

The value of genetic data from 665,460 individuals in managing iron deficiency anemia and suitability to donate blood

Short title: Genetic data in managing donation suitability

Authors: Jarkko Toivonen¹, Elias Allara^{2,3,4}, FINNGEN, Johanna Castrén¹, Emanuele di Angelantonio^{2,3,4,5,6,7}, Mikko Arvas¹

¹Finnish Red Cross Blood Service, Helsinki, Finland.

²British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.

³Heart and Lung Research Institute, University of Cambridge, Cambridge UK

⁴National Institute for Health and Care Research Blood and Transplant Research Unit in Donor Health and Behaviour, University of Cambridge, Cambridge, UK

⁵British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge, UK

⁶Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge, UK

⁷Health Data Science Research Centre, Human Technopole, Milan, Italy 20157

Corresponding author:

Jarkko Toivonen

Address:

Finnish Red Cross Blood Service

Biomedicum 1

Haartmaninkatu 8

00290 Helsinki

Finland

Email: Jarkko.toivonen@veripalvelu.fi

Phone: +358 444344384

Fax: +358 9 454 6253

Conflict of interest: no conflicts identified

Funding: E.A. is funded by a BHF Programme Grant (RG/18/13/33946).

Data Availability Statement: The Hb-deferral GWAS results are available at GWAS catalogue under accession number GCST90104113. The iron deficiency anemia GWAS results are available through the website https://www.finngen.fi/en/access_results. The PRS weights are available at the PRS catalog under ID PGS003426 for ferritin and PGS003425 for hemoglobin.

Abstract: 236 words

Figures: 3

Tables: 3

References: 30

Main text: 5005 words

Acknowledgements

J.T. performed the analysis and wrote the first draft of the manuscript. E.A. performed the validation of the IDA logistic regression on the INTERVAL cohort. FinnGen performed the meta-analysis of iron deficiency anemia. J.C. helped in defining the phenotypes and offered medical expertise throughout the research. E.d.A. supervised the validation on the INTERVAL cohort. M.A. supervised the research and participated in writing the first draft. All authors reviewed and edited the manuscript. A full list of FinnGen members and their affiliations appears in the Supplementary Information in Supplementary File 1.

Abstract

Background and objectives: While the genetic determinants of hemoglobin and ferritin have been widely studied, those of the clinically and globally relevant iron deficiency anemia (IDA) and deferral due to hypohemoglobinemia (Hb-deferral) are unclear. In this investigation, we aimed to quantify the value of genetic information in predicting IDA and Hb-deferral.

Materials and methods: We analyzed genetic data from up to 665,460 participants of the FinnGen, Blood Service Biobank and UK Biobank, and used INTERVAL (N=39,979) for validation.

We performed genome-wide association studies (GWASs) of IDA and Hb-deferral and utilized publicly available genetic associations to compute polygenic scores for IDA, ferritin and hemoglobin. We fitted models to estimate effect sizes of these polygenic risk scores (PRSs) on IDA and Hb-deferral risk while accounting for individual's age, sex, weight, height, smoking status and blood donation history.

Results: Significant variants in GWASs of IDA and Hb-deferral appear to be a small subset of variants associated with ferritin and hemoglobin. Effect sizes of genetic

predictors of IDA and Hb-deferral are similar to those of age and weight, which are typically used in blood donor management.

A total genetic score for Hb-deferral was estimated for each individual. The OR estimate, between first decile against that at ninth decile of total genetic score distribution, ranged from 1.4 to 2.2.

Conclusion: The value of genetic data in predicting IDA or suitability to donate blood appears to be on a practically useful level.

Keywords: iron deficiency anemia, Hb-deferral, genetic risk, statistical inference, GWAS, PRS

Highlights

1. Genetic information could be as valuable as age and weight to blood donation management.
2. Genetic variation predisposing significantly to iron deficiency anemia seems to be a small subset of genetic variation effecting hemoglobin and ferritin.

Introduction

Iron deficiency is a substantial contributor of global disease burden, and the World Health Organization (WHO) has set its reduction as a global health priority [1]. Iron deficiency (ID) can lead to iron deficiency anemia (IDA). Especially in 15 to 49 years old women anemia prevalence varied from 7.3% to 66.1% globally, of which half was estimated to be IDA, according to WHO statistics of 2019 [2]. Although IDA is most common in Southern Asia and Central and Western Africa, it has been reported to significantly decrease overall survival and health-related quality of life in over 60 year old Dutch [3]. The WHO defines anemia as blood Hb below 130 g/L in men and 120 g/L in women. To separate IDA from other anemias various molecular markers have been developed. Serum or plasma ferritin, although difficult to interpret as it is elevated by inflammation, is commonly used with a WHO cut-off of <15 µg/L to signify ID [4].

