Journal of the American Heart Association

ORIGINAL RESEARCH

# Impact of Genetically Predicted Red Blood Cell Traits on Venous Thromboembolism: Multivariable Mendelian Randomization Study Using UK Biobank

Shan Luo, MPH; Shiu Lun Au Yeung, PhD; Verena Zuber, PhD; Stephen Burgess, PhD; Catherine Mary Schooling, PhD

**BACKGROUND:** Red blood cell (RBC) transfusion and erythropoiesis-stimulating agent administration are cornerstones of clinical practice, yet concerns exist as to potential increased risk of thrombotic events. This study aims to identify RBC traits most relevant to venous thromboembolism (VTE) and assess their genetically predicted effects on VTE in the general population.

**METHODS AND RESULTS:** We used multivariable mendelian randomization with bayesian model averaging for exposure selection. We obtained genetic variants predicting any of 12 RBC traits from the largest genome-wide association study of hematological traits (173 480 participants of European ancestry) and applied them to the UK Biobank (265 424 white British participants). We used univariable mendelian randomization methods as sensitivity analyses for validation. Among 265 424 unrelated participants in the UK Biobank, there were 9752 cases of VTE (4490 men and 5262 women). Hemoglobin was selected as the plausible important RBC trait for VTE (marginal inclusion probability=0.91). The best-fitting model across all RBC traits contained hemoglobin only (posterior probability=0.46). Using the inverse variance–weighted method, genetically predicted hemoglobin was positively associated (odds ratio, 1.21 per g/dL unit of hemoglobin; 95% CI, 1.05–1.41) with VTE. Sensitivity analyses (mendelian randomization–Egger, weighted median, and mendelian randomization pleiotropy residual sum and outlier test) gave consistent estimates.

**CONCLUSIONS:** Endogenous hemoglobin is the key RBC trait causing VTE, with a detrimental effect in the general population on VTE. Given men have higher hemoglobin than women, this finding may help explain the sexual disparity in VTE rates. The benefits of therapies and other factors that raise hemoglobin need to be weighed against their risks.

**Key Words:** hemoglobin ■ mendelian randomization ■ venous thromboembolism



Red blood cell (RBC) transfusion is the most readily available method to alleviate anemia and bleeding resulting from a variety of clinical conditions, yet concerns exist as to the risk of adverse effects, with several trials ongoing to establish the optimal transfusion threshold in patients.[1] Erythropoiesis-stimulating agents (ESAs) are widely used in clinical practise to increase hemoglobin concentration by mimicking endogenous erythropoietin and stimulating erythropoiesis in the bone marrow in response to cellular hypoxia.[2] Systematic reviews and meta-analyses of randomized controlled trials of ESAs for treatment of anemia in patients have found increased risk of thrombotic vascular events.[3,4] In 2007, the US Food and Drug Administration issued a public health advisory about the increased risk on ESAs of blood clots and venous thromboembolism (VTE), and required a label warning suggesting more caution when using ESAs,[5] as reflected in recent clinical practice guidelines.[6] Notably, similar warnings have also been issued about specific ESAs,[7] which induce

Correspondence to: Catherine Mary Schooling, PhD, 55 W 125th St, New York, NY 10027. E-mails: mary.schooling@sph.cuny.edu; cms1@hku.hk

Supplementary Materials for this article are available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.120.016771

For Sources of Funding and Disclosures, see page 8.



*JAHA* is available at: www.ahajournals.org/journal/jaha

## CLINICAL PERSPECTIVE

### What Is New?

- Red blood cell attributes are highly correlated genetically and phenotypically, making it difficult to disentangle causal drivers of disease.
- This study used multivariable mendelian randomization with bayesian model averaging to select and prioritize between 12 red blood cell traits, which suggested endogenous hemoglobin is the key factor for venous thromboembolism, with a detrimental effect in the general population.

### What Are the Clinical Implications?

- This study is consistent with randomized controlled trials showing that increasing hemoglobin in patients with anemia, via blood products, erythropoiesis-stimulating agents, or blood transfusion, increases the risk of thromboembolic events.
- This study also suggests relevance to the general population as well as to patients.
- The benefits of therapies and other factors that raise hemoglobin need to be weighed against their risks.

### Nonstandard Abbreviations and Acronyms

| | |
|---|---|
| **ESA** | erythropoiesis-stimulating agent |
| **GWAS** | genome-wide association study |
| ***ICD-9*** | *International Classification of Diseases, Ninth Revision* |
| ***ICD-10*** | *International Classification of Diseases, Tenth Revision* |
| **IVW** | inverse variance weighted |
| **MAF** | minor allele frequency |
| **MCH** | mean corpuscular hemoglobin |
| **MIP** | marginal inclusion probability |
| **MR** | Mendelian randomization |
| **MR-BMA** | multivariable mendelian randomization based on bayesian model averaging |
| **MR-PRESSO** | mendelian randomization pleiotropy residual sum and outlier test |
| **PP** | posterior probability |
| **RBC** | red blood cell |
| **VTE** | venous thromboembolism |

erythropoiesis[8] and may drive the association with VTE.[9] Correspondingly, targeting lower erythroid cells for polycythemia vera reduced the rate of major thrombosis.[10] However, it is unclear whether these findings are the result of specific interventions in patients, and whether they extend to the general population across the normal range is yet to be determined.

Several observational studies have assessed the role of RBC attributes in thrombosis in the general population.[11,12] A study of hemoglobin concentration and its changes in healthy young women to avoid bias from confounding by ill health and selection bias suggested hemoglobin concentration increased thrombosis.[11] Several RBC attributes are altered simultaneously by RBC transfusion and ESAs, and RBC attributes are highly correlated both genetically and phenotypically,[13] making it difficult to disentangle the causal drivers of disease risk. As RBC transfusion and ESA administration are cornerstones of clinical practice, better understanding of the causal determinants of thrombosis has critical clinical importance and public health implications in the general population.

Mendelian randomization (MR), using genetic variants randomly allocated during conception as instrumental variables, is less prone to confounding than traditional observational studies, and can help ascertain causal effects.[14] MR, at the interface of experimental and observational studies, provides a distinct strand of genetic evidence on potential targets of interventions. Multivariable MR models multiple exposures simultaneously, accounting for measured pleiotropic effects via any of the observed exposures.[15] Previous studies have used univariable and multivariable approaches to assess the effects of genetically predicted blood cell traits on disease risk[13,16]; however, these analyses have been limited in statistical power and their ability to consider high-dimensional highly correlated attributes, such as all 12 RBC traits. To address these limitations, we used a novel approach for multivariable MR based on bayesian model averaging (MR-BMA), which scales to the high-throughput candidate exposures and enables exposure prioritization in a bayesian framework.[17] MR-BMA performs well even when the exposures considered are highly correlated because of biological processes.[17] In the UK Biobank, we used MR-BMA to select the RBC traits most relevant to VTE both on average and individually. We then assessed the effects of the top-ranking exposure(s) on VTE in univariable MR.

## METHODS

The UK Biobank received ethical approval from the research ethics committee (11/NW/0382), and participants provided written informed consent. Summary statistics were generated from publicly available data that had previously received appropriate ethics and

institutional review board approvals, and further sanction was therefore not required. The individual-level data in the UK Biobank are available by application directly to the UK Biobank. The data that support the findings of this study are available from the corresponding author on reasonable request. The statistical code in R for implementing MR-BMA can be obtained from the open-source code from Github (https://github.com/verena-zuber/demo_AMD).

## Study Design

This is a 2-sample multivariable MR study, which relies on 3 instrumental variable assumptions (Figure S1). First, the genetic variant is associated with at least one of the exposures. Second, the variant is independent of all confounders of each of the exposure-outcome associations. Third, the variant is independent of the outcome conditional on the exposures and confounders.

## Genetic Predictors of Endogenous RBC Traits

Genetic predictors of RBC traits were extracted from summary statistics generated from an existing publicly available genome-wide association study (GWAS) of hematological traits conducted in 173 480 participants of European ancestry without any blood cancer or other major blood disorder.[13] Participants were from the UK Biobank (132 959, 52% women) and the INTERVAL (Efficiency and safety of varying the frequency of whole blood donation) studies (40 521, 50% women).[18] Blood samples for full blood count analysis were collected by venipuncture in EDTA tubes, and measured by a combination of fluorescence and impedance flow cytometry at the centralized processing laboratory of UK Biocentre (Stockport, UK) within 36 hours.[13] Genotyping was undertaken with Affymetrix Axiom 2.0 Array, and variants were excluded if they deviated from Hardy-Weinberg equilibrium ($P<5\times10^{-6}$), the within-batch call rate was <97%, the across-batch call rate was <75%, or they were nonautosomal biallelic.[13] Imputation was performed using a combined 1000 Genomes phase 3 and UK 10K imputation panel.[13] Univariable associations of each RBC trait with 29.5 million variants (with imputation information score >0.4 and minor allele frequency [MAF] >0.01%) were obtained from linear mixed model using BOLT-LMM v2.2,[19] adjusted for the top 10 principal components of ancestry and adjusted for recruitment center.[13]

## Selection of Genetic Variants

We obtained genetic variants that robustly (genome-wide significance $P<8.31\times10^{-9}$, a recent threshold for genome-wide analyses of common, low-frequency, and rare variants) and independently ($r^2<0.001$) predicted any of the 12 RBC traits (ie, RBC count, mean corpuscular volume, hematocrit, hemoglobin concentration, mean corpuscular hemoglobin [MCH], MCH concentration, red cell distribution width, reticulocyte count, reticulocyte fraction of red cells, immature fraction of reticulocytes, high light scatter reticulocyte count, and high light scatter reticulocyte percentage of red cells). These variants were checked for imputation quality and validity as instrumental variables using individual data from the UK Biobank, with the following exclusion criteria: (1) imputation information score <0.3 for MAF >3%, information score <0.6 for MAF 1% to 3%, information score <0.8 for MAF 0.5% to 1%, and information score <0.9 for MAF 0.1% to 0.5%; (2) departure from Hardy-Weinberg equilibrium at Bonferroni-corrected significance; (3) associated with potential confounders (described below) of the variant-outcome relation at Bonferroni-corrected significance; (4) in the *ABO* gene, which is well known to be highly pleiotropic;[20] or (5) were equivocally palindromic (allele frequency close to 0.5).

## Genetic Association With VTE

The UK Biobank recruited ≈500 000 participants intended to be aged 40 to 69 years from 2006 to 2010 at 22 recruitment centers across Scotland, Wales, and England in the United Kingdom.[21] Participants provided samples, completed questionnaires, including self-reported diseases and regular prescription medications, underwent assessments, and had nurse-led interviews. Longitudinal follow-up via record linkage to all health service encounters and deaths is ongoing. Prevalent and incident diseases were defined using *International Classification of Diseases, Ninth Revision* (*ICD-9*), and *International Classification of Diseases, Tenth Revision* (*ICD-10*), codes. Causes of death were classified using *ICD-10* codes. Genotyping was undertaken with 2 similar arrays, the UK Biobank Lung Exome Variant Evaluation Axiom array (49 979 participants) and the UK Biobank Axiom array (438 398 participants).[21] Genotype imputation was to a reference set combining the UK10K haplotype and the Haplotype Reference Consortium reference panels.[21] To reduce confounding by a hereditary tendency to thrombophilia[22] and latent population structure,[23] we restricted the analysis to genetically verified white British participants and further excluded participants with (1) withdrawn consent, (2) sex mismatch (genetic sex differs from reported sex), (3) aneuploidy of sex chromosomes, (4) low-quality genotyping (missing rate >1.5%), or (5) relatedness (greater than putative third-degree

relatives in the kinship table).[21] We used genotype and phenotype data from the UK Biobank provided in March 7 and November 6, 2018, updates.

## Exposures

The exposures were 12 genetically predicted RBC traits (ie, RBC, mean corpuscular volume, hematocrit, hemoglobin, MCH, MCH concentration, red cell distribution width, reticulocyte count, reticulocyte fraction of red cells, immature fraction of reticulocytes, high light scatter reticulocyte count, and high light scatter reticulocyte percentage of red cells).

## Outcome

We developed classification algorithms for VTE following the recommendations of the UK Biobank.[24] We defined VTE on the basis of self-report at baseline (internal UK Biobank codes 1068, 1093, and 1094) or subsequent primary or secondary diagnosis of hospital episodes (*ICD-9* 415.1, 416.2, and 451–453 and *ICD-10* I26 and I80–I82) or underlying and contributory causes of death (*ICD-10* I26 and I80–I82). Incident and prevalent cases of VTE were combined to maximize statistical power, under the implicit assumption that all events occur incident to a genetic exposure.

## Potential Confounders

To check the randomization, we assessed the association of each genetic variant with potential confounders (ie, established risk factors substantially affecting both hematological traits[13] and higher risk of VTE[25]) in the UK Biobank. Body mass index was calculated as weight divided by height squared (kg/m$^2$). Smoking and alcohol drinking status were categorized as never, previous, current smoker/drinker, and prefer not to answer. Educational level was categorized into degree/professional, nondegree, none of the above, and prefer not to answer, derived from the questionnaire. Townsend deprivation index (a composite indicator of socioeconomic status) was based on preceding census data for area of residential postcode at the baseline visit.

## Statistical Analysis

Analysis of variance (continuous) and $\chi^2$ tests (categorical) were used to assess whether each genetic variant was associated with the potential confounders. The association of each variant with VTE was obtained using an additive genetic model, adjusted for sex, age, genotyping array, and 40 principal components of genetic ancestry.

## Multivariable MR Based on Bayesian Model Averaging

Exposure selection was performed using MR-BMA. On the assumption that one of the models considered is true, MR-BMA ranks all these submodels from the larger model where all 12 RBC traits could have a causal effect on VTE (ie, a single RBC trait or a combination of multiple RBC traits on VTE),[15] according to the posterior probability (PP) of their associations with the outcome. PP is the probability, given the larger model and a set of priors, that a submodel is true. PP is derived from a bayesian model fit criterion, which assesses how well a linear combination of genetic associations with RBC traits predicts the genetic associations with VTE. To aggregate the evidence for individual trait, we combine evidence across all models that include the particular RBC trait(s). The marginal inclusion probability (MIP) is the sum of the PP over all models, including the RBC trait. Outliers were quantified by Q statistic, and influential variants were identified by Cook distance. We repeated the analyses excluding outliers (Q >10) or influential variants (d > median variant of the relevant F-distribution) consistently detected in all the best models (PP >0.02). The flow of MR-BMA is depicted in the Figure.

As recommended,[15] with 12 RBC traits, we initially set prior probability $P$=0.1, corresponding to a priori expecting 1.2 causal factor (p×d). On the basis of a simulation study,[15] we fixed the prior variance $\delta^2$=0.25, corresponding to the priori for the variance of RBC traits. To check the impact of the prior selection, we varied the prior probability of selecting a causal factor from $P$=0.2 to 0.4, reflecting 2.4 to 4.8 expected causal factors.

We also excluded the top-ranking exposure from each model to check if any alternative exposure had equally strong probability of causality. Finally, as platelets may play a role in the development of VTE, we additionally included 4 platelet traits (platelet count, mean platelet volume, platelet distribution width, and plateletcrit) as alternative potential exposures to assess if these play a role.

## Sensitivity Analyses

To verify our finding from MR-BMA, we used several univariable MR methods. We used an inverse variance–weighted (IVW) multiplicative random effects meta-analysis of the genetic variant–specific Wald estimates. IVW provides unbiased estimates as long as all genetic variants are valid instruments. The weighted median provides valid estimates if at least 50% of the weight comes from valid variants.[26] MR-Egger is an extension of IVW but captures horizontal pleiotropy as long as the instrument strength is independent of the direct effect.[27] The MR-Egger intercept, with $P$<0.05, indicates presence of a pleiotropic effect, suggesting the IVW estimate is invalid. MR-Egger can have low statistical power, so we concentrated on the direction and effect size rather

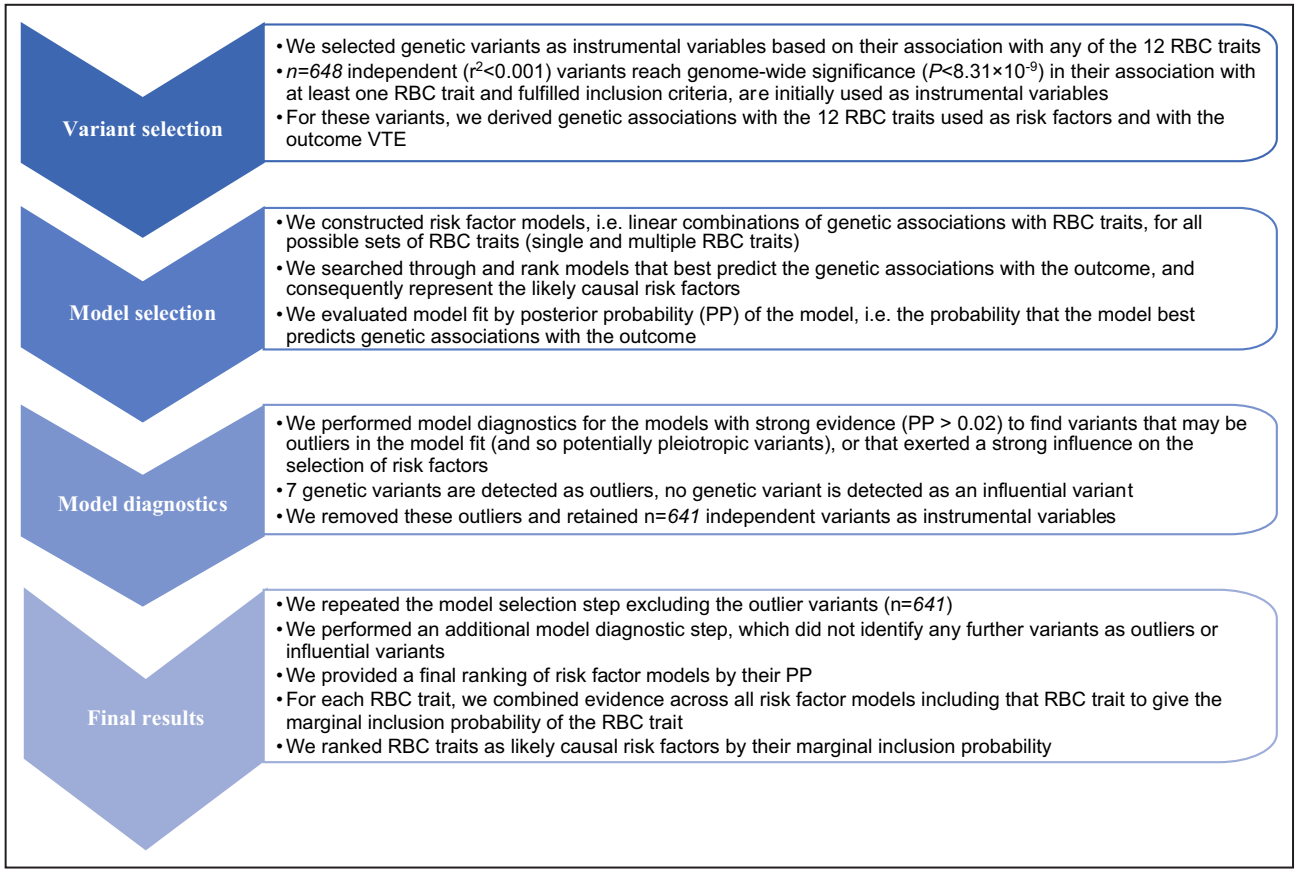



**Figure. The arrow diagram of the multivariable mendelian randomization based on bayesian model averaging.**
PP indicates posterior probability; RBC, red blood cell; and VTE, venous thromboembolism.



than statistical significance. MR pleiotropy residual sum and outlier test (MR-PRESSO), which assumes instrument strength is independent of the direct effect and at least 50% of the variants are valid, is a statistical method for detecting and, if necessary, correcting for horizontal pleiotropic outliers.[28] An MR-PRESSO global test $P$<0.05 (on the basis of 10 000 simulations) indicates horizontal pleiotropy. The MR-PRESSO uses the difference between the observed and expected distribution of RSS residual sum of squares for each variant to identify potentially horizontal pleiotropic outliers, and provides a corrected estimate by removing these outliers.[28] Finally, we excluded any variants associated with cholesterol and/or ischemic heart disease identified from PhenoScanner ($P$<5×$10^{-8}$).[29]

In 2-sample MR, sample overlap may introduce bias and inflate type I error rate, when weak instrument bias is present.[30] The bias depends on the proportion of overlap and the instrument strength.[30] The bias can be estimated as the product of the bias of the observational estimate, the proportion of overlap, and the reciprocal of instrument strength.[30] As the hematological GWAS included 132 959 participants who were randomly selected from the UK Biobank, we estimated the bias using an online tool (https://sb452.shinyapps.io/overlap/).

Genetic associations with the outcome were estimated using the *SNPTEST v2.5.4* program. MR-BMA was performed using the open-source code from Github (https://github.com/verena-zuber/demo_AMD). Univariable MR analyses were performed using the *TwoSampleMR* and *MR-PRESSO* packages in the R version 3.4.4 software platform (R Development Core Team, Vienna, Austria). Two-sided $P$ values are reported throughout.

## RESULTS

Of the 731 genetic variants independently predicting any of the 12 RBC traits at genome-wide significance, 648 remained after excluding on imputation quality, Hardy-Weinberg equilibrium, association with potential confounders, being in the *ABO* gene, or being equivocally palindromic (Figure S2). After applying extensive exclusion criteria, the mean age of 265 424 unrelated participants (123 809 men and 141 615 women) was 56.9 years, with 9752 cases

of VTE (4490 men and 5262 women) used in the analysis.

When including all genetic variants available for the RBC traits (n=648), the exposure most relevant to VTE on the basis of MIP was hemoglobin (MIP=0.90); all other RBC traits had MIP <0.24 (Table S1). To check model fit, we used the best individual models with PP >0.02 (Table S1). Seven outlying variants were identified with high Q statistics (Q >10) consistently in these best models (Table S2 and Figure S3). No influential variant was identified by Cook distance (Table S3 and Figure S3).

We repeated the analysis without the 7 outlying variants (n=641). Again, the most relevant RBC trait was hemoglobin (MIP=0.91), which was followed in relevance by hematocrit (MIP=0.28) (Table 1). Genetic associations with hemoglobin and hematocrit were strongly correlated ($r$=0.91), and models including both had relatively low probability (PP=0.09; Table 2). Figure S4 shows the scatterplots of the genetic associations with each of hemoglobin and hematocrit individually against the genetic associations with VTE risk. We selected the 5 best individual models with PP >0.02 and verified the model fit (Figure S5); no variant with consistently large Q statistics or Cook distance was observed (Tables S4 and S5). We tested the robustness of the results with respect to different initial prior probability parameters that did not alter the ranking of RBC traits (Table S6). As a further sensitivity analysis, we repeated the analysis with several sets of RBC traits. Hemoglobin was still selected with the highest MIP when removing highly correlated hematocrit. MCH (ie, hemoglobin/RBC) was the top exposure when hemoglobin was removed (Table S7). Hemoglobin was also selected with the highest MIP when also considering platelet traits. All of which suggest the effect of hemoglobin is insensitive to the specific selection of RBC traits.

**Table 1. Ranking of RBC Traits According to Their MIP for VTE in the UK Biobank After Exclusion of Outlying Variants (n=641) Using MR-BMA**

| | Exposure | MIP | Model-Averaged Causal Estimate (OR) |
|---|---|---|---|
| 1 | Hemoglobin | 0.912 | 1.22 |
| 2 | Hematocrit | 0.275 | 0.95 |
| 3 | HLSR | 0.154 | 0.98 |
| 4 | MCHC | 0.108 | 1.01 |
| 5 | RET% | 0.104 | 0.99 |
| 6 | IRF | 0.084 | 1.01 |
| 7 | HLSR% | 0.076 | 1.00 |
| 8 | RET | 0.070 | 1.00 |
| 9 | RBC | 0.067 | 1.00 |
| 10 | MCH | 0.060 | 1.01 |

HLSR indicates high light scatter reticulocyte count; HLSR%, high light scatter reticulocyte fraction of red cells; IRF, immature fraction of reticulocytes; MCH, mean corpuscular hemoglobin; MCHC, MCH concentration; MIP, marginal inclusion probability; MR-BMA, multivariable mendelian randomization based on bayesian model averaging; OR, odds ratio; RBC, red blood cell; RET, reticulocyte count; RET%, reticulocyte fraction of red cells; and VTE, venous thromboembolism.

**Table 2. Ranking of Models (ie, Sets of Exposures) According to Their PP for VTE in the UK Biobank After Exclusion of Outlying Variants (n=641) Using MR-BMA**

| | Exposure(s) | PP | Model-Specific Causal Estimate (OR) |
|---|---|---|---|
| 1 | Hemoglobin | 0.461 | 1.16 |
| 2 | Hematocrit, hemoglobin | 0.085 | 0.82, 1.38 |
| 3 | Hemoglobin, HLSR | 0.034 | 1.19, 0.94 |
| 4 | Hematocrit, hemoglobin, HLSR | 0.030 | 0.77, 1.52, 0.92 |
| 5 | Hemoglobin, RBC | 0.024 | 1.21, 0.94 |
| 6 | Hemoglobin, MCHC | 0.019 | 1.14, 1.07 |
| 7 | Hematocrit, hemoglobin, RET% | 0.018 | 0.74, 1.54, 0.93 |
| 8 | MCH, RBC | 0.017 | 1.15, 1.16 |
| 9 | Hemoglobin, MCH | 0.017 | 1.14, 1.04 |
| 10 | Hematocrit, hemoglobin, HLSR%, IRF | 0.015 | 0.67, 1.73, 0.81, 1.23 |

HLSR indicates high light scatter reticulocyte count; HLSR%, high light scatter reticulocyte fraction of red cells; IRF, immature fraction of reticulocytes; MCH, mean corpuscular hemoglobin; MCHC, MCH concentration; MR-BMA, multivariable mendelian randomization based on bayesian model averaging; OR, odds ratio; PP, posterior probability; RBC, red blood cell; RET%, reticulocyte fraction of red cells; and VTE, venous thromboembolism.

Estimates for hemoglobin using univariable MR are shown in Table 3 on the basis of 81 and 72 variants (after exclusion for known potential pleiotropy) (Table S8). Using IVW, hemoglobin was consistently positively associated with VTE. The weighted median, MR-Egger, and MR-PRESSO estimates were of similar magnitude and were also directionally consistent (Table 3), suggesting that bias caused by horizontal pleiotropy is unlikely, assuming the genetic instruments do not directly affect a confounder of RBCs on VTE.

## DISCUSSION

This MR study using MR-BMA to choose between 12 correlated RBC traits suggests hemoglobin is the RBC trait most relevant to VTE. This finding is consistent with randomized controlled trials in patients showing that increasing hemoglobin in anemia, via blood products, ESAs, or blood transfusion, increases thromboembolic events.[31] Our study also suggests relevance to the general population as well as previously seen in patients.

Hemoglobin is a functional protein released from RBCs into the circulation when RBCs are removed by phagocytic activity or hemolysis. Increases in total intracellular hemoglobin or excessive extracellular hemoglobin in chronic and acute anemia can clog blood vessels.[32] Experimental evidence shows hemoglobin and associated stasis augment platelet adhesion

**Table 3. Effect of Genetically Predicted Hemoglobin Concentration on the Risk of VTE in the UK Biobank Using Univariable MR**

| Variant* | Method | OR (95% CI) | *P* Value | Intercept | *P* Value† | *P* Value‡ |
|---|---|---|---|---|---|---|
| 81 | IVW | 1.21 (1.05–1.41) | 0.01 | | | |
| | MR-PRESSO | 1.21 (1.06–1.38) | 0.01 | | | <0.001 |
| | MR-Egger | 1.39 (1.04–1.87) | 0.03 | −0.006 | 0.30 | |
| | Weighted median | 1.18 (0.99–1.41) | 0.06 | | | |
| 72 | IVW | 1.20 (1.05–1.37) | 0.01 | | | |
| | MR-PRESSO | No significant outliers | | | | |
| | Weighted median | 1.18 (1.00–1.40) | 0.05 | | | |
| | MR-Egger | 1.28 (0.98–1.67) | 0.08 | −0.003 | 0.59 | |

IVW indicates inverse variance weighted; MR, mendelian randomization; MR-PRESSO, MR pleiotropy residual sum and outlier test; OR, odds ratio; and VTE, venous thromboembolism.

*Variant indicates number of genetic variants.

†*P* value for MR-Egger intercept.

‡*P* value for global test, indicates horizontal pleiotropy.

reactivity (a well-established coagulator) in vivo and in vitro, even with a low platelet count, independent of hematocrit.[33] Hemoglobin inhibits the a disintegrin and metalloproteinase with a thrombospondin type 1 motif, member 13, cleavage of von Willebrand factor proteolysis.[34] Increasing a disintegrin and metalloproteinase with a thrombospondin type 1 motif, member 13, activity reduces ischemic heart disease,[35] but effects on VTE have not been assessed. Hemoglobin also increases endothelin-1 to rapidly and irreversibly scavenge NO, favoring systemic vasoconstriction and platelet activation, creating conditions that lead to intravascular thrombosis.[36] Our study suggests higher endogenous hemoglobin protein in RBCs relates to thromboembolic events, assessed from hemoglobin or from hematocrit as the ratio of the number of RBCs/total blood cells. Further investigation of the mechanistic role of hemoglobin protein in thrombosis is warranted.

The main strength of this study is the implementation of MR-BMA to select and prioritize potential drivers of VTE from 12 RBC traits, accounting for widespread pleiotropy of highly correlated RBC traits, with validation and provision of precise estimation of causal effects using univariable MR. Other strengths include rigorous selection of genetic instruments that robustly and independently predicted the 12 RBC traits and examination of confounding and sensitivity analyses to identify pleiotropic violations of the exclusion-restriction assumption using one of the largest biobanks globally with sufficient statistical power (Figure S5).

Some limitations of our study should be acknowledged. First, although hemoglobin is correlated with other RBC traits, our findings in the univariable MR are unlikely caused by pleiotropic effects of other RBC traits, as MR-BMA suggests they do not appear to be causal or only have minor direct effects on VTE. Second, MR-BMA is a statistical variable selection method; as is common for variable selection methods, it does not provide unbiased effect estimates. MR-BMA effect estimates were shrunk toward the null because of accounting for selection across a large number of traits. Bias in the effect estimates is traded for reduced variance to stabilize and improve the selection of causal risk factors from several RBC traits. Instead, we provided unbiased estimates using standard univariable MR. Third, the hematological GWAS implemented high-quality procedures to maximize the precision of blood cell traits, but complete measurement accuracy is impossible. A more accurately measured trait would inevitably be prioritized over a less accurately measured trait; we cannot exclude the possibility that hemoglobin is easier to measure accurately than hematocrit. Fourth, although ≈30% of participants overlapped, with strong instruments (explaining 4.2% of the variance in hemoglobin, with F statistic of 93; Figure S6), bias caused by sample overlap is likely to be negligible (bias, 0.004).[30] Fifth, we cannot rule out selection bias in the UK Biobank and the INTERVAL study, resulting from the recruitment of generally healthier participants and survivors, which might bias toward the null.[37] Sixth, because of the lack of sex-specific GWAS of RBC traits, we did not assess the sex-specific associations of hemoglobin with VTE, although the reference range for hemoglobin[38] and the VTE incident rate[39] are higher in men. So, our findings may go some way toward explaining higher VTE rates in men than women. Seventh, use of summary statistics precluded examination of nonlinear association; examination of threshold effects, when possible, would be clinically relevant. Eighth, 23% (40 521/173 480) of participants in the studies providing genetic associations with RBC traits were healthy blood donors from the INTERVAL study, but these GWASs did not adjust for blood donation frequency, which may impair

precision of these estimates. Ninth, our estimates represent average causal effects in the general population, so they may not apply to all subgroups or translate into the optimal level of hemoglobin in at-risk populations. Finally, our study compares groups of people with genetically predicted lower and higher hemoglobin to infer the effects of raising hemoglobin via ESAs and/or blood transfusion. However, several qualitative and quantitative differences between these comparisons may limit the applicability to intervening on hemoglobin. Specifically, small but lifelong changes in endogenous hemoglobin were determined by the genetic variants, via modulating a particular biological pathway, compared with large changes in hemoglobin within a short time.[32]

From a public health and clinical perspective, this study draws attention to factors that modulate hemoglobin, as potential targets of intervention to prevent thromboembolic events. This may include therapeutic phlebotomy[40] and angiotensin II blockage.[41] In contrast, testosterone induces erythrocytosis and substantially increases hemoglobin.[8] Our findings may provide a potential mechanisms by which testosterone could increase the risk of VTE.[9]

In conclusion, the present MR study suggests hemoglobin could be the trait most relevant to VTE, and suggests a detrimental impact on VTE in the general population. Whether other factors that drive hemoglobin could be targets of intervention might bear consideration.

## ARTICLE INFORMATION

Received March 27, 2020; accepted May 28, 2020.

### Affiliations

From the School of Public Health, Li Ka Shing Faculty of Medicine (S.L., S.L.A.Y., C.M.S.), Medical Research Council Biostatistics Unit, School of Clinical Medicine University of Cambridge, United Kingdom (V.Z., S.B.); Department of Epidemiology and Biostatistics, Imperial College London, London, United Kingdom (V.Z.); Medical Research Council/ British Heart Foundation Cardiovascular Epidemiology Unit, School of Clinical Medicine, University of Cambridge, United Kingdom (S.B.); and School of Public Health and Health Policy, City University of New York, NY (C.M.S.).

### Acknowledgments

This research has been conducted using the UK Biobank Resource (http://www.ukbiobank.ac.uk) under application number 14864. Open-source codes for facilitating multivariable mendelian randomization based on bayesian model averaging were obtained from Github (https://github.com/verena-zuber/demo_AMD). The summary statistics of genetic association of hematological traits were downloaded from a genome-wide association study catalog https://www.ebi.ac.uk/gwas/.

### Sources of Funding

This work was supported by the Small Project Funding from the University of Hong Kong (grant 201409176231 to Dr Au Yeung); Dr Luo is supported by the Bau Tsu Zung Bau Kwan Yeun Hing Research and Clinical Fellowship from the University of Hong Kong (grant *200008682.920006.20006.400.01 ); Drs Burgess and Zuber are supported by Sir Henry Dale fellowship, jointly funded by the Wellcome Trust and the Royal Society (grant 204623/Z/16/Z). The funders have no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Disclosures

None.

### Supplementary Materials

**Tables S1–S8**
**Figures S1–S6**

## REFERENCES

1. Carson JL, Triulzi DJ, Ness PM. Indications for and adverse effects of red-cell transfusion. *N Engl J Med*. 2017;377:1261–1272.
2. McMahon FG, Vargas R, Ryan M, Jain AK, Abels RI, Perry B, Smith IL. Pharmacokinetics and effects of recombinant human erythropoietin after intravenous and subcutaneous injections in healthy volunteers. *Blood*. 1990;76:1718–1722.
3. Tonia T, Bohlius J. Ten years of meta-analyses on erythropoiesis-stimulating agents in cancer patients. *Cancer Treat Res*. 2011;157:217–238.
4. Lindquist DE, Cruz JL, Brown JN. Use of erythropoiesis-stimulating agents in the treatment of anemia in patients with systolic heart failure. *J Cardiovasc Pharmacol Ther*. 2015;20:59–65.
5. US Food and Drug Administration. FDA drug safety communication: modified dosing recommendations to improve the safe use of erythropoiesis-stimulating agents (ESAs) in chronic kidney disease. Published June 24, 2011. https://www.fda.gov/drugs/drug-safety-and-availability/fda-drug-safety-communication-modified-dosing-recommendations-improve-safe-use-erythropoiesis. Accessed May 29, 2020.
6. Carson JL, Guyatt G, Heddle NM, Grossman BJ, Cohn CS, Fung MK, Gernsheimer T, Holcomb JB, Kaplan LJ, Katz LM, et al. Clinical practice guidelines from the AABB: red blood cell transfusion thresholds and storage. *JAMA*. 2016;316:2025–2035.
7. US Food and Drug Administration. Testosterone products: FDA/CDER statement—risk of venous blood clots. Published June 20, 2014. https://wayback.archive-it.org/7993/20170406123836/https://www.fda.gov/Safety/MedWatch/SafetyInformation/SafetyAlertsforHumanMedicalProducts/ucm402054.htm, Accessed May 29, 2020.
8. Bachman E, Travison TG, Basaria S, Davda MN, Guo W, Li M, Connor Westfall J, Bae H, Gordeuk V, Bhasin S. Testosterone induces erythrocytosis via increased erythropoietin and suppressed hepcidin: evidence for a new erythropoietin/hemoglobin set point. *J Gerontol A Biol Sci Med Sci*. 2014;69:725–735.
9. Luo S, Au Yeung SL, Zhao JV, Burgess S, Schooling CM. Association of genetically predicted testosterone with thromboembolism, heart failure, and myocardial infarction: mendelian randomisation study in UK Biobank. *BMJ*. 2019;364:l476.
10. Marchioli R, Finazzi G, Specchia G, Cacciola R, Cavazzina R, Cilloni D, De Stefano V, Elli E, Iurlo A, Latagliata R, et al. Cardiovascular events and intensity of treatment in polycythemia vera. *N Engl J Med*. 2013;368:22–33.
11. Lee G, Choi S, Kim K, Yun JM, Son JS, Jeong SM, Kim SM, Kim YY, Park SY, Koh Y, et al. Association between changes in hemoglobin concentration and cardiovascular risks and all-cause mortality among young women. *J Am Heart Assoc*. 2018;7:e008147. 10.1161/JAHA.117.008147.
12. Braekkan SK, Mathiesen EB, Njolstad I, Wilsgaard T, Hansen JB. Hematocrit and risk of venous thromboembolism in a general population: the Tromso study. *Haematologica*. 2010;95:270–275.
13. Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, Mead D, Bouman H, Riveros-Mckay F, Kostadima MA, et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell*. 2016;167:1415–1429.e19.
14. Burgess S, Butterworth A, Malarstig A, Thompson SG. Use of mendelian randomisation to assess potential benefit of clinical intervention. *BMJ*. 2012;345:e7325.
15. Burgess S, Thompson SG. Multivariable mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol*. 2015;181:251–260.
16. Zhong Y, Lin SL, Schooling CM. The effect of hematocrit and hemoglobin on the risk of ischemic heart disease: a mendelian randomization study. *Prev Med*. 2016;91:351–355.
17. Zuber V, Colijn JM, Klaver C, Burgess S. Selecting likely causal risk factors from high-throughput experiments using multivariable mendelian randomization. *Nat Commun*. 2020;11:29.

18. Moore C, Sambrook J, Walker M, Tolkien Z, Kaptoge S, Allen D, Mehenny S, Mant J, Di Angelantonio E, Thompson SG, et al. The INTERVAL trial to determine whether intervals between blood donations can be safely and acceptably decreased to optimise blood supply: study protocol for a randomised controlled trial. *Trials*. 2014;15:363.
19. Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjalmsson BJ, Finucane HK, Salem RM, Chasman DI, Ridker PM, Neale BM, Berger B, et al. Efficient bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet*. 2015;47:284–290.
20. Jick H, Slone D, Westerholm B, Inman WH, Vessey MP, Shapiro S, Lewis GP, Worcester J. Venous thromboembolic disease and abo blood type: a cooperative study. *Lancet*. 1969;1:539–542.
21. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562:203–209.
22. Martinelli I, De Stefano V, Mannucci PM. Inherited risk factors for venous thromboembolism. *Nat Rev Cardiol*. 2014;11:140–156.
23. Haworth S, Mitchell R, Corbin L, Wade KH, Dudding T, Budu-Aggrey A, Carslake D, Hemani G, Paternoster L, Smith GD, et al. Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nat Commun*. 2019;10:333.
24. Schnier C, Sudlow C; with input from members of the UK Biobank Follow-up and Outcomes Adjudication Group. Algorithmically-defined health outcomes. Published January 2017. https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/alg_outcome_main.pdf. Accessed May 29, 2020.
25. Gregson J, Kaptoge S, Bolton T, Pennells L, Willeit P, Burgess S, Bell S, Sweeting M, Rimm EB, Kabrhel C, et al.; Emerging Risk Factors Consortium. Cardiovascular risk factors associated with venous thromboembolism. *JAMA Cardiol*. 2019;4:163–173.
26. Bowden J, Smith GD, Haycock PC, Burgess S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol*. 2016;40:304–314.
27. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. *Int J Epidemiol*. 2016;45:1961–1974.
28. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases. *Nat Genet*. 2018;50:693–698.
29. Staley JR, Blackshaw J, Kamat MA, Ellis S, Surendran P, Sun BB, Paul DS, Freitag D, Burgess S, Danesh J, et al. Phenoscanner: a database of human genotype-phenotype associations. *Bioinformatics*. 2016;32:3207–3209.
30. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol*. 2016;40:597–608.
31. Coyne DW. The health-related quality of life was not improved by targeting higher hemoglobin in the normal hematocrit trial. *Kidney Int*. 2012;82:235–241.
32. Rother RP, Bell L, Hillmen P, Gladwin MT. The clinical sequelae of intravascular hemolysis and extracellular plasma hemoglobin: a novel mechanism of human disease. *JAMA*. 2005;293:1653–1662.
33. Silvain J, Pena A, Cayla G, Brieger D, Bellemain-Appaix A, Chastre T, Vignalou JB, Beygui F, Barthelemy O, Collet JP, et al. Impact of red blood cell transfusion on platelet activation and aggregation in healthy volunteers: results of the transfusion study. *Eur Heart J*. 2010;31:2816–2821.
34. Studt JD, Kremer Hovinga JA, Antoine G, Hermann M, Rieger M, Scheiflinger F, Lammle B. Fatal congenital thrombotic thrombocytopenic purpura with apparent ADAMTS13 inhibitor: in vitro inhibition of ADAMTS13 activity by hemoglobin. *Blood*. 2005;105:542–544.
35. Schooling CM, Luo S, Johnson G. ADAMTS-13 activity and ischemic heart disease: a mendelian randomization study. *J Thromb Haemost*. 2018;16:2270–2275.
36. Radomski MW, Palmer RM, Moncada S. Endogenous nitric oxide inhibits human platelet adhesion to vascular endothelium. *Lancet*. 1987;2:1057–1058.
37. Gkatzionis A, Burgess S. Contextualizing selection bias in mendelian randomization: how bad is it likely to be? *Int J Epidemiol*. 2019;48:691–701.
38. Murphy WG. The sex difference in haemoglobin levels in adults—mechanisms, causes, and consequences. *Blood Rev*. 2014;28:41–47.
39. Bleker SM, Coppens M, Middeldorp S. Sex, thrombosis and inherited thrombophilia. *Blood Rev*. 2014;28:123–133.
40. Barbui T, De Stefano V, Ghirardi A, Masciulli A, Finazzi G, Vannucchi AM. Different effect of hydroxyurea and phlebotomy on prevention of arterial and venous thrombosis in polycythemia vera. *Blood Cancer J*. 2018;8:124.
41. Senchenkova EY, Russell J, Almeida-Paula LD, Harding JW, Granger DN. Angiotensin II-mediated microvascular thrombosis. *Hypertension*. 2010;56:1089–1095.

# SUPPLEMENTAL MATERIAL

**Table S1. Ranking of red blood cell traits for venous thromboembolism in the UK Biobank top panel) ranking of exposures according to their marginal inclusion probability and bottom panel) ranking of models (i.e. sets of exposures) according to their posterior probability**

Ranking of exposures

| | Exposure | Marginal inclusion probability | Model-averaged causal estimate (OR) |
|---|---|---|---|
| 1 | HGB | 0.90 | 1.21 |
| 2 | HCT | 0.24 | 0.96 |
| 3 | HLSR | 0.18 | 0.98 |
| 4 | RET% | 0.11 | 0.99 |
| 5 | MCHC | 0.09 | 1.01 |
| 6 | RBC | 0.09 | 1.00 |
| 7 | MCH | 0.08 | 1.01 |
| 8 | HLSR% | 0.07 | 1.00 |
| 9 | RET | 0.07 | 1.00 |
| 10 | IRF | 0.07 | 1.01 |

Ranking of models (i.e. sets of exposures)

| | Exposure(s) | Posterior probability | Model-specific causal estimate (OR) |
|---|---|---|---|
| 1 | HGB | 0.45 | 1.16 |
| 2 | HCT, HGB | 0.07 | 0.82, 1.39 |
| 3 | HGB, HLSR | 0.04 | 1.20, 0.94 |
| 4 | HGB, RBC | 0.03 | 1.21, 0.94 |
| 5 | HCT, HGB, HLSR | 0.03 | 0.76, 1.53, 0.92 |
| 6 | HGB, MCH | 0.02 | 1.14, 1.04 |
| 7 | MCH, RBC | 0.02 | 1.15, 1.16 |
| 8 | HCT, HGB, RET% | 0.02 | 0.74, 1.56, 0.92 |
| 9 | HGB, MCHC | 0.02 | 1.14, 1.07 |
| 10 | HGB, HLSR, RET% | 0.01 | 1.25, 0.70, 1.32 |

HGB, haemoglobin concentration; HCT, haematocrit; MCHC, mean corpuscular haemoglobin concentration; HLSR, high light scatter reticulocyte count; RBC, red blood cell count; MCH, mean corpuscular haemoglobin; RET%, reticulocyte fraction of red cells; RET, reticulocyte count; HLSR%, high light scatter reticulocyte fraction of red cells; MCV, mean corpuscular volume; IRF, immature fraction of reticulocytes; RDW, red cell distribution width. Calculation is based on 648 genetic variants, using $\sigma^2 = 0.25$ as prior variance and $p = 0.1$ as prior probability, corresponding to *a priori* expecting one causal factor.

**Table S2. Q statistics using n = 648 genetic variants for the best individual models and the maximum Q of each variant among these models for diagnostics.**

| No | Variant | Gene | Q M1 | Q M2 | Q M3 | Q M4 | Q M5 | max Q |
|---|---|---|---|---|---|---|---|---|
| 1 | rs77542162 | ABCA6 | **20.79** | **20.82** | **20.43** | **20.34** | **20.36** | **20.82** |
| 2 | rs174533 | MYRF | **16.28** | **16.31** | **15.94** | **15.72** | **15.88** | **16.31** |
| 3 | rs11187938 | TBC1D12 | **15.92** | **15.90** | **15.40** | **15.69** | **15.22** | **15.92** |
| 4 | rs738408 | PNPLA3 | **13.94** | **14.09** | **14.08** | **14.17** | **14.33** | **14.33** |
| 5 | rs3747207 | PNPLA3 | **13.49** | **13.67** | **13.63** | **13.73** | **13.91** | **13.91** |
| 6 | rs139974673 | CATSPER2P1 | **11.39** | **11.15** | **11.19** | **11.15** | **10.79** | **11.39** |
| 7 | rs147233090 | CATSPER2P1 | **11.00** | **10.76** | **10.77** | **10.77** | **10.39** | **11.00** |
| 8 | rs78378222 | TP53 | 7.99 | 8.29 | 8.37 | 8.44 | 8.89 | 8.89 |
| 9 | rs11122449 | GALNT2 | 8.74 | 8.58 | 8.71 | 8.75 | 8.48 | 8.75 |
| 10 | rs41282676 | EIF2AK1 | 7.49 | 7.44 | 8.42 | 7.45 | 8.63 | 8.63 |
| 11 | rs2835349 | AP000695.6 | 8.42 | 7.91 | 8.43 | 8.23 | 7.74 | 8.43 |
| 12 | rs35979828 | NFE2 | 8.08 | 8.34 | 7.90 | 7.68 | 8.20 | 8.34 |
| 13 | rs9535495 | DLEU7 | 7.71 | 7.83 | 7.04 | 7.69 | 7.00 | 7.83 |
| 14 | rs17248895 | PLEK2 | 7.29 | 7.36 | 7.53 | 7.20 | 7.69 | 7.69 |
| 15 | rs1339847 | TRIM58 | 7.65 | 7.34 | 5.16 | 7.69 | 4.19 | 7.69 |
| 16 | rs4859682 | SHROOM3 | 7.34 | 6.85 | 7.46 | 7.06 | 6.83 | 7.46 |
| 17 | rs6880621 | CTD-2197M16.1 | 6.71 | 7.12 | 6.65 | 6.83 | 7.19 | 7.19 |
| 18 | rs972761 | CTD-2197M16.1 | 6.45 | 6.83 | 6.39 | 6.55 | 6.90 | 6.90 |
| 19 | rs6712203 | COBLL1 | 6.44 | 6.73 | 5.99 | 6.43 | 6.22 | 6.73 |
| 20 | rs4434553 | TFR2 | 6.08 | 5.10 | 6.67 | 5.13 | 5.45 | 6.67 |
| 21 | rs61750953 | EGLN2 | 6.27 | 6.55 | 5.99 | 6.33 | 6.28 | 6.55 |
| 22 | rs73652622 | MIR4289 | 6.01 | 6.55 | 5.37 | 5.93 | 5.88 | 6.55 |
| 23 | rs78415359 | CTIF | 6.16 | 6.27 | 6.17 | 6.37 | 6.33 | 6.37 |
| 24 | rs833805 | RP5-1120P11.1 | 5.74 | 6.32 | 5.51 | 5.99 | 6.20 | 6.32 |
| 25 | rs72996113 | RN7SL222P | 5.67 | 6.12 | 5.01 | 5.56 | 5.39 | 6.12 |
| 26 | rs13389219 | COBLL1 | 5.82 | 6.07 | 5.40 | 5.79 | 5.61 | 6.07 |
| 27 | rs964184 | ZNF259 | 5.51 | 5.02 | 6.07 | 5.57 | 5.53 | 6.07 |
| 28 | rs12548939 | PVT1 | 5.89 | 5.73 | 5.99 | 5.51 | 5.79 | 5.99 |
| 29 | rs1569419 | PRDM16 | 5.97 | 5.68 | 5.99 | 5.55 | 5.60 | 5.99 |
| 30 | rs56235845 | RGS14 | 5.02 | 5.28 | 5.08 | 5.15 | 5.46 | 5.46 |

Q, Q statistics; M: Model. M1 (HGB), M2 (HCT and HGB), M3 (HGB and HLSR), M4 (HGB and RBC), M5 (HCT, HGB and HLSR). Variants with Q statistics > 10 are given in bold. This table displays the 30 variants with the largest maximum Q and the gene region they fall in.

**Table S3. Cook's distance using n = 648 genetic variants for the best individual models and the maximum Cook's distance of each variant among these models for diagnostics.**

| | Variant | Gene | Cd M1 | Cd M2 | Cd M3 | Cd M4 | Cd M5 | max Cd |
|---|---|---|---|---|---|---|---|---|
| 1 | rs77542162 | ABCA6 | 0.079 | 0.04 | 0.041 | 0.042 | 0.027 | 0.079 |
| 2 | rs1339847 | TRIM58 | 0 | 0.002 | 0.069 | 0 | 0.044 | 0.069 |
| 3 | rs73728279 | PRKAG2 | 0.05 | 0.028 | 0.027 | 0.025 | 0.019 | 0.05 |
| 4 | rs10224210 | PRKAG2 | 0.049 | 0.027 | 0.027 | 0.024 | 0.019 | 0.049 |
| 5 | rs198851 | HIST1H1T | 0.035 | 0.043 | 0.017 | 0.046 | 0.031 | 0.046 |
| 6 | rs1799945 | HFE | 0.034 | 0.042 | 0.016 | 0.045 | 0.031 | 0.045 |
| 7 | rs174533 | MYRF | 0.04 | 0.02 | 0.021 | 0.024 | 0.014 | 0.04 |
| 8 | rs833805 | RP5-1120P11.1 | 0.034 | 0.025 | 0.017 | 0.019 | 0.016 | 0.034 |
| 9 | rs10168349 | PRKCE | 0.03 | 0.014 | 0.017 | 0.014 | 0.01 | 0.03 |
| 10 | rs738408 | PNPLA3 | 0.028 | 0.015 | 0.015 | 0.015 | 0.01 | 0.028 |
| 11 | rs10495928 | PRKCE | 0.028 | 0.013 | 0.016 | 0.013 | 0.009 | 0.028 |
| 12 | rs147233090 | CATSPER2P1 | 0.027 | 0.014 | 0.014 | 0.014 | 0.01 | 0.027 |
| 13 | rs3747207 | PNPLA3 | 0.027 | 0.014 | 0.014 | 0.015 | 0.01 | 0.027 |
| 14 | rs2968478 | PIEZO1 | 0.026 | 0.019 | 0.016 | 0.013 | 0.014 | 0.026 |
| 15 | rs139974673 | CATSPER2P1 | 0.024 | 0.013 | 0.012 | 0.013 | 0.009 | 0.024 |
| 16 | rs2106786 | SPPL2C | 0.022 | 0.014 | 0.012 | 0.014 | 0.012 | 0.022 |
| 17 | rs17563683 | LINC02210-CRHR1 | 0.021 | 0.013 | 0.012 | 0.013 | 0.011 | 0.021 |
| 18 | rs4606752 | KANSL1 | 0.02 | 0.013 | 0.012 | 0.013 | 0.012 | 0.02 |
| 19 | rs4434553 | TFR2 | 0.012 | 0.019 | 0.011 | 0.017 | 0.015 | 0.019 |
| 20 | rs2923411 | RP11-503E24.3 | 0.01 | 0.017 | 0.005 | 0.005 | 0.011 | 0.017 |
| 21 | rs551238 | EPO | 0.01 | 0.016 | 0.006 | 0.014 | 0.011 | 0.016 |
| 22 | rs2835349 | AP000695.6 | 0.016 | 0.012 | 0.008 | 0.008 | 0.008 | 0.016 |
| 23 | rs11970772 | CCND3 | 0 | 0 | 0.005 | 0.016 | 0.004 | 0.016 |
| 24 | rs12548939 | PVT1 | 0.014 | 0.007 | 0.007 | 0.009 | 0.005 | 0.014 |
| 25 | rs972761 | CTD-2197M16.1 | 0.014 | 0.01 | 0.007 | 0.007 | 0.007 | 0.014 |
| 26 | rs6880621 | CTD-2197M16.1 | 0.013 | 0.01 | 0.007 | 0.007 | 0.007 | 0.013 |
| 27 | rs837763 | PIEZO1 | 0.013 | 0.009 | 0.008 | 0.007 | 0.006 | 0.013 |
| 28 | rs34164109 | HBS1L | 0.012 | 0.013 | 0.009 | 0.013 | 0.008 | 0.013 |
| 29 | rs72805692 | HK1 | 0.009 | 0.009 | 0.008 | 0.004 | 0.013 | 0.013 |
| 30 | rs592423 | AL356739.1 | 0.002 | 0.001 | 0.009 | 0.013 | 0.005 | 0.013 |
| | threshold | | 0.455 | 0.694 | 0.694 | 0.694 | 0.789 | |

Cd Cook distance; M: Model. M1 (HGB), M2 (HCT and HGB), M3 (HGB and HLSR), M4 (HGB and RBC), M5 (HCT, HGB and HLSR). The final line gives the suggested cut-off for Cook's distance. This table displays the 30 variants with the largest maximum Cook's distance and the gene region they fall in.

**Table S4. Q statistics using n = 641 genetic variants after exclusion of outlying variants, for the best individual models and the maximum Q of each variant among these models for diagnostics.**

| No | Variant | Gene | Q M1 | Q M2 | Q M3 | Q M4 | Q M5 | max Q |
|---|---|---|---|---|---|---|---|---|
| 1 | rs78378222 | TP53 | 8.00 | 8.29 | 8.34 | 8.85 | 8.40 | 8.85 |
| 2 | rs11122449 | GALNT2 | 8.75 | 8.59 | 8.72 | 8.49 | 8.76 | 8.76 |
| 3 | rs41282676 | EIF2AK1 | 7.50 | 7.44 | 8.34 | 8.54 | 7.45 | 8.54 |
| 4 | rs2835349 | AP000695.6 | 8.40 | 7.90 | 8.42 | 7.75 | 8.24 | 8.42 |
| 5 | rs35979828 | NFE2 | 8.08 | 8.33 | 7.92 | 8.20 | 7.72 | 8.33 |
| 6 | rs9535495 | DLEU7 | 7.72 | 7.83 | 7.11 | 7.06 | 7.70 | 7.83 |
| 7 | rs1339847 | TRIM58 | 7.65 | 7.35 | 5.36 | 4.40 | 7.69 | 7.69 |
| 8 | rs17248895 | PLEK2 | 7.30 | 7.36 | 7.51 | 7.67 | 7.21 | 7.67 |
| 9 | rs4859682 | SHROOM3 | 7.33 | 6.84 | 7.44 | 6.83 | 7.08 | 7.44 |
| 10 | rs6880621 | CTD-2197M16.1 | 6.72 | 7.13 | 6.67 | 7.19 | 6.83 | 7.19 |
| 11 | rs972761 | CTD-2197M16.1 | 6.46 | 6.84 | 6.41 | 6.90 | 6.55 | 6.90 |
| 12 | rs6712203 | COBLL1 | 6.43 | 6.71 | 6.02 | 6.25 | 6.42 | 6.71 |
| 13 | rs4434553 | TFR2 | 6.07 | 5.10 | 6.60 | 5.42 | 5.20 | 6.60 |
| 14 | rs61750953 | EGLN2 | 6.29 | 6.56 | 6.03 | 6.31 | 6.34 | 6.56 |
| 15 | rs73652622 | MIR4289 | 6.01 | 6.54 | 5.42 | 5.92 | 5.93 | 6.54 |
| 16 | rs78415359 | CTIF | 6.17 | 6.28 | 6.18 | 6.33 | 6.36 | 6.36 |
| 17 | rs833805 | RP5-1120P11.1 | 5.77 | 6.33 | 5.55 | 6.22 | 5.99 | 6.33 |
| 18 | rs72996113 | RN7SL222P | 5.67 | 6.12 | 5.07 | 5.44 | 5.57 | 6.12 |
| 19 | rs13389219 | COBLL1 | 5.81 | 6.06 | 5.43 | 5.63 | 5.78 | 6.06 |
| 20 | rs964184 | ZNF259 | 5.51 | 5.02 | 6.02 | 5.50 | 5.57 | 6.02 |
| 21 | rs12548939 | PVT1 | 5.91 | 5.74 | 5.99 | 5.79 | 5.55 | 5.99 |
| 22 | rs1569419 | PRDM16 | 5.97 | 5.68 | 5.98 | 5.61 | 5.59 | 5.98 |
| 23 | rs56235845 | RGS14 | 5.03 | 5.28 | 5.09 | 5.45 | 5.14 | 5.45 |
| 24 | rs17006441 | MITF | 5.28 | 5.03 | 5.38 | 5.08 | 5.11 | 5.38 |
| 25 | rs2106786 | SPPL2C | 4.52 | 4.89 | 4.73 | 5.31 | 4.87 | 5.31 |
| 26 | rs17563683 | LINC02210-CRHR1 | 4.45 | 4.80 | 4.68 | 5.25 | 4.76 | 5.25 |
| 27 | rs159058 | NOL4L | 4.85 | 4.95 | 4.65 | 4.72 | 5.24 | 5.24 |
| 28 | rs4606752 | KANSL1 | 4.36 | 4.73 | 4.62 | 5.22 | 4.69 | 5.22 |
| 29 | rs717662 | RN7SL222P | 4.72 | 5.13 | 4.20 | 4.54 | 4.65 | 5.13 |
| 30 | rs3812049 | SLC12A2 | 4.95 | 5.09 | 4.91 | 5.09 | 4.93 | 5.09 |

Q, Q statistics; M: model. M1 (HGB), M2 (HCT and HGB), M3 (HGB and HLSR), M4 (HCT, HGB and HLSR), M5 (HGB and MCHC). This table displays the 30 variants with the largest maximum Q and the gene region they fall in.

**Table S5. Cook's distance using n = 641 genetic variants after exclusion of outlying variants, the best individual models and the maximum Cook's distance of each variant among these models for diagnostics.**

| No | Variant | Gene | Cd M1 | Cd M2 | Cd M3 | Cd M4 | Cd M5 | max Cd |
|---|---|---|---|---|---|---|---|---|
| 1 | rs1339847 | TRIM58 | 0.001 | 0.002 | 0.08 | 0.052 | 0 | 0.08 |
| 2 | rs73728279 | PRKAG2 | 0.057 | 0.032 | 0.031 | 0.022 | 0.028 | 0.057 |
| 3 | rs10224210 | PRKAG2 | 0.056 | 0.031 | 0.03 | 0.021 | 0.027 | 0.056 |
| 4 | rs198851 | HIST1H1T | 0.039 | 0.048 | 0.019 | 0.035 | 0.05 | 0.05 |
| 5 | rs1799945 | HFE | 0.039 | 0.048 | 0.019 | 0.035 | 0.049 | 0.049 |
| 6 | rs833805 | RP5-1120P11.1 | 0.04 | 0.028 | 0.02 | 0.019 | 0.022 | 0.04 |
| 7 | rs10168349 | PRKCE | 0.034 | 0.016 | 0.019 | 0.011 | 0.016 | 0.034 |
| 8 | rs10495928 | PRKCE | 0.032 | 0.015 | 0.018 | 0.01 | 0.015 | 0.032 |
| 9 | rs2968478 | PIEZO1 | 0.03 | 0.022 | 0.018 | 0.016 | 0.015 | 0.03 |
| 10 | rs2106786 | SPPL2C | 0.025 | 0.016 | 0.014 | 0.013 | 0.016 | 0.025 |
| 11 | rs17563683 | LINC02210-CRHR1 | 0.024 | 0.015 | 0.014 | 0.013 | 0.015 | 0.024 |
| 12 | rs4606752 | KANSL1 | 0.023 | 0.015 | 0.013 | 0.013 | 0.015 | 0.023 |
| 13 | rs4434553 | TFR2 | 0.013 | 0.022 | 0.013 | 0.016 | 0.02 | 0.022 |
| 14 | rs2923411 | RP11-503E24.3 | 0.011 | 0.019 | 0.006 | 0.012 | 0.006 | 0.019 |
| 15 | rs551238 | EPO | 0.011 | 0.019 | 0.007 | 0.012 | 0.016 | 0.019 |
| 16 | rs2835349 | AP000695.6 | 0.018 | 0.013 | 0.009 | 0.009 | 0.01 | 0.018 |
| 17 | rs11970772 | CCND3 | 0 | 0 | 0.006 | 0.004 | 0.017 | 0.017 |
| 18 | rs34164109 | HBS1L | 0.014 | 0.014 | 0.01 | 0.009 | 0.017 | 0.017 |
| 19 | rs12548939 | PVT1 | 0.016 | 0.008 | 0.008 | 0.006 | 0.01 | 0.016 |
| 20 | rs972761 | CTD-2197M16.1 | 0.016 | 0.011 | 0.008 | 0.008 | 0.008 | 0.016 |
| 21 | rs6880621 | CTD-2197M16.1 | 0.015 | 0.011 | 0.008 | 0.008 | 0.008 | 0.015 |
| 22 | rs9376090 | HBS1L | 0.012 | 0.013 | 0.008 | 0.008 | 0.015 | 0.015 |
| 23 | rs837763 | PIEZO1 | 0.015 | 0.01 | 0.009 | 0.007 | 0.008 | 0.015 |
| 24 | rs4859682 | SHROOM3 | 0.014 | 0.011 | 0.008 | 0.007 | 0.008 | 0.014 |
| 25 | rs72805692 | HK1 | 0.011 | 0.01 | 0.009 | 0.014 | 0.005 | 0.014 |
| 26 | rs5758896 | A4GALT | 0.014 | 0.007 | 0.007 | 0.005 | 0.008 | 0.014 |
| 27 | rs592423 | AL356739.1 | 0.002 | 0.001 | 0.01 | 0.006 | 0.014 | 0.014 |
| 28 | rs7541039 | PROX1 | 0.014 | 0.01 | 0.007 | 0.007 | 0.007 | 0.014 |
| 29 | rs12548864 | PVT1 | 0.013 | 0.007 | 0.007 | 0.005 | 0.008 | 0.013 |
| 30 | rs41282676 | EIF2AK1 | 0.001 | 0 | 0.013 | 0.01 | 0 | 0.013 |
|  | threshold |  | 0.455 | 0.694 | 0.694 | 0.790 | 0.694 |  |

CD Cook distance; M: model. M1 (HGB), M2 (HCT and HGB), M3 (HGB and HLSR), M4 (HCT, HGB and HLSR), M5 (HGB and MCHC). The final line gives the suggested cut-off for Cook's distance. This table displays the 30 variants with the largest maximum Cook's distance and the gene region they fall in.

**Table S6. Parameter check for the prior probability p, ranging from p=0.2 to 0.4.**

| p = 0.2 | | | |
|---|---|---|---|
| No. | Exposure | Marginal inclusion probability | model-averaged causal effect (OR) |
| 1 | HGB | 0.874 | 1.32 |
| 2 | HCT | 0.539 | 0.88 |
| 3 | HLSR | 0.306 | 0.96 |
| 4 | IRF | 0.298 | 1.05 |
| 5 | RET% | 0.260 | 0.97 |
| 6 | HLSR% | 0.244 | 1.00 |
| 7 | MCHC | 0.231 | 1.03 |
| 8 | RET | 0.223 | 0.99 |
| 9 | RBC | 0.087 | 1.00 |
| 10 | MCH | 0.078 | 1.01 |
| p = 0.3 | | | |
| No. | Exposure | Marginal inclusion probability | model-averaged causal effect (OR) |
| 1 | HGB | 0.855 | 1.38 |
| 2 | HCT | 0.660 | 0.85 |
| 3 | IRF | 0.451 | 1.08 |
| 4 | HLSR | 0.368 | 0.95 |
| 5 | HLSR% | 0.363 | 0.99 |
| 6 | RET% | 0.347 | 0.95 |
| 7 | RET | 0.335 | 0.99 |
| 8 | MCHC | 0.305 | 1.04 |
| 9 | RBC | 0.105 | 1.00 |
| 10 | MCH | 0.093 | 1.01 |
| p = 0.4 | | | |
| No. | Exposure | Marginal inclusion probability | model-averaged causal effect (OR) |
| 1 | HGB | 0.842 | 1.40 |
| 2 | HCT | 0.706 | 0.83 |
| 3 | IRF | 0.541 | 1.11 |
| 4 | HLSR% | 0.445 | 0.98 |
| 5 | RET | 0.415 | 0.98 |
| 6 | HLSR | 0.408 | 0.96 |
| 7 | RET% | 0.403 | 0.95 |
| 8 | MCHC | 0.352 | 1.04 |
| 9 | RBC | 0.145 | 1.01 |
| 10 | MCH | 0.130 | 1.02 |

p = 0.2 to 0.4 reflects 2.4 to 4.8 expected causal exposures. OR, odds, ratio.

**Table S7. Ranking of blood cell traits for venous thromboembolism with different selections of exposures, according to their marginal inclusion probability.**

| Cell lineage: 12 red blood cell and 4 platelet traits, n = 961 | | | |
|---|---|---|---|
| No. | Exposure | Marginal inclusion probability | model-averaged causal effect (OR) |
| 1 | HGB | 0.915 | 1.22 |
| 2 | HCT | 0.211 | 0.96 |
| 3 | HLSR | 0.146 | 0.98 |
| 4 | RBC | 0.124 | 1.00 |
| 5 | MCH | 0.101 | 1.01 |
| 6 | IRF | 0.088 | 1.02 |
| 7 | RET% | 0.080 | 1.00 |
| 8 | HLSR% | 0.076 | 1.00 |
| 9 | RET | 0.071 | 0.99 |
| 10 | MCHC | 0.055 | 1.01 |
| Cell lineage: 11 red blood cell traits (Exclude hematocrit), n = 578 | | | |
| No. | Exposure | Marginal inclusion probability | model-averaged causal effect (OR) |
| 1 | HGB | 0.527 | 1.07 |
| 2 | MCHC | 0.329 | 1.04 |
| 3 | HLSR | 0.189 | 0.98 |
| 4 | RET% | 0.176 | 0.98 |
| 5 | MCH | 0.160 | 1.01 |
| 6 | HLSR% | 0.103 | 1.00 |
| 7 | RET | 0.092 | 1.00 |
| 8 | IRF | 0.089 | 1.01 |
| 9 | RBC | 0.071 | 1.00 |
| 10 | MCV | 0.061 | 1.00 |
| Cell lineage: Red blood cell traits (Exclude hemoglobin), n = 566 | | | |
| No. | Exposure | Marginal inclusion probability | model-averaged causal effect (OR) |
| 1 | MCH | 0.436 | 1.04 |
| 2 | MCHC | 0.343 | 1.04 |
| 3 | HCT | 0.177 | 1.02 |
| 4 | RET% | 0.158 | 0.97 |
| 5 | RBC | 0.154 | 1.02 |
| 6 | MCV | 0.150 | 1.01 |
| 7 | HLSR | 0.125 | 1.00 |
| 8 | HLSR% | 0.065 | 1.00 |
| 9 | RET | 0.064 | 1.01 |
| 10 | IRF | 0.048 | 1.01 |

OR, odds ratio. We used p = 0.1 as prior probability and excluded outlying variants in the above analyses.

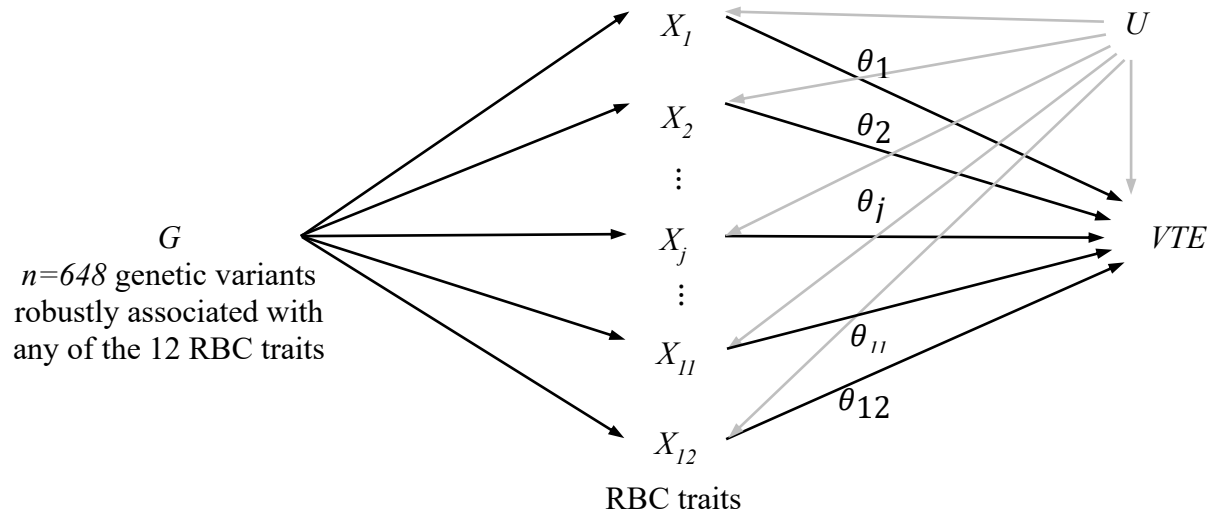
**Table S8. Genetic variants predicting haemoglobin concentration used in the univariable Mendelian randomization analyses.**

| | | | | | | Haemoglobin | | VTE | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Variant | Info | EA | OA | EAF | $R^2$ (%) | Beta | SE | Beta | SE | Exclude | PhenoScanner |
| rs1010269 | 0.98 | A | G | 0.17 | 0.028 | 0.031 | 0.005 | 0.014 | 0.020 | | |
| rs10495928 | 1.00 | A | G | 0.66 | 0.240 | 0.073 | 0.004 | -0.009 | 0.015 | | |
| rs10899133 | 1.00 | T | C | 0.11 | 0.025 | 0.036 | 0.006 | 0.003 | 0.024 | | |
| rs11072567 | 1.00 | A | G | 0.49 | 0.066 | 0.036 | 0.004 | 0.000 | 0.015 | | |
| rs11122272 | 1.00 | G | A | 0.63 | 0.033 | 0.027 | 0.004 | 0.019 | 0.015 | | |
| rs115986297 | 1.00 | A | G | 0.46 | 0.038 | 0.028 | 0.004 | 0.005 | 0.015 | | |
| rs11749327 | 0.98 | A | C | 0.31 | 0.022 | 0.023 | 0.004 | 0.012 | 0.016 | | |
| rs11772705 | 1.00 | C | T | 0.29 | 0.036 | 0.030 | 0.004 | 0.004 | 0.016 | | |
| rs1181870 | 0.95 | A | C | 0.24 | 0.037 | 0.032 | 0.004 | -0.019 | 0.018 | | |
| rs1182933 | 1.00 | T | C | 0.30 | 0.024 | 0.024 | 0.004 | -0.027 | 0.016 | √ | CHD, cholesterol |
| rs123698 | 1.00 | C | G | 0.60 | 0.036 | 0.027 | 0.004 | -0.017 | 0.015 | | |
| rs12548874 | 1.00 | C | A | 0.53 | 0.021 | 0.021 | 0.004 | -0.019 | 0.015 | | |
| rs1256061 | 1.00 | G | T | 0.52 | 0.030 | 0.024 | 0.004 | 0.023 | 0.015 | | |
| rs12811512 | 1.00 | C | T | 0.85 | 0.021 | 0.029 | 0.005 | 0.006 | 0.020 | | |
| rs128494 | 0.97 | C | T | 0.77 | 0.036 | 0.032 | 0.004 | -0.018 | 0.017 | | |
| rs12945870 | 1.00 | C | T | 0.43 | 0.026 | 0.023 | 0.004 | 0.030 | 0.015 | | |
| rs1340818 | 0.99 | C | T | 0.61 | 0.025 | 0.023 | 0.004 | -0.002 | 0.015 | | |
| rs144861591 | 0.98 | T | C | 0.08 | 0.466 | 0.181 | 0.007 | 0.033 | 0.027 | | |
| rs147233090 | 0.99 | C | T | 0.98 | 0.041 | 0.093 | 0.012 | 0.181 | 0.050 | | |
| rs17006441 | 0.99 | A | C | 0.42 | 0.025 | 0.023 | 0.004 | -0.031 | 0.015 | | |
| rs174533 | 1.00 | A | G | 0.35 | 0.036 | 0.028 | 0.004 | -0.058 | 0.015 | √ | Cholesterol |
| rs17476364 | 1.00 | C | T | 0.11 | 0.450 | 0.151 | 0.006 | 0.026 | 0.023 | | |
| rs17563683 | 1.00 | G | A | 0.23 | 0.068 | 0.043 | 0.004 | 0.043 | 0.017 | √ | Blood pressure |
| rs17773190 | 0.98 | G | A | 0.48 | 0.028 | 0.024 | 0.004 | 0.021 | 0.015 | | |
| rs184088518 | 0.93 | G | T | 0.98 | 0.047 | 0.101 | 0.012 | -0.004 | 0.048 | | |
| rs1997595 | 0.98 | A | C | 0.66 | 0.030 | 0.031 | 0.004 | 0.002 | 0.015 | | |
| rs218264 | 0.98 | A | T | 0.75 | 0.044 | 0.034 | 0.004 | 0.014 | 0.017 | | |
| rs2186037 | 1.00 | G | A | 0.52 | 0.044 | 0.024 | 0.004 | 0.003 | 0.015 | | |
| rs2230657 | 1.00 | G | A | 0.53 | 0.033 | 0.026 | 0.004 | 0.004 | 0.015 | | |
| rs2246363 | 0.99 | G | A | 0.25 | 0.023 | 0.025 | 0.004 | -0.004 | 0.017 | | |
| rs2269188 | 0.94 | G | C | 0.72 | 0.031 | 0.028 | 0.004 | 0.014 | 0.017 | | |
| rs228917 | 1.00 | T | C | 0.57 | 0.025 | 0.044 | 0.004 | 0.007 | 0.015 | | |
| rs2519796 | 0.99 | G | A | 0.33 | 0.021 | 0.024 | 0.004 | -0.006 | 0.016 | | |
| rs261332 | 1.00 | G | A | 0.79 | 0.031 | 0.025 | 0.004 | 0.030 | 0.018 | √ | Cholesterol |
| rs2870238 | 1.00 | T | C | 0.50 | 0.024 | 0.025 | 0.004 | -0.014 | 0.015 | | |
| rs2878889 | 0.99 | A | G | 0.55 | 0.031 | 0.022 | 0.004 | 0.000 | 0.015 | | |
| rs2923411 | 1.00 | C | T | 0.59 | 0.031 | 0.025 | 0.004 | 0.036 | 0.015 | | |
| rs2928166 | 0.99 | C | T | 0.13 | 0.127 | 0.037 | 0.005 | 0.029 | 0.022 | | |
| rs2968478 | 0.96 | T | G | 0.42 | 0.024 | 0.051 | 0.004 | 0.034 | 0.015 | | |
| rs35060063 | 1.00 | G | A | 0.50 | 0.023 | 0.022 | 0.004 | -0.006 | 0.015 | | |
| rs3811444 | 1.00 | T | C | 0.34 | 0.022 | 0.023 | 0.004 | -0.025 | 0.016 | | |
| rs3996993 | 1.00 | C | T | 0.53 | 0.024 | 0.021 | 0.004 | -0.022 | 0.015 | | |
| rs4073770 | 0.99 | A | T | 0.75 | 0.046 | 0.025 | 0.004 | 0.001 | 0.017 | | |
| rs442177 | 1.00 | G | T | 0.41 | 0.023 | 0.031 | 0.004 | 0.030 | 0.015 | √ | Cholesterol |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| rs447735 | 1.00 | C | T | 0.42 | 0.071 | 0.022 | 0.004 | -0.002 | 0.015 | | |
| rs4760682 | 1.00 | C | A | 0.19 | 0.026 | 0.048 | 0.005 | -0.007 | 0.019 | | |
| rs4791641 | 1.00 | C | T | 0.50 | 0.027 | 0.023 | 0.004 | -0.009 | 0.015 | √ | Cholesterol |
| rs4951074 | 0.99 | A | G | 0.10 | 0.021 | 0.039 | 0.006 | 0.044 | 0.024 | | |
| rs4957325 | 0.99 | C | T | 0.12 | 0.022 | 0.032 | 0.006 | -0.016 | 0.023 | | |
| rs554019 | 1.00 | C | T | 0.58 | 0.023 | 0.021 | 0.004 | 0.017 | 0.015 | | |
| rs56235845 | 0.99 | G | T | 0.33 | 0.026 | 0.023 | 0.004 | 0.038 | 0.016 | | |
| rs56262900 | 0.99 | A | G | 0.10 | 0.096 | 0.037 | 0.006 | 0.016 | 0.024 | | |
| rs5758896 | 1.00 | C | T | 0.59 | 0.046 | 0.031 | 0.004 | 0.034 | 0.015 | | |
| rs57908212 | 0.99 | C | T | 0.47 | 0.037 | 0.027 | 0.004 | 0.012 | 0.015 | | |
| rs58017093 | 1.00 | A | C | 0.37 | 0.031 | 0.026 | 0.004 | 0.006 | 0.015 | | |
| rs59901009 | 1.00 | T | C | 0.76 | 0.088 | 0.049 | 0.004 | 0.029 | 0.017 | | |
| rs6064559 | 1.00 | A | C | 0.60 | 0.024 | 0.022 | 0.004 | -0.003 | 0.015 | | |
| rs61750953 | 1.00 | C | T | 0.98 | 0.024 | 0.088 | 0.014 | 0.166 | 0.061 | | |
| rs62435145 | 0.94 | G | T | 0.31 | 0.036 | 0.029 | 0.004 | -0.014 | 0.016 | | |
| rs6459467 | 1.00 | G | A | 0.62 | 0.025 | 0.023 | 0.004 | -0.002 | 0.015 | | |
| rs662735 | 0.99 | A | T | 0.80 | 0.026 | 0.029 | 0.004 | -0.012 | 0.018 | | |
| rs66561647 | 0.99 | C | T | 0.66 | 0.037 | 0.029 | 0.004 | 0.031 | 0.015 | | |
| rs6665764 | 1.00 | A | G | 0.26 | 0.047 | 0.035 | 0.004 | 0.036 | 0.016 | | |
| rs66782572 | 1.00 | A | G | 0.46 | 0.021 | 0.021 | 0.004 | 0.026 | 0.015 | | |
| rs67145503 | 0.99 | A | T | 0.12 | 0.050 | 0.049 | 0.006 | 0.043 | 0.023 | | |
| rs6841433 | 1.00 | T | G | 0.82 | 0.023 | 0.028 | 0.005 | -0.018 | 0.019 | | |
| rs6967414 | 1.00 | A | G | 0.11 | 0.025 | 0.036 | 0.006 | -0.041 | 0.024 | | |
| rs7045087 | 1.00 | T | C | 0.70 | 0.023 | 0.023 | 0.004 | 0.009 | 0.016 | | |
| rs73728279 | 0.99 | G | T | 0.72 | 0.167 | 0.064 | 0.004 | -0.024 | 0.016 | | |
| rs753381 | 1.00 | C | T | 0.55 | 0.021 | 0.021 | 0.004 | 0.004 | 0.015 | √ | Cholesterol |
| rs7560180 | 0.98 | T | A | 0.22 | 0.047 | 0.037 | 0.004 | 0.013 | 0.018 | | |
| rs768090 | 0.99 | A | T | 0.32 | 0.027 | 0.025 | 0.004 | -0.029 | 0.016 | | |
| rs77542162 | 1.00 | G | A | 0.02 | 0.044 | 0.099 | 0.012 | 0.213 | 0.043 | √ | Cholesterol |
| rs7875291 | 0.99 | G | A | 0.64 | 0.047 | 0.032 | 0.004 | -0.005 | 0.015 | | |
| rs7945705 | 0.99 | G | A | 0.55 | 0.046 | 0.030 | 0.004 | -0.017 | 0.015 | | |
| rs8055546 | 1.00 | T | C | 0.07 | 0.030 | 0.048 | 0.007 | -0.031 | 0.029 | | |
| rs833805 | 0.91 | G | A | 0.89 | 0.098 | 0.070 | 0.006 | 0.069 | 0.024 | | |
| rs8887 | 0.98 | C | T | 0.57 | 0.034 | 0.027 | 0.004 | -0.019 | 0.015 | | |
| rs9376090 | 1.00 | T | C | 0.74 | 0.095 | 0.050 | 0.004 | -0.013 | 0.017 | √ | Cholesterol |
| rs9472135 | 0.99 | T | C | 0.69 | 0.056 | 0.036 | 0.004 | 0.000 | 0.016 | | |
| rs972761 | 0.98 | T | C | 0.53 | 0.032 | 0.025 | 0.004 | 0.041 | 0.015 | | |

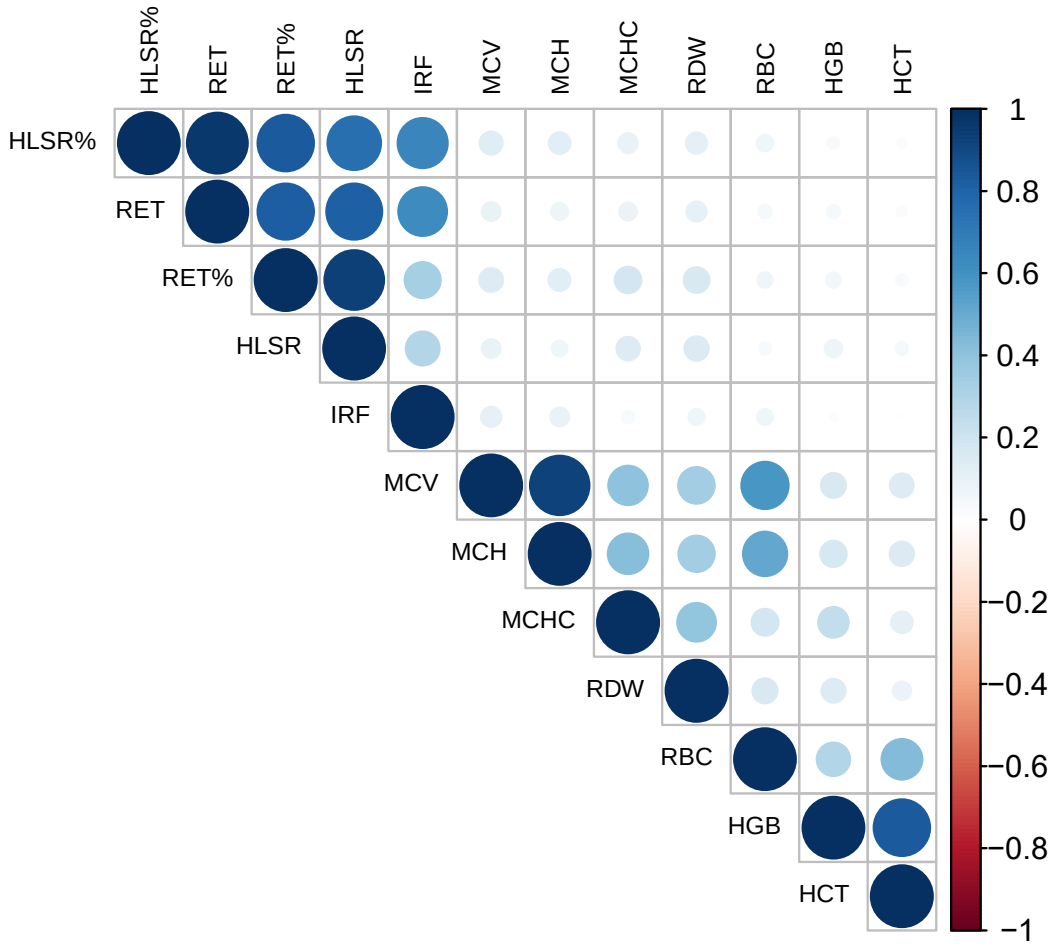
Beta (standard error) with haemoglobin concentration/ venous thromboembolism (VTE) are the changes in haemoglobin concentration (g/dL)/ log-transformed venous thromboembolism per additional copy of the effect allele; Estimates of variant on haemoglobin concentration are taken from Astle et al. Estimates of variant on venous thromboembolism are derived with individual data in the UK Biobank. EA, effect allele; OA, other allele; EAF, effect allele frequency; SE, standard error; $R^2$, the proportion of variance explained for the association between variant and haemoglobin concentration, presented in percentage. Exclude, a tick indicates the variant is associated with potential causes of venous thromboembolism ($P < 5\times10^{-8}$) based on PhenoScanner, excluded in the analysis. CHD, coronary heart disease

**Figure S1. Directed acyclic graph of instrumental variable assumptions made in multivariable Mendelian randomization.**

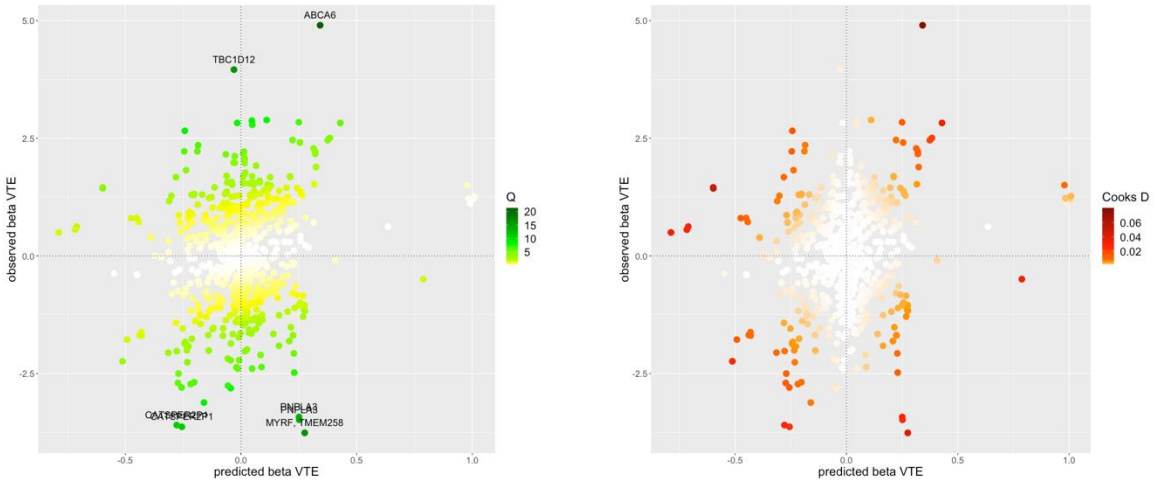



G = Genetic variants, $X_j$ = risk factor *j* for *j* = 1, …,12 red blood cell traits, U = confounders, $\theta_j$ = causal effect of risk factor *j* for *j* = 1, …,12 red blood cell traits.

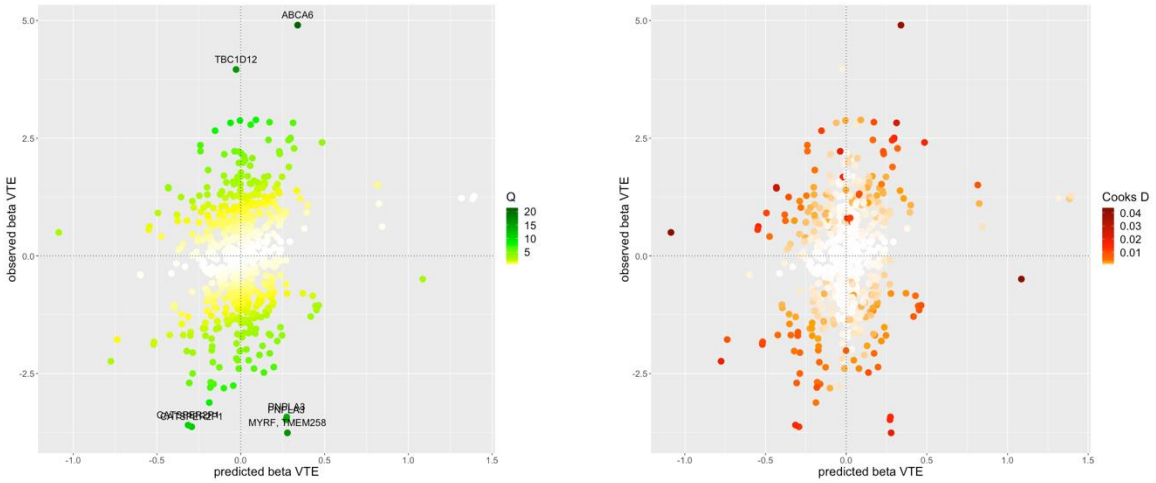
**Figure S2. Genetic correlation between 12 red blood cell traits based on the n = 648 genetic variants used as instrumental variables.**



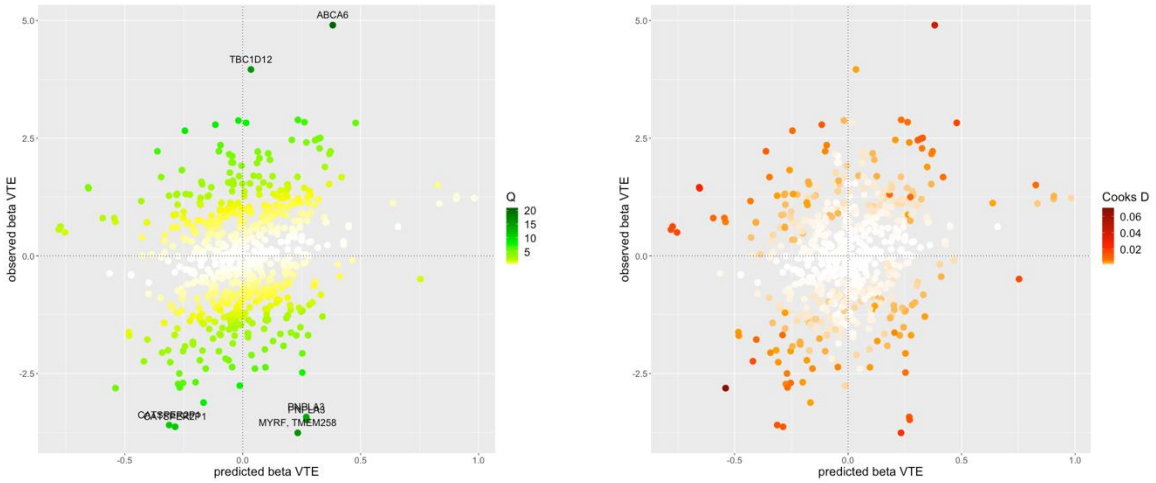
1) HGB



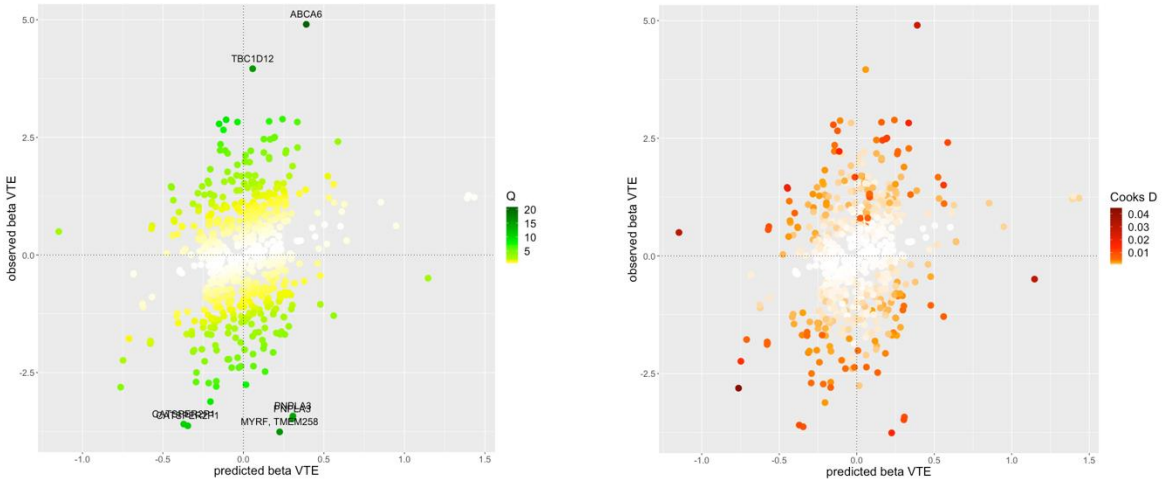

2) HCT and HGB




3) HGB and HLSR



4) HGB and RBC




5) HCT, HGB and HLSR

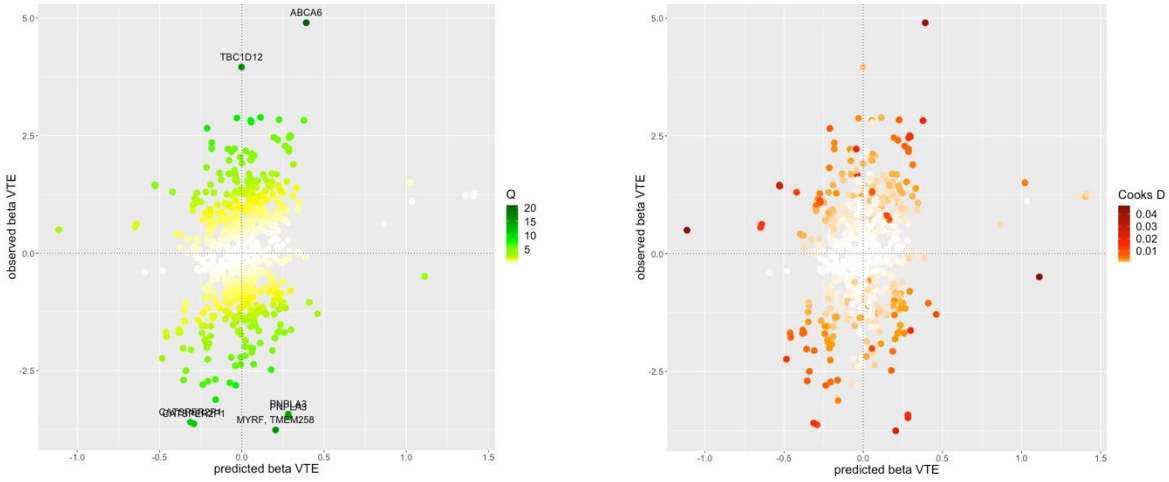
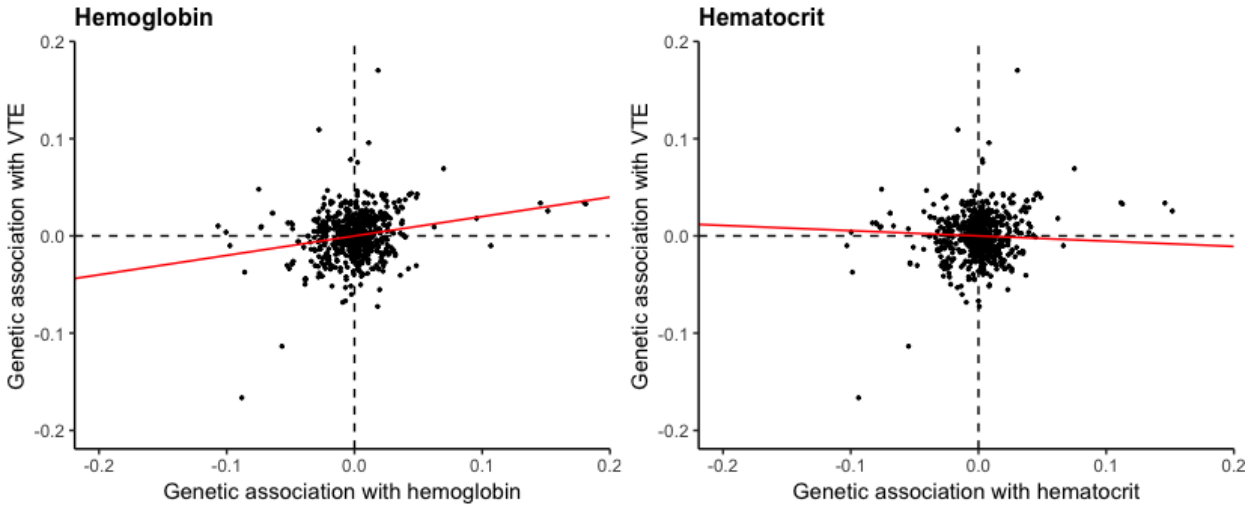


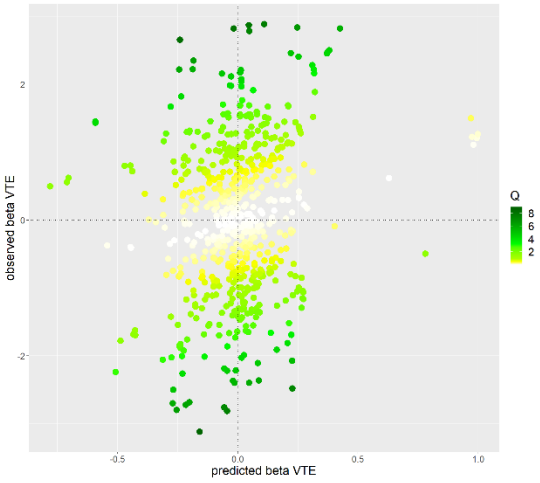

**Figure S3. Diagnostic plots of the predicted associations with venous thromboembolism (VTE) (x-axis) based on the best individual models 1 (HGB), model 2, (HCT and HGB), model 3 (HGB and HLSR), model 4 (HGB and RBC), model 5 (HCT, HGB and HLSR), against the observed associations with VTE (y-axis).** The colour code shows left) the Q-statistics for outliers and right) Cook's distance for influential points. Any genetic variant with Q-statistic larger than 10 or Cook's distance larger than median is marked by a label indicating the gene region.

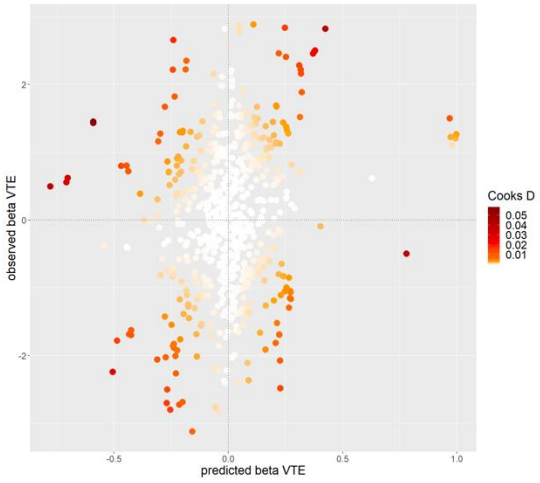
**Figure S4. Scatterplot of associations with A) haemoglobin and B) haematocrit on the x-axis against the association with venous thromboembolism (VTE) y-axis after excluding the outlying variants.**



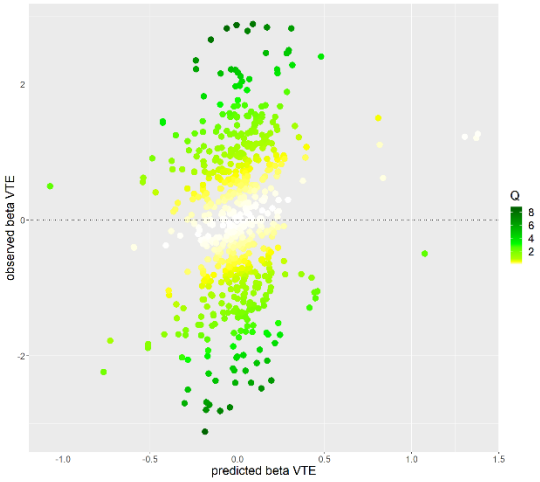

The mode-averaged causal effect of each exposure on VTE was marked in red.

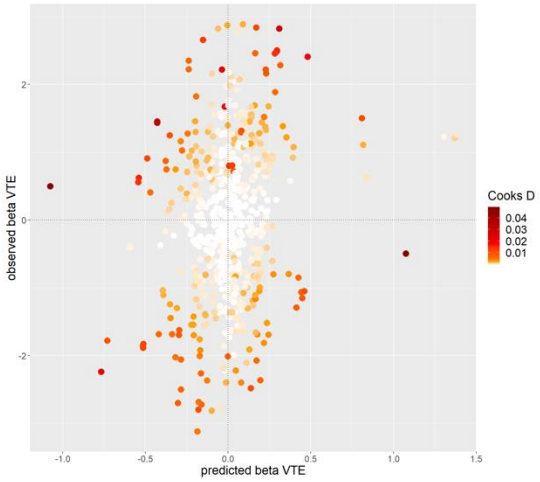
1) HGB



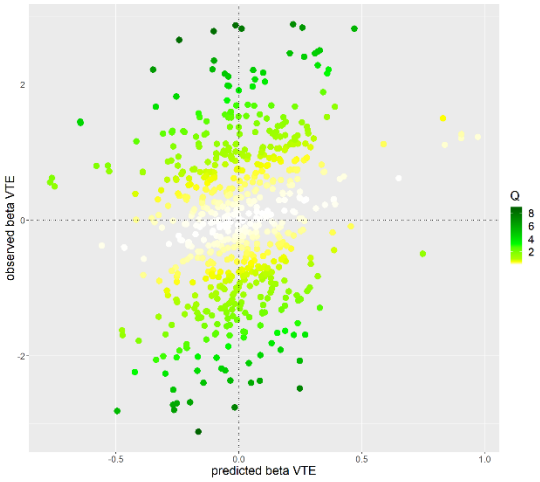



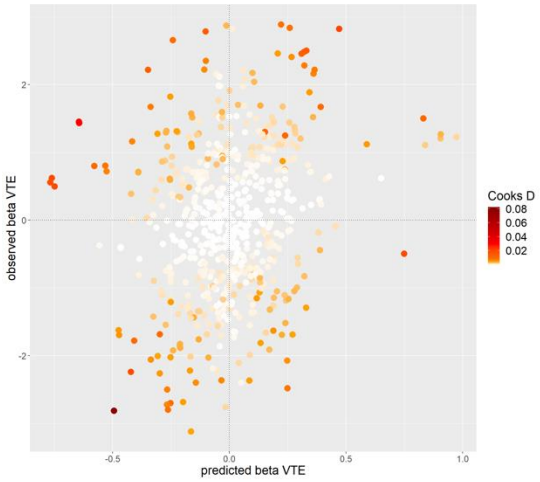

2) HCT and HGB



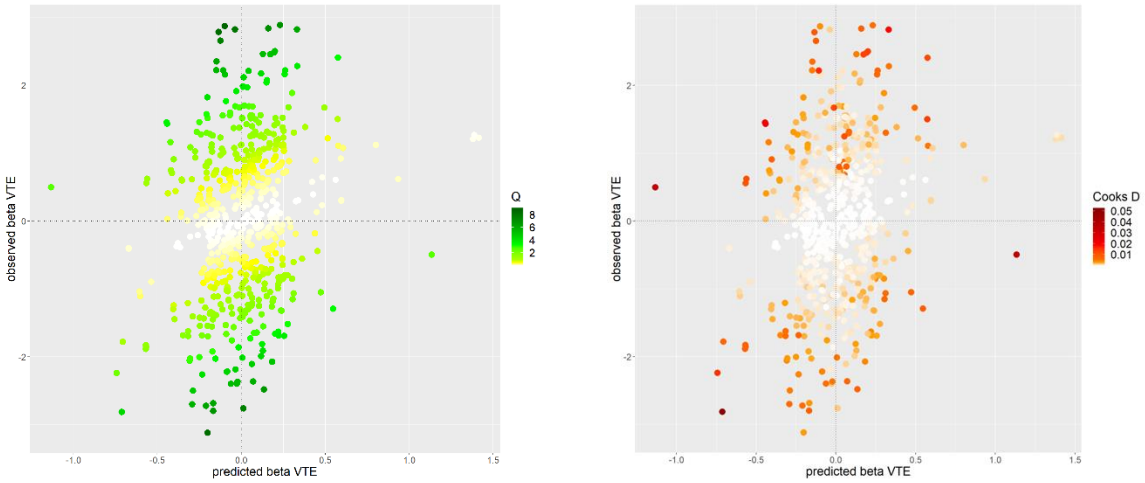




3) HGB and HLSR



4) HCT, HGB and HLSR




5) HGB and RBC

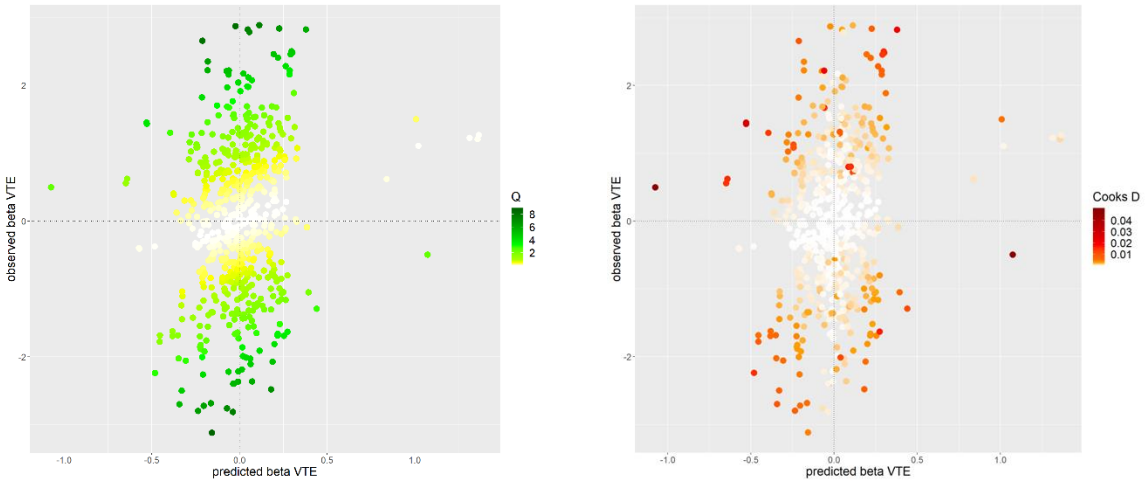
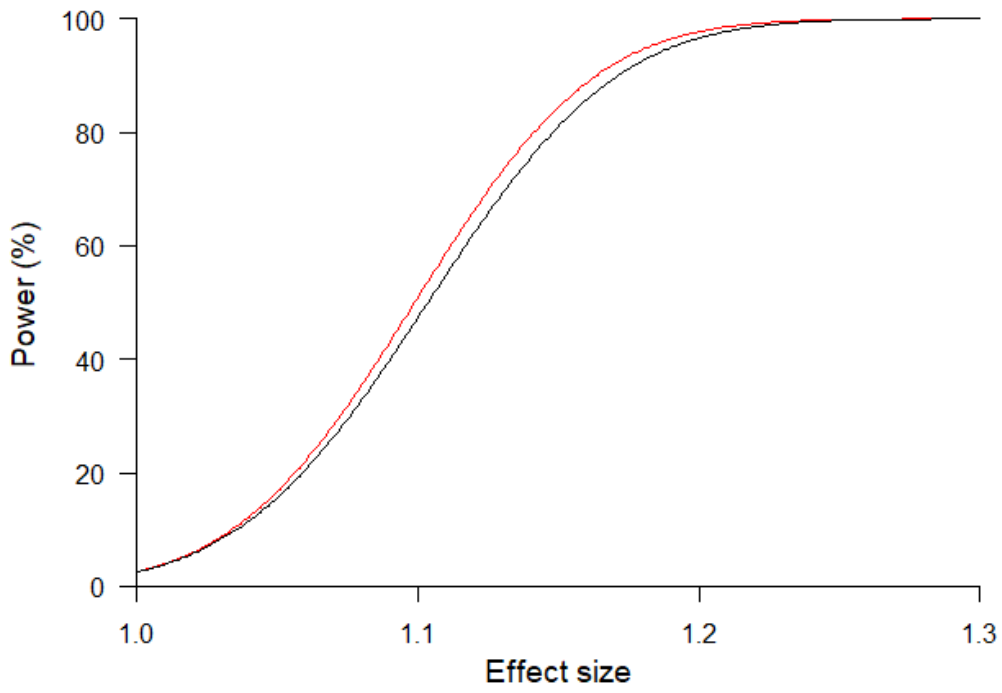



**Figure S5. Diagnostic plots of the predicted associations with venous thromboembolism (VTE) (x-axis) based on the best individual models 1 (HGB), model 2, (HCT and HGB), model 3 (HGB and HLSR), model 4 (HCT, HGB and HLSR) and model 5 (HGB and RBC), against the observed associations with VTE (y-axis), after excluding the outlying variants.** The colour code shows left) the Q-statistics for outliers and right) Cook's distance for influential points. Any genetic variant with Q-statistics large than 10 or Cook's distance large than median is marked by a label indicating the gene region.

**Figure S6. Power curves for venous thromboembolism with a sample size of 265 424 with 9 752 venous thromboembolism cases in the UK Biobank.**




The red line represents power in the venous thromboembolism analysis using 81 genetic variants explaining 4.2% of the variance in haemoglobin, with F-statistics of 93; the black line represents power in the venous thromboembolism analysis using 72 genetic variants explaining 3.8% of the variance in haemoglobin, with F-statistics of 95. Power is reasonable (above 80%) for effect sizes of 1.14 and 1.15 for venous thromboembolism in the UK Biobank.