury SK, Canault M, Greene D, et al. Expanded repertoire of RASGRP2 variants responsible for platelet dysfunction and severe bleeding. *Blood.* 2017;130(8):1026-1030.
29. Sivapalaratnam S, Westbury SK, Stephens JC, et al. Rare variants in GP1BB are responsible for autosomal dominant macrothrombocytopenia. *Blood.* 2017;129(4):520-524.
30. Inaba H, Shinozawa K, Amano K, Fukutake K. Identification of deep intronic individual variants in patients with hemophilia A by next-generation sequencing of the whole factor VIII gene. *Res Pract Thromb Haemost.* 2017;1(2):264-274.
31. Nuzzo F, Bulato C, Nielsen BI, et al. Characterization of an apparently synonymous F5 mutation causing aberrant splicing and factor V deficiency. *Haemophilia.* 2015;21(2):241-248.
32. Daidone V, Gallinaro L, Grazia Cattini M, et al. An apparently silent nucleotide substitution (c.7056C>T) in the von Willebrand factor gene is responsible for type 1 von Willebrand disease. *Haematologica.* 2011;96(6):881-887.
33. Xie J, Pabon D, Jayo A, Butta N, Gonzalez-Manchon C. Type I Glanzmann thrombasthenia caused by an apparently silent beta3 mutation that results in aberrant splicing and reduced beta3 mRNA. *Thromb Haemost.* 2005;93(5):897-903.

34. Pippucci T, Savoia A, Perrotta S, et al. Mutations in the 5' UTR of ANKRD26, the ankirin repeat domain 26 gene, cause an autosomal-dominant form of inherited thrombocytopenia, THC2. *Am J Hum Genet.* 2011;88(1):115-120.
35. Sabater-Lleal M, Chillón M, Howard TE, et al. Functional analysis of the genetic variability in the F7 gene promoter. *Atherosclerosis.* 2007;195(2):262-268.
36. Petersen R, Lambourne JJ, Javierre BM, et al. Platelet function is modified by common sequence variation in megakaryocyte super enhancers. *Nat Commun.* 2017;8:16058.
37. Rehm HL, Berg JS, Brooks LD, et al. ClinGen--the Clinical Genome Resource. *N Engl J Med.* 2015;372(23):2235-2242.
38. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018;562(7726):203-209.
39. Gaziano JM, Concato J, Brophy M, et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol.* 2016;70:214-223.
40. Turnbull C, Scott RH, Thomas E, et al. The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. *BMJ.* 2018;361:k1687.
41. Greinacher A, Eekels JJM. Diagnosis of hereditary platelet disorders in the era of next-generation sequencing: "primum non nocere". *J Thromb Haemost.* 2019.
42. Abraham G, Havulinna AS, Bhalala OG, et al. Genomic prediction of coronary heart disease. *Eur Heart J.* 2016;37(43):3267-3278.
43. Khera AV, Emdin CA, Drake I, et al. Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. *N Engl J Med.* 2016;375(24):2349-2358.
44. Fuchsberger C, Flannick J, Teslovich TM, et al. The genetic architecture of type 2 diabetes. *Nature.* 2016;536(7614):41-47.
45. Maas P, Barrdahl M, Joshi AD, et al. Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. *JAMA Oncol.* 2016;2(10):1295-1302.
46. Wenger AM, Gudur H, Bernstein JA, Bejerano G. Systematic reanalysis of clinical exome data yields additional diagnoses: implications for providers. *Genet Med.* 2017;19(2):209-214.
47. Wright CF, West B, Tuke M, et al. Assessing the Pathogenicity, Penetrance, and Expressivity of Putative Disease-Causing Variants in a Population Setting. *Am J Hum Genet.* 2019;104(2):275-286.
48. Costanzo M, Kuzmin E, van Leeuwen J, et al. Global Genetic Networks and the Genotype-to-Phenotype Relationship. *Cell.* 2019;177(1):85-100.

Table 1

The ThromboGenomics TG.V2 and TG.V3 high throughput sequencing test content, experimental methods, analysis methods and variant interpretation guidelines.