Blood donation and availability of blood products is a critical part of modern health care. However, blood donation can lead to anemia, due to iron loss, if not properly managed [5]. Pre-donation point-of-care hemoglobin (Hb) measurement is a common prerequisite for blood donation. In Europe a Hb level of ≥ 125 g/L for women and ≥ 135 g/L for men is typically required before donation. If the donor's Hb is below that level, they are deferred from donating due to low Hb (shortly, "Hb-deferred"). Deferral is demotivating for blood donors and leads to costs for both the donor and blood establishment [6]. Sufficient minimum donation intervals and iron supplementation have been found to be effective ways to keep Hb-deferral rates of blood services low [7]. As Hb thresholds for anemia and Hb-deferral are very similar, they are expected to be closely related. In addition to the minimum Hb requirement, a minimum weight of 50 kg and minimum age of 18 is required in the EU [8].

FinnGen is a public-private partnership project combining electronic health record (EHR) data from six regional and three national Finnish biobanks [9]. FinnGen aims to collect EHR data and genotypes for ~0.5 million Finns, of which >10% will be active blood donors from the Blood Service Biobank. Genotyping in FinnGen is carried out with an array of ~0.5 million genetic variants, after which up to 17 million variants were imputed. The isolated nature of Finnish population allows identification of novel protective or harmful genetic variants. In FinnGen, genome-wide association studies (GWASs) were carried out for ~2000 clinical endpoints, including IDA. The UK biobank [10] provides a similar resource.

From GWAS results, a polygenic risk score (PRS) can be calculated for all individuals. A PRS is the weighted sum of the number of alternative alleles of each genetic variant included in the PRS. The weights are the effect sizes of the variants' individual association with the phenotype of interest, e.g. IDA. In this paper, the PRS represents the total effect of a person's genotype to the risk of exhibiting the said phenotype, e.g. risk of developing anemia [11]. The clinical value of PRSs of blood donation traits is still under active research [12]. The heritability explained by currently known genetic variants is typically smaller than the heritability estimated by family-based studies.

However, variant discovery by sequencing and increasing of sample size is expected to make PRSs far more powerful in future [13].

EHRs often contain Hb measurements, and hence multiple well-powered GWASs have been carried out for Hb recently [14–16]. For ferritin, the summary statistics of one well powered meta-analysis is publicly available [17]. Variants in or near HFE and TMPRSS6 genes have been associated to Hb in several studies [18] and recently also to ferritin [17]. HFE (Human homeostatic iron regulator protein) is thought to regulate iron absorption through an interaction with the transferrin receptor in the cellular iron import pathway. TMPRSS6 (Transmembrane protease, serine 6) inhibits the production of the iron regulatory protein hepcidin [19]. Numerous other genes reported to be associated to Hb and ferritin at genome-wide significant level cover additional biological pathways involved in iron homeostasis, such as iron sensing and storage, inflammation and blood clotting, intestinal iron absorption, iron recycling, erythropoiesis and menstruation.

Although GWAS of IDA has been reported before, embedded in large scale analysis [20, 21], we present, to our knowledge, the first dedicated analysis. In addition, using large scale electronic health care and biobank data, we quantify for the first time the value of genotyping data at genome-wide scale in assessing the blood donation suitability of a donor.

Methods

The FinnGen data release 6 has 230,000 participants from Finnish hospital biobanks and 30,000 blood donors from the Finnish Red Cross Blood Service biobank, see Table 1 for subgroup counts. The basic information of each participant comprises the individual's age, sex, smoking information, weight and height (see Figures S1 and S2 and Tables S1-S3 for the variable distributions). In addition to the genome-wide SNP genotyping data of 16.7 million variants, we have access to the electronic health record (EHR) information of the participants. In this paper, we have only extracted the IDA

events (defined by the ICD-10 code D50; information partially available since year 1969 and fully available since 1998) from the EHR database, and the followup ends at the end of year 2019. The D50 code is given in Finland, when the laboratory measurements indicate low Hb (below 130 g/L in men and 120 g/L in women) and the treating medical doctor decides that the patient has signs of iron deficiency, for instance low ferritin and/or high transferrin receptor. Our models for IDA utilize FinnGen data only. Since weight, height and smoking information is not available for all individuals, we are left with 136,168 individuals with full information.

