American Society of Hematology
2021 L Street NW, Suite 900,
Washington, DC 20036
Phone: 202-776-0544 | Fax 202-776-0545
editorial@hematology.org

The effects of pathogenic variants for inherited hemostasis disorders in 140,214 UK Biobank participants

Tracking no: BLD-2023-020118R2 - CORRECTION

Luca Stefanucci (National Health Service (NHS) Blood and Transplant, Cambridge Biomedical, Campus, Cambridge, UK, United Kingdom) Janine Collins (University of Cambridge, United Kingdom) Matthew Sims (Sheffield Teaching Hospitals NHS Foundation Trust, United Kingdom) Iñigo Barrio-Hernandez (European Molecular Biology Laboratory, European Bioinformatics Institute, United Kingdom) Luanluan Sun (BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, United Kingdom) Oliver Burren (Centre for Genomics Research, Discovery Sciences, BioPharmaceuticals R&D, United Kingdom) Livia Perfetto (Sapienza University of Rome, Italy) Isobel Bender (University of Oxford, United Kingdom) Tiffany Callahan (Department of Biomedical Informatics, United States) Kathryn Fleming (School of Cellular and Molecular Medicine, United Kingdom) Jose Guerrero (University of Cambridge, United Kingdom) Henning Hermjakob (European Molecular Biology Laboratory, European Bioinformatics Institute, United Kingdom) Maria Martin (European Molecular Biology Laboratory, European Bioinformatics Institute, United Kingdom) James Stephenson (European Molecular Biology Laboratory, European Bioinformatics Institute, United Kingdom) Kalpana Paneerselvam (European Molecular Biology Laboratory, European Bioinformatics Institute, United Kingdom) Slavé Petrovski (AstraZeneca, United Kingdom) Pablo Porras (European Molecular Biology Laboratory, European Bioinformatics Institute, United Kingdom) Peter Robinson (,) Quanli Wang (Centre for Genomics Research, Discovery Sciences, BioPharmaceuticals R&D, United Kingdom) Xavier Watkins (European Molecular Biology Laboratory, European Bioinformatics Institute, United Kingdom) Mattia Frontini (University of Exeter, United Kingdom) Roman Laskowski (European Molecular Biology Laboratory, European Bioinformatics Institute, United Kingdom) Pedro Beltrao (Institute of Molecular Systems Biology,, Switzerland) Emanuele Di Angelantonio (University of Cambridge, United Kingdom) Keith Gomez (Royal Free London NHS Foundation Trust, United Kingdom) Michael Laffan (Imperial College, United Kingdom) Willem Ouwehand (University of Cambridge, United Kingdom) Andrew Mumford (University of Bristol, UK, United Kingdom) Kathleen Freson (KULeuven, Belgium) Keren Carss (AstraZeneca, United Kingdom) Kate Downes (University of Cambridge, United Kingdom) Nicholas Gleadall (University of Cambridge, United Kingdom) Karyn Megy (University of Cambridge, United Kingdom) Elspeth Bruford (University of Cambridge, United Kingdom) Dragana Vuckovic (Department of Epidemiology and Biostatistics, United Kingdom)

Abstract:

Rare genetic diseases affect millions, and identifying causal DNA variants is essential for patient care. Therefore, it is imperative to estimate the effect of each independent variant and improve their pathogenicity classification. Our study of 140,214 unrelated UK Biobank (UKB) participants found each carries a median of 7 variants previously reported as pathogenic or likely pathogenic. We focused on 967 diagnostic-grade genes (DGGs) variants for rare bleeding, thrombotic, and platelet disorders (BTPDs) observed in 12,367 UKB participants. By association analysis, for a subset of these variants, we estimated effect sizes for platelet count and volume, and odds ratios for bleeding and thrombosis. Variants causal of some autosomal recessive platelet disorders revealed phenotypic consequences in carriers. Loss-of-function variants in *MPL*, which cause chronic amegakaryocytic thrombocytopenia if biallelic, were unexpectedly associated with increased platelet counts in carriers. We also demonstrated that common variants identified by genome-wide association studies (GWAS) for platelet count or thrombosis risk may influence the penetrance of rare variants in BTPD DGGs on their associated hemostasis disorders. Network-propagation analysis applied to an interactome of 18,410 nodes and 571,917 edges showed that GWAS variants with large effect sizes are enriched in DGGs and their first-order interactors. Finally, we illustrate the modifying effect of polygenic scores for platelet count and thrombosis risk on disease severity in participants carrying rare variants in *TUBB1*, or *PROC* and *PROS1*, respectively. Our findings demonstrate the power of association analyses using large population datasets in improving pathogenicity classifications of rare variants.

Conflict of interest: COI declared - see note

COI notes: O.B., Q.W., K.C., P.P., K.M. and S.P. are current employees and/or stockholders of AstraZeneca.

Preprint server: No;

Author contributions and disclosures: L.Stefanucci, J.C. and M.C.S. wrote the manuscript, analyzed data, participated in MDTs and oversight analysis pipeline. I.B-H., L.Sun, O.B., L.P., I.B., N.G., T.J.C., R.A.L. and P.B analyzed data and oversight analysis pipeline. J.A.G., H.H., K.P, S.P., P.P., Q.W., K.C., X.W., E.d.A and M.J.M. oversight analysis pipeline. M.F. reviewed the manuscript. K.G., M.L. and K.D. participated in MDTs. A.D.M. and K.F. reviewed the manuscript and participated in MDTs. W.H.O. conceptualized the study and reviewed the manuscript. K.M. project administration, data curation, wrote the manuscript and participated in MDTs. E.B. wrote the manuscript, conceptualized the study, project administration and data curation. D.V. wrote the manuscript, analyzed data and participated in MDT.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: Genotype and phenotypes data are accessible at UK Biobank (https://www.ukbiobank.ac.uk/) and require an active project and application.

Clinical trial registration information (if any):

Figure 1

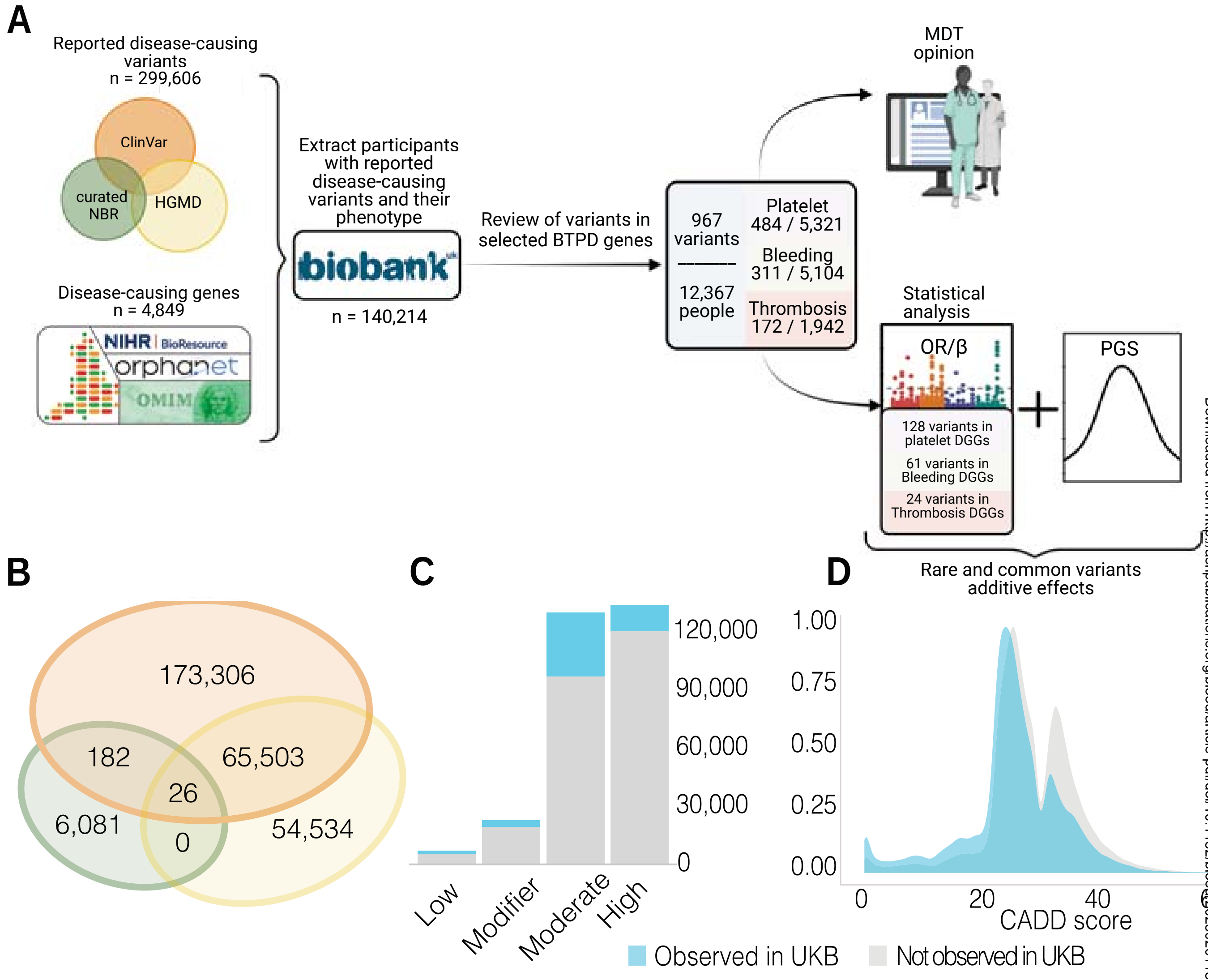
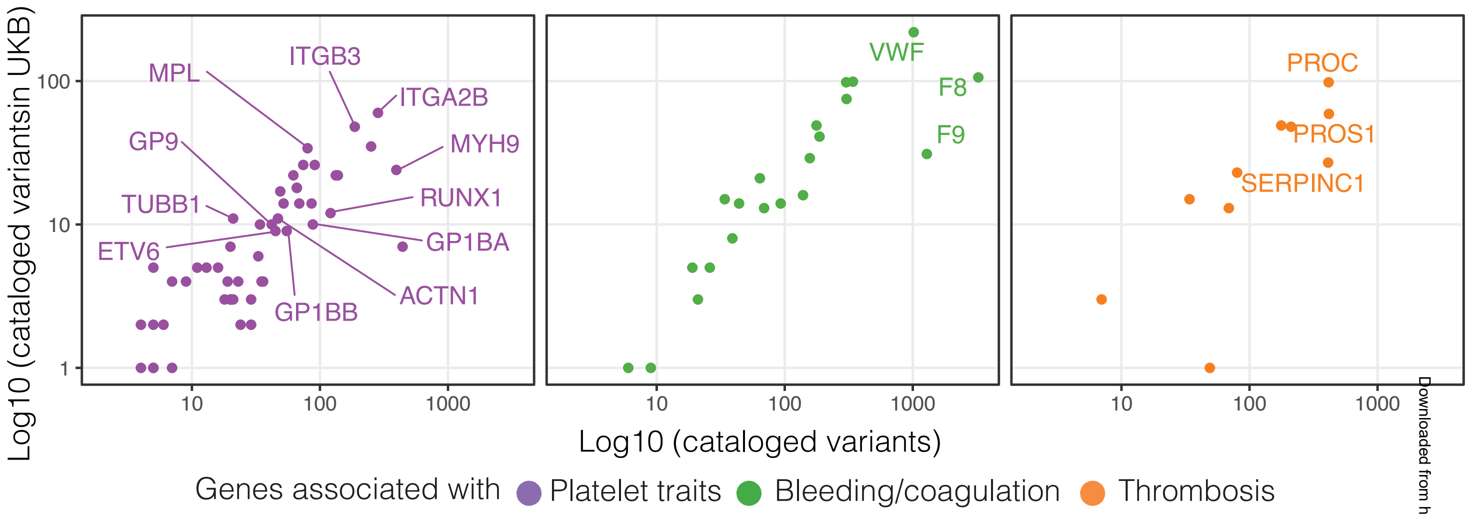

