Author(s): Popovic, Relja; Fabre, Margarete; Schutzman, Jennifer; Kulkarni, Diptee; Porello, Alessandro; Loboda, Andrey; Lehtonen, Heli; McDonough, Stefan; Vuoti, Sauli; Kaarniranta, Kai; Turunen, Joni A.; Ollila, Terhi; Uusitalo, Hannu; Karjalainen, Juha; Liu, Mengzhen; Loomis, Stephanie; Strauss, Erich; Chen, Hao; Tasanen, Kaisa; Huilaja, Laura; Hannula-Jouppi, Katariina; Salmi, Teea; Peltonen, Sirkku;

Title: Novel loci and biomedical consequences of iron homoeostasis variation

Year: 2024

Version: Published version

communications biology

Article

https://doi.org/10.1038/s42003-024-07115-3

Novel loci and biomedical consequences of iron homoeostasis variation

Elias Allara[1,2,3,146] ✉, Steven Bell[4,5,146], Rebecca Smith[1,3], Spencer J. Keene[1,3], Dipender Gill[6], Liam Gaziano[1,3,7,8], Deisy Morselli Gysi[7,9,10], Feiyi Wang[11], Vinicius Tragante[12], Amy Mason[1,3], Savita Karthikeyan[1,3], R. Thomas Lumbers[13], Emmanuela Bonglack[1,3,14], Willem Ouwehand[15,16,17,18], David J. Roberts[19,20], Joseph Dowsett[21], Sisse Rye Ostrowski[21,22], Margit Hørup Larsen[21], Henrik Ullum[23], Ole Birger Pedersen[22,24], Søren Brunak[25], Karina Banasik[25], Christian Erikstrup[26], DBDS Genomic Consortium*, Jonathan Mitchell[27], Christian Fuchsberger[27], Cristian Pattaro[27], Peter P. Pramstaller[27,28], Domenico Girelli[29], Mikko Arvas[30], Jarkko Toivonen[30], Sophie Molnos[31,32], Annette Peters[32,33,34], Ozren Polasek[35], Igor Rudan[36], Caroline Hayward[37], Ciara McDonnell[36,38], Nicola Pirastu[36,39], James F. Wilson[36,37], Katja van den Hurk[40,41], Franke Quee[40,41], Luigi Ferrucci[42], Stefania Bandinelli[43], Toshiko Tanaka[42], Giorgia Girotto[44,45], Maria Pina Concas[44], Alessandro Pecori[44], Niek Verweij[46,47], Pim van der Harst[46,48], Yordi J. van de Vegte[46], Lambertus A. Kiemeney[49,50], Fred C. Sweep[51], Tessel E. Galesloot[49], Patrick Sulem[12], Daniel Gudbjartsson[12,52], Egil Ferkingstad[12], FinnGen Consortium*, Luc Djousse[7,8,10,53], Kelly Cho[7,8,10], Michael Inouye[1,3,14,54,55,56], Stephen Burgess[1,3,14,57], Beben Benyamin[58,59], Konrad Oexle[60,61], Dorine Swinkels[51,62], Kari Stefansson[12,63], Magnus Magnusson[12,63], Andrea Ganna[11], Michael Gaziano[7,8,10], Kerry Ivey[7,8,10], John Danesh[1,2,3,14,55,64,147], Alexandre Pereira[7,8,10,147], Angela M. Wood[1,2,3,14,55,65,147], Adam S. Butterworth[1,2,3,14,55,147] & Emanuele Di Angelantonio[1,2,3,14,55,66,147] ✉

Iron homoeostasis is tightly regulated, with hepcidin and soluble transferrin receptor (sTfR) playing significant roles. However, the genetic determinants of these traits and the biomedical consequences of iron homoeostasis variation are unclear. In a meta-analysis of 12 cohorts involving 91,675 participants, we found 43 genomic loci associated with either hepcidin or sTfR concentration, of which 15 previously unreported. Mapping to putative genes indicated involvement in iron-trait expression, erythropoiesis, immune response and cellular trafficking. Mendelian randomisation of 292 disease outcomes in 1,492,717 participants revealed associations of iron-related loci and iron status with selected health outcomes across multiple domains. These associations were largely driven by *HFE*, which was associated with the largest iron variation. Our findings enhance understanding of iron homoeostasis and its biomedical consequences, suggesting that lifelong exposure to higher iron levels is likely associated with lower risk of anaemia-related disorders and higher risk of genitourinary, musculoskeletal, infectious and neoplastic diseases.

Iron is essential for various biological functions, including respiration, energy production, DNA synthesis, and cell proliferation[1,2]. Iron homoeostasis in healthy individuals is tightly regulated, with hepcidin and soluble transferrin receptor (sTfR) playing significant roles. Hepcidin, a liver-produced peptide hormone, regulates systemic iron levels by suppressing dietary iron absorption and recycling in response to elevated iron levels[1,3]. sTfR, the circulating extracellular part of transferrin receptor 1, serves as a biomarker indicating iron demand relative to supply, although its biological function is largely unknown[1,4]. Despite their relevance to iron homoeostasis and potential clinical utility for assessing iron status in adults[1], the genetic

A full list of affiliations appears at the end of the paper. *Lists of authors and their affiliations appear at the end of the paper. ✉e-mail: ea431@medschl.cam.ac.uk; ed303@medschl.cam.ac.uk

determinants of hepcidin and sTfR concentrations remain poorly understood[5,6], with previous large-scale genome-wide association studies (GWASs) primarily focusing on conventional clinical biomarkers such as serum iron, ferritin, transferrin saturation (TSAT), and either transferrin or total iron-binding capacity (TIBC)[7–9].

Disruption in iron homoeostasis can cause iron deficiency and iron overload. Iron deficiency affects over two billion people worldwide[1], which underscores the need to understand its long-term consequences on population health. Although iron overload is less prevalent, its extreme form—hemochromatosis—can lead to severe clinical manifestations[10]. Previous research on iron-regulating pathways has primarily focused on exploring the long-term biomedical consequences of perturbations in *HFE*[11], a genetic locus involved in the aetiology of hemochromatosis. The long-term clinical associations of systemic iron status have been assessed in multiple observational studies[12–19], Mendelian randomisation (MR) investigations[20–25], and randomised trials[26–30], with uncertainty mainly arising from residual bias in observational studies, limited statistical power and pleiotropy in MR investigations, and the breadth of health outcomes analysed in randomised trials.

To enhance the understanding of the genetic regulation of hepcidin and sTfR, we combined data from 12 original GWASs. We identified and described 43 genomic loci, including 2 new loci associated with hepcidin and 13 new loci associated with sTfR that had not been reported in any previous GWAS of iron-related biomarkers. To address the uncertainties related to the long-term consequences of individual iron-regulating pathways and systemic iron status on health outcomes, we performed locus-based and polygenic phenome-wide MR analyses on 292 clinical outcomes in up to 1,492,717 participants from deCODE, FinnGen, the Million Veteran Programme (MVP) and UK Biobank (UKBB), and 47 biomedical traits in up to 860,060 participants from MVP and UKBB.

Results

Genetic predictors of hepcidin and sTfR

We included 12 cohorts with imputed genotype array data comprising up to 91,675 participants and 16,261,412 variants with assessments of hepcidin concentration and up to 45,330 participants and 13,606,859 variants with measurements of sTfR concentration (Fig. 1; Supplementary Data 1). Across participating cohorts, the mean age ranged between 40 and 67 years, and the percentage of female participants ranged between 47% and 61% (Supplementary Data 1). All studies included admixed European-ancestry participants. Using LD Score regression and the 1000 G EUR reference panel, common SNP-based heritability estimates were 4.1% for hepcidin and 16.5% for sTfR (by comparison, heritability ranged between 15–48% in recent GWASs of conventional iron biomarkers[8,9]), and genetic associations were typically weaker for hepcidin compared to sTfR (Fig. 2A). Please note that we provide definitions of common genetic terminology in Table 1. Sensitivity analyses only adjusted for age and sex show similar results (Supplementary Information, page 9). Sensitivity analyses adjusted for C-reactive protein, in addition to the other covariates included in the main analysis, also show results similar to the main model (Supplementary Information, pages 9–10). Genetic and phenotypic correlations between hepcidin, sTfR and other iron traits (serum iron, ferritin, TSAT, TIBC) were broadly concordant (Fig. 2B; Supplementary Data 2).

We found 52 genome-wide significant ($P < 5 \times 10^{-8}$), conditionally independent and uncorrelated ($r^2 < 0.01$) signals mapped to 43 loci (Table 2; Supplementary Data 3; Supplementary Data 4). Of these, we found 20 associations with hepcidin (mapped to 16 loci) and 32 associations with sTfR (27 loci). All these 16 loci are associated with hepcidin for the first time, and two of them have not been reported in previous GWASs of iron biomarkers[7–9]. Twenty-four loci are associated with sTfR for the first time and 13 are previously unreported in GWASs of iron biomarkers. In Supplementary Data 3, we annotate the studies where the loci were previously reported. Three loci (*DUOX2*, *HFE*, *TMPRSS6*) contained signals for both hepcidin and sTfR, suggesting some shared genetic aetiology for these two biomarkers. We found 13 out of 52 sentinel variants with the strongest evidence for association ($P < 1 \times 10^{-15}$) in known loci such as *DUOX2*, *HFE* and *PCSK7*, and two variants in new loci (rs116816795, $P = 3.13 \times 10^{-46}$, nearest gene: *NDFIP1;* rs885122, $P = 2.28 \times 10^{-18}$, *LVRN*) (Table 2; Supplementary Data 3; Supplementary Data 4), suggesting potential involvement of immune response (*LVRN*) in affecting hepcidin and a potential connection between iron import regulation (*NDFIP1*) and sTfR.

Among 17 of the 52 sentinel variants, or their strong proxies ($r^2 > 0.7$) at novel loci (Supplementary Data 3), 7 were missense, 6 were intronic, 2 were intergenic and 2 were downstream (Supplementary Data 5). Two of these variants at novel loci had a minor allele frequency (MAF) of <0.01 and both were associated with sTfR (rs143437464, intronic, and rs200307986, missense, near *TMEM181*).

Phenome-wide scans using PhenoScanner v.2 showed associations of multiple variants with a wide array of phenotypic traits across multiple domains. In addition to associations with haematologic traits (e.g., haemoglobin concentration, erythrocyte count), we also observed strong associations with traits relating to the cardiovascular system, autoimmune activity, infectious diseases, respiratory and hepatorenal function. Taken together, these results indicate involvement in multiple biological functions across several human body systems for nearly all the genetic variants associated with hepcidin and/or sTfR concentrations (Supplementary Data 6).

We mapped the 52 sentinel variants to 43 non-overlapping loci based on the nearest gene, of which 16 were associated with hepcidin and 27 with sTfR. We used colocalization with expression and protein quantitative trait loci to guide the selection of putative causal genes, in combination with evidence from functional studies (Supplementary Information, pp 7–8). Among the 16 candidate genes associated with hepcidin, we were able to annotate 14 putative causal genes based on either biology or a combination of colocalization and biology ('biologically plausible genes'), one gene based on colocalization only and one gene based on vicinity to the sentinel variant (Supplementary Data 7, Supplementary Data 8). Biologically plausible genes were involved in hepcidin synthesis (*HAMP*), iron-sensing and hepcidin modulation (*AXIN1*, *HFE*, *TMPRSS6*), iron absorption and recycling (*DUOX2*, *FUT2*, *SLC11A2*, *SLC40A1*), reaction to hypoxia and haematopoiesis (*ARHGAP9/R3HDM2*, *EGLN3*, *IARS2*, *SOX7*), and immune reaction to pathogens (*LVRN*, *MPO*) (Fig. 3A). Of these, two putative causal genes (*ARHGAP9/R3HDM2* and *LVRN*) were not previously associated with iron traits. Among the 27 candidate genes annotated for sTfR, we were able to identify 19 biologically plausible genes (Supplementary Data 8), including genes involved in transferrin receptor synthesis, modulation, transport, degradation, recycling and shedding (*GALNT6*, *MARCH8*, *PCSK7*, *PGS1*, *RPS6KB1*, *TFRC*, *TFR2*, *UBXN6*), iron-sensing and hepcidin modulation (*HFE*, *TMPRSS6*, *ZFPM1*), intestinal iron absorption (*DUOX2*, *NDFIP1*), erythropoiesis (*CPS1*, *HBS1L/MYB*, *HK1*, *IRS2*, *SLC22A5*), and immune response (*MFSD6*) (Fig. 3A, B). Of these, 8 putative causal genes (*IRS2*, *MARCH8*, *MFSD6*, *NDFIP1*, *PGS1*, *SLC22A5*, *UBXN6* and *ZFPM1*) were at loci previously not reported in GWASs of iron traits.