For 28,901 blood donors of the Finnish Red Cross Blood Service (FRCBS) we have also at our disposal all the blood donation histories stored in the eProgesa database (MAKSYSTEM, Paris, France) between years 2000 and 2020. For every donation event, information about the Hb value (pre-donation point-of-care capillary finger-prick sample), time of day, donation location, type of donation and whether it was successful is included. This dataset was subsequently preprocessed to derive new variables, described in Tables S1, S4 and S5 and Figures S3 and S4, that are used in our models. Importantly, our second endpoint, Hb-deferral, is defined as whether the pre-donation Hb measurement is below a threshold (125 g/L for females, and 135 g/L for males), and hence the donation is Hb-deferred.

These two datasets along with the external data utilized in this study, ferritin meta-analysis from Bell et al [17] and hemoglobin GWAS from Vuckovic et al [15], are depicted in Figure 1 and Table 3. Note that currently, due to regulations and contracts, we can only combine the eProgesa information with genotype and basic data (weight, height, and smoking status) of the blood donors, not with the EHR data. We develop our models for Hb-deferral on this combined dataset for blood donors.

We use Saige [22] for finding genetic variants that are associated with IDA and Hb-deferral. As covariates we use sex, age (age at last donation for Hb-deferral; age at first event, or at death, or at the end of follow-up, whichever comes first, for IDA), weight, height, smoking status and the first ten principal components of the genetic relationship matrix (and the genotyping batch when available). Fixed-effect inverse-variance IDA meta-analysis of FinnGen and UKBB [9] and the Hb-deferral GWAS on blood donors

were used in selecting important SNP variants as predictors in our models. To allow use of genetic information outside these selected SNPs as well, we derive polygenic risk scores for three related endpoints: IDA, ferritin and hemoglobin. The PRS of IDA was based on a GWAS for FinnGen participants who are not blood donors. To avoid overlap with the selected four SNPs, we excluded areas of ± 3 Mb around them from the 16.7 million variants while performing the PRS weight computation with the PRS-CS [11] software. The cohorts used to derive the PRS weights were essentially separate from the cohorts for which we computed the polygenic scores, and the IDA PRS was not used as a predictor in the IDA models, to avoid overfitting.

We use logistic regression models to evaluate the value of genetic variables in predicting IDA and Hb-deferral. For Hb-deferral we predict the most recent donation attempt using most recent covariate values. For IDA we predict whether a donor has ever been diagnosed with IDA using baseline covariate values, except age which is defined as in GWAS. We also fitted Cox proportional hazards (Cox PH) models for both endpoints. For Hb-deferral we predict the time-to-event using time-dependent values for the Hb and donation count covariates, and baseline values for other covariates. For IDA the Cox model uses only baseline covariate values. The details of model fitting are in the supplemental methods, and source code used in this research is publicly available at https://github.com/FRCBS/anemia_and_hb_deferral_prediction.

Results

The Manhattan plots of the external hemoglobin and ferritin GWASs, that were used in building the PRS weights for the respective phenotypes, are shown for completeness in Figures 2A and 2B, respectively. To have more detection power we carried out a meta-analysis of IDA with the FinnGen data and the UK Biobank data (18,076 cases, 647,384 controls). The Manhattan plot of this IDA meta-analysis, shown in Figure 2C, reveals four genome-wide significant SNPs, which are independent by SuSiE fine-mapping analysis [9]. We selected these four SNPs, listed in Table 2, to be included in our

models for IDA and Hb-deferral. We used the minor allele in FinnGen as the effect allele. Note that this is different from the alternative allele in case of SNP rs199138 in chromosome 15. The Manhattan plot of the Hb-deferral GWAS (6,714 cases, 21,189 controls) is shown in Figure 2D. Even though the GWAS found only one significant locus (lead SNP rs199598395), there were several peaks nearly reaching the significance threshold of $p < 5e-8$. The respective quantile-quantile plots and the genomic inflation factors of the four GWASs or meta-analyses are shown in Figure S5. Detailed genotype distributions of the four SNPs by the case-control status for both Hb-deferral and IDA are shown in Figures S6-S7 and Tables S10-S11.