A
Reported disease-causing variants
n = 299,606
ClinVar
curated NBR
HGMD
Disease-causing genes
n = 4,849
NIHR | BioResource
orphanet
OMIM
Extract participants with reported disease-causing variants and their phenotype
biobank
n = 140,214
Review of variants in selected BTPD genes
967 variants
12,367 people
Platelet 484 / 5,321
Bleeding 311 / 5,104
Thrombosis 172 / 1,942
MDT opinion
Statistical analysis
OR/β
128 variants in platelet DGGs
61 variants in Bleeding DGGs
24 variants in Thrombosis DGGs
+
PGS
Rare and common variants additive effects
B
173,306
182
65,503
26
6,081
0
54,534
C
120,000
90,000
60,000
30,000
0
Low
Modifier
Moderate
High
D
1.00
0.75
0.50
0.25
0.00
0
20
40
60
CADD score
Observed in UKB
Not observed in UKB

Figure 2



Log10 (cataloged variantsin UKB)
100
10
1
10
100
1000
Log10 (cataloged variants)
MPL
ITGB3
ITGA2B
GP9
MYH9
TUBB1
RUNX1
ETV6
GP1BA
ACTN1
GP1BB
VWF
F8
F9
PROC
PROS1
SERPINC1
Genes associated with
● Platelet traits
● Bleeding/coagulation
● Thrombosis

Figure 3

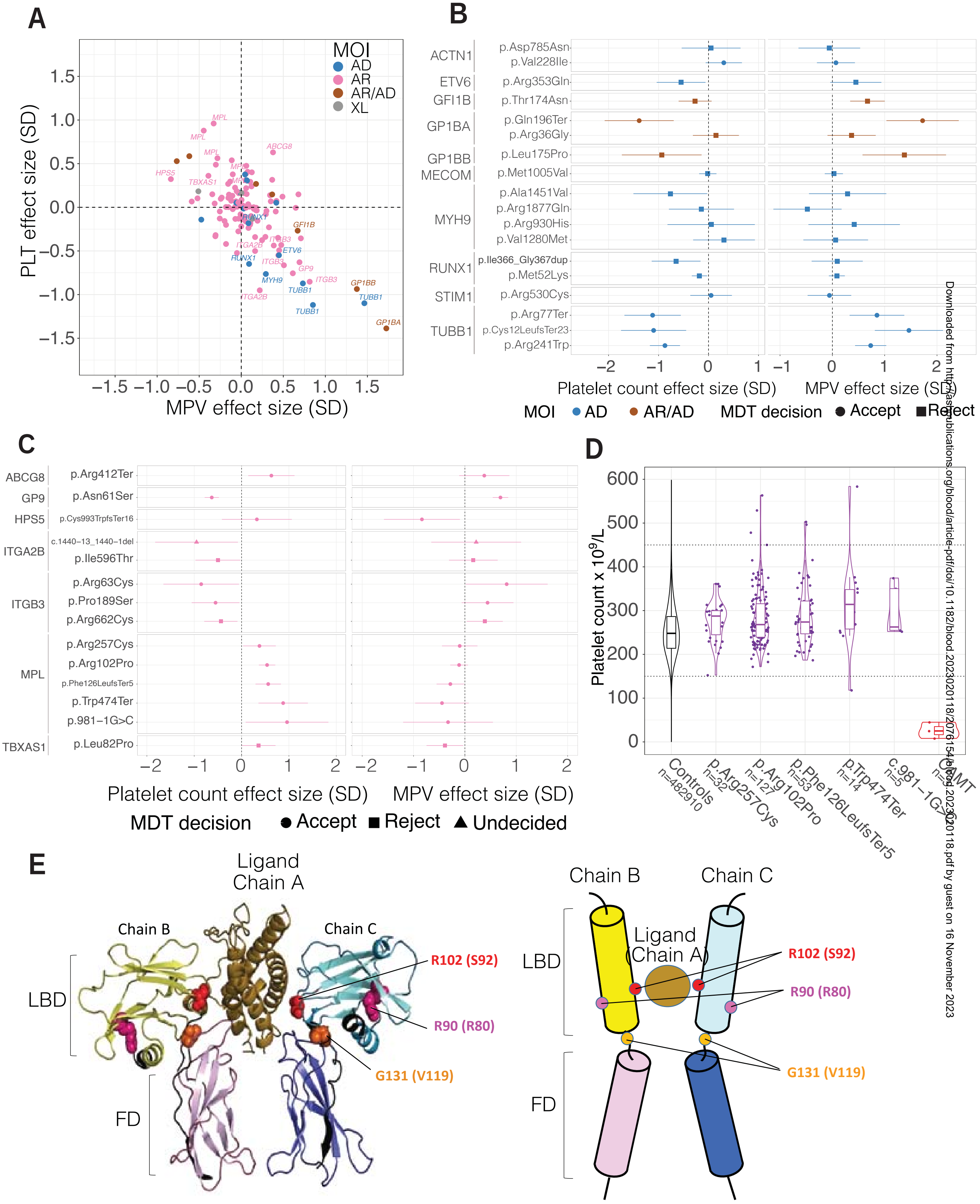

A
MOI
AD
AR
AR/AD
XL
PLT effect size (SD)
MPV effect size (SD)
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.5 −1.0 −0.5 0.0 0.5 1.0 1.5
MPL
ABCG8
HPS5
TBXAS1
RUNX1
GFI1B
ITGB3
ITGA2B
ETV6
GP9
MYH9
TUBB1
GP1BB
GP1BA
B
ACTN1
p.Asp785Asn
p.Val228Ile
ETV6
p.Arg353Gln
GFI1B
p.Thr174Asn
GP1BA
p.Gln196Ter
p.Arg36Gly
GP1BB
p.Leu175Pro
MECOM
p.Met1005Val
MYH9
p.Ala1451Val
p.Arg1877Gln
p.Arg930His
p.Val1280Met
RUNX1
p.Ile366_Gly367dup
p.Met52Lys
STIM1
p.Arg530Cys
TUBB1
p.Arg77Ter
p.Cys12LeufsTer23
p.Arg241Trp
−2 −1 0 1
−1 0 1 2
Platelet count effect size (SD)
MPV effect size (SD)
MOI ● AD ● AR/AD
MDT decision ● Accept ■ Reject
C
ABCG8
p.Arg412Ter
GP9
p.Asn61Ser
HPS5
p.Cys993TrpfsTer16
ITGA2B
c.1440−13_1440−1del
p.Ile596Thr
ITGB3
p.Arg63Cys
p.Pro189Ser
p.Arg662Cys
MPL
p.Arg257Cys
p.Arg102Pro
p.Phe126LeufsTer5
p.Trp474Ter
p.981−1G>C
TBXAS1
p.Leu82Pro
−2 −1 0 1 2
−2 −1 0 1 2
Platelet count effect size (SD)
MPV effect size (SD)
MDT decision ● Accept ■ Reject ▲ Undecided
D
Platelet count x 10^9/L
600
500
400
300
200
100
0
Controls n=482910
p.Arg257Cys n=32
p.Arg102Pro n=127
p.Phe126LeufsTer5 n=53
p.Trp474Ter n=14
c.981−1G>C n=5
CAMT n=9
E
Ligand Chain A
Chain B
Chain C
LBD
FD
R102 (S92)
R90 (R80)
G131 (V119)
Chain B
Chain C
Ligand (Chain A)
LBD
FD
R102 (S92)
R90 (R80)
G131 (V119)