Putative causal effects of iron-related loci and iron status on disease outcomes and biomedical traits

To mitigate pleiotropy and collider bias when defining instruments for MR analysis, we first collated 197 genetic variants associated with either hepcidin or sTfR in this study, and with serum iron, ferritin, TSAT or TIBC in a previous study[9] (Fig. 1). We then removed variants affected by horizontal pleiotropy (i.e. influencing non-iron traits via pathways not mediated by iron traits, such as *ABO*), indirect vertical pleiotropy (i.e. influencing iron traits via pathways not mediated by iron traits, such as *F5*) and affected by collider bias (which can result in spurious genetic associations and invalid MR instruments) (Supplementary Data 9). For each non-pleiotropic and non-collider-biased variant associated with iron traits, we defined a 200 Kb region around its putative causal gene, selected conditionally independent variants using GCTA-COJO with summary statistics for the most strongly associated iron trait (Supplementary Data 10; Supplementary Data 11), and then performed *cis*-MR (i.e. locus-specific MR rescaled by genetic

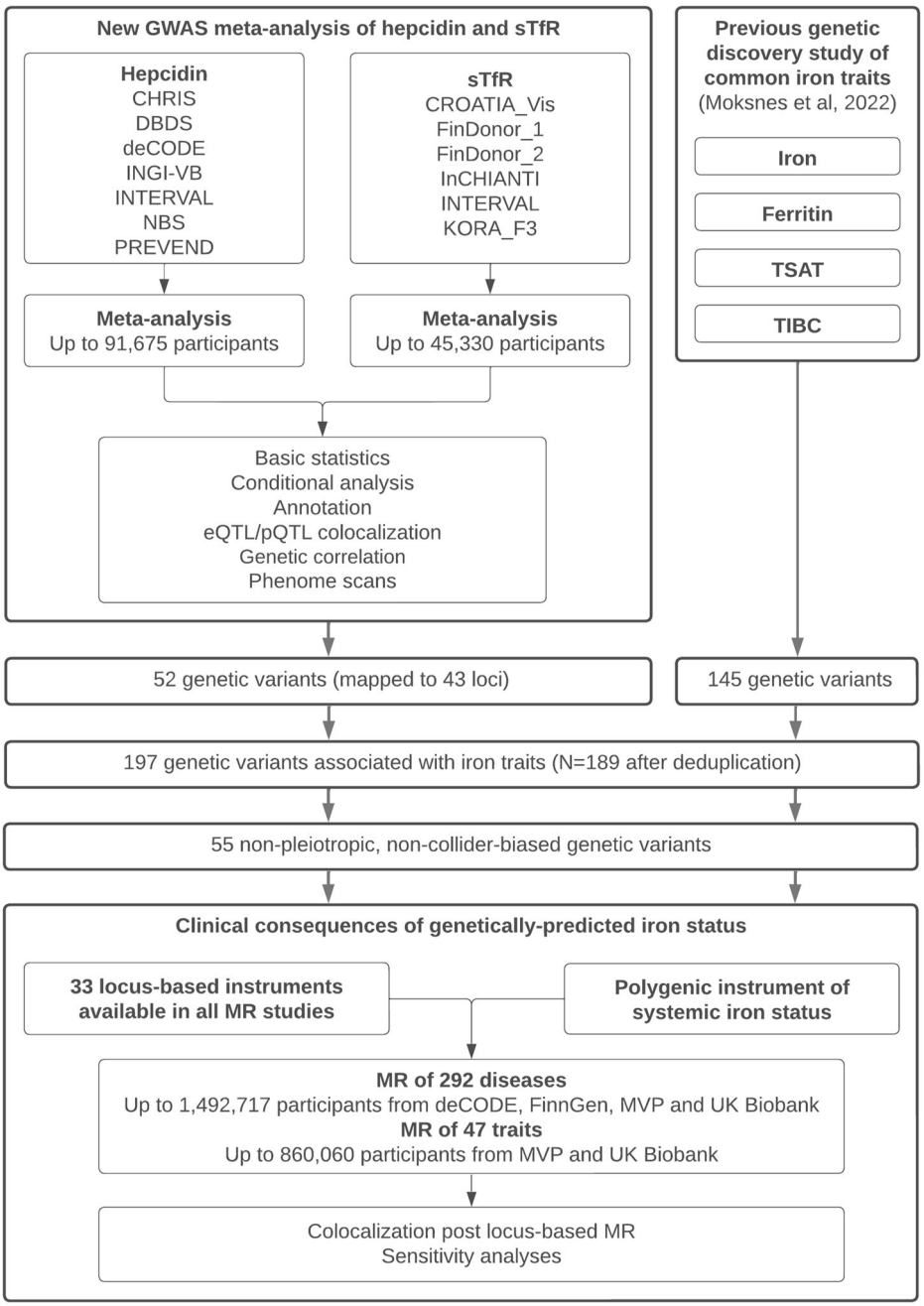


Fig. 1 | Study overview. GWAS, genome-wide association study. sTfR soluble transferrin receptor, TSAT transferrin saturation, TIBC total iron-binding capacity, eQTL expression quantitative trait loci, pQTL protein quantitative trait loci, MR Mendelian randomisation, MVP Million Veteran Programme. The genetic variants from Moksnes 2022 were obtained from the paper's Supplementary Data 1.

associations with iron traits) and colocalization. We found Bonferroni-significant log-linear associations of 19 loci with 47 diseases in 1,492,717 deCODE, FinnGen, MVP, and UKBB participants (Supplementary Fig. 1A; Supplementary Data 12). Of these, we found evidence of colocalization for four loci and six diseases (Fig. 4A; Supplementary Fig. 1B; Supplementary Data 12; Supplementary Data 13), highlighting the usefulness of this method in addressing residual genetic confounding. *HFE* (rescaled by TSAT) and *TMPRSS6* (iron) were strongly associated with inverse risk of iron-deficiency anaemia. We also found that *EPAS1* (TIBC) was inversely associated with hypertension and that *SLC25A28* (ferritin) was positively associated with colorectal cancer and benign neoplasm of colon; however, no other iron-related loci were associated and colocalized with these diseases, suggesting that these effects may be driven by horizontal pleiotropy. These four loci were associated with multiple biomedical traits, showing evidence of positive association of *HFE*, *TMPRSS6* and *EPAS1* with haemoglobin and inverse associations of *EPAS1* and *HFE* with total cholesterol, suggesting that iron may play a role in affecting these traits via these loci, as well as several other associations of isolated loci with glycaemic, inflammatory, hepatorenal and other traits (Fig. 4B; Supplementary Fig. 2; Supplementary Data 12).

We then generated a polygenic instrument of systemic iron status composed by six variants mapped to *ERFE*, *HAMP*, *HFE*, *SLC25A37*, *TFR2* and *TMPRSS6* (Supplementary Data 11), that were not affected by horizontal pleiotropy, indirect vertical pleiotropy, or collider bias and that: (i) were associated ($P < 5 \times 10^{-8}$) with at least one trait; (ii) were nominally associated ($P < 0.05$) with all the other iron traits except for hepcidin (as its levels are influenced by systemic iron status); and (iii) displayed a direction of association consistent across all traits (e.g., positive for iron, ferritin, TSAT and negative for TIBC and sTfR). To reduce the impact of study-specific estimates that may disproportionately affect meta-analytic estimates, we present Bonferroni-significant and nominal results for diseases and traits having MR estimates with the same direction (regardless of their *p* value) in

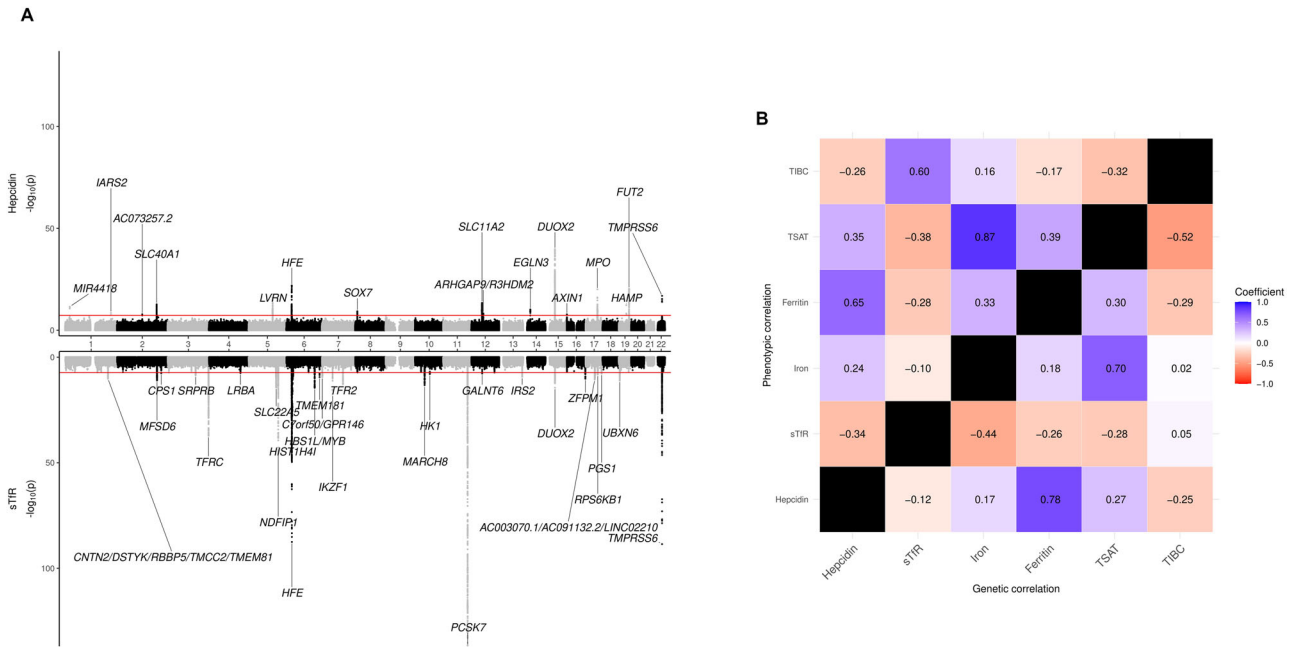


Fig. 2 | Results of GWAS meta-analysis of hepcidin and sTfR and correlations with common iron traits. A Miami plot for hepcidin (upper plot, $N = 91,675$ participants) and sTfR (lower plot, $N = 45,330$ participants). For each locus ($N = 16$ loci for hepcidin, $N = 27$ for sTfR), we show the candidate gene name for the sentinel variant with the lowest p value. **B** Genetic and phenotypic correlations between the iron traits analysed in this study (hepcidin, sTfR) and those investigated in previous studies (ferritin, iron, TIBC, TSAT). Phenotypic correlations were estimated in the INTERVAL study (up to 40,197 participants). Genetic correlations were estimated using associations from the present study (hepcidin, sTfR; up to 91,675 participants) and Moksnes et al. 2022 (ferritin, iron, TIBC, TSAT; up to 257,953 participants).