	TG.V2	TG.V3
Panel content	19 Coagulation genes 8 Thrombotic genes 53 Platelet genes HGMD Pro 2015.2 variants	21 Coagulation genes 9 Thrombotic genes 66 Platelet genes HGMD Pro 2016.4 variants 10,000 SNVs
Region of interest	0.222 Mb	0.275 Mb
Methods	DNA fragmentation – 220 bp Multiplex – 48 samples	DNA fragmentation – 350 bp Multiplex – 96 samples
Known variants	HGMD_PRO_2017.2	HGMD_PRO_2017.4 Curated known variants
Variant interpretation	MDT criteria (Supplemental Table 4)	ACMG guidelines ¹⁰

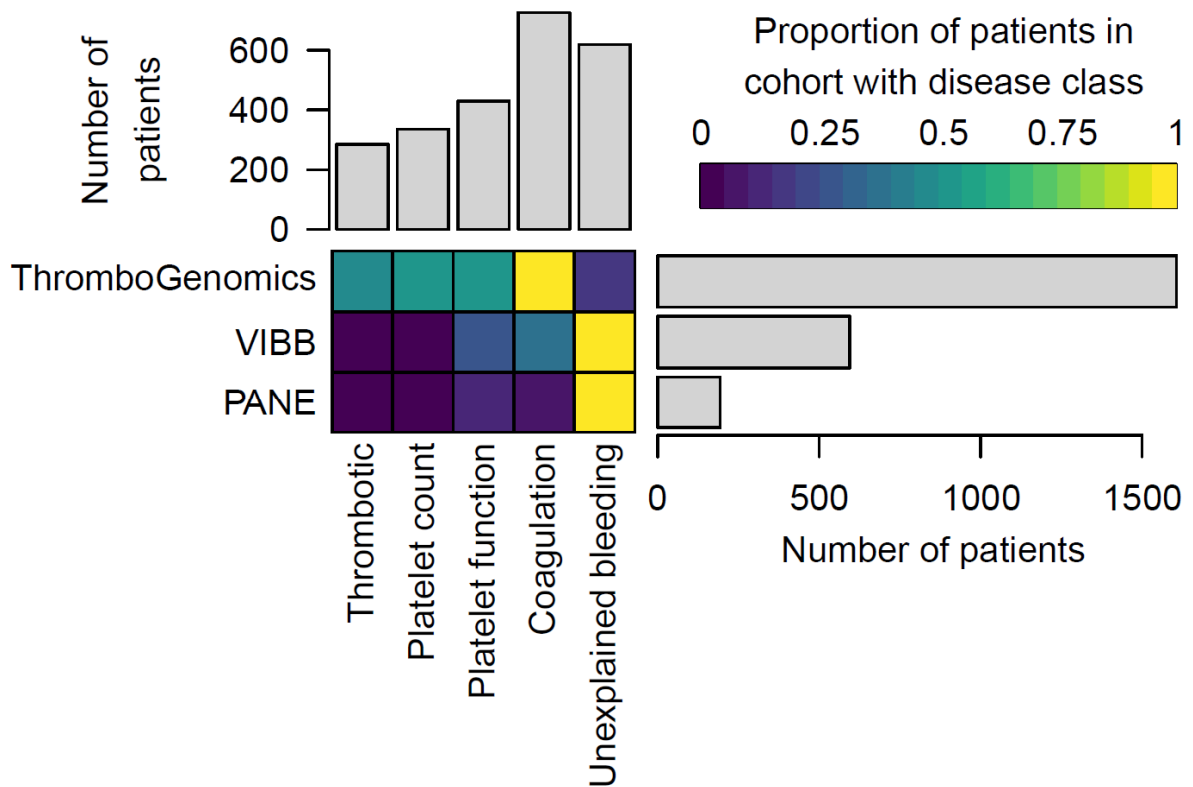
Figure 1

Classification of patients using clinical and laboratory phenotypes.

(A) Classification of 2,396 index patients from the ThromboGenomics, VIBB and PANE cohorts into one of five disease classes; thrombotic, platelet count, platelet function, coagulation and unexplained bleeding.

(B) Representative HPO codes for patients characterized in each of the five disease classes.