Figure 4

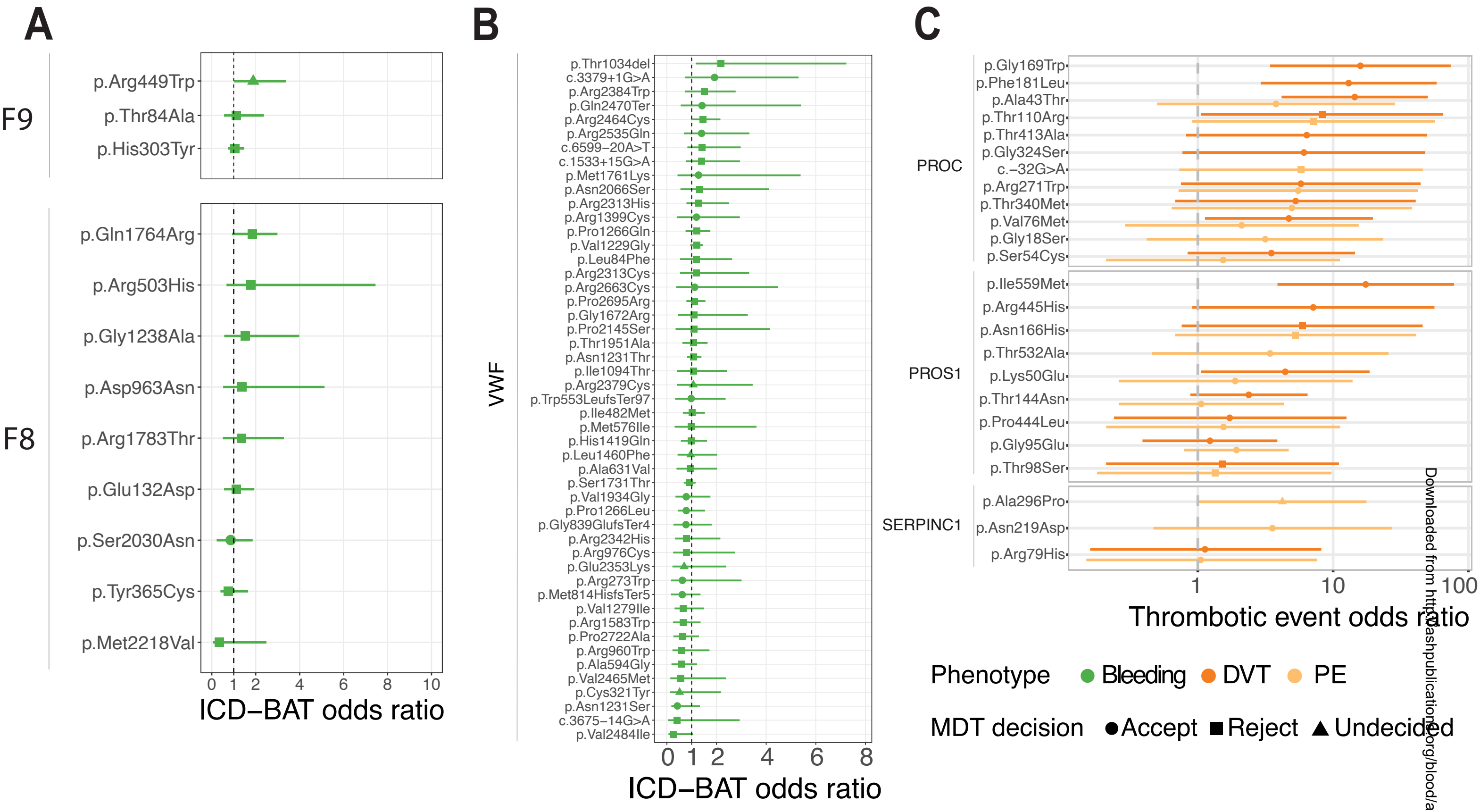
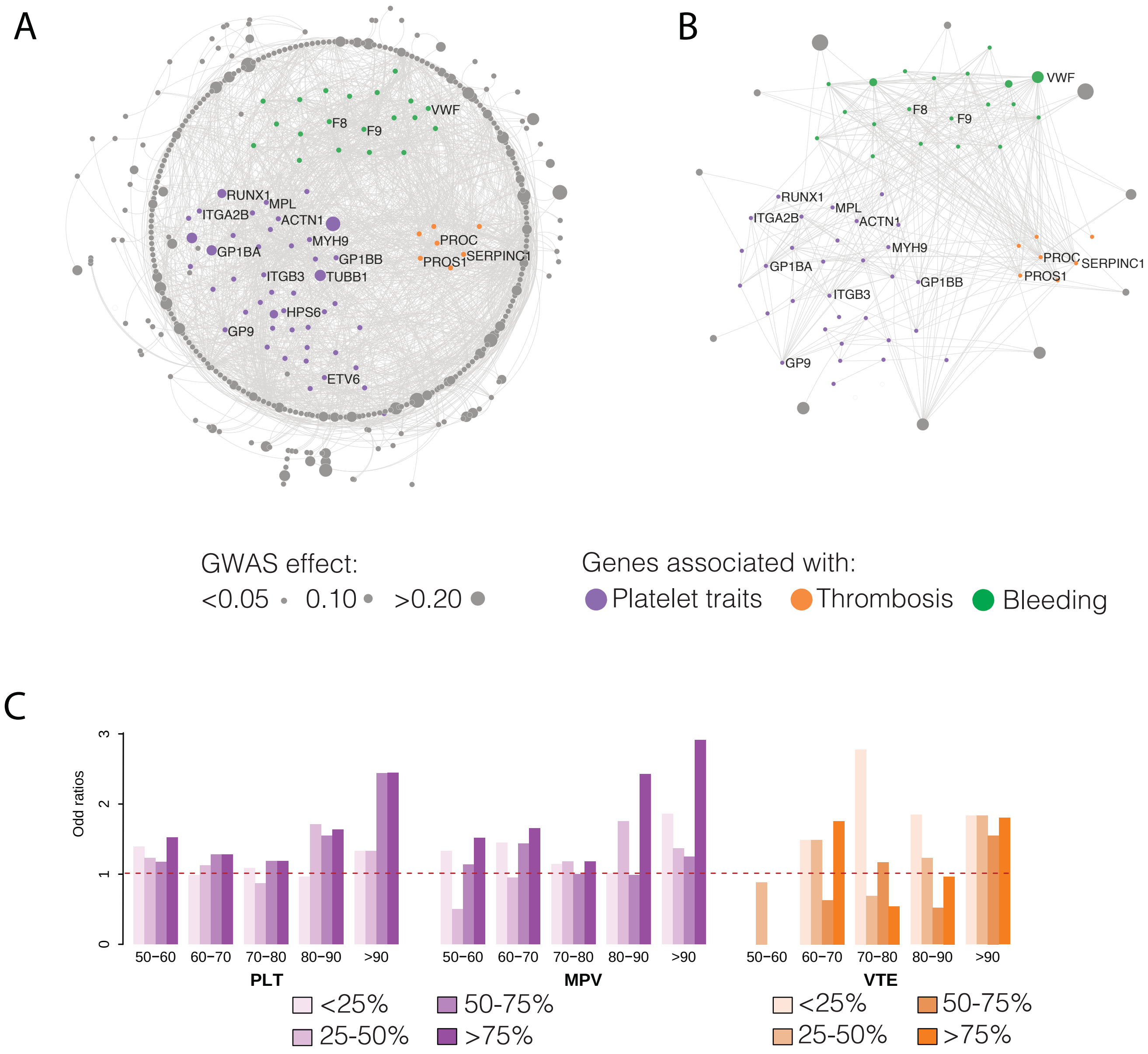
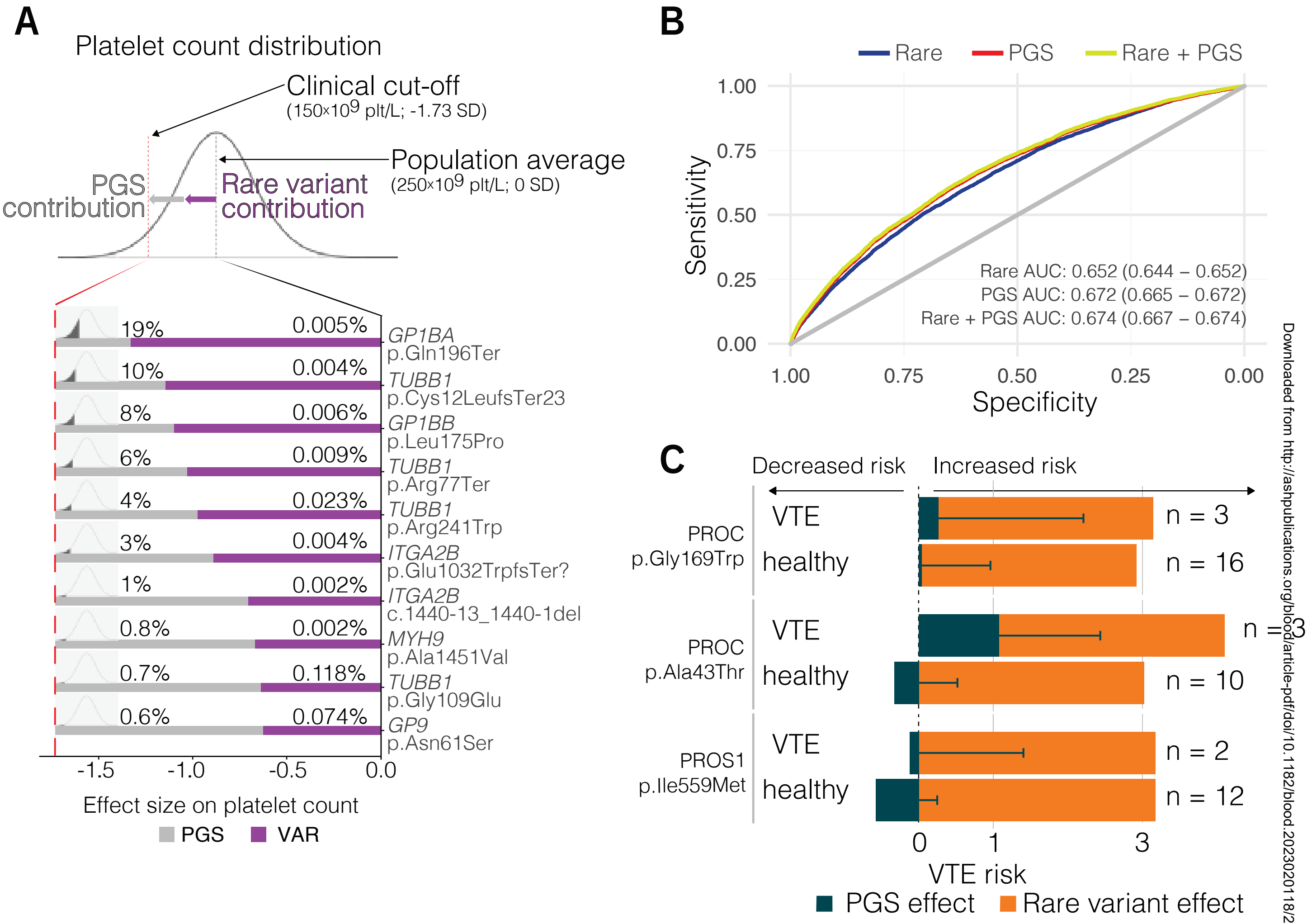

A
F9
p.Arg449Trp
p.Thr84Ala
p.His303Tyr
F8
p.Gln1764Arg
p.Arg503His
p.Gly1238Ala
p.Asp963Asn
p.Arg1783Thr
p.Glu132Asp
p.Ser2030Asn
p.Tyr365Cys
p.Met2218Val
0 1 2 4 6 8 10
ICD−BAT odds ratio
B
VWF
p.Thr1034del
c.3379+1G>A
p.Arg2384Trp
p.Gln2470Ter
p.Arg2464Cys
p.Arg2535Gln
c.6599−20A>T
c.1533+15G>A
p.Met1761Lys
p.Asn2066Ser
p.Arg2313His
p.Arg1399Cys
p.Pro1266Gln
p.Val1229Gly
p.Leu84Phe
p.Arg2313Cys
p.Arg2663Cys
p.Pro2695Arg
p.Gly1672Arg
p.Pro2145Ser
p.Thr1951Ala
p.Asn1231Thr
p.Ile1094Thr
p.Arg2379Cys
p.Trp553LeufsTer97
p.Ile482Met
p.Met576Ile
p.His1419Gln
p.Leu1460Phe
p.Ala631Val
p.Ser1731Thr
p.Val1934Gly
p.Pro1266Leu
p.Gly839GlufsTer4
p.Arg2342His
p.Arg976Cys
p.Glu2353Lys
p.Arg273Trp
p.Met814HisfsTer5
p.Val1279Ile
p.Arg1583Trp
p.Pro2722Ala
p.Arg960Trp
p.Ala594Gly
p.Val2465Met
p.Cys321Tyr
p.Asn1231Ser
c.3675−14G>A
p.Val2484Ile
0 1 2 4 6 8
ICD−BAT odds ratio
C
PROC
p.Gly169Trp
p.Phe181Leu
p.Ala43Thr
p.Thr110Arg
p.Thr413Ala
p.Gly324Ser
c.−32G>A
p.Arg271Trp
p.Thr340Met
p.Val76Met
p.Gly18Ser
p.Ser54Cys
PROS1
p.Ile559Met
p.Arg445His
p.Asn166His
p.Thr532Ala
p.Lys50Glu
p.Thr144Asn
p.Pro444Leu
p.Gly95Glu
p.Thr98Ser
SERPINC1
p.Ala296Pro
p.Asn219Asp
p.Arg79His
1 10 100
Thrombotic event odds ratio
Phenotype ● Bleeding ● DVT ● PE
MDT decision ● Accept ■ Reject ▲ Undecided

Figure 5



A
B
C
VWF
F8
F9
RUNX1
MPL
ITGA2B
ACTN1
MYH9
GP1BA
GP1BB
ITGB3
TUBB1
HPS6
GP9
ETV6
PROC
PROS1
SERPINC1
GWAS effect:
<0.05 • 0.10 ● >0.20 ●
Genes associated with:
● Platelet traits ● Thrombosis ● Bleeding
Odd ratios
0
1
2
3
50−60
60−70
70−80
80−90
>90
PLT
MPV
VTE
<25%
25-50%
50-75%
>75%

Figure 6


A
Platelet count distribution
Clinical cut-off
(150×10^9 plt/L; -1.73 SD)
Population average
(250×10^9 plt/L; 0 SD)
PGS contribution
Rare variant contribution
19% 0.005% GP1BA p.Gln196Ter
10% 0.004% TUBB1 p.Cys12LeufsTer23
8% 0.006% GP1BB p.Leu175Pro
6% 0.009% TUBB1 p.Arg77Ter
4% 0.023% TUBB1 p.Arg241Trp
3% 0.004% ITGA2B p.Glu1032TrpfsTer?
1% 0.002% ITGA2B c.1440-13_1440-1del
0.8% 0.002% MYH9 p.Ala1451Val
0.7% 0.118% TUBB1 p.Gly109Glu
0.6% 0.074% GP9 p.Asn61Ser
-1.5
-1.0
-0.5
0.0
Effect size on platelet count
PGS
VAR
B
Rare
PGS
Rare + PGS
1.00
0.75
0.50
0.25
0.00
Sensitivity
1.00
0.75
0.50
0.25
0.00
Specificity
Rare AUC: 0.652 (0.644 – 0.652)
PGS AUC: 0.672 (0.665 – 0.672)
Rare + PGS AUC: 0.674 (0.667 – 0.674)
C
Decreased risk
Increased risk
PROC p.Gly169Trp
VTE
n = 3
healthy
n = 16
PROC p.Ala43Thr
VTE
n = 3
healthy
n = 10
PROS1 p.Ile559Met
VTE
n = 2
healthy
n = 12
0
1
3
VTE risk
PGS effect
Rare variant effect

The effects of pathogenic and likely pathogenic variants for inherited hemostasis disorders in 140,214 UK Biobank participants

Luca Stefanucci[1,2,3,*], Janine Collins[1,2,4,*], Matthew C Sims[1,2,5,†], Iñigo Barrio-Hernandez[6,†], Luanluan Sun[7,†], Oliver Burren[8], Livia Perfetto[6,9], Isobel Bender[10], Tiffany J Callahan[11], Kathryn Fleming[12], Jose A Guerrero[13,2,4], Henning Hermjakob[6,14], Maria J Martin[6], James Stephenson[6], NIHR Bioresource, Kalpana Paneerselvam[6], Slavé Petrovski[15,16], Pablo Porras[6,‡], Peter N Robinson[17,18], Quanli Wang[8], Xavier Watkins[6], Mattia Frontini[1,2,3,19], Roman A Laskowski[6], Pedro Beltrao[20], Emanuele Di Angelantonio[7,3,21,22,23,24], Keith Gomez[25], Mike Laffan[26,27], Willem H Ouwehand[1,2,28], Andrew D Mumford[12], Kathleen Freson[29], Keren Carss[8], Kate Downes[1,2,30], Nick Gleadall[1,2], Karyn Megy[1,‡], Elspeth Bruford[6,1,#], Dragana Vuckovic[31,#]