We compared the similarity of association peaks in IDA meta-analysis to hemoglobin and ferritin GWASs and found that only rs6025 and rs199138 were clearly associated to ferritin and rs3129761 to hemoglobin (see supplemental results).

After the PRSs were computed for IDA, ferritin and hemoglobin phenotypes, we fitted multivariable logistic models for IDA (events ever) and Hb-deferral (latest event), and multivariable Cox PH models for time to first IDA or Hb-deferral events. Note that even though the IDA and hemoglobin phenotypes are partly related, we still included the PRSs of both in the Hb-deferral model as they are not highly correlated according to Figure S9, and also the Manhattan plots look dissimilar. The odds ratios of logistic models and hazard ratios of Cox PH models with their 95% confidence intervals (CIs) are visualized in Figure 3 and Tables S13-S16, and the Kaplan-Meier plots of Hb-deferral and IDA are shown in Figures S14-S15. Overall, the CIs for predictors are smaller in the Cox PH models, suggesting greater power of the Cox analysis. Subsequently, we considered the predictors whose CI does not cross 1 to have an effect in the model in question (Figure 3). When a predictor was available for both phenotypes, the signs of the effect seemed to agree, except for the weight and smoking variables. These differences are most likely explained by the interactions between weight and smoking with the “Is blood donor” variable. At least for women in the Cox PH model of IDA the hazard ratio of the interaction with weight is statistically significant, see Figure 3 and Table S16. The fact that being blood donor seems to prevent IDA is due to inverse causation: individuals with low Hb are not allowed to donate.

Although the effect sizes of genetic data had similar directions of effect, their magnitudes varied between models. The largest effect of the genetic data in IDA model was by the SNP in chromosome 17 for pre-menopausal females 2.9 (2.1 – 4.0) and for Hb-deferral again by the same SNP for pre-menopausal females 3.3 (2.0 – 5.3). In addition to individual variants, ferritin PRS was found to be inversely associated with IDA in females according to the Cox model and in premenopausal females according to the logistic model, while Hb PRS was found to be inversely associated with Hb-deferral in both sexes by the Cox PH model.

We then used the INTERVAL cohort [23] to validate our findings about the importance of genetic predictors. We were only able to fit the multivariable logistic regression model of IDA, due to lack of similar response or predictor variables. Some of the four SNPs and both the ferritin and Hb PRSs were found to be significant predictors of IDA, for details, see Figure S17 and supplemental results.

As an initial attempt to get a more quantitative measure for the total genetic effect on the Hb-deferral, we computed for each donor the weighted sum of genetic variables (SNPs and IDA, ferritin and hemoglobin PRSs), here called the total genetic score, using the Bayesian logistic regression and the Cox PH model for the Hb-deferral, see the Supplemental results for the formula. The medians of the genetic variables in each bin defined by the deciles of the total genetic score in the Cox PH model with time-dependent covariates for Hb-deferral are shown in Figure S21 (logistic model omitted for brevity). Even though the effect size of the SNP in chromosome 17 is high, the SNP in chromosome 15, for instance, is far more important, since it is more common. Next, we compared the total genetic score at the first decile against that at the ninth decile. This gave us an estimate for the lower limit of odds ratios of Hb deferral between individuals who are genetically ill-suited for donation versus those who are well-suited for donation. This lower limit for the odds ratio ranged from 1.4 to 2.2 over different model types and demographic groups, see Table S17.

Discussion

The Manhattan plot for hemoglobin from [15] (Figure 2A) shows 388 genome wide significant peaks ($p < 5e-8$) and for ferritin from [17] (Figure 2B) 58 peaks. In contrast for IDA, with 18,076 cases and 647,384 controls we could detect 4 peaks and for Hb-deferral with 6714 cases and 21,189 controls 1 peak. Most likely this is due to the lack of power in IDA and Hb-deferral GWASs relative to the Hb and ferritin GWASs. However, we had enough power to detect more peaks of the hemoglobin GWAS in the IDA GWAS, but we didn't (Figure S22). This might indicate that the SNPs detected for Hb, but not for IDA, are more related to higher iron values. Because for each peak the causal gene is not known, we compared the set of genes around in each peak (see supplemental results), and it appears that the genetic variation predisposing significantly to IDA is a small subset of both the genetic variation affecting levels of ferritin and hemoglobin.