Table 1 | Glossary of genetic terms

Term	Acronym	Description
Colocalization	-	Statistical method used to determine whether two traits share a causal variant within a specific locus.
Expression quantitative trait locus	eQTL	Genetic loci that explain variation in mRNA expression levels.
Genome-wide association study	GWAS	Research approach for identifying genomic variants that are statistically linked to a particular trait.
Linkage disequilibrium	LD	Non-random association of alleles at different loci. It can be estimated statistically using the correlation coefficient between pairs of loci (r^2).
Locus	-	Specific position on a chromosome where a particular gene or genetic marker is located.
Mendelian randomization	MR	Statistical method for assessing causal relationships that tests whether genetic variants associated with an exposure (e.g., iron status) are also associated with an outcome (e.g., diseases or intermediate traits).
Minor allele frequency	MAF	Frequency at which the second most common allele occurs in a given population. Common alleles typically have a MAF ≥ 0.1.
Protein quantitative trait locus	pQTL	Genetic loci that explain variation in protein levels.
Single nucleotide polymorphism	SNP	Genomic variant occurring at a single nucleotide position in the DNA sequence.
Variant	-	A change in the most common DNA nucleotide sequence.

all the studies included in the meta-analysis. In an agnostic analysis of 292 disease outcomes in 1,492,717 deCODE, FinnGen, MVP and UKBB participants, we found four expected Bonferroni-significant log-linear associations of genetically predicted higher iron status with lower risk of mineral deficiency (a cluster of conditions that includes iron-deficiency), iron-deficiency anaemia and other deficiency anaemia, and with higher risk of disorders of mineral metabolism (a cluster of conditions that includes haemochromatosis). We also found six Bonferroni-significant associations with higher risk of cystitis and urethritis, dermatophytosis/dermatomycosis, postoperative infection, acquired foot deformities, arthropathy associated with other disorders, and liver cancer (Fig. 5A; Supplementary Data 14). Genetically predicted systemic iron status was also nominally associated with multiple clinical outcomes spanning various domains: circulatory, dermatologic, digestive, endocrine/metabolic, genitourinary, haematopoietic, infectious-disease, musculoskeletal, neoplasms, respiratory, sense organs and symptoms. Sensitivity analyses showed general robustness of findings when using MR Egger regression and the weighted median estimator (Supplementary Fig. 3). Although some between-variant heterogeneity was present for specific outcomes (e.g., disorders of mineral metabolism and other anaemias; Supplementary Data 14), MR Egger intercepts generally showed no evidence of residual horizontal pleiotropy (Supplementary Fig. 3). However, when removing the pC282Y variant in *HFE* from the polygenic instrument, most of these associations did not reach significance, except for iron-deficiency anaemia and iron-metabolism disorders, suggesting that these associations may be largely driven by *HFE*, although reduced statistical power might play a role in widening confidence intervals (Fig. 5A; Supplementary Data 14). We also found ten Bonferroni-corrected log-linear associations of genetically predicted iron concentration with multiple biomedical traits in up to 860,060 MVP and UKBB participants across the following domains: glycaemic indices, haematologic,

https://doi.org/10.1038/s42003-024-07115-3

Table 2 | Conditionally independent and uncorrelated sentinel variants associated with hepcidin (*N* = 20) and sTfR (*N* = 32)

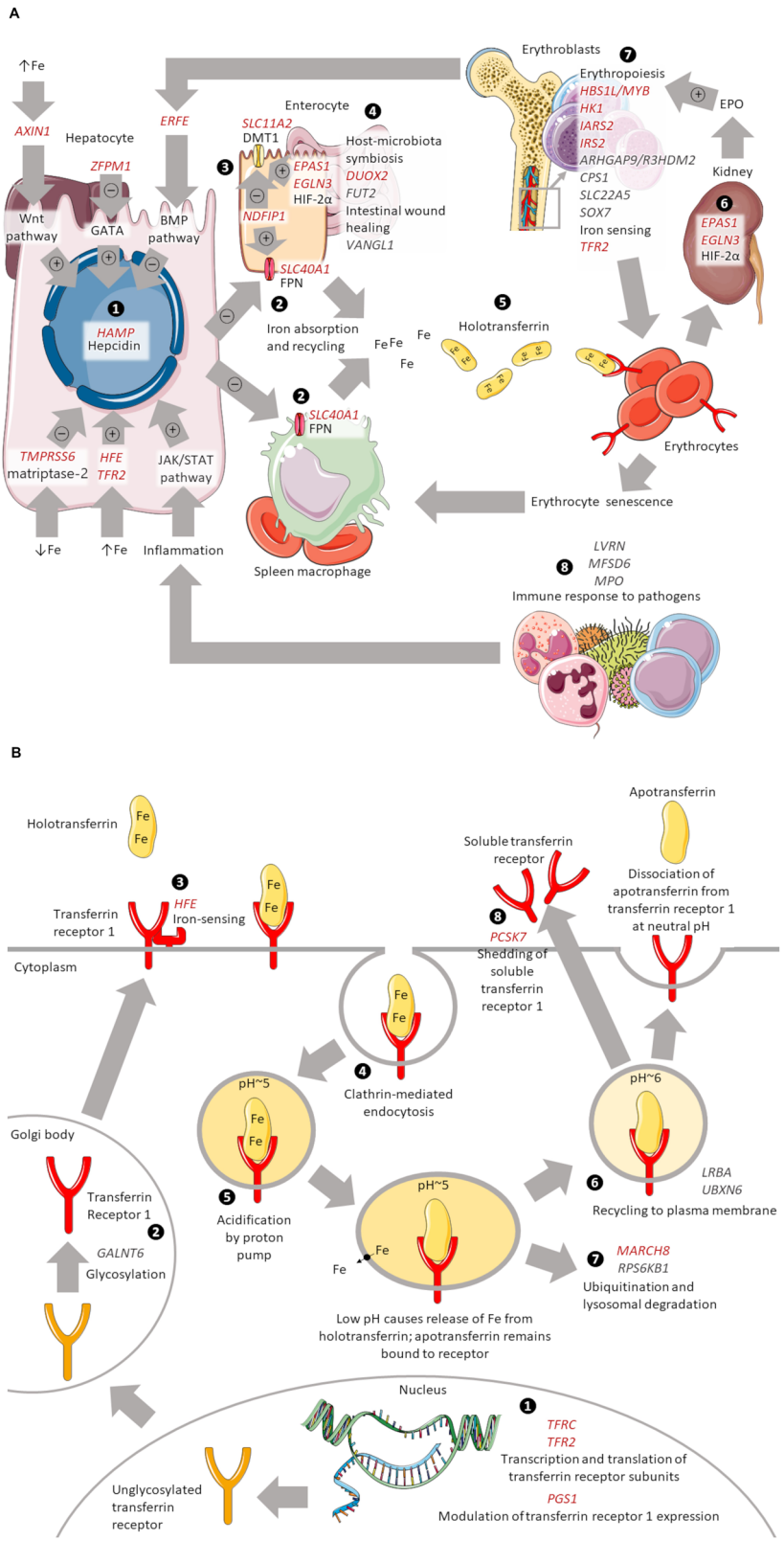
Trait	Chr	Position	RsID	EA	NEA	EAF	Beta	SE	*P* value	Het. *P* value	*N*	Nearest gene	Candidate gene	Locus novelty (hepcidin and sTfR)	Variant novelty (all iron traits)	Locus novelty (all iron traits)
Hepcidin	1	22,584,002	rs75965181	A	T	0.023	−0.118	0.017	2.76E-12	2.13E-01	91675	*MIR4418*	*MIR4418*	Novel	Known	Known
	1	220,291,414	rs201255457	A	ATC	0.989	−0.156	0.026	3.15E-09	3.77E-01	75972	*IARS2*	*IARS2*	Novel	Known	Known
	2	121,310,269	rs6706968	A	C	0.424	0.030	0.005	1.43E-08	6.11E-01	82690	*AC073257.1*	*AC073257.2*	Novel	Known	Known
	2	190,390,963	rs149682241	C	G	0.026	0.101	0.015	4.33E-11	8.47E-01	91675	*SLC40A1*	*SLC40A1*	Novel	Novel	Known
	2	190,521,054	rs7568449	T	C	0.762	0.043	0.006	1.88E-13	2.24E-01	89099	*ASNSD1*	*SLC40A1*	Novel	Novel	Known
	5	115,331,335	rs885122	A	G	0.598	−0.043	0.005	2.28E-18	8.07E-09	91675	*LVRN*	*LVRN*	Novel	Novel	Novel
	6	26,098,474	rs79220007	T	C	0.930	0.094	0.010	1.12E-22	1.25E-03	91675	*HFE*	*HFE*	Novel	Known	Known
	8	10,577,987	rs10090558	A	G	0.918	−0.060	0.010	4.64E-10	2.32E-03	78548	*SOX7*	*SOX7*	Novel	Known	Known
	12	51,439,858	rs9739943	T	C	0.059	0.079	0.010	3.78E-14	2.04E-01	91675	*LETMD1*	*SLC11A2*	Novel	Known	Known
	12	57,735,045	rs11609805	A	G	0.270	−0.032	0.006	5.24E-09	4.48E-01	91675	*R3HDM2*	*ARHGAP9/R3HDM2*	Novel	Novel	Novel
	14	34,410,892	rs996347	T	C	0.645	−0.033	0.005	6.06E-11	2.04E-02	91675	*EGLN3*	*EGLN3*	Novel	Known	Known
	15	45,274,530	rs549876436	T	C	0.907	0.082	0.011	7.30E-15	8.99E-02	83099	*TERB2*	*DUOX2*	Novel	Novel	Known
	15	45,319,959	rs2600895	A	G	0.913	0.082	0.012	2.73E-11	7.53E-02	88174	*SORD*	*DUOX2*	Novel	Novel	Known
	15	45,320,493	rs113000093	C	G	0.089	−0.078	0.013	7.01E-10	4.66E-02	85598	*SORD*	*DUOX2*	Novel	Novel	Known
	15	45,387,550	rs199138	A	G	0.076	−0.122	0.009	1.55E-40	3.41E-03	91675	*DUOX2*	*DUOX2*	Novel	Known	Known
	16	353,122	rs71378510	A	G	0.907	0.050	0.009	1.62E-08	5.31E-01	91675	*AXIN1*	*AXIN1*	Novel	Novel	Known
	17	56,436,109	rs34523089	T	C	0.162	0.063	0.007	8.15E-21	8.61E-01	89817	*RNF43*	*MPO*	Novel	Known	Known
	19	35,775,902	rs104894696	A	G	0.003	−0.387	0.054	4.90E-13	2.09E-04	76690	*HAMP*	*HAMP*	Novel	Novel	Known
	19	49,207,255	rs485073	A	G	0.453	0.046	0.005	4.64E-21	5.91E-01	91675	*FUT2*	*FUT2*	Novel	Known	Known
	22	37,462,936	rs855791	A	G	0.429	0.042	0.005	1.34E-17	8.89E-02	91675	*TMPRSS6*	*TMPRSS6*	Novel	Known	Known
sTfR	1	205,041,952	rs6696846	T	C	0.484	−0.043	0.007	6.22E-11	7.35E-01	45330	*CNTN2*	*CNTN2/DSTYK/ RBBP5/TMCC2/ TMEM81*	Novel	Novel	Novel
	2	191,357,694	rs35095338	A	T	0.556	−0.046	0.007	9.67E-12	8.77E-01	44341	*TMEM194B*	*MFSD6*	Novel	Novel	Novel
	2	211,543,055	rs715	T	C	0.688	−0.041	0.007	2.28E-08	4.12E-01	45330	*CPS1*	*CPS1*	Novel	Known	Known
	3	133,539,500	rs11371594	G	GA	0.845	0.058	0.010	2.54E-09	3.95E-01	40837	*SRPRB*	*SRPRB*	Novel	Novel	Known
	3	195,795,618	rs112856048	A	G	0.243	−0.100	0.008	5.93E-38	5.28E-01	45330	*TFRC*	*TFRC*	Novel	Known	Known
	3	195,921,311	rs9325434	A	G	0.141	−0.062	0.010	2.27E-10	8.47E-01	45330	*ZDHHC19*	*TFRC*	Novel	Novel	Known
	4	151,199,080	rs2290846	A	G	0.277	−0.041	0.007	3.11E-08	9.61E-01	45330	*LRBA*	*LRBA*	Novel	Novel	Novel
	5	131,784,393	rs12521868	T	G	0.427	0.052	0.007	4.41E-15	7.83E-01	45330	*C5orf56*	*SLC22A5*	Novel	Novel	Novel
	5	141,482,333	rs116816795	T	C	0.168	0.127	0.009	3.13E-46	7.55E-01	45330	*NDFIP1*	*NDFIP1*	Novel	Novel	Novel
	5	141,602,204	rs2906082	T	C	0.783	0.057	0.008	5.34E-12	5.65E-01	44584	*NDFIP1*	*NDFIP1*	Novel	Novel	Novel
	6	25,957,426	rs72832593	T	C	0.897	0.130	0.011	8.09E-33	5.91E-01	45330	*TRIM38*	*HIST1H4I*	Novel	Known	Known
	6	26,093,141	rs1800562	A	G	0.075	−0.253	0.013	2.88E-88	2.95E-02	45330	*HFE*	*HFE*	Known	Known	Known
	6	135,427,159	rs9389269	T	C	0.729	0.059	0.007	3.10E-15	6.84E-01	45330	*HBS1L*	*HBS1L/MYB*	Novel	Known	Known
	6	159,020,121	rs143437464	A	G	0.009	−0.234	0.037	3.81E-10	5.08E-01	43595	*TMEM181*	*TMEM181*	Novel	Novel	Novel