A



B

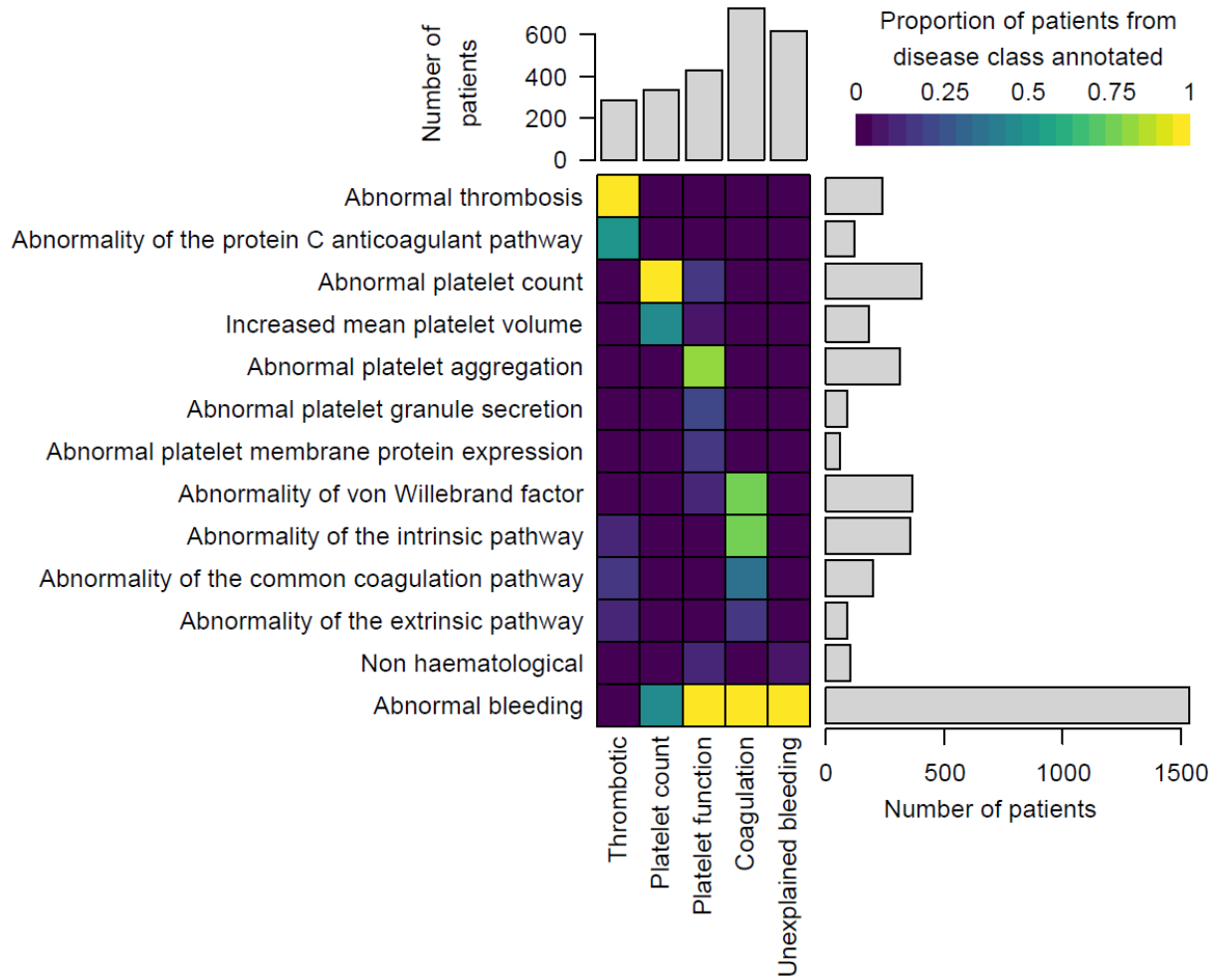


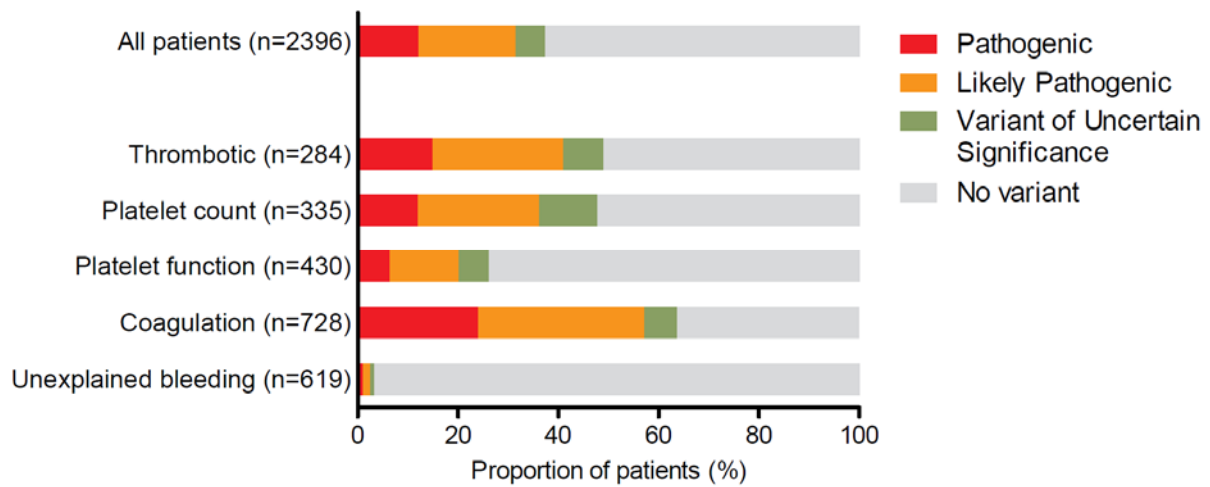
Figure 2

Diagnostic yield and proportion of novel variants by disease class.

(A) Diagnostic yield of reported variants for 2,396 index patients for each of the five disease classes; thrombotic, platelet count, platelet function, coagulation and unexplained bleeding. For patients with more than one reported variant, the most pathogenic variant was used in this analysis (n = number of index patients).