The TMPRSS6 and HFE genes have been shown to be associated with markers of iron metabolism many times [17, 24]. Curiously, these were not detected in our Hb-deferral GWAS and IDA meta-analysis from Finnish and UK population. The HFE SNPs rs1800562 and rs179945 and the TMPRSS6 SNP rs855791 were detected in our own GWAS for blood donor hemoglobin (not shown) suggesting that their lack of association with IDA is not specific to Finnish population. The reason these three SNPs were not significantly associated with either IDA or Hb-deferral is either due to lack of power or because the SNPs are more related to iron overload i.e. hemochromatosis instead of low iron. In particular, FinnGen hemochromatosis GWAS detects only the HFE gene, whose effect size is much larger than in the IDA, see Figure S23. Accordingly, even though Mast et al. showed the TMPRSS6 and HFE SNPs to be associated with iron related variables, they did not find any significant associations in their donation tolerance GWAS [24].

We assumed that the effect of smoking and weight on the IDA may vary by blood donor status, and hence we included interactions between these, see supplemental methods for details. This highlights the difference between the healthy blood donors and rest of the FinnGen participants, which are hospital patients. The rs199598395 (17_58358769_C_T) SNP in the RNF43 gene has larger effect than age or weight.

According to Open Targets [25] it has an allele frequency of 0.013 in Finnish population and in Non-Finnish European population 0.0034, i.e. it is Finnish enriched. According to variant effect predictor (VEP) [26] rs199598395 is a missense variant in gene RNF43 (ENSG00000108375.13) and AC004687.2 (ENSG00000285897.1) and intron variant in TSOAP1-AS1 (ENSG00000265148.6). Although this Finnish enriched SNP has not been previously reported to be associated to IDA, in the 1 Mb area around RNF43, ferritin associations have been reported in genes MRPS23 [27] in Dutch blood donor population, MTMR4 in UK, Danish and Icelandic population [17] and TEX14 in European populations [28]. Bell et al performed [17] a literature review that showed the gene MTMR4 to be associated with hepcidin. In addition, their pQTL study revealed a variant in gene MTMR4 to be associated with hepcidin levels. Associations to various blood counts have been reported to this same locus [29]. Genetically isolated populations such as Finns have been thought to provide special opportunities to discover rare causal variants [9, 30], hence the missense variant in gene RNF43 could represent the true causal variant behind all these associations, but further molecular analysis is required to resolve this.

We quantify the effect of various explanatory variables to the risk of becoming anemic or Hb-deferred with two different methods, multivariable Bayesian logistic regression and Cox proportional hazards models and get very similar results. Overall, Cox PH models appear to be more powerful in quantifying effects as exemplified by their smaller CIs. Apart from whether a person is a blood donor or their previous pre-blood donation Hb measurement, genetic effects are found to be larger or in similar range than other variables available to us. Age and weight are typically used for donor selection. Considering our results, genetic information could be of equal value to blood donor management. The potential value of genetic data is confirmed by our INTERVAL cohort validation, although exact effect sizes can differ, possibly due to differences in genetics, donation policies and availability of data.

Both ferritin and Hb PRSs were found to have significant effects. Their effect sizes might reflect the fact that much of their heritability is yet to be discovered ($R^2=0.03$ for variance of Hb explained by Hb PRS).

Price of array genotyping varies but can be currently expected to be for institutional use tens of euros. A genotyping result is valid for a lifetime and allows blood group and HLA imputation as well, while for example a ferritin measurement would be expected to cost some euros per measurement.