https://doi.org/10.1038/s42003-024-07115-3

Table 2 (continued) | Conditionally independent and uncorrelated sentinel variants associated with hepcidin (*N* = 20) and sTfR (*N* = 32)

Trait	Chr	Position	RsID	EA	NEA	EAF	Beta	SE	*P* value	Het. *P* value	*N*	Nearest gene	Candidate gene	Locus novelty (hepcidin and sTfR)	Variant novelty (all iron traits)	Locus novelty (all iron traits)
	6	159,026,327	rs200307986	A	G	0.003	0.358	0.065	2.69E-08	1.29E-01	42671	*TMEM181*	*TMEM181*	Novel	Novel	Novel
	7	1,080,897	rs186044114	T	C	0.852	−0.061	0.010	4.19E-10	4.27E-01	41188	*C7orf50*	*C7orf50/GPR146*	Novel	Novel	Novel
	7	50,427,982	rs6592965	A	G	0.458	−0.058	0.007	2.43E-18	9.60E-02	45330	*IKZF1*	*IKZF1*	Novel	Known	Known
	7	100,235,970	rs7385804	A	C	0.627	−0.051	0.007	1.49E-13	8.85E-01	45330	*TFR2*	*TFR2*	Novel	Known	Known
	10	45,953,767	rs7908745	A	G	0.687	−0.054	0.007	5.61E-14	1.38E-01	45330	*MARCH8*	*MARCH8*	Novel	Novel	Novel
	10	71,093,392	rs16926246	T	C	0.133	0.056	0.010	1.05E-08	9.71E-01	45330	*HK1*	*HK1*	Novel	Known	Known
	11	117,021,097	rs187669805	C	G	0.994	0.407	0.046	1.94E-18	3.86E-01	43487	*PCSK7*	*PCSK7*	Known	Novel	Known
	11	117,081,500	rs11216316	A	C	0.892	−0.273	0.011	1.77E-137	5.68E-02	45330	*PCSK7*	*PCSK7*	Known	Novel	Known
	11	117,088,082	rs2238005	T	C	0.061	0.118	0.014	2.71E-17	7.23E-01	45330	*PCSK7*	*PCSK7*	Known	Novel	Known
	12	51,783,420	rs10876169	A	T	0.410	0.039	0.007	1.85E-08	1.15E-01	44584	*GALNT6*	*GALNT6*	Novel	Novel	Known
	13	110,401,304	rs76944188	T	C	0.942	−0.090	0.015	5.45E-10	4.07E-01	45330	*IRS2*	*IRS2*	Novel	Novel	Novel
	15	45,395,901	rs75922593	G	GT	0.066	0.112	0.014	3.64E-15	5.67E-01	40837	*DUOX2*	*DUOX2*	Novel	Known	Known
	16	88,567,333	rs74035509	T	C	0.083	0.083	0.013	6.36E-11	6.39E-01	44584	*ZFPM1*	*ZFPM1*	Novel	Novel	Novel
	17	43,556,807	rs55925547	T	C	0.801	−0.057	0.009	2.46E-11	7.60E-01	43244	*PLEKHM1*	*AC003070.1/ AC091132.2/ LINC02210*	Novel	Novel	Novel
	17	57,925,649	rs1292072	A	G	0.794	−0.048	0.008	7.13E-09	9.06E-01	45330	*RNU6-450P*	*RPS6KB1*	Novel	Known	Known
	17	76,401,328	rs1976703	T	C	0.493	0.039	0.007	1.57E-08	1.08E-01	41826	*PGS1*	*PGS1*	Novel	Novel	Novel
	19	4,502,282	rs13041	T	C	0.518	0.048	0.007	8.53E-12	8.38E-01	44584	*PLIN4*	*UBXN6*	Novel	Novel	Novel
	22	37,462,936	rs855791	A	G	0.434	0.134	0.007	2.70E-89	8.92E-01	45330	*TMPRSS6*	*TMPRSS6*	Known	Known	Known

Chromosomal positions are in GRCh37 assembly. Beta estimates are per-allele dosage increase of the effect allele (EA). NEA indicates non-effect allele. EAF is EA frequency. SE, standard error. 'Locus novelty' indicates whether the variant is within a 500-Kb window from a previously reported genetic variant associated with other iron traits[7–9]; loci that are novel with respect to previous GWASs of hepcidin and sTfR[5,6] are indicated in the 'Locus novelty (hepcidin and sTfR)' column. 'Variant novelty' indicates whether a variant is in moderate or high linkage disequilibrium ($r^2 \geq 0.2$) with a previously reported genetic variant associated with iron traits[7–9]. Variants that are novel according to both criteria across all iron traits are highlighted in red on grey background. Candidate genes are assigned based on colocalization, manual curation and proximity (see Supplementary Methods, Supplementary Data 7 and Supplementary Data 8).



hepatorenal function and respiratory (Fig. 5B; Supplementary Data 14). Sensitivity analyses showed general robustness of findings when using MR Egger regression and the weighted median estimator (Supplementary Fig. 4). Three associations persisted after removing the pC282Y variant in *HFE*: an inverse association of genetically predicted iron concentration with glycated haemoglobin (HbA1c), and positive associations with direct bilirubin and total bilirubin.

Discussion

In this meta-analysis of 12 original GWASs of over 90,000 participants, we identify 43 loci associated with hepcidin and sTfR concentrations, including 15 novel loci not previously associated with iron biomarkers. Through manual curation and colocalization, we mapped the new loci to several putative genes, suggesting potential roles in iron-trait expression (*PGS1*, *ZFPM1*), erythropoiesis (*ARHGAP9/R3HDM2*, *IRS2*, *SLC22A5*), immune

Fig. 3 | Established and potential candidate genes mapped to variants associated with hepcidin or sTfR: summary of their role and contextual information. A This figure summarises the genes mentioned in Table 2 of this study, as well as other iron-homoeostasis genes provided for contextual information. Genes with an established role in iron homoeostasis are shown in red and italic; genes with a potential role are presented in dark grey and italic. Relevant references to other studies are included in Supplementary Data 8 ❶ Hepcidin is tightly regulated by several pathways. *TMPRSS6, ERFE* (via the BMP pathway), and *ZFPM1* suppress hepcidin expression in hepatocytes. *HFE*, *TFR2*, the Wnt pathway, and the JAK/STAT pathway increase hepcidin expression. Activation of the Wnt pathways is observed in iron overload, with involvement of *AXIN1*. Activation of JAK/STAT signalling has been proposed as a possible link between inflammation and iron homoeostasis. ❷ In presence of iron abundance, hepcidin suppresses function of ferroportin (FPN), an iron transporter coded by *SLC40A1* that mediates dietary intestinal iron uptake and iron recycling by macrophages from senescent erythrocytes. *NDFIP1* prevents degradation of ferroportin in vitro. ❸ Hypoxia-inducible factor 2α (HIF-2 α), coded by *EPAS1* and regulated by *EGLN3*, also controls duodenal iron absorption by promoting the expression of divalent metal transporter 1 (DMT1), coded by *SLC11A2*, on the luminal side of enterocytes. *NDFIP1* regulates DMT1 expression in mice. *EGLN3* hydroxylates key prolyl residues on HIF-2α, providing a recognition motif for its degradation. ❹ Several genes appear relevant to intestinal iron absorption: (i) *DUOX2* regulates interactions between the intestinal microbiota and the mucosa to maintain immune homoeostasis in mice, which likely enables intestinal iron absorption; (ii) *FUT2* codes for fucosyltransferase 2, an enzyme responsible for maintaining host-microbiota symbiosis via fucosylation of intestinal epithelial cells; (iii) *VANGL1* encodes a protein involved in mediating intestinal trefoil factor-induced wound healing in the intestinal mucosa. ❺ Iron released through ferroportin is bound to iron carrier transferrin (referred to as apotransferrin when not bound to iron), forming iron-loaded transferrin (holotransferrin), which delivers iron to most cells, especially erythrocytes. ❻ In presence of hypoxia, raised levels of HIF-2 α result in increased erythropoietin (EPO) production. ❼ EPO stimulates erythropoiesis, which is also modulated by several genes involved in erythroblast proliferation and differentiation: (i) the *HBS1L/MYB* intergenic region regulates erythroid cell proliferation, maturation, and foetal haemoglobin expression; (ii) *HK1* mutations lead to haemolytic anaemia via hexokinase deficiency, which in turn likely affects erythropoiesis; (iii) *IRS2* expression plays a role in erythroid cell differentiation through binding to cellular receptors involved in normal haematopoiesis; (iv) *ARHGAP9* regulates adhesion of haematopoietic cells to the extracellular matrix, which can influence their localisation and differentiation potential, and *R3HDM2* has been mapped to haemoglobin and red blood cell traits in large-scale GWASs; (v) *CPS1* is directly related to glycine, which is an essential requirement for haem synthesis; (vi) *SLC22A5* is involved in the active cellular uptake of carnitine, which stimulates erythropoiesis; (vii) *SOX7* blocks differentiation of hematopoietic progenitors to erythroid and myeloid lineages. In erythroblasts, *TFR2* is a sensor of holotransferrin, and is thought to protect against excessive erythrocytosis in the presence of iron deficiency. ❽ Finally, the immune response to external pathogens, which compete for iron, may also influence overall iron availability. Among the genes identified, *LVRN* may play a role in the synthesis of defensins and defensin-like peptides such as hepcidin, potentially contributing to iron homoeostasis via immune response; (ii) *MFSD6* recognises major histocompatibility complex type I (MHC-I) molecules and mediates MHC-I restricted killing by macrophages; (iii) *MPO* catalyses the production of hypohalous acids, primarily hypochlorous acid in physiologic situations, and other toxic intermediates that greatly enhance microbicidal activity. Images from Servier Medical Art (https://smart.servier.com), licensed under a Creative Commons Attribution 3.0 Unported (CC BY 3.0) Licence. **B** This figure summarises the genes mentioned in Table 2 of this study, as well as other iron-homoeostasis genes provided for contextual information. Genes with an established role in transferrin receptor synthesis, recycling, or degradation are shown in red and italic; genes with a potential role are presented in dark grey and italic. Relevant references to other studies are included in Supplementary Data 8. ❶ *TFRC* codes for transferrin receptor 1, which is constitutively expressed in most cells, especially erythrocytes. *TFR2* codes for transferrin receptor 2, linked to iron sensing and maintenance of body iron homoeostasis. *PGS1* is involved in the synthesis of cardiolipin, a phospholipid of mitochondrial membranes implicated in the regulation of transferrin receptor expression. ❷ After O-linked glycosylation, possibly mediated by the protein product of *GALNT6*, transferrin receptor 1 is expressed on the external surface of the cytoplasmic membrane. ❸ *HFE* interactswith transferrin receptor 1, facilitating cellular iron-sensing function and playing an important part in the regulation of hepcidin expression in response to body iron status. ❹ Iron-loaded transferrin (holotranferrin) binds to the receptor and the complex is internalised through clathrin-mediated endocytosis. ❺ A proton pump acidifies the endosome, which causes release of iron from holotransferrin; iron-deprived transferrin (apotransferrin) remains bound to its receptor. ❻ The endosome is usually recycled to the plasma membrane, a process likely regulated by (i) *LRBA*, known to influence recycling of cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) via the classical recycling pathway used by receptors such as transferrin and (ii) *UBXN6*, which negatively regulates the adenosine triphosphate (ATP) hydrolytic activity of valosin containing protein (VCP), an ATP-driven segregase; VCP depletion delays transferrin receptor recycling. ❼ At neutral pH, apotransferrin dissociates from transferrin receptor and is ready to bind to free iron. The transferrin receptor may also be ubiquitinated and directed to lysosomal degradation, which is mediated by *MARCH8*, a membrane-associated zinc-finger factor, and, possibly, also by *RPS6KB1*, a protein kinase involved in the mammalian target of rapamycin-protein S6 kinase (mTOR-S6K) pathway, which is implicated in the degradation of transferrin receptor 1. ❽ Finally, *PCSK7* mediates the shedding of soluble transferrin receptor (sTfR) from the transferrin receptor. When iron availability is limited, sTfR levels increase at least in part by downregulating expression of *PCSK7* or neighbouring genes. Images from Servier Medical Art (https://smart.servier.com), licensed under a Creative Commons Attribution 3.0 Unported (CC BY 3.0) Licence.