(B) Proportion of reported variants that were novel or known for patients in each disease class (n = number of variants).

A



B

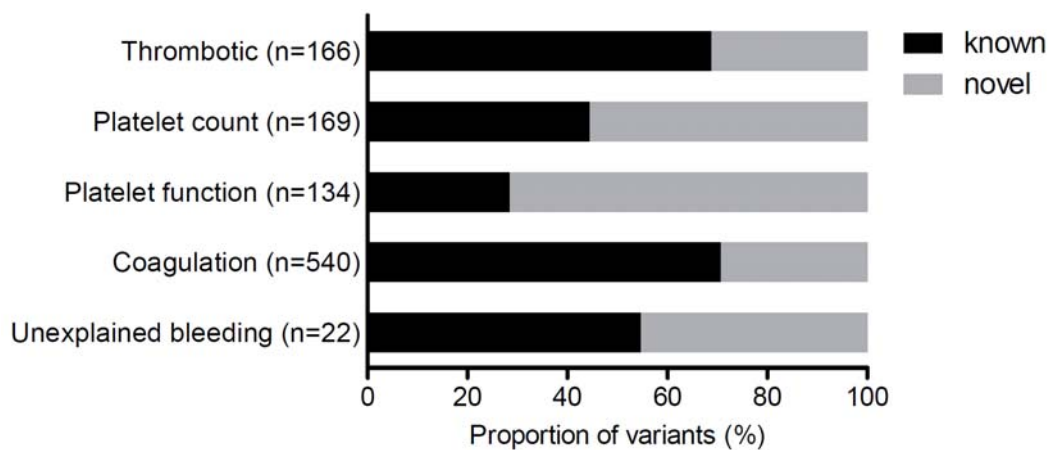


Figure 3

Gene ranking according to the number of reports per disease class.

For each disease class, genes were ranked according to number of times they were reported. Dashed lines represents, from top to bottom, the 25, 50 and 75 quantiles. Recently discovered genes and changes of mode of inheritance are in blue.

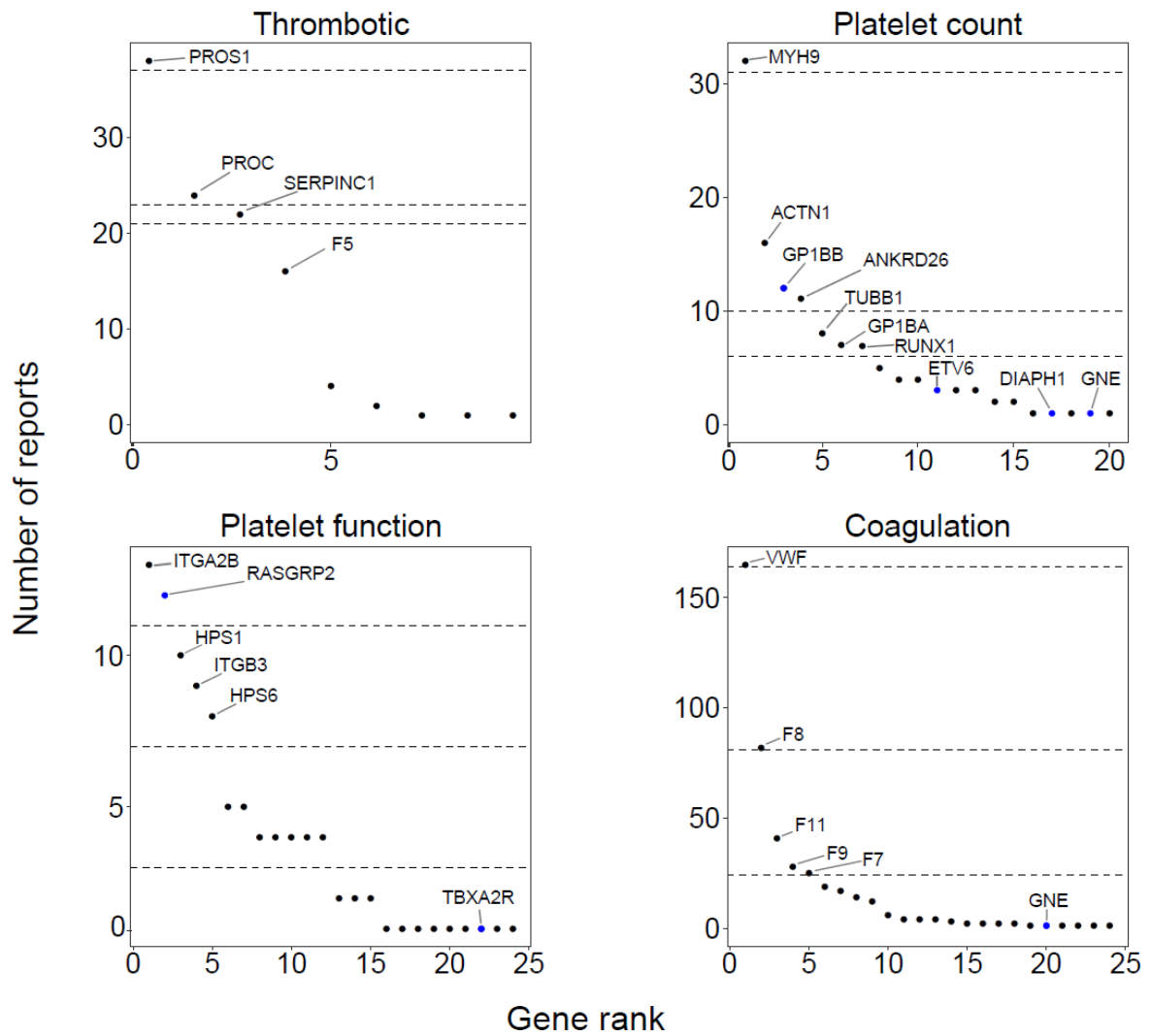


Figure 4
Oligogenic variants in patients with thrombotic (red) and coagulation (blue) disorders.

From outside to inside, track 1: amino acid numbering for thrombotic (red) and coagulation (blue) genes; lighter shade denote untranslated regions of the F2 gene and the 5' of the PROS1 gene; track 2: amino acid conservation scores; track 3: variant frequency in gnomAD (minor allele frequency normalized scale to 1/106); track 4: disease causing (red) and questionable disease causing (orange) Human Gene Mutation Database variants; track 5 and arcs: reported variants in the 11 thrombotic (red) and 13 coagulation (blue) patients with the arcs representing oligogenic findings.