[1]Department of Haematology, University of Cambridge, Cambridge Biomedical Campus,, Cambridge, UK, [2]National Health Service (NHS) Blood and Transplant, Cambridge Biomedical Campus, Cambridge, UK, [3]British Heart Foundation, BHF Centre of Research Excellence, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK, [4]Department of Haematology, Barts Health NHS Trust, London, UK, [5]Oxford Haemophilia and Thrombosis Centre, Oxford University Hospitals NHS Foundation Trust, NIHR Oxford Biomedical Research Centre, Oxford, UK, [6]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Cambridge, UK, [7]BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK, [8]Centre for Genomics Research, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK, [9]Department of Biology and Biotechnology "C.Darwin", Sapienza University of Rome, Rome, Italy, [10]Department of Biochemistry, University of Oxford, Oxford Science Area, Oxford, UK, [11]Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, New York, USA, [12]School of Cellular and Molecular Medicine, University of Bristol, Bristol, UK, [13]MRC Toxicology Unit, University of Cambridge, Cambridge Biomedical Campus,, Cambridge, UK, [14]National Center for Protein Sciences Beijing, Beijing Institute of Life Omics, Beijing, China, [15]Centre for Genomics Research, Discovery Sciences, Discovery Sciences, AstraZeneca, Cambridge, UK, [16]Department of Medicine, Austin Health, University of Melbourne, Melbourne, Australia, [17]Genomic Medicine, The Jackson Laboratory, Farmington, CT, USA,

[18]Institute for Systems Genomics, University of Connecticut, Farmington, CT, USA, [19]Department of Clinical and Biomedical Sciences, University of Exeter Medical School, Faculty of Health and Life Sciences RILD Building, Barrack Road, Exeter, UK, [20]Institute of Molecular Systems Biology, ETH Zürich, 8093, Zürich, Switzerland, [21]Heart and Lung Research Institute, University of Cambridge, Cambridge, UK, [22]NIHR Blood and Transplant Research Unit in Donor Health and Behaviour, Cambridge, UK, [23]Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge, UK, [24]Health Data Science Centre, Human Technopole, Milan, Italy, [25]Haemophilia Centre and Thrombosis Unit, Royal Free London NHS Foundation Trust, London, UK, [26]Department of Haematology, Imperial College Healthcare NHS Trust, London, UK, [27]Centre for Haematology, Department of Immunology and Inflammation, Imperial College London, London, UK, [28]Department of Haematology, University College London Hospitals NHS Trust, London, UK, [29]Department of Cardiovascular Sciences, Center for Molecular and Vascular Biology, KULeuven, Leuven, Belgium, [30]Cambridge Genomics Laboratory, Cambridge University Hospitals National Health Service Foundation Trust, Cambridge Biomedical Campus, Cambridge, UK, [31]Department of Epidemiology and Biostatistics, Imperial College London, London, UK

* Co-first authors

† Contributed equally

# Shared last authors

≠ Current affiliation: Centre for Genomics Research, Discovery Sciences,

BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK

Corresponding author: Dr. Dragana Vuckovic, d.vuckovic@imperial.ac.uk

Text word count: 4,378
Abstract word count: 250
Number of figures and tables: 6
Number of references: 83

Short title for the running head

Inherited hemostasis disorder variants in UKB

Key points

1. Rare variants causal of recessive hemostasis disorders have clinical consequences in carriers
2. Common variants modify these consequences and are one of the reasons for different phenotypic expressivity

Abstract

Rare genetic diseases affect millions, and identifying causal DNA variants is essential for patient care. Therefore, it is imperative to estimate the effect of each independent variant and improve their pathogenicity classification. Our study of 140,214 unrelated UK Biobank (UKB) participants found each carries a median of 7 variants previously reported as pathogenic or likely pathogenic. We focused on 967 diagnostic-grade genes (DGGs) variants for rare bleeding, thrombotic, and platelet disorders (BTPDs) observed in 12,367 UKB participants. By association analysis, for a subset of these variants, we estimated effect sizes for platelet count and volume, and odds ratios for bleeding and thrombosis. Variants causal of some autosomal recessive platelet disorders revealed phenotypic consequences in carriers. Loss-of-function variants in *MPL*, which cause chronic amegakaryocytic thrombocytopenia if biallelic, were unexpectedly associated with increased platelet counts in carriers. We also demonstrated that common variants identified by genome-wide association studies (GWAS) for platelet count or thrombosis risk may influence the penetrance of rare variants in BTPD DGGs on their associated hemostasis disorders. Network-propagation analysis applied to an interactome of 18,410 nodes and 571,917 edges showed that GWAS variants with large effect sizes are enriched in DGGs and their first-order interactors. Finally, we illustrate the modifying effect of polygenic scores for platelet count and thrombosis risk on disease severity in participants carrying rare variants in *TUBB1*, or *PROC* and *PROS1*, respectively. Our findings demonstrate the power of association analyses using large population datasets in improving pathogenicity classifications of rare variants.

Introduction

More than 9,000 rare diseases have been described, affecting over 400 million people worldwide.[1] High-throughput sequencing (HTS) has enabled the resolution of the genetic etiology of over 50% of rare diseases.[2] However, pathogenic variant identification for many suspected inherited diseases, including hemostasis disorders, remains challenging — in part, because there are often no reliable metrics to distinguish between loss- or gain- of function variants (LoF, GoF, respectively).[3] Moreover, individuals carry many pathogenic variants without any obvious clinical sequelae, indicating either incomplete penetrance or incorrect variant classification.[4–6]

To improve classification of candidate variants for rare diseases, most diagnostic laboratories adopt standardized reporting practices where pathogenicity evidence is considered by a multidisciplinary team (MDT), using knowledge from variant catalogs (e.g. ClinVar, HGMD) and the ACMG guidelines.[7–12] Variant pathogenicity classification within these catalogs is primarily based on published evidence, for which there are several important constraints. Firstly, most studies of rare inherited disorders are based on few pedigrees or genetically independent cases.[2] Secondly, reliable information on the minor allele frequency (MAF) remains inadequate for many variants, especially for non-European ancestry populations.[2,13–16] Thirdly, genetic admixture remains a significant cause of inflation for variant pathogenicity.[17–19] Finally, the predicted consequence of non-synonymous single nucleotide variants (SNV) on protein function, inferred from *in vitro* models or structural studies, may not reflect human physiological processes.[20–23]

Challenges to reliable variant classification adversely impact reporting in all rare diseases, and are well illustrated by hemostasis disorders. Clinical and laboratory phenotypes are well-characterized, but systematic variant reporting in large cohorts of patients with bleeding, thrombotic and platelet disorders (BTPDs) such as the NIHR rare disease BioResource has yielded unequivocal identification of pathogenic or likely pathogenic variants in almost 50% of cases.[24] Initiatives like ClinGen aim to improve the accuracy of pathogenicity assignment through application of refined and disease-specific ACMG/AMP rules; this immense manual curation task has so far been completed for three BTPD DGGs: *RUNX1* and the Glanzmann Thrombasthenia (GT) genes *ITGA2B* and *ITGB3*.[25–28]

GWAS offers an additional approach to understanding the genetic architecture of BTPDs, as effect size estimates for thousands of variants with MAF ≥ 0.1% are now available (GWAS-variants hereafter). These GWAS studies include complete blood count (CBC) parameters, such as platelet count and mean platelet volume (MPV), and have also identified hundreds of variants conferring risk for venous thromboembolism (VTE).[29–31] However, using imputed genotypes reduces the power to determine the effect size of rare variants when compared to direct genotyping.[15,31–33] With the release of whole exome sequencing (WES) genotypes for UKB participants, accurate rare variant counts became available for use in association studies.[15,32,34,35]

Using a GWAS-like statistical framework and electronic health record (EHR) data, we calculated the clinical associations of rare variants in DGGs for inherited BTPDs, including over 100 variants for autosomal recessive (AR) platelet disorders, to improve the current knowledge about carrier phenotypes.[36–43] Additionally, using an interactome of 18,410 proteins and 571,917 interactions, we illustrated how the interplay between rare variants and hundreds of GWAS-variants explains, at least partially, the variable penetrance of rare variants.[2,24,44] Ultimately, these findings narrow the distinction between dominant and recessive modes of inheritance (MOI), suggesting variant effects are additive, and highlight how statistical-genomic approaches can be used to improve variant classification in clinical genetic reporting.[45]

Methods

Rare disease gene list

We compiled a list of 4,849 genes implicated in rare Mendelian diseases, including 93 BTPD DGGs (supplemental Methods, supplemental Table1).

Catalog of pathogenic and likely pathogenic variants

A list of 299,606 previously reported pathogenic and likely pathogenic variants (hereafter cataloged-variants), including SNVs and insertion/deletions <50 bp (indels), was compiled from: (i) ClinVar “pathogenic” or “likely pathogenic” variants, that are not also “benign“ or “likely benign” (in cases of conflict of interpretation); (ii) Human Gene Mutation Database (HGMD) Pro 2019.4, disease-causing (DM) or probable/possible disease-causing (DM?) variants; (iii) NBR curated variants for BTPDs and European Association for Haemophilia and Allied Disorders (EAHAD) resources.[2,24,46] The NBR curated variants that mapped to GRCh37 were re-mapped to GRCh38 using AssemblyConverter (https://www.ensembl.org/Homo_sapiens/Tools/AssemblyConverter, Ensembl v.100). Using the Ensembl Variant Effect Predictor (VEP; ensembl-vep: 100.2), we extracted the impact, transcript effect and the Combined Annotation-Depletion Dependent (CADD) scores for each variant.[47,48]

The UKB cohort

UKB is a prospective cohort of 500,000 British individuals aged 40-69 years when recruited between 2006 and 2010.[49,50] Genotype and EHR data were accessed under UKB application 13745. Variant call data derived from WES results of 200,000 participants were downloaded from data-field 23151 (https://biobank.ndph.ox.ac.uk/ukb/field.cgi?id=23151; 11/2020) and EHR data from “Category 2002” (03/2021). Quality control and filtering steps of UKB genotypes are described in supplemental Methods. The benefit of our analysis relies on extensive manual curation and interpretation of pathogenic or likely pathogenic variants. This began after the release of the 200,000 WES, limiting the analyses to the observed variants in this subset. After checking a subset of variants in the entire cohort, we confirmed that the expanded cohort did not affect variant interpretation or the manuscript message.

Curation of variants by MDTs

Three MDTs, each comprising a clinician, geneticist and bioinformatician, manually curated the rare variants observed in UKB in 1) platelet, 2) thrombosis (*PROC*/*PROS1*/*SERPINC1*), and 3) bleeding/coagulation (*VWF*/*F8*/*F9*) DGGs. MDTs were blinded to EHR data and review was performed on a variant-by-variant basis. After considering factors such as the MAF compared to the prevalence of the associated disorder, each variant received a decision of “accept”, “undecided”, or “reject” as pathogenic or likely pathogenic (supplemental Table3**;** supplemental Methods).

Statistical analyses

Single-variant linear and logistic regression analyses were used to calculate effect sizes (in standard deviations=SD) and odds ratios (ORs) for relevant phenotypes in unrelated UKB participants of European ancestry (n=131,022). Phenotype and EHR selection are described in supplemental Methods. Analyses of the continuous traits platelet count and MPV were carried out for BTPD-variants in platelet disorder genes with at least 5 carriers (i.e. participants with the variant under analysis), after adjusting for relevant covariates and genomic principal components, and excluding participants with major blood disorders (supplemental Methods).[29] For calculating ORs for bleeding and VTE, BTPD-variants in *VWF/F8/F9* and *PROC/PROS1/SERPINC1* respectively were included, if present in at least 10 carriers (for *F8/F9*, females only), with at least one recorded event. Covariates used in the OR estimate are in supplemental Methods. Nominal *P*-values are reported, with significance at $P<5 \times 10^{-2}$. Our decision was guided by the fact that we do not present a new discovery analysis here, instead testing previously reported pathogenic and likely pathogenic variants in what is methodologically more similar to a replication analysis.