response (*LVRN*, *MFSD6*) and cellular trafficking (*MARCH8*, *NDFIP1*, *UBXN6*), although functional confirmation of candidate loci and variants is required. In the first large-scale MR study, involving over 1.4 million participants from four studies, we investigated the causal effects of iron-related pathways and systemic iron status on over 292 major health outcomes and conditions. Our findings showed that higher genetically predicted systemic iron status was inversely associated with mineral deficiency and anaemia-related disorders, suggesting that, aside from anaemia, iron deficiency is unlikely to be associated with major diseases explored in this analysis. Conversely, higher systemic iron status was positively associated with a range of conditions, including genitourinary, musculoskeletal, infectious, and neoplastic diseases. These associations attenuated after using the polygenic instrument without the pC282Y variant in *HFE*, which increases the risk for iron overload and, in its homozygous form, accounts for the majority of hemochromatosis cases[10], suggesting that they might be driven by very high iron levels.

We identified two new putative causal loci associated with hepcidin, *ARHGAP9/R3HDM2,* and *LVRN*, whose role in hepcidin metabolism is not yet fully understood. *ARHGAP9/R3HDM2* is involved in regulating adhesion of hematopoietic cells to the extracellular matrix, which can influence their localisation and differentiation potential[31]. Erythropoietic expansion, in turn, depresses hepcidin production[1]. *LVRN* codes for laeverin, an aminopeptidase cleaving N-terminal amino acids of peptides[32]. β-defensins, implicated in innate immunity, feature two sites under positive selection in the N-terminal region that may contribute to their functional diversity in primates[33]. *LVRN* may play a role in the synthesis of defensins and could affect hepcidin through an inflammation-mediated pathway. We also found biologically plausible candidate genes for 8 of the 13 new loci mapped to sTfR-associated sentinel variants. Of these, *IRS2*, *MARCH8* and *NDFIP1* appear to have a more established role in iron biology. *IRS2* is involved in erythroid cell differentiation[34], which in turn affects iron availability and transferrin receptor presentation[1], *MARCH8* mediates the lysosomal degradation of the transferrin receptor[35], and *NDFIP1* regulates iron import[36,37]. Five additional genes mapped to new sTfR-associated variants likely play a role in iron homoeostasis. *MFSD6* contributes to shaping the gut microbiome[38], possibly increasing iron availability; additionally, as it is an MHC-I receptor homologue[39], could potentially compete with transferrin receptor 1 for interacting with *HFE*, an MHC-I homologue[40]. *PGS1* could affect expression of transferrin receptor 1 via cardiolipin[41]. *SLC22A5* is involved in the cellular uptake of carnitine[42], which stimulates erythropoiesis[43]. *UBXN6* regulates endosome recycling to the plasma membrane[44], likely mediating transferrin receptor presentation and sTfR

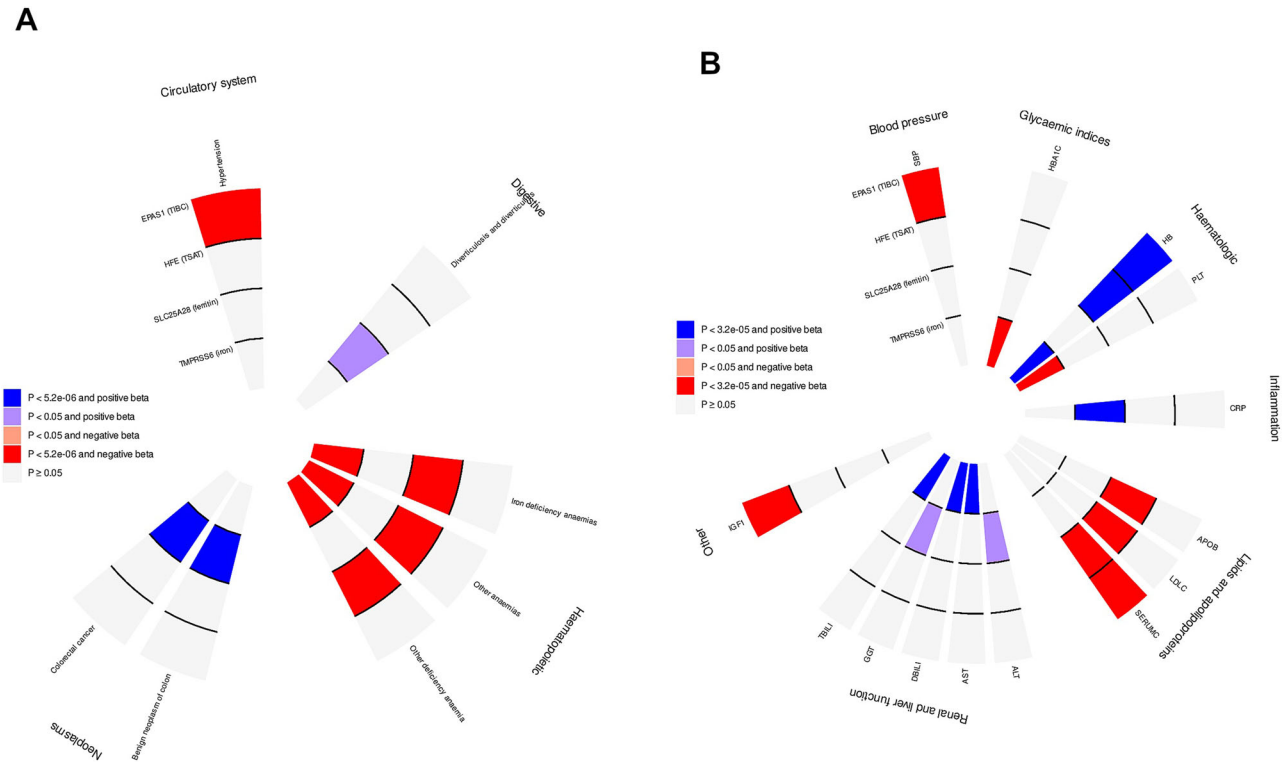


Fig. 4 | Putative causal effects of genetically predicted iron-related loci: Bonferroni-significant and nominal associations (null findings not presented). **A** Locus-based MR associations with disease outcomes in up to 1,469,361 deCODE, FinnGen, MVP, and UK Biobank participants. Only loci that are associated ($P < 5.2 \times 10^{-6}$) with at least one disease and have suggestive evidence of colocalization are shown. The terms in parenthesis indicate the trait that has been used for rescaling. **B** Locus-based MR associations with biomedical traits in up to 854,977 MVP and UK Biobank participants. Only loci that are associated with at least one disease outcome are shown. The terms in parenthesis indicate the trait that has been used for rescaling.

concentration. Finally, *ZFPM1* suppresses GATA-mediated activation of hepcidin expression[45], although its connection with sTfR remains unclear. It is worth noting that GWASs rely on population-level natural variation, which can lead to both overstatement and understatement of the role of individual modulators due to their natural variants being over- or under-represented in human genomes. At the population level, the impact of common variants that have a relatively minor role in iron biology (e.g., *HFE* variants) may be overstated, whereas the impact of rarer variants with a major effect on iron homoeostasis (e.g., *HAMP* variants) may be understated.

In addition to the expected association with anaemia-related phenotypes, the only other Bonferroni-significant associations of individual biological pathways that persisted in colocalization were *EPAS1* with hypertension and *SLC25A28* with colorectal cancer and benign neoplasm of colon. However, no other loci were associated with these diseases, suggesting that mediation through pathways specific to these loci (rather than through iron-related pathways) is more likely. For example, *EPAS1* codes for hypoxia-inducible factor 2-alpha, a transcription factor that contributes to maintaining oxygen homoeostasis in response to hypoxia through activation of several biological pathways, such as raising norepinephrine levels[46], in addition to iron absorption and transport.

The finding that multiple positive associations of systemic iron status with diseases attenuated after removing the pC282Y variant in *HFE* constitutes one of the key results of this study, suggesting that these associations may be driven by extreme iron overload and that moderate iron overload may be unlikely to affect health outcomes other than mineral metabolism disorders. In keeping with this interpretation, the strongest association with non-haematologic and non-metabolic disease outcomes was with greater risk of liver cancer, which is consistent with reports showing associations of the pC282Y variant with liver cancer[11], and mentioning hepatocellular carcinoma as a common manifestation of hemochromatosis[10]. The second-strongest association with non-haematologic and non-metabolic disease outcomes was with arthropathy, which is also consistent with reports of associations of pC282Y with osteoarthritis[11] and mentioning joint pain as a common symptom of hemochromatosis[10]. It is also possible, however, that the wider confidence intervals observed after removing the pC282Y variant in *HFE* may be due to reduced statistical power.