References

1. WHO. Anaemia Policy Brief. 2012. Cited 2023 Jan 24. Available from http://www.who.int//iris/bitstream/10665/148556/1/WHO_NMH_NHD_14.4_eng.pdf.
2. WHO. Prevalence of anaemia in women of reproductive age. THE GLOBAL HEALTH OBSERVATORY. 2021. Cited 2023 Jan 24. Available from [https://www.who.int/data/gho/data/indicators/indicator-details/GHO/prevalence-of-anaemia-in-women-of-reproductive-age\(-\)](https://www.who.int/data/gho/data/indicators/indicator-details/GHO/prevalence-of-anaemia-in-women-of-reproductive-age(-)).
3. Wouters HJCM, van der Klauw MM, de Witte T, Stauder R, Swinkels DW, Wolffenbuttel BHR, et al. Association of anemia with health-related quality of life and survival: A large population-based cohort study. *Haematologica*. 2019;104:468–76.
4. Pasricha SR, Tye-Din J, Muckenthaler MU, Swinkels DW. Iron deficiency. *Lancet*. 2021;397:233–48.
5. Kiss JE. How do we manage iron deficiency after blood donation? *Br J Haematol*. 2018;181:590–603.
6. Custer B, Chinn A, Hirschler N V., Busch MP, Murphy EL. The consequences of temporary deferral on future whole blood donation. *Transfusion*. 2007;47:1514–23.
7. Zalpuri S, Romeijn B, Allara E, Goldman M, Kamel H, Gorlin J, et al. Variations in hemoglobin measurement and eligibility criteria across blood donation services are associated with differing low-hemoglobin deferral rates: a BEST Collaborative study. *Transfusion*. 2020;60:544–52.

8. Commission Directive 2004/33/EC implementing Directive 2002/98/EC of the European Parliament and of the Council as Regards Certain Technical Requirements for Blood and Blood Components. *Off J Eur Union*. 2004;L 91/25.
9. Kurki, M.I., Karjalainen, J., Palta P et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature*. 2023;613:508–518.
10. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562:203–9.
11. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* 2019 101. 2019;10:1–10.
12. Lewis ACF, Green RC. Polygenic risk scores in the clinic: new perspectives needed on familiar ethical issues. *Genome Med*. 2021;13:1–10.
13. López-Cortegano E, Caballero A. Inferring the Nature of Missing Heritability in Human Traits Using Data from the GWAS Catalog. *Genetics*. 2019;212:891–904.
14. Chen MH, Raffield LM, Mousas A, Sakaue S, Huffman JE, Moscati A, et al. Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. *Cell*. 2020;182:1198-1213.e14.
15. Vuckovic D, Bao EL, Akbari P, Lareau CA, Mousas A, Jiang T, et al. The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell*. 2020;182:1214-1231.e11.
16. Hu Y, Stilp AM, McHugh CP, Rao S, Jain D, Zheng X, et al. Whole-genome sequencing association analysis of quantitative red blood cell phenotypes: The NHLBI TOPMed program. *Am J Hum Genet*. 2021;108:874–93.
17. Bell S, Rigas AS, Magnusson MK, Ferkingstad E, Allara E, Bjornsdottir G, et al. A genome-wide meta-analysis yields 46 new loci associating with biomarkers of iron homeostasis. *Commun Biol*. 2021;4:156.

18. Timmer T, Tanck MWT, Huis in 't Veld EMJ, Veldhuisen B, Daams JG, de Kort WLAM, et al. Associations between single nucleotide polymorphisms and erythrocyte parameters in humans: A systematic literature review. *Mutat Res - Rev Mutat Res*. 2019;779:58–67.
19. Wang CY, Meynard D, Lin HY. The role of TMPRSS6/matriptase-2 in iron regulation and anemia. *Front Pharmacol*. 2014;5.
20. Backman JD, Li AH, Marcketta A, Sun D, Mbatchou J, Kessler MD, et al. Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature*. 2021;599:628–34.
21. Guindo-Martínez M, Amela R, Bonàs-Guarch S, Puiggròs M, Salvoro C, Miguel-Escalada I, et al. The impact of non-additive genetic associations on age-related complex diseases. *Nat Commun*. 2021;12:2436.
22. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet*. 2018;50:1335–41.
23. Di Angelantonio E, Thompson SG, Kaptoge S, Moore C, Walker M, Armitage J, et al. Efficiency and safety of varying the frequency of whole blood donation (INTERVAL): a randomised trial of 45 000 donors. *Lancet*. 2017;390:2360–71.
24. Mast AE, Langer JC, Guo Y, Bialkowski W, Spencer BR, Lee T, et al. Genetic and behavioral modification of hemoglobin and iron status among first-time and high-intensity blood donors. *Transfusion*. 2020;60:747–58.
25. Ghousaini M, Mountjoy E, Carmona M, Peat G, Schmidt EM, Hercules A, et al. Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res*. 2021;49:D1311–20.
26. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17:1–14.
27. Timmer T, Tanck M, Penkett C, Stirrups K, Gleadall N, Kort W, et al. Genetic

- determinants of ferritin, haemoglobin levels and haemoglobin trajectories: results from Donor InSight. *Vox Sang.* 2021;116:755–65.
28. Benyamin B, Esko T, Ried JS, Radhakrishnan A, Vermeulen SH, Traglia M, et al. Novel loci affecting iron homeostasis and their effects in individuals at risk for hemochromatosis. *Nat Commun.* 2014;5:4926.
 29. Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, et al. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell.* 2016;167:1415-1429.e19.
 30. Zuk O, Schaffner SF, Samocha K, Do R, Hechter E, Kathiresan S, et al. Searching for missing heritability: Designing rare variant association studies. *Proc Natl Acad Sci U S A.* 2014;111:E455-64.

Tables

Table 1. FinnGen participant subgroup sizes.

Group	n	Male	Premenopausal female	Postmenopausal female
FinnGen	260,405	113,344	37,750	109,311
Blood donors	28,901	11,807	9,242	7,852
Non-blood donor	231,504	101,537	28,508	101,459

Table 2. The four genome-wide significant lead SNPs from the IDA meta-analysis.

CHR	POS	SNP	REF	EA	Nearest gene	FG EAF	FG OR	FG P-value	UK EAF	UKBB OR	UKBB P-value	All OR	All P-value
1	169549811	rs6025	C	T	F5	0.02	0.74	1.01E-06	0.02	0.82	2.44E-05	0.79	2.54E-10
6	32617727	rs3129761	G	C	HLA-DQA1	0.48	1.07	9.40E-05	0.46	1.10	4.61E-12	1.09	5.32E-15
15	45095352	rs199138	A	G	DUOX2	0.93	0.87	2.69E-05	0.92	0.87	1.39E-07	0.87	1.65E-11
17	58358769	rs199598395	C	T	RNF43	0.01	3.00	1.49E-40				3.00	1.49E-40

Table 3. Number of individuals in each GWAS or meta-analysis.

Phenotype	n	Cases	Controls	Cohorts
hemoglobin	408,112			UKBB (UK)

ferritin	246,139			deCODE (Iceland), INTERVAL (UK) and Danish Blood Donor Study (Denmark)
IDA	665,460	18,076	647,384	UKBB and FinnGen
Hb-deferral	27,903	6,714	21,189	FinnGen blood donors

Figure Legends

Figure 1. The initial and derived datasets used to fit the models. Boxes filled with red denote blood donors, with blue the rest of FinnGen cohort, with green external and white boxes are results computed in this paper. **A** Dataflow of the Hb-deferral models. **B** Dataflow of the IDA models. FRCBS, Finnish Red Cross Blood Service; BD, blood donor; BDH, Blood donation history; FG, FinnGen; FG Basic, FinnGen basic data includes weight, height and smoking status of donors; EHR, Electronic health record; GWAS, Genome-wide association studies; SNP, Single nucleotide polymorphism; PRS, Polygenic risk score.

Figure 2. Manhattan plots of four phenotypes obtained from GWAS or meta-analysis on different datasets. The lead SNPs are marked with a black circle, and their corresponding gene names (given by VEP [26]) are shown if space permits. Only p-values smaller than $1e-4$ are plotted for performance reasons. **A** GWAS of Hemoglobin (n=408,112) [15]. **B** Meta-analysis of ferritin (n=246,139) combining GWAS results from Iceland, the UK and Denmark [17]. **C** Meta-analysis of IDA (18,076 cases, 647,384 controls) combining GWAS results from FinnGen R6 and UKBB cohorts. **D** GWAS of Hb-deferral phenotype on FinnGen blood donors (6,714 cases, 21 189 controls).

Figure 3. The odds ratios of the multivariable Bayesian logistic regression and hazard ratios of the multivariable Cox proportional hazards regression for both IDA and Hb-deferral. The explanatory variables, with the exception of binary variables and the allele dosages from the four SNPs, were standardized to zero mean and standard deviation of one. The horizontal bars indicate the 95% confidence intervals of the odds or hazard ratio estimates. The estimates are drawn as hollow points, if the odds/hazard ratio 1 is contained in the corresponding confidence interval. In the Cox PH models females were not split into pre- and postmenopausal groups as that would have meant changing the strata at the age of 45. In the Cox PH model of Hb-deferral the variables “Donation count” and “Previous Hb” were modeled as time-dependent variables.