Interactome and polygenic scores

We generated a protein-protein interaction network by combining STRING (v11.0, score >0.75) with the interactome developed by the Open Targets project (www.opentargets.org; version November 2019)[51,52], a human-focused compilation of physical interactions (IntAct; https://www.ebi.ac.uk/intact/home) with causal relationships and pathways (Reactome https://reactome.org/; SIGNOR https://signor.uniroma2.it/).[53–55] All nodes were mapped to Ensembl Gene Identifiers, and duplicated edges and self-loops removed. The network propagation analysis is described in supplemental Methods. Polygenic score for platelet count was calculated as previously validated (PGS Catalog PGP000078); the method to estimate additive effect sizes is explained in supplemental Methods.[31]

Data sharing statement

Genotype and phenotype data are accessible at UK Biobank (https://www.ukbiobank.ac.uk/) and require an active project and application. The Data analysis scripts will be shared by contacting the corresponding author.

Results

UKB participants carry rare pathogenic and likely pathogenic variants

Of our 299,606 cataloged-variants, less than a quarter (n=65,503) were in both ClinVar and HGMD (Figure1A-B), highlighting their differing deposition strategies (supplemental Methods). The majority of cataloged-variants (89.6%) had high or moderate VEP-classified impact (Figure1C) with high CADD scores (median PHRED CADD=24.9, Figure1D).[48] A total of 82,415 cataloged-variants were observed in at least one of the 140,214 unrelated UKB participants; each participant had a median of 7 variants (interquartile range 5-10) (supplemental Figure1). These 82,415 variants were located in 4,150 (85.6%) of the 4,849 rare disease genes; they were significantly depleted of high-impact variants (Figure1C; Fisher's exact test, OR=0.342, $P<2.2\text{x}10^{-16}$) and had lower CADD scores than the 217,191 cataloged-variants not observed in our study population (Figure1D; Kolmogorov-Smirnov Test, D=0.137, $P<2.2\text{x}10^{-16}$; supplemental Methods).

We next performed a detailed analysis of the 12,765 cataloged-variants in DGGs for BTPDs, a subset of rare diseases in which we have thoroughly characterized the phenotypic features and genetic architecture (supplemental Figure2)[2,24,56]. There was a positive correlation between the number of variants observed in UKB participants and the number of cataloged-variants in BTPD genes (Pearson's correlation, estimate=0.642, $P=2.76\text{x}10^{-11}$; Figure2A-C). Following variant filtering (supplemental Methods), 1,465 rare variants in 79 of the 93 BTPD DGGs (hereafter BTPD-variants) were observed in 18,300 (13.1%) of the 140,214 unrelated UKB participants (supplemental Table3). Similar to our observations for all cataloged-variants, these BTPD-variants were depleted of high CADD score variants compared to all 12,765 cataloged-variants in BTPD DGGs (Kolmogorov-Smirnov test, D=0.153, $P <2.2\text{x}10^{-16}$; supplemental Figure3).

MDTs considered the pathogenicity likelihood for 967 of the 1,465 BTPD-variants, comprising all BTPD-variants in platelet disorder DGGs (n=484), and those in *F8/F9/VWF* (n=311) and *PROC/PROS1/SERPINC1* (n=172), the most commonly represented DGGs for the coagulation and thrombotic disorders, respectively. In 12,367 UKB participants, 967 BTPD-variants were observed (supplemental Figure2): 12,129 (98.1%) were heterozygous, 205 were males carrying a variant in an X-linked gene (*F8*, *F9*, *FLNA*, *WAS*), and 33 had a variant in homozygosity. The MDTs accepted the pathogenic or likely pathogenic label for 67% of BTPD-variants. The main reason for rejection of a pathogenic label was a MAF in any of the main ancestries of UKB participants that was incompatible with the prevalence of the disorder (supplemental Figure2,4).

Variant effect sizes on platelet count and volume

There was a negative correlation between platelet count and MPV, as expected, in the 3,359 UKB carriers of 128 platelet disorder variants included in the association analysis (Figure3A).[29,31] We detected significant associations for 24 variants (P<0.05) with effect sizes ranging from -1.4 to +1.0 SD, equating to a change in platelet count of -83 to +59 x10^9/L, and MPV effect sizes ranging from -0.8 to +1.7 SD (-0.9 to +1.8 fL) (Figure3A, supplemental Table4).

Eighteen of the 128 analyzed variants were in DGGs implicated in autosomal dominant (AD) thrombocytopenia disorders. Ten had significant effects on platelet count and/or volume in carriers, including 3 variants in *TUBB1*, 2 in *RUNX1*, one in *ETV6,* and one in *MYH9* (Figure3B). Rare variants in *GP1BA* and *GP1BB* cause both AD macrothrombocytopenia and AR Bernard Soulier Syndrome (BSS).[36,37,57,58] There were no UKB carriers of BSS-variants in homozygosity or compound heterozygosity. The *GP1BA* premature stop p.Gln196Ter had the largest effect sizes on both platelet count (□=-1.4, P=8.3x10^{-5}) and volume (□=1.7 SD, P=1.0x10^{-6}) in heterozygotes, equating to an average reduction in count of 82x10^9/L and 1.8 fL increase in MPV (Figure3B, supplemental Figure5). We detected effect sizes >0.9 or >0.7 SD for platelet count with 5 or 10 carriers, respectively. Therefore, the effect size for the remaining 8 variants in AD thrombocytopenia disorder genes was either too modest to be detected or, contrary to their pathogenicity labels, these do not significantly affect platelet count or volume (Figure3B).

Interestingly, we also observed significant effects on platelet count and/or volume for 14 variants causal of AR platelet disorders (Figure3A,C). First we confirmed our previous finding that carriers of *GP9* p.Asn61Ser had a reduced platelet count (Figure3C).[31] Monoallelic variants in *GP9* are not currently deemed causal for AD macrothrombocytopenia, however our association analysis identified variants in all 3 BSS DGGs (*GP1BA, GP1BB, GP9*) that reduce the count and increase the volume of platelets in carriers. The impact of these BSS-variants was, however, generally insufficient to diagnose macrothrombocytopenia (supplemental Figure5).

We performed a similar association analysis for 13 LoF variants in the genes for GT (GT-variants) that were heterozygous in 148 individuals in our study population (no participants had >1 GT-variant).[59,60] Patients with GT typically have normal platelet counts, therefore it was interesting to observe 3 variants in *ITGB3* and 2 in *ITGA2B* with significant effects on platelet count in carriers (□ range -0.4 to -1.0), with an average reduction in the range 26-

$56x10^9$/L (Figure3C, supplemental Figure6*)*. These LoF variants are distinct from the GoF variants in *ITGB3* and *ITGA2B*.[61,62]

Our analysis also revealed significant effects on platelet count for 5 out of 13 monoallelic LoF variants in *MPL*, which in homozygosity or compound heterozygosity cause congenital amegakaryocytic thrombocytopenia (CAMT), a disorder characterized by profound thrombocytopenia and progression to aplastic anemia (Figure3C).[2,63,64] These 5 CAMT-variants were collectively carried by 274 UKB participants. Unexpectedly, and in sharp contrast to the reduction in counts of carriers of some BSS- and GT-variants, we observed increased platelet counts, with effect sizes between 0.4 and 1.0 SD (Figure3C), equating to an average increase of 22-$57x10^9$/L (Figure3D). This resulted in thrombocytosis (platelet count >$450x10^9$/L) for 7 carriers. For 4 of these 5 CAMT-variants, the association with increased platelet count was corroborated in an extended analysis of 383,000 UKB participants (supplemental Methods). This analysis also revealed an association between increased platelet count and heterozygosity for an additional 8 CAMT-variants (supplemental Figure7). Of the 17 CAMT-variants on which we reported the outcome of association analysis, 4 were premature stops, 4 splice-site variants and 9 missense variants. Structural data was available for 8 missense variants and assuming that carrying one CAMT-variant allows expression of the mutant MPL receptor, we predicted that 5 of the 6 amino acid changes, which significantly increased platelet count in carriers, have functional consequences (Figure3E, supplemental Figure7). None of the CAMT-variants reported on were GoF variants, which confer an increased risk of myeloproliferative disorders.[2,63,64]

Subsequently, we replicated the relevant platelet count variants in the full UKB cohort and observed a good agreement between the effect sizes (Pearson R=0.71, $P=6x10^{-6}$; supplemental Figure 8).

The risk of bleeding and venous thromboembolism due to rare BTPD variants

Bleeding is a more heterogeneous and less-standardized phenotype than CBC-measured platelet count and volume. Therefore, to assess the association between BTPD-variants and bleeding, we used International Statistical Classification of Diseases and Related Health Problems 10th revision (ICD-10) codes to capture hospital episodes associated with bleeding over 23.5 years. An additive ICD-BAT score was developed, indicating the number of bleeding episodes across 19 domains (supplemental Methods; supplemental Table2; supplemental Figure9). We first investigated female carriers of *F8* or *F9* variants, to resolve uncertainty about the extent to which these cause abnormal bleeding.[65,66] Of 12 variants amenable to analysis, one (*F9* p.Arg449Trp) significantly increased the risk for a higher ICD-

BAT score (OR=1.89, P=4x10^{-2}, Figure4A). For *VWF* variants, 1,151 male and 1,302 female carriers were analyzed together. In contrast to *F8/F9* there was no clear directionality in the ORs for bleeding, excepting the inframe indel p.Thr1034del which was associated with increased bleeding in 21 carriers, none of whom had a second VWD-variant (OR=2.17, P=1x10^{-2}; Figure4B). Ten (47.6%) of the p.Thr1034del carriers, which since the MDT has been reported to cause AR type 3 VWD, presented to hospital with bleeding.[67]

We observed a novel bleeding risk in carriers of *HPS6* premature stop variants (p.Ter776ArgextTer38, p.Leu22ArgfsTer33) and an *ANO6* splice-acceptor variant (supplemental Figure10).[60,68] We detected no increased risk of bleeding in carriers of variants for BSS, GT, CAMT or AD thrombocytopenia disorders.

VTE is a leading cause of death worldwide, with an estimated 25,000 cases annually in the UK.[69] Of the 257 BTPD-variants in thrombosis DGGs, 172 (66.9%) were in genes encoding antithrombin (*SERPINC1)*, protein C (*PROC*) and protein S (*PROS1*).[70] Single variant analysis of 24 variants in these 3 genes showed increased risk for deep vein thrombosis (DVT) for 7 variants (OR=4.43-17.42, P<5x10^{-2}), and for pulmonary embolism (PE) for one variant (OR=4.22, P=4.8x10^{-2}) in carriers; 4 had ORs >10 (Figure4C).

The interplay between common and rare variants

Rare variants are embedded in a complex genetic architecture that also affects traits and diseases, as shown by GWAS studies. As this architecture may alter the penetrance and effect of rare variants, we explored the interplay between GWAS-variants and rare variants, using an interactome of 18,410 nodes and 571,917 edges (Methods). This interactome allowed us to evaluate whether common variants exert an effect on clinical traits using the same pathways altered by rare variants. Platelet count was regulated by 658 common GWAS-variants (MAF >0.01), explaining ~19% of the variance.[31] Summing the weighted allele counts for these variants provided a polygenic score (PGS) for platelet count.[31,71] A network-propagation analysis with the 93 BTPD DGGs as seed nodes showed that the GWAS-variants with the largest effect sizes (i.e. top quartile) are enriched in nodes encoded by the 93 DGGs and their first-order interactors (Figure5A-C). In contrast, effect sizes for GWAS-variants in genes encoding nodes on the interactome periphery were smaller (Figure5C). A similar observation was made for the 297 GWAS-variants used to calculate the PGS for VTE risk (Figure5B-C). We conclude that GWAS-variants with large effect sizes for platelet traits and VTE were strongly enriched in BTPD DGGs or their immediate functional interactors.

This prompted us to explore the interplay between PGS and rare BTPD-variants. Effect sizes for rare platelet gene variants on platelet count (Figure3A-C) and ORs for rare thrombosis gene variants on VTE risk (Figure4C) are at least two orders of magnitude higher than those observed for GWAS-variants (supplemental Figure11).[72] To explore whether there was an additive effect between rare BTPD- and GWAS-variants, we first tested for interaction (i.e. synergistic effect) between each of the 128 platelet gene variants included in the association analysis and the platelet PGS. Within the power limitation of this sample size, there were no significant interaction effects, indicating these contributions are independent and additive (supplemental Table4). For 10 rare variants with the largest platelet count effect sizes we combined their effect and frequency with the PGS distribution, to calculate the additive PGS contribution required to reduce platelet count below the clinical cut-off for thrombocytopenia (150 x 10^9/L) (Figure6A). We estimate at least 3,242 UK individuals have a count <150x10^9/L due to the combination of one of these rare variants plus an unfavorable PGS (supplemental Methods). This interplay is illustrated by the *TUBB1* premature termination (p.Cys12LeufsTer23), which reduced platelet count by -1.1 SD in UKB carriers. Carriers with counts >150x10^9/L (n=9) had favorable PGS values, while the one with thrombocytopenia (platelet count 126 x10^9/L) had an unfavorable PGS that lowered the count by -1.29 SD.

Similarly, we reasoned that given a constant shared risk for VTE due to inheritance of the same rare variant, PGS could improve the prediction of VTE events. We estimated this improvement using a binary classifier and showed increased sensitivity and specificity when including PGS compared to a model based only on rare variants (DeLong's test, Z=11.31, P=2.2x10^{-6}; Figure6B).[30] The minor improvement in the PGS model predictive capacity due to the inclusion of rare variants is expected, as rare variants are not widely shared among individuals, which hampers their use in population-scale prediction. Then, we explored whether inclusion of the PGS improved interpretation of the clinical impact of 3 rare variants with the largest ORs for VTE. Considering the 46 carriers (8 with a VTE event, 38 without) of these 3 variants, those with favorable PGS had fewer events (Fisher’s exact test; P=1.26x10^{-1}). When reviewed per variant, the PGS contribution improved the distinction between carriers with and without VTE (Figure6C). Therefore, individuals carrying rare variants with a large OR who also have an unfavorable PGS have a higher risk of VTE.

Discussion

Rare diseases collectively affect hundreds of millions worldwide.[73,74] Incorporating genetic testing into clinical care has proved crucial in diagnosing patients with rare diseases, increasing the diagnostic rate to 50.8% for BTPDs.[2,16,24,75,76] It can inform treatment decisions, identify affected relatives and influence families reproductive choices. Here we describe an approach to support and accelerate the generation of evidence to define the pathogenicity of rare variants in BTPDs. This approach leveraged WES genotypes and linked EHR data from UKB to remove possible bias inherent to discoveries and pathogenicity classifications from extreme clinical cases. To generate evidence about pathogenicity we have, for the first time, estimated effect sizes and ORs of rare BTPD-variants on platelet count and volume, bleeding and VTE risk in UKB participants of European ancestry. There were sufficient UKB carriers for association analysis for 91 out of 3,068 (3.0%) cataloged-variants in DGGs for AD BTPDs, including AD thrombocytopenia disorders, VWD, and deficiencies in antithrombin, protein C and S (Figure3-4). There was no enrichment of variants accepted by the MDT as pathogenic or likely pathogenic amongst those with significant effect sizes. We observed non-significant effects for accepted variants and significant associations for rejected variants (Figure 3, 4). For example, *VWF* p.Thr1034del was rejected because of its high MAF in individuals of African ancestry (MAF=0.015).[77] This suggests that synergising the results of association analyses with traditional MDT decision approaches can assist with pathogenicity classifications during clinical variant reporting.

We also systematically explored the phenotypic consequences from carrying a single BTPD-variant for disorders with a recessive MOI. We showed a significant increase in bleeding in female carriers of *F9* variant p.Arg449Trp (Figure4). Our association analysis provides compelling evidence that LoF variants in several AR platelet disorder DGGs are associated with altered platelet count or increased risk of bleeding when monoallelic. We show this both in DGGs with an established mixed MOI (e.g. *GP1BA, GP1BB*) and in those for which a carrier phenotype has not previously been described (e.g. *ITGA2B*, *ITGB3*). These findings support an additive effect of rare variants in disorders traditionally understood to be recessively inherited and narrows the distinction between dominant and recessive MOIs.

Unexpectedly, LoF variants in *MPL*, which cause the AR disorder CAMT, increased platelet counts of UKB carriers, sufficient in some cases to result in thrombocytosis (Figure3). One CAMT-variant (p.Arg102Pro) has been reported in heterozygosity in a family with thrombocytosis.[78] When biallelic, MPL cell surface translocation is blocked, therefore despite high circulating TPO levels, a lack of MPL-TPO signaling suppresses megakaryopoiesis, explaining the profound thrombocytopenia in patients with CAMT.[79] However, monoallelic p.Arg102Pro only reduces cell surface expression and TPO clearance; through a negative feedback loop this increases megakaryocyte proliferation and platelet production.[78] TPO measurements are unavailable in UKB EHRs to support reduced clearance as the mechanism for increased platelet counts in CAMT-variant carriers.

The incomplete penetrance of pathogenic variants is recognized across rare diseases including BTPDs and may partially be explained by the modifying effect of common GWAS-variants. The effects of GWAS-variants for platelet traits and VTE are dispersed across hundreds of proteins in the interactome, but can be enumerated by the PGS value. Our expansion analysis showed that GWAS-variants with the largest effect sizes are enriched in the proximity of proteins encoded by the 93 BTPD DGGs (Figure5). This is consistent with the omnigenic model recently proposed[80] and supports the idea that traits, phenotypes and diseases are a continuum regulated by common and rare variants, and interplay between them. We illustrated that the effects of the PGS for platelet count, and of rare platelet gene variants, are independent and additive in causing thrombocytopenia (Figure6).[31,81] For VTE, we observed that incorporation of both the PGS for VTE and the ORs for rare variants in thrombosis genes improved predictive models compared with the model using BTPD-variants alone (Figure 6). These observations confirm the interplay between BTPD-variants and PGS, and verify similar observations for other complex traits such as type 2 diabetes and hemoglobin A1C levels, familial hypercholesterolemia, and some hereditary cancers, such as Lynch syndrome.[82,83] Incorporating the effect of the relevant PGS when considering the clinical impact of rare BTPD-variants can be readily achieved if WGS analysis is used in the diagnostic workup.[2,76] There are other possible explanations for altered variant effect sizes, for example interaction across different genes, however, these were beyond the scope of this manuscript.

To summarize, we have reported on effect sizes and odds ratios of rare pathogenic and likely pathogenic variants considered causal of rare inherited hemostasis disorders. Our analysis further challenges the dogma that rare variants causal of autosomal recessive disorders are silent, shrinking the distinction between dominant and recessive inheritance. We also demonstrate that PGS for platelet count and VTE risk modify the clinical penetrance of rare BTPD-variants causal of autosomal dominant thrombocytopenia and VTE, respectively. UKB is a representative cohort of the general UK population of sufficient size to estimate rare variant effects in individuals of European ancestry. As non-European population cohorts increase and ancestry-specific PGS become available, studies of this kind can be replicated. Many variants considered in genomics MDTs will be amenable to association analyses. The integration of these results into clinical variant reporting will assist with the pathogenicity classification of rare variants implicated in hemostasis disorders and other rare diseases.

Acknowledgements

This study uses genotype and phenotype data generated by the National Institute for Health and Care Research (NIHR) BioResource for the rare disease programme - these data are available from relevant publications. This research has been conducted using the UK Biobank Resource under Application Number 13745. The authors gratefully acknowledge the participation of all NIHR BioResource volunteers and thank the NIHR BioResource centre and staff for their contribution. They also thank the NHS Blood and Transplant. We are grateful to Dr. Catherine Snow (Genomics England) and MSc Arina Puzriakova (Genomics England) for their assistance with PanelApp and for sharing their knowledge about rare disease domains, to Professor Jim Crawley (Imperial College, London) for commenting on the discussion and to Dr Ernest Turro (Mount Sinai), Dr. Daniel Greene (Mount Sinai) and Dr. William Astle (University of Cambridge) for their feedback and knowledge on UK biobank data. For the AstraZeneca UK Biobank work, we thank the participants and investigators in the UK Biobank study who made this work possible (Resource Application Number 26041); the UKB Exome Sequencing Consortium (UKB-ESC) members AbbVie, Alnylam Pharmaceuticals, AstraZeneca, Biogen, Bristol-Myers Squibb, Pfizer, Regeneron and Takeda for funding the generation of the data and Regeneron Genetics Center for completing the sequencing and initial quality control of the exome sequencing data; the AstraZeneca Centre for Genomics Research Analytics and Informatics team for processing and analysis of sequencing data. Research in the Ouwehand laboratory received funding from the British Heart Foundation (BHF), the International Society on Thrombosis and Haemostasis, Medical Research Council (MRC), National Health Service (NHS) Blood and Transplant, and the NIHR. L.S. was for his PhD period supported by the BHF grant for the Cambridge BHF Centre of Research Excellence (RE/18/1/34212); J.C. and M.S. were awarded Medical Research Council Clinical Research Training Fellowships (MR/P02002X/1 and MR/R002363/1, respectively). K.F. was supported by the Research Council of the University of Leuven (Special Research Fund [BOF] KU Leuven, Belgium, C14/19/096) and by an unrestricted grant of Sobi. L.P., H.H., K.P, and P.P. received funding from EMBL core funding, Open Targets (grant agreements OTAR-044, OTAR02-048, OTAR02-066) and the Wellcome Trust grant INVAR (grant ref: 212925/Z/18/Z). P.B. is supported by the Helmut Horten Stiftung and the ETH Zurich Foundation. W.H.O. is an NIHR senior investigator. D.V. is a member of the Health Protection Research Unit in Chemical and Radiation Threats and Hazards, a partnership between Public Health England and Imperial College London which is funded by the National Institute for Health Research (NIHR). MF is supported by the British Heart Foundation (FS/18/53/33863) and by the National Institute for Health and Care Research Exeter Biomedical Research Centre. EB is supported by National Human Genome Research Institute (NHGRI) grant U24HG003345 & Wellcome Trust grant 208349/Z/17/Z. The views expressed are those of the author(s) and not necessarily those of the funders, NIHR or the Department of Health and Social Care. Support for title page creation and format was provided by AuthorArranger, a tool developed at the National Cancer Institute. Figure 1.A and visual abstract were created with BioRender.com.

Authorship and conflict-of-interest statements

L.Stefanucci, J.C. and M.C.S. wrote the manuscript, analyzed data, participated in MDTs and oversight analysis pipeline. I.B-H., L.Sun, O.B., L.P., I.B., N.G., T.J.C., R.A.L. and P.B analyzed data and oversight analysis pipeline. J.A.G., H.H., K.P, S.P., P.P., Q.W., K.C., X.W., E.d.A and M.J.M. oversight analysis pipeline. M.F. reviewed the manuscript. K.G., M.L. and K.D. participated in MDTs. A.D.M. and K.F. reviewed the manuscript and participated in MDTs. W.H.O. conceptualized the study and reviewed the manuscript. K.M. project administration, data curation, wrote the manuscript and participated in MDTs. E.B. wrote the manuscript, conceptualized the study, project administration and data curation. D.V. wrote the manuscript, analyzed data and participated in MDT. O.B., Q.W., K.C., P.P., K.M. and S.P. are current AstraZeneca employees and/or stockholders.

Bibliography

1 Ferreira CR. The burden of rare diseases. *Am J Med Genet A* 2019; **179**: 885–92.

2 Turro E, Astle WJ, Megy K, *et al.* Whole-genome sequencing of patients with rare diseases in a national health system. *Nature* 2020; **583**: 96–102.

3 Dudley JT, Kim Y, Liu L, *et al.* Human genomic disease variants: a neutral evolutionary explanation. *Genome Res* 2012; **22**: 1383–94.

4 MacArthur DG, Tyler-Smith C. Loss-of-function variants in the genomes of healthy humans. *Hum Mol Genet* 2010; **19**: R125–30.

5 Forrest IS, Chaudhary K, Vy HMT, *et al.* Population-Based Penetrance of Deleterious Clinical Variants. *JAMA* 2022; **327**: 350–9.

6 MacArthur DG, Manolio TA, Dimmock DP, *et al.* Guidelines for investigating causality of sequence variants in human disease. *Nature* 2014; **508**: 469–76.

7 Krawczak M, Cooper DN. The human gene mutation database. *Trends Genet* 1997; **13**: 121–2.

8 Landrum MJ, Lee JM, Riley GR, *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* 2014; **42**: D980–5.

9 Richards S, ; on behalf of the ACMG Laboratory Quality Assurance Committee, Aziz N, *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. Genetics in Medicine. 2015; **17**: 405–23.

10 Kalia SS, Adelman K, Bale SJ, *et al.* Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med* 2017; **19**: 249–55.

11 McVey JH, Rallapalli PM, Kemball-Cook G, *et al.* The European Association for Haemophilia and Allied Disorders (EAHAD) Coagulation Factor Variant Databases: Important resources for haemostasis clinicians and researchers. *Haemophilia* 2020; **26**: 306–13.

12 Megy K, Downes K, Morel-Kopp M-C, *et al.* GoldVariants, a resource for sharing rare genetic variants detected in bleeding, thrombotic, and platelet disorders: Communication from the ISTH SSC Subcommittee on Genomics in Thrombosis and Hemostasis. *J Thromb Haemost* 2021; **19**: 2612–7.

13 Lek M, Karczewski KJ, Minikel EV, *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016; **536**: 285–91.

14 Karczewski KJ, Francioli LC, Tiao G, *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020; **581**: 434–43.

15 Van Hout CV, Tachmazidou I, Backman JD, *et al.* Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature* 2020; **586**: 749–56.

16 100,000 Genomes Project Pilot Investigators, Smedley D, Smith KR, *et al.* 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care - Preliminary Report. *N Engl J Med* 2021; **385**: 1868–80.

17 Keen-Kim D, Mathews CA, Reus VI, *et al.* Overrepresentation of rare variants in a specific ethnic group may confuse interpretation of association analyses. *Hum Mol Genet* 2006; **15**: 3324–8.

18 Simonti CN, Vernot B, Bastarache L, *et al.* The phenotypic legacy of admixture between modern humans and Neandertals. *Science* 2016; **351**: 737–41.

19 Popejoy AB, Ritter DI, Crooks K, *et al.* The clinical imperative for inclusivity: Race, ethnicity, and ancestry (REA) in genomics. *Hum Mutat* 2018; **39**: 1713–20.

20 Surolia I, Pirnie SP, Chellappa V, *et al.* Functionally defective germline variants of sialic acid acetylesterase in autoimmunity. *Nature* 2010; **466**: 243–7.

21 Hunt KA, Smyth DJ, Balschun T, *et al.* Rare and functional SIAE variants are not associated with autoimmune disease risk in up to 66,924 individuals of European ancestry. *Nat Genet* 2011; **44**: 3–5.

22 van der Meer AD, van den Berg A. Organs-on-chips: breaking the in vitro impasse. *Integr Biol* 2012; **4**: 461–70.

23 Duval K, Grover H, Han L-H, *et al.* Modeling Physiological Events in 2D vs. 3D Cell Culture. *Physiology* 2017; **32**: 266–77.

24 Downes K, Megy K, Duarte D, *et al.* Diagnostic high-throughput sequencing of 2396 patients with bleeding, thrombotic, and platelet disorders. *Blood* 2019; **134**: 2082–91.

25 Rehm HL, Berg JS, Brooks LD, *et al.* ClinGen--the clinical genome resource. *N Engl J Med* 2015; **372**: 2235–42.

26 Luo X, Feurstein S, Mohan S, *et al.* ClinGen Myeloid Malignancy Variant Curation Expert Panel recommendations for germline RUNX1 variants. *Blood Adv* 2019; **3**: 2962–79.

27 Ross JE, Zhang BM, Lee K, *et al.* Specifications of the variant curation guidelines for ITGA2B/ITGB3: ClinGen Platelet Disorder Variant Curation Panel. *Blood Adv* 2021; **5**: 414–31.

28 Preston CG, Wright MW, Madhavrao R, *et al.* ClinGen Variant Curation Interface: a variant classification platform for the application of evidence criteria from ACMG/AMP guidelines. *Genome Med* 2022; **14**: 6.

29 Astle WJ, Elding H, Jiang T, *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* 2016; **167**: 1415–29.e19.

30 Klarin D, Busenkell E, Judy R, *et al.* Genome-wide association analysis of venous thromboembolism identifies new risk loci and genetic overlap with arterial vascular disease. *Nat Genet* 2019; **51**: 1574–9.

31 Vuckovic D, Bao EL, Akbari P, *et al.* The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell* 2020; **182**: 1214–31.e11.

32 Bycroft C, Freeman C, Petkova D, *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018; **562**: 203–9.

33 Taliun D, Harris DN, Kessler MD, *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* 2021; **590**: 290–9.

34 Szustakowski JD, Balasubramanian S, Kvikstad E, *et al.* Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. *Nat Genet* 2021; : 1–7.

35 Wang Q, Dhindsa RS, Carss K, *et al.* Rare variant contribution to human disease in 281,104 UK Biobank exomes. *Nature* 2021; published online Aug 10. DOI:10.1038/s41586-021-03855-y.

36 Noris P, Perrotta S, Bottega R, *et al.* Clinical and laboratory features of 103 patients from 42 Italian families with inherited thrombocytopenia derived from the monoallelic Ala156Val mutation of GPIbα (Bolzano mutation). *Haematologica* 2012; **97**: 82–8.

37 Sivapalaratnam S, Westbury SK, Stephens JC, *et al.* Rare variants in GP1BB are responsible for autosomal dominant macrothrombocytopenia. *Blood* 2017; **129**: 520–4.

38 Noris P, Marconi C, De Rocco D, *et al.* A new form of inherited thrombocytopenia due to monoallelic loss of function mutation in the thrombopoietin gene. *Br J Haematol* 2018; **181**: 698–701.

39 Cornish N, Aungraheeta MR, FitzGibbon L, *et al.* Monoallelic loss-of-function THPO variants cause heritable thrombocytopenia. *Blood Adv* 2020; **4**: 920–4.

40 Hou Y-CC, Yu H-C, Martin R, *et al.* Precision medicine integrating whole-genome sequencing, comprehensive metabolomics, and advanced imaging. *Proc Natl Acad Sci U S A* 2020; **117**: 3053–62.

41 Shi Z, Wei J, Na R, *et al.* Cystic fibrosis F508del carriers and cancer risk: Results from the UK Biobank. *Int J Cancer* 2021; **148**: 1658–64.

42 Barton AR, Hujoel MLA, Mukamel RE, Sherman MA, Loh P-R. A spectrum of recessiveness among Mendelian disease variants in UK Biobank. *Am J Hum Genet* 2022; **109**: 1298–307.

43 Hellerbrand C, Pöppl A, Hartmann A, Schölmerich J, Lock G. HFE C282Y heterozygosity in hepatocellular carcinoma: evidence for an increased prevalence. *Clin Gastroenterol Hepatol* 2003; **1**: 279–84.

44 Megy K, Downes K, Simeoni I, *et al.* Curated disease-causing genes for bleeding, thrombotic, and platelet disorders: Communication from the SSC of the ISTH. *J Thromb Haemost* 2019; **17**: 1253–60.

45 Zschocke J, Byers PH, Wilkie AOM. Gregor Mendel and the concepts of dominance and recessiveness. *Nat Rev Genet* 2022; **23**: 387–8.

46 Simeoni I, Stephens JC, Hu F, *et al.* A high-throughput sequencing test for diagnosing inherited bleeding, thrombotic, and platelet disorders. *Blood* 2016; **127**: 2791–803.

47 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.

48 McLaren W, Gil L, Hunt SE, *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* 2016; **17**: 122.

49 Palmer LJ. UK Biobank: bank on it. *Lancet* 2007; **369**: 1980–2.

50 Collins R. What makes UK Biobank special? *Lancet* 2012; **379**: 1173–4.

51 Szklarczyk D, Kirsch R, Koutrouli M, *et al.* The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* 2023; **51**: D638–46.

52 Ochoa D, Hercules A, Carmona M, *et al.* The next-generation Open Targets Platform: reimagined, redesigned, rebuilt. *Nucleic Acids Res* 2023; **51**: D1353–9.

53 Del Toro N, Shrivastava A, Ragueneau E, *et al.* The IntAct database: efficient access to fine-grained molecular interaction data. *Nucleic Acids Res* 2022; **50**: D648–53.

54 Gillespie M, Jassal B, Stephan R, *et al.* The reactome pathway knowledgebase 2022. *Nucleic Acids Res* 2022; **50**: D687–92.

55 Lo Surdo P, Iannuccelli M, Contino S, *et al.* SIGNOR 3.0, the SIGnaling network open resource 3.0: 2022 update. *Nucleic Acids Res* 2023; **51**: D631–7.

56 Shovlin CL, Simeoni I, Downes K, *et al.* Mutational and phenotypic characterization of hereditary hemorrhagic telangiectasia. *Blood* 2020; **136**: 1907–18.

57 Miller JL, Lyle VA, Cunningham D. Mutation of leucine-57 to phenylalanine in a platelet glycoprotein Ib alpha leucine tandem repeat occurring in patients with an autosomal dominant variant of Bernard-Soulier disease. *Blood* 1992; **79**: 439–46.

58 Savoia A, Kunishima S, De Rocco D, *et al.* Spectrum of the mutations in Bernard-Soulier syndrome. *Hum Mutat* 2014; **35**: 1033–45.

59 Nurden AT, Fiore M, Nurden P, Pillois X. Glanzmann thrombasthenia: a review of ITGA2B and ITGB3 defects with emphasis on variants, phenotypic variability, and mouse models. *Blood* 2011; **118**: 5996–6005.

60 Nurden P, Stritt S, Favier R, Nurden AT. Inherited platelet diseases with normal platelet count: phenotypes, genotypes and diagnostic strategy. *Haematologica* 2021; **106**: 337–50.

61 Ghevaert C, Salsmann A, Watkins NA, *et al.* A nonsynonymous SNP in the ITGB3 gene disrupts the conserved membrane-proximal cytoplasmic salt bridge in the alphaIIbbeta3 integrin and cosegregates dominantly with abnormal proplatelet formation and macrothrombocytopenia. *Blood* 2008; **111**: 3407–14.

62 Kunishima S, Kashiwagi H, Otsu M, *et al.* Heterozygous ITGA2B R995W mutation inducing constitutive activation of the αIIbβ3 receptor affects proplatelet formation and causes congenital macrothrombocytopenia. *Blood* 2011; **117**: 5479–84.

63 Ballmaier M, Germeshausen M. Advances in the understanding of congenital amegakaryocytic thrombocytopenia. *Br J Haematol* 2009; **146**: 3–16.

64 Germeshausen M, Ballmaier M. CAMT-MPL: congenital amegakaryocytic thrombocytopenia caused by MPL mutations - heterogeneity of a monogenic disorder - a comprehensive analysis of 56 patients. *Haematologica* 2021; **106**: 2439–48.

65 Staber J, Croteau SE, Davis J, Grabowski EF, Kouides P, Sidonio RF Jr. The spectrum of bleeding in women and girls with haemophilia B. *Haemophilia* 2018; **24**: 180–5.

66 Puetz J, Cheng D. Descriptive analysis of bleeding symptoms in haemophilia carriers enrolled in the ATHNdataset. *Haemophilia* 2021; **27**: 1045–50.

67 Baronciani L, Peake I, Schneppenheim R, *et al.* Genotypes of European and Iranian patients with type 3 von Willebrand disease enrolled in 3WINTERS-IPS. *Blood Adv* 2021; **5**: 2987–3001.

68 Millington-Burgess SL, Harper MT. Gene of the issue: ANO6 and Scott Syndrome. *Platelets* 2020; **31**: 964–7.

69 Baglin T. Venous thromboembolism in hospitalised patients: a public health crisis? *Br J Haematol* 2008; **141**: 764–70.

70 Middeldorp S. Inherited thrombophilia: a double-edged sword. *Hematology Am Soc Hematol Educ Program* 2016; **2016**: 1–9.

71 Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet* 2013; **9**: e1003348.

72 Flint J. GWAS. *Curr Biol* 2013; **23**: R265–6.

73 Boycott KM, Rath A, Chong JX, *et al.* International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases. *Am J Hum Genet* 2017; **100**: 695–705.

74 Nguengang Wakap S, Lambert DM, Olry A, *et al.* Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *Eur J Hum Genet* 2020; **28**: 165–73.

75 Simeoni I, Shamardina O, Deevi SVV, Thomas M. GRID–Genomics of Rare Immune Disorders: a highly sensitive and specific diagnostic gene panel for patients with primary immunodeficiencies. *bioRxiv* 2019. https://www.biorxiv.org/content/10.1101/431544v3.abstract.

76 Thaventhiran JED, Lango Allen H, Burren OS, *et al.* Whole-genome sequencing of a sporadic primary immunodeficiency cohort. *Nature* 2020; **583**: 90–5.

77 Bellissimo DB, Christopherson PA, Flood VH, *et al.* VWF mutations and new sequence variations identified in healthy controls are more frequent in the African-American population. *Blood* 2012; **119**: 2135–40.

78 Bellanné-Chantelot C, Mosca M, Marty C, Favier R, Vainchenker W, Plo I. Identification of MPL R102P Mutation in Hereditary Thrombocytosis. *Front Endocrinol* 2017; **8**: 235.

79 Varghese LN, Zhang J-G, Young SN, *et al.* Functional characterization of c-Mpl ectodomain mutations that underlie congenital amegakaryocytic thrombocytopenia. *Growth Factors* 2014; **32**: 18–26.

80 Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* 2017; **169**: 1177–86.

81 Collins J, Astle WJ, Megy K, Mumford AD, Vuckovic D. Advances in understanding the pathogenesis of hereditary macrothrombocytopenia. *Br J Haematol* 2021; published online March 30. DOI:10.1111/bjh.17409.

82 Fahed AC, Wang M, Homburger JR, *et al.* Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. *Nat Commun* 2020; **11**: 3635.

83 Dornbos P, Koesterer R, Ruttenburg A, *et al.* A combined polygenic score of 21,293 rare and 22 common variants improves diabetes diagnosis based on hemoglobin A1C levels. *Nat Genet* 2022; **54**: 1609–14.

Figure legends

Figure 1 Diagram of the workflow and a high-level summary of variants analyzed. A) Visual depiction of project workflow showing the selection and filtering steps adopted to generate the catalog of BTPD-variants used in the MDT review and the statistical analysis of UKB exome dataset (Methods); from left to right: the online sources used to retrieve gene and variant information; the number of UKB participants available for inclusion in the study; the number of BTPD-variants and the number of UKB participants carrying a BTPD variant; review of BTPD variants for pathogenicity by MDTs and used for association analysis to estimate effect sizes and ORs. Of the 967 BTPD-variants, 213 (22%) had sufficient UKB carriers to perform an association analysis (Methods). Altogether, we estimated effect sizes for 128 variants in platelet disorder DGGs associated with platelet count and volume, ORs for 67 of these variants and for 61 variants in the coagulation genes *F8*/*F9/VWF* on bleeding, and ORs for 24 variants in thrombotic disorder genes *PROC/PROS1/SERPINC1* on VTE (supplemental Table4). DGGs: diagnostic grade genes B) Venn diagram showing the overlap for variant pathogenicity labels between the resources with ClinVar (orange), HGMD (yellow), and NIHR BioResource (green). C) VEP-calculated functional impacts of the 299,606 pathogenic and likely pathogenic variants, subclassified according to whether they were observed or not in the UKB study population. X-axis: variants grouped according to their impact on protein function (Methods), ranging from low (e.g. synonymous SNV), modifier, moderate to high impact (e.g. premature stop). Y-axis: number of unique variants in each functional effect category. In gray are the variants not observed in UKB, and in blue those observed. D) CADD (PHRED) scores for the pathogenic and likely pathogenic variants: CADD score distribution for all cataloged-variants in gray and for the subset observed in UKB participants in blue. Y-axis: relative density.

Figure 2 Number of pathogenic and likely pathogenic variants per BTPD gene. Scatter plots show the number of cataloged-variants per BTPD gene retrieved from the resources (X-axis) versus the ones observed in UKB participants (Y-axis). The BTPD genes are categorized according to whether they are associated with platelet (purple), bleeding and coagulation (green) and thrombotic (orange) disorders. HGNC gene symbols label key genes flagged in the results. X- and Y-axis are logarithmic scaled.

Figure 3 Effect sizes for variants in platelet disorder genes and MPL structure. A) Effect sizes (in standard deviations=SD) for the platelet count (PLT) and mean platelet volume (MPV) in carriers of 128 BPTD-variants present in at least 5 unrelated European UKB participants; 24 variants labeled by HGNC gene symbol have a significant effect on PLT and/or MPV (P $<5x10^{-2}$). Variants are color-coded by mode of inheritance (MOI) for the

associated platelet disorder: AD, autosomal dominant; AR, autosomal recessive; XL, X-linked inheritance. B) Effect sizes in SD with 95% confidence intervals (CI) for platelet count and MPV associated with 19 cataloged-variants for AD thrombocytopenia disorders, of which 10 have significant effects (P $<5x10^{-2}$). Variants with AD MOI are in blue and AD/AR in brown. MDT decision is indicated by circles and squares for accept and reject, respectively. C) Effect sizes in SD with 95% CI values for platelet count and MPV that are significantly associated with 14 cataloged-variants for AR platelet disorders (P $<5x10^{-2}$). Circles, squares and triangles indicate MDT decisions for accept, reject and undecided, respectively. D) Violin plots with platelet count distributions of UKB controls (black), carriers of one of 5 congenital amegakaryocytic thrombocytopenia (CAMT)-causing *MPL* variants that were associated with a significant increase in platelet count (purple) and patients with CAMT (red); each point represents a unique UKB individual, except for the CAMT cases for whom platelet count values were retrieved from the NIHR BioResource study database. E) Probable structure of the MPL receptor and its ligand thrombopoietin, as represented by the 3D structure of the highly homologous erythropoietin receptor (chains B and C) and bound erythropoietin (chain A) from PDB entry 1eer, which is the best available model for the impact of MPL residue changes. Left: PyMOL image of the 1eer structure with three variants shown in spacefill on chains B and C. Two are possible LoF variants – Arg102Pro, labeled R102 on chain C and shown in red, and Gly131Ser, labeled G131, orange – and one predicted as benign – Arg90Gln, labeled R90, magenta. In brackets are the residue numbers in the 1eer structure. Right: Schematic representation of the complex, with the same colors for the domains and variants (small, colored circles). Additional variants with possible functional consequences and which, like the LoF variant Gly131Ser, are highly conserved and occur in the linker region between the domains are: Pro136Arg, Pro136His and Gly131Ser (not shown). LBD: Ligand-binding domain; FD: Fibronectin Type III domain

Figure 4 Odds ratios of hematological phenotypes for coagulation and thrombotic genes. A) Risk of increased ICD-BAT score, as a measure of bleeding, in female UKB carriers of BPTD-variants in *F9*(NM_000133.3) (n=3) and *F8*(NM_000132.3) (n=9). B) Risk of increased ICD-BAT score in UKB carriers 49 BPTD-variants in *VWF*(NM_000552.3). C) Risk of deep vein thrombosis (DVT - dark orange), or pulmonary embolism (PE - yellow) in UKB carriers of BPTD-variants in *PROC*(NM_000312.3) (n=12), *PROS1*(NM_000313.3) (n=9), and *SERPINC1*(NM_000488.3) (n=3). The risk is given as an odds ratio, with 95% CIs. MDT decision is indicated by circles, squares and triangles for accept, reject and undecided, respectively.

Figure 5 Interactomes and omnigenic model of complex polygenic hematological phenotypes. A) An interactome of 366 nodes and 1559 edges was generated using the

proteins encoded by the 93 BTPD DGGs and the 658 proteins encoded by the genes harboring GWAS-variants for platelet count as ‘seeds’ for retrieving their first-order interactors. B) A similar interactome of 73 nodes and 374 edges was generated for venous thrombotic events (VTE) using the 93 DGG-encoded proteins and 297 proteins encoded by genes harboring GWAS-variants for VTE as seeds.[30] For (A) and (B), only interaction from the IntAct database[31] are shown, in order to simplify the network visualization. Nodes and edges were arranged using Cytoscape software circular layout. Seed genes (i.e. DGG genes) were positioned in the center of the circles. The nodes in the outer rings are 1st-order interactors of the seed genes. Although the algorithm used for platelet traits and thrombosis is the same, the number of nodes is much larger in platelet genes, which led to a better resolution of the outer circle. The outer circle highlights genes that interact with BTPD genes, but are not BTPD genes themselves. The radius of nodes are proportional to the estimated effect size, in standard deviations, of the GWAS-variant residing in the gene. Nodes have been colored purple, green and orange for genes/proteins implicated in platelet, bleeding, and thrombotic disorders or in gray if the gene/protein does not belong to one of these DGG domains. C) Barplots showing the results of the expansion analysis using the entire human interactome of 18,410 nodes and 571,917 edges showing the enrichment in effect sizes of GWAS-variants as a function of the distance from the core seed genes. X-axis shows the odds ratio of the proximity to the core seed genes/proteins, with >90 to 50-60 groups representing the nodes (proteins) most proximal and most distal from seed proteins (Figure 5A, 5B). Group “>90” consists of the seed genes/proteins and their close protein interactors estimated via propagation score. The reported odds ratios are calculated using the most distant proteins (<50%) as a reference. The effect sizes of GWAS-variants for platelets and VTE (Figure 5C) are split into 4 quartile effects described for the PGS analysis for VTE and platelets.[30,31] The top quartile (i.e. 75%) contains the variants that have been associated with the largest effect sizes in the relevant GWAS. The y-axis shows the enrichment (in odds ratio) for a set of effect-size quartile bins, in a given distance from the center of the expansion network (in comparison to the periphery of the interaction network). For example, the top quartile of large effect variants for PLT has an odds ratio of >2 of being in close proximity to seed genes (bin group “>90”). Results of the expansion analyses for the count (PLT) and mean volume (MPV) of platelets are in purple and for VTE in orange.

Figure 6 Interplay between BTPD-variant and polygenic scores. A) Each line represents the interplay between the effect of a unique variant causal of autosomal dominant (AD) thrombocytopenia (the top 10 BTPD-variants were selected, see Figure 3). The estimated effect size of the variants causal of AD thrombocytopenia is represented by the purple segment of the bar and the PGS contribution is represented by the gray segment of the bar.

The percentages given above the bars represent the frequencies of UKB participants carrying the BTPD-variant and the predicted percentage of the population having a given PGS value for platelet count. The combination of the rare-variant effects and PGS ones is the effect required to drop platelet count below the clinical threshold. The X-axis reports the effect size on platelet count in standard deviation (SD) required to reduce the platelet count below the 150x10^9/L threshold. B) Receiver operating characteristic curve showing the prediction of VTE phenotypes using a predictive model based solely on rare BTPD-variants for thrombosis (blue), a second model using only the PGS common variants (yellow), and a third one integrating rare BTPD- and common GWAS-variants (yellow). The area under the curve (AUC) indicates performance in variant classification. P/LPVs: pathogenic and likely pathogenic BTPD-variants. C) Additive effect of the polygenic score (PGS) for VTE derived from common GWAS-variants and 2 rare BTPD-variants in *PROC* and one in *PROS1*. The X-axis shows the effects and directionalities of PGS effect estimates in SD (in blue; i.e. increased versus decreased risk) and the odds ratio for the rare BTPD variant in OR (in orange). The contribution to VTE risk given by the 3 rare BTPD-variants is constant, per variant, in carriers with VTE and "healthy" carriers without VTE (the orange portion of the bars). The distribution of PGS values differs significantly between the carriers with VTE and the "healthy" carriers (blue portion of the bars).