We also found positive associations of systemic iron status with greater risk of dermatophytosis/dermatomyositis, postoperative infection and cystitis/urethritis, broadly consistent with previous research that showed associations with skin[20] and bacterial[24] infections. It is worth noting that we did not observe associations with heart failure, which is consistent with a recent randomised trial[30] but in disagreement with previous trials[26–29]. We did find an inverse nominal association with ischaemic heart disease, in keeping with previous MR studies[20,21] but in disagreement with some observational evidence[12–15]. We also found a strong inverse association of genetically predicted systemic iron status with HbA1c, which may reflect greater erythrocyte turnover driven by iron excess[47]. Our findings reinforce previous warnings about interpreting HbA1c concentration in patients with iron-status imbalances[47], leading to potential underestimation of type-2 diabetes in individuals with iron overload.

Our investigation has several strengths. Firstly, the GWASs of hepcidin and sTfR have the largest sample size collected to date for genomic studies of these traits, enabling the discovery of the first genetic loci associated with hepcidin and multiple new loci associated with sTfR. Secondly, to assess the biomedical consequences of iron-altering biological pathways and systemic iron status, we employed an MR design on a wide range of major clinical outcomes, which reduces the impact of common sources of bias present in observational studies, such as confounding and reverse causality. It is, however, possible that this analysis may not capture rarer conditions and diseases not included in the curated list of health outcomes. Finally, by

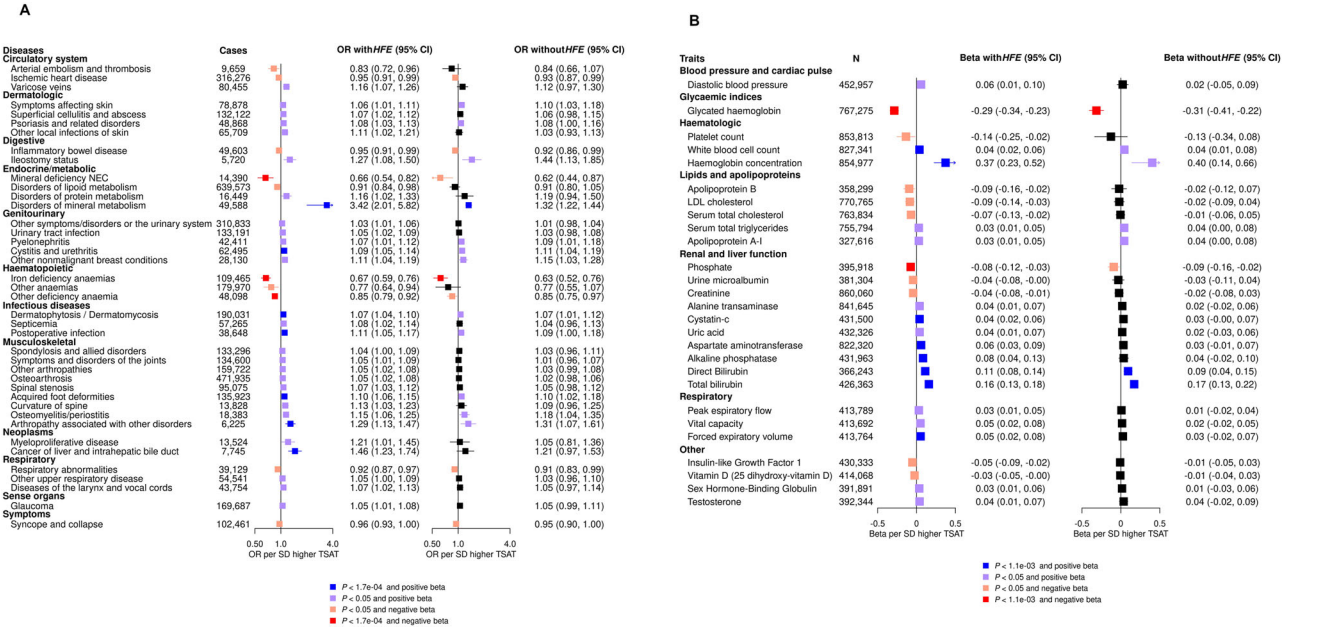


Fig. 5 | Putative causal effects of genetically predicted systemic iron status: Bonferroni-significant and nominal associations (null findings not presented). **A** MR associations of systemic iron status with disease outcomes in up to 1,492,717 deCODE, FinnGen, MVP, and UK Biobank participants. The estimates are expressed in odds ratio per one standard deviation (SD) higher transferrin saturation (TSAT) with confidence intervals shown between brackets. The plot shows estimates with the pC282Y variant in *HFE* (left-hand Forest plot) and without that variant (right-hand Forest plot), presenting diseases that have MR point estimates with the same direction in all the biobanks included in the meta-analysis. The instrument was generated using six variants mapped to *ERFE*, *HAMP*, *HFE*, *SLC25A37*, *TFR2* and *TMPRSS6* not affected by horizontal pleiotropy, indirect vertical pleiotropy, or collider bias and that: (i) were associated ($P < 5 \times 10^{-8}$) with at least one trait; (ii) were nominally associated ($P < 0.05$) with all the other iron traits except for hepcidin (as its levels are influenced by systemic iron status); and (iii) displayed a direction of association consistent across all traits. **B** MR associations of systemic iron status, using the same instrument, with biomedical traits in up to 860,060 MVP and UK Biobank participants. The estimates are expressed in mean change (beta) per one SD higher TSAT. The plot shows estimates with the pC282Y variant in *HFE* (left-hand Forest plot) and without that variant (right-hand Forest plot), presenting traits that have MR point estimates with the same direction in all the biobanks included in the meta-analysis.

leveraging the largest sample size in an iron MR conducted to date, we had very good (>90%) statistical power for the majority of the 292 outcomes included in our analysis.

However, there are also some limitations. Firstly, in our GWASs we focused on variants with MAF ≥ 0.001. Despite identifying some associations with rare variants, the role of rarer variants remains to be fully investigated. Secondly, the focus on European ancestry participants limits the generalisability of these findings, particularly in countries and ethnicities where the majority of the burden of iron deficiency lies. Thirdly, methodological differences in the GWASs, such as diverse adjustments for covariates and varying limits of detection, may have reduced homogeneity in meta-analysis, despite all GWASs adhering to the same analysis plan. Fourthly, our phenome scans demonstrated the extensive influence of genetic pleiotropy on iron traits. This study, however, utilises a systematic approach to reduce its impact on MR analyses by selecting only variants that are likely non-pleiotropic, and complementing locus-based MR with colocalization analysis to further reduce the impact of genetic confounding. Finally, this study assumes additive genetic associations of instrumental variables with iron traits, potentially missing between-variant interactions, and focuses on the linear effects of iron, potentially overlooking non-linear associations.

Taken together, this study increases knowledge of iron homoeostasis and its biomedical consequences in humans, suggesting that long-term exposure to higher iron levels is likely associated with lower risk of anaemia-related disorders and higher risk of genitourinary, musculoskeletal, infectious and neoplastic diseases.

Methods

Genetic discovery study of emerging iron traits

All studies included in the GWASs of hepcidin and sTfR followed the same analysis plan, described in the Supplementary Information, p 2. The characteristics of the cohorts included in this study are described in the Supplementary Information, pp 2–6 and in Supplementary Data 1.

We established a data-management and quality-check pipeline for study-specific GWAS results (Supplementary Information, p 6). We performed fixed-effect meta-analysis in METAL using the SCHEME STDERR command for all variants with MAF ≥ 0.001. After removing variants available in only one study and with a combined sample size lower than 20,000 participants, we estimated SNP-based heritability and genomic inflation factor using LDSC v. v1.0.1 (Supplementary Information, p 6).

To identify genetic variants independently associated with either hepcidin or sTfR concentration, we performed approximate conditional analysis using stepwise algorithm ('--cojo-slct') in gcta64 v. 1.26.0 on the whole genome. We selected all single nucleotide polymorphisms (SNPs) with $P < 5 \times 10^{-8}$ in the meta-analysis of each trait and we specified the same *p* value for the '--cojo p' argument, to ensure that conditionally independent SNPs were still genome-wide significant. Consistently with a previous study[48], we then clumped all resulting GWAS variants using PLINK v1.9 to include only independent variants not in linkage disequilibrium (LD) with one another within a 1 Mb window ($r^2 < 0.01$). We performed both GCTA and clumping using LD information from 41,845 unrelated participants in the INTERVAL study.

We provisionally mapped conditionally independent (sentinel) variants to their nearest gene using PhenoScanner v.2, a phenome scan tool that includes mapping to nearest gene retrieved from BEDOPS v. 2.4.26, with additional manual verification using the Ensembl genome browser (https://grch37.ensembl.org/). We assessed the novelty of association using two approaches. Firstly, we defined a 'novel variant' as any SNP (or its $r^2 \geq 0.7$ proxy) not associated with any iron traits in previous genome-wide studies[6–9]. Secondly, we defined as 'novel locus' any genomic locus (within 500 Kb window from each independent variant) not including one or more variants discovered in previous studies[6–9].

We used Ensembl Variant Effect Prediction to obtain information for several measures of functional consequence for each sentinel variant and their proxy variants ($r^2 > 0.7$) (Supplementary Information, pp 6–7)[49]. We conducted phenome scans drawing on the curated database of >65 billion genetic summary statistics available in PhenoScanner v.2 (Supplementary Information, p 7). We estimated genetic correlation using summary statistics from the present study (meta-analysis of hepcidin and sTfR concentration) and from a previous GWAS of conventional iron traits[9], using LDSC with the '--rg' argument. We estimated phenotypic Pearson correlation and its precision in up to 40,197 INTERVAL participants.

To map sentinel variants to candidate genes, we used a combination of manual curation and colocalization with expression and protein quantitative data. We first mapped the above-defined conditionally independent and uncorrelated GWAS signals to their nearest gene and then collapsed overlapping genes within 200 Kb from each other. The process was performed independently for hepcidin- and sTfR-associated variants. This led to the definition of 43 non-overlapping loci. Of these, 21 already had a biologically plausible candidate gene (e.g., *HFE*, *TMPRSS6*, *HAMP*, *TFRC*). For the remaining 22 loci, we performed conditional colocalization in Sum of Single Effects (SuSiE) v. 0.11.92 and Coloc v. 5.1.0, following the procedure described in Supplementary Information, pp 7–8. Locus-specific information on our candidate gene mapping process, including a summary of our manual curation, is available in Supplementary Data 8.

Locus-based and polygenic phenome-wide MR analysis

We collated 197 genetic variants (189 after deduplication) associated with iron traits, of which 52 associated with either hepcidin or sTfR (in the present study) and 145 associated with serum iron, ferritin, TSAT and TIBC (reported previously[9]). Because iron is involved in multiple biological processes, genetic variants associated with iron traits are often associated also with other traits (pleiotropy). This may lead to biased MR associations if genetic associations with iron traits are distinct (horizontal pleiotropy) or mediated by a non-iron trait (indirect vertical pleiotropy). To reduce the impact of horizontal and indirect vertical pleiotropy in our analysis, we performed phenome scans in MR Base and retained 57 genetic variants mapped to genes that (i) included only sentinel variants associated with iron traits or iron-related traits (such as haemoglobin concentration and erythrocyte count); (ii) affected iron homoeostasis directly and not via a non-iron phenotype (e.g., variants mapped to *HFE*, *TMPRSS6* and *HAMP*). We then assessed potential collider bias by comparing the genetic associations with and without adjustment for covariates that may result in collider bias (e.g., body mass index, smoking and others) in up to $N = 40{,}197$ INTERVAL participants (Supplementary Information, p 8). This analysis showed very high correlation ($r^2 \approx 1.00$) between estimates of these two approaches. All effect estimates had the same direction in the two models, apart from two variants (rs79694859 and rs10804630) that we removed from our list of MR instruments, leaving 55 variants for further analysis (Supplementary Data 9).

To select locus-based MR instruments, firstly, we mapped these 55 variants to their most plausible or nearest genes as defined in their source GWAS, leading to 39 non-overlapping loci. To ensure better generalisability of associations with clinical outcomes, we further selected 33 (out of 39) loci including at least one variant available in all the studies involved in the MR (Supplementary Data 10). We performed stepwise approximate conditional analysis for the 200 Kb region around each variant's mapped candidate gene ($P < 10^{-5}$, $r^2 < 0.1$) in gcta64 v. 1.26.0 using genetic summary statistics for the most strongly associated iron trait at each locus. This returned locus-based instruments for 33 loci with variance explained between <0.1%–4.1% (Supplementary Information, p 8; Supplementary Data 10; Supplementary Data 11). To select the polygenic MR instrument of systemic iron status, we filtered the above-mentioned 55 variants and included those that were: (i) associated ($P < 5 \times 10^{-8}$) with at least one iron trait, (ii) nominally associated ($P < 0.05$) with all the other iron traits considered; and (iii) with a direction consistent across all traits (e.g., positive for iron, ferritin, TSAT and negative for TIBC and sTfR; or the other way round) (Supplementary Fig. 5; Supplementary Data 11). Because hepcidin is influenced by systemic iron status and therefore it is difficult to disentangle whether genetic associations with hepcidin affect this trait directly or through other iron traits, we did not consider hepcidin associations in the definition of the polygenic instrument. We selected six variants for the polygenic instrument of systemic iron status mapped to *ERFE*, *HAMP*, *HFE*, *SLC25A37*, *TFR2* and *TMPRSS6*. In MR analysis, we rescaled the polygenic instrument by TSAT as it had the highest variance explained, 4.4%. We performed sensitivity analyses for key MR results utilising more liberal sets of polygenic instruments, illustrating the value of the 6-variant instrument in mitigating pleiotropy and heterogeneity (Supplementary Data 15). We estimated statistical power for this instrument, showing ≥90% power for ≥50% disease outcomes while assuming an OR of 1.5 (Supplementary Information, p 8; Supplementary Fig. 6). Because variants in *HFE* had the strongest genetic associations across all traits analysed (Supplementary Fig. 7), we performed a sensitivity analysis using a polygenic instrument without the *HFE* variant to identify MR association driven by *HFE*.

Before performing MR analysis, we estimated genetic associations of all instruments with health outcomes from deCODE, FinnGen data freeze 10 (R10), MVP and UKBB and with biomedical traits from MVP and UKBB in European-ancestry participants. We adjusted for age, sex (for non-sex-specific outcomes) and either the first 10 principal components of ancestry (FinnGen, MVP and UKBB) or county (deCODE). Information on the deCODE[50], FinnGen[51], MVP[52] and UKBB[53] cohorts is available elsewhere. We meta-analysed study-specific genetic associations using fixed-effects models in the 'metafor' R package. We defined 292 binary disease outcomes available in all four studies using a curated list of major phecodes available in the 'PheWAS' R package. To restrict our analysis to major health outcomes of interest, we discarded any sub-categories (i.e. phecodes with four or more characters), removed hereditary/poisoning-related/accident-related outcomes and those with less than 100 events in each study. The disease outcomes were grouped in the following domains: circulatory system, dermatologic, digestive, endocrine/metabolic, genitourinary, haematopoietic, infectious diseases, mental disorders, musculoskeletal, neoplasms, neurological, pregnancy complications, respiratory, sense organs, symptoms. We grouped biomedical traits in the following domains: blood pressure and cardiac pulse, glycaemic indices, haematologic, inflammation, lipids and apolipoproteins, renal and liver function, respiratory, other.

We performed univariable MR using the inverse-variance weighted method for each locus-based and polygenic instrument while accounting for between-variant correlation estimated in INTERVAL. We performed sensitivity analyses using MR Egger regression and weighted median estimator. We used fixed-effect models in locus-based analyses and random-effects models in polygenic analyses. We quantified between-variant heterogeneity using the I-squared statistic. To account for multiple testing, we used Bonferroni-corrected thresholds for all analyses. For locus-based analyses, these were $P < 0.05/(33 \times 292)$ (5.2×10^{-6}) for diseases and $P < 0.05/(33 \times 47)$ (3.2×10^{-5}) for traits. For polygenic analyses, the thresholds were $P < 0.05/292$ (1.7×10^{-4}) for diseases and $P < 0.05/47$ (1.1×10^{-3}) for traits. To reduce the impact of individual study-specific estimates that may disproportionately affect meta-analytic estimates, in the main figures we present Bonferroni-significant and nominal results for diseases and traits that have MR estimates in the same direction (regardless of their p value) in all the biobanks included in the meta-analysis, although all results are available in the Supplementary Data. Associations with p values below 0.05 but above the Bonferroni thresholds are described as 'nominal associations'. For locus-based MR nominal associations, we performed colocalization analysis to remove associations chiefly driven by genetic confounding (Supplementary Information, p 8).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

GWAS summary statistics are publicly available through the NHGRI-EBI GWAS Catalogue (hepcidin: accession number GCST90451683; soluble transferrin receptor: accession number GCST90451684).

Received: 22 March 2024; Accepted: 21 October 2024;
Published online: 06 December 2024

References

1. Camaschella, C. Iron-deficiency anemia. *N. Engl. J. Med.* **372**, 1832–1843 (2015).
2. Hentze, M. W., Muckenthaler, M. U., Galy, B. & Camaschella, C. Two to tango: regulation of Mammalian iron metabolism. *Cell* **142**, 24–38 (2010).
3. Nemeth, E. & Ganz, T. Hepcidin and iron in health and disease. *Annu. Rev. Med.* **74**, 261–277 (2023).
4. Beguin, Y. Soluble transferrin receptor for the evaluation of erythropoiesis and iron status. *Clin. Chim. Acta* **329**, 9–22 (2003).
5. Galesloot, T. E. et al. Meta-GWAS and meta-analysis of exome array studies do not reveal genetic determinants of serum hepcidin. *PLoS One* **11**, e0166628 (2016).
6. Oexle, K. et al. Novel association to the proprotein convertase PCSK7 gene locus revealed by analysing soluble transferrin receptor (sTfR) levels. *Hum. Mol. Genet.* **20**, 1042–1047 (2011).
7. Benyamin, B. et al. Novel loci affecting iron homeostasis and their effects in individuals at risk for hemochromatosis. *Nat. Commun.* **5**, 4926 (2014).
8. Bell, S. et al. A genome-wide meta-analysis yields 46 new loci associating with biomarkers of iron homeostasis. *Commun. Biol.* **4**, 156 (2021).
9. Moksnes, M. R. et al. Genome-wide meta-analysis of iron status biomarkers and the effect of iron on all-cause mortality in HUNT. *Commun. Biol.* **5**, 591 (2022).
10. Brissot, P. et al. Haemochromatosis. *Nat. Rev. Dis. Prim.* **4**, 18016 (2018).
11. Pilling, L. C. et al. Common conditions associated with hereditary haemochromatosis genetic variants: cohort study in UK Biobank. *BMJ* **364**, k5222 (2019).
12. Sempos, C. T., Looker, A. C., Gillum, R. F. & Makuc, D. M. Body iron stores and the risk of coronary heart disease. *N. Engl. J. Med.* **330**, 1119–1124 (1994).
13. Danesh, J. & Appleby, P. Coronary heart disease and iron status: meta-analyses of prospective studies. *Circulation* **99**, 852–854 (1999).
14. Sempos, C. T. Do body iron stores increase the risk of developing coronary heart disease? *Am. J. Clin. Nutr.* **76**, 501–503 (2002).
15. Munoz-Bravo, C., Gutierrez-Bedmar, M., Gomez-Aracena, J., Garcia-Rodriguez, A. & Navajas, J. F. Iron: protector or risk factor for cardiovascular disease? Still controversial. *Nutrients* **5**, 2384–2404 (2013).
16. Das De, S., Krishna, S. & Jethwa, A. Iron status and its association with coronary heart disease: systematic review and meta-analysis of prospective studies. *Atherosclerosis* **238**, 296–303 (2015).
17. Liu, J., Li, Q., Yang, Y. & Ma, L. Iron metabolism and type 2 diabetes mellitus: a meta-analysis and systematic review. *J. Diabetes Investig.* **11**, 946–955 (2020).
18. Fonseca-Nunes, A., Jakszyn, P. & Agudo, A. Iron and cancer risk-a systematic review and meta-analysis of the epidemiological evidence. *Cancer Epidemiol. Biomark. Prev.* **23**, 12–31 (2014).
19. Chang, V. C., Cotterchio, M. & Khoo, E. Iron intake, body iron status, and risk of breast cancer: a systematic review and meta-analysis. *BMC Cancer* **19**, 543 (2019).
20. Gill, D. et al. Associations of genetically determined iron status across the phenome: a mendelian randomization study. *PLoS Med.* **16**, e1002833 (2019).
21. Gill, D. et al. The effect of iron status on risk of coronary artery disease: a mendelian randomization study-brief report. *Arterioscler. Thromb. Vasc. Biol.* **37**, 1788–1792 (2017).
22. Gill, D. et al. Effects of genetically determined iron status on risk of venous thromboembolism and carotid atherosclerotic disease: a mendelian randomization study. *J. Am. Heart Assoc.* **8**, e012994 (2019).
23. Galesloot, T. E. et al. Iron and hepcidin as risk factors in atherosclerosis: what do the genes say? *BMC Genet.* **16**, 79 (2015).
24. Butler-Laporte, G. et al. Increasing serum iron levels and their role in the risk of infectious diseases: a Mendelian randomization approach. *Int. J. Epidemiol.* **52**, 1163–1174 (2023).
25. Yuan, S. et al. Iron status and cancer risk in UK biobank: a two-sample mendelian randomization study. *Nutrients* **12**, https://doi.org/10.3390/nu12020526 (2020).
26. von Haehling, S., Jankowska, E. A., van Veldhuisen, D. J., Ponikowski, P. & Anker, S. D. Iron deficiency and cardiovascular disease. *Nat. Rev. Cardiol.* **12**, 659–669 (2015).
27. Anker, S. D. et al. Ferric carboxymaltose in patients with heart failure and iron deficiency. *N. Engl. J. Med.* **361**, 2436–2448 (2009).
28. Ponikowski, P. et al. Beneficial effects of long-term intravenous iron therapy with ferric carboxymaltose in patients with symptomatic heart failure and iron deficiencydagger. *Eur. Heart J.* **36**, 657–668 (2015).
29. Kalra, P. R. et al. Intravenous ferric derisomaltose in patients with heart failure and iron deficiency in the UK (IRONMAN): an investigator-initiated, prospective, randomised, open-label, blinded-endpoint trial. *Lancet* **400**, 2199–2209 (2022).
30. Mentz, R. J. et al. Ferric carboxymaltose in heart failure with iron deficiency. *N. Engl. J. Med.* **389**, 975–986 (2023).
31. Klamer, S. & Voermans, C. The role of novel and known extracellular matrix and adhesion molecules in the homeostatic and regenerative bone marrow microenvironment. *Cell Adh. Migr.* **8**, 563–577 (2014).
32. Maruyama, M. et al. Laeverin/aminopeptidase Q, a novel bestatin-sensitive leucine aminopeptidase belonging to the M1 family of aminopeptidases. *J. Biol. Chem.* **282**, 20088–20096 (2007).
33. Semple, C. A. et al. The complexity of selection at the major primate beta-defensin locus. *BMC Evol. Biol.* **5**, 32 (2005).
34. Machado-Neto, J. A. et al. Downregulation of IRS2 in myelodysplastic syndrome: a possible role in impaired hematopoietic cell differentiation. *Leuk. Res.* **36**, 931–935 (2012).
35. Fujita, H., Iwabu, Y., Tokunaga, K. & Tanaka, Y. Membrane-associated RING-CH (MARCH) 8 mediates the ubiquitination and lysosomal degradation of the transferrin receptor. *J. Cell Sci.* **126**, 2798–2809 (2013).
36. Foot, N. J. et al. Ndfip1-deficient mice have impaired DMT1 regulation and iron homeostasis. *Blood* **117**, 638–646 (2011).
37. Traeger, L. et al. UBA6 and NDFIP1 regulate the degradation of ferroportin. *Haematologica* **107**, 478–488 (2022).
38. Yoshida, R. MHC class I recognition by monocyte-/macrophage-specific receptors. *Adv. Immunol.* **124**, 207–247 (2014).
39. Boldison, J. et al. Gene expression profiling in NOD mice reveals that B cells are highly educated by the pancreatic environment during autoimmune diabetes. *Diabetologia* **66**, 551–566 (2023).
40. Reuben, A., Chung, J. W., Lapointe, R. & Santos, M. M. The hemochromatosis protein HFE 20 years later: an emerging role in antigen presentation and in the immune system. *Immun. Inflamm. Dis.* **5**, 218–232 (2017).
41. Li, Y. et al. Cardiolipin-deficient cells have decreased levels of the iron-sulfur biogenesis protein frataxin. *J. Biol. Chem.* **295**, 11928–11937 (2020).
42. Tamai, I. Pharmacological and pathophysiological roles of carnitine/organic cation transporters (OCTNs: SLC22A4, SLC22A5 and Slc22a21). *Biopharm. Drug Dispos.* **34**, 29–44 (2013).

43. Kitamura, Y., Satoh, K., Satoh, T., Takita, M. & Matsuura, A. Effect of L-carnitine on erythroid colony formation in mouse bone marrow cells. *Nephrol. Dial. Transplant.* **20**, 981–984 (2005).
44. Trusch, F. et al. The N-terminal region of the ubiquitin regulatory X (UBX) domain-containing protein 1 (UBXD1) modulates interdomain communication within the valosin-containing protein p97. *J. Biol. Chem.* **290**, 29414–29427 (2015).
45. Bagu, E. T. & Santos, M. M. Friend of GATA suppresses the GATA-induced transcription of hepcidin in hepatocytes through a GATA-regulatory element in the HAMP promoter. *J. Mol. Endocrinol.* **47**, 299–313 (2011).
46. Peng, Y. J. et al. Hypoxia-inducible factor 2alpha (HIF-2alpha) heterozygous-null mice exhibit exaggerated carotid body sensitivity to hypoxia, breathing instability, and hypertension. *Proc. Natl Acad. Sci. USA* **108**, 3065–3070 (2011).
47. Ahmad, J. & Rafat, D. HbA1c and iron deficiency: a review. *Diabetes Metab. Syndr.* **7**, 118–122 (2013).
48. Codd, V. et al. Polygenic basis and biomedical consequences of telomere length variation. *Nat. Genet.* **53**, 1425–1433 (2021).
49. McLaren, W. et al. The ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).
50. Hakonarson, H., Gulcher, J. R. & Stefansson, K. deCODE genetics, Inc. *Pharmacogenomics* **4**, 209–215 (2003).
51. Kurki, M. I. et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
52. Gaziano, J. M. et al. Million veteran program: a mega-biobank to study genetic influences on health and disease. *J. Clin. Epidemiol.* **70**, 214–223 (2016).
53. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

Acknowledgements

We thank Professor Clara Camaschella for her contribution to the definition of non-pleiotropic loci. This research was supported by the British Heart Foundation (RG/18/13/33946), National Institute for Health and Care Research (NIHR) Cambridge Biomedical Research Centre (BRC-1215-20014; NIHR203312) [*] and Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome. S.Be. is supported by Cancer Research UK (A27657). A.M. is funded by the NIHR Blood and Transplant Research Unit (BTRU) in Donor Health and Behaviour (NIHR203337) [*]. E.B. is supported by Schmidt Science Fellows, in partnership with the Rhodes Trust. S.Bu. is supported by the Wellcome Trust (225790/Z/22/Z) and the United Kingdom Research and Innovation Medical Research Council (MC_UU_00002/7). This research was supported by the National Institute for Health Research Cambridge Biomedical Research Centre (NIHR203312). J.Da. holds a British Heart Foundation Professorship and a NIHR Senior Investigator Award [*]. A.M.W. is part of the BigData@-Heart Consortium, funded by the Innovative Medicines Initiative-2 Joint Undertaking under grant agreement No 116074. *The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. The CHRIS study is conducted in collaboration between the Eurac Research Institute for Biomedicine and the Healthcare System of the Autonomous Province of Bolzano-South Tyrol. Investigators thank all study participants, general practitioners of the Vinschgau/Val Venosta district, and Schlanders/Silandro and Autonomous Province of Bolzano-South Tyrol Healthcare System staff for their support and collaboration. They also thank the study team in Silandro/Schlanders, CHRIS Biobank personnel, and all Institute for Biomedicine colleagues who contributed to the study. Extensive acknowledgment is reported at DOI: 10.1186/s12967-015-0704-9. CHRIS Bioresource Research Impact Factor (BRIF) code: BRIF6107. The CHRIS study is funded by the Department of Innovation, Research, and the University of the Autonomous Province of Bolzano-South Tyrol and supported by the European Regional Development Fund (FESR1157). We thank the LURM Research Facility at the University Hospital of Verona for supporting in measuring hepcidin. CROATIA_Vis. This study was funded by the Medical Research Council (UK), European Commission Framework 6 project EUROSPAN (Contract No. LSHG-CT-2006-018947), and Republic of Croatia Ministry of Science, Education, and Sports research grants to I.R. (108-1080315-0302). CH was supported by an MRC Human Genetics Unit programme grant, 'Quantitative traits in health and disease' (U. MC_UU_00007/10). We would like to acknowledge the staff of several institutions in Croatia that supported the fieldwork, including but not limited to The University of Split and Zagreb Medical Schools, the Institute for Anthropological Research in Zagreb, and the Croatian Institute for Public Health. DBDS. This study was supported by the Danish Council for Independent Research (09-069412 and 0602-02634B) and the Bio- and Genome Bank Denmark. The DBDS genetic infrastructure was supported by the Novo Nordisk Foundation (NNF17OC0027594). We thank the Danish blood donors for their valuable participation in the Danish Blood Donor Study, as well as the staff at the blood centres for making this study possible. deCODE. We would like to thank the individuals who participated in the study and whose contribution made this work possible. FinDonor_1 & FinDonor_2. These studies were supported by the Finnish Funding Agency for Technology and Innovation (Tekes) to the Salwe GID (Personalised Diagnostics and Care) programme (ID 3982/31/2013) and by the VTR funding from the Finnish Government. FinnGen. We want to acknowledge the participants and investigators of FinnGen study. The FinnGen project is funded by two grants from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and the following industry partners: AbbVie Inc., AstraZeneca UK Ltd, Biogen MA Inc., Bristol Myers Squibb (and Celgene Corporation & Celgene International II Sàrl), Genentech Inc., Merck Sharp & Dohme LCC, Pfizer Inc., GlaxoSmithKline Intellectual Property Development Ltd., Sanofi US Services Inc., Maze Therapeutics Inc., Janssen Biotech Inc, Novartis AG, and Boehringer Ingelheim International GmbH. Following biobanks are acknowledged for delivering biobank samples to FinnGen: Auria Biobank (www.auria.fi/biopankki), THL Biobank (www.thl.fi/biobank), Helsinki Biobank (www.helsinginbiopankki.fi), Biobank Borealis of Northern Finland (https://www.ppshp.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-English.aspx), Finnish Clinical Biobank Tampere (www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere), Biobank of Eastern Finland (www.ita-suomenbiopankki.fi/en), Central Finland Biobank (www.ksshp.fi/fi-FI/Potilaalle/Biopankki), Finnish Red Cross Blood Service Biobank (www.veripalvelu.fi/verenluovutus/biopankkitoiminta), Terveystalo Biobank (www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/) and Arctic Biobank (https://www.oulu.fi/en/university/faculties-and-units/faculty-medicine/northern-finland-birth-cohorts-and-arctic-biobank). All Finnish Biobanks are members of BBMRI.fi infrastructure (www.bbmri.fi). Finnish Biobank Cooperative -FINBB (https://finbb.fi/) is the coordinator of BBMRI-ERIC operations in Finland. The Finnish biobank data can be accessed through the Fingenious® services (https://site.fingenious.fi/en/) managed by FINBB. The FinnGen study is approved by Finnish Institute for Health and Welfare (permit numbers: THL/2031/6.02.00/2017, THL/1101/5.05.00/2017, THL/341/6.02.00/2018, THL/2222/6.02.00/2018, THL/283/6.02.00/2019, THL/1721/5.05.00/2019 and THL/1524/5.05.00/2020), Digital and population data service agency (permit numbers: VRK43431/2017-3, VRK/6909/2018-3, VRK/4415/2019-3), the Social Insurance Institution (permit numbers: KELA 58/522/2017, KELA 131/522/2018, KELA 70/522/2019, KELA 98/522/2019, KELA 134/522/2019, KELA 138/522/2019, KELA 2/522/2020, KELA 16/522/2020), Findata permit numbers THL/2364/14.02/2020, THL/4055/14.06.00/2020,,THL/3433/14.06.00/2020, THL/4432/14.06/2020, THL/5189/14.06/2020, THL/5894/14.06.00/2020, THL/6619/14.06.00/2020, THL/209/14.06.00/2021, THL/688/14.06.00/2021, THL/1284/14.06.00/2021, THL/1965/14.06.00/2021, THL/5546/14.02.00/2020, THL/2658/14.06.00/2021, THL/4235/14.06.00/202, Statistics Finland (permit numbers: TK-53-1041-17

and TK/143/07.03.00/2020 (earlier TK-53-90-20) TK/1735/07.03.00/2021, TK/3112/07.03.00/2021) and Finnish Registry for Kidney Diseases permission/extract from the meeting minutes on 4th July 2019. The Biobank Access Decisions for FinnGen samples and data utilised in FinnGen Data Freeze 10 include: THL Biobank BB2017_55, BB2017_111, BB2018_19, BB_2018_34, BB_2018_67, BB2018_71, BB2019_7, BB2019_8, BB2019_26, BB2020_1, Finnish Red Cross Blood Service Biobank 7.12.2017, Helsinki Biobank HUS/359/2017, HUS/248/2020, Auria Biobank AB17-5154 and amendment #1 (August 17 2020), AB20-5926 and amendment #1 (April 23 2020) and it's modification (22 Sep 2021), Biobank Borealis of Northern Finland_2017_1013, Biobank of Eastern Finland 1186/2018 and amendment 22 § /2020, Finnish Clinical Biobank Tampere MH0004 and amendments (21.02.2020 & 06.10.2020), Central Finland Biobank 1-2017, and Terveystalo Biobank STB 2018001 and amendment 25th Aug 2020. InCHIANTI. The study baseline (1998-2000) was supported by the Italian Ministry of Health (ICS110.1/RF97.71) and by the U.S. National Institute on Aging (Contracts: 263 MD 9164 and 263 MD 821336). INGI-VB. The research was supported by funds from Compagnia di San Paolo, Torino, Italy; Fondazione Cariplo, Italy and Ministry of Health, Ricerca Finalizzata 2008 and CCM 2010, and Telethon, Italy to Daniela Toniolo, Italian Ministry of Health, through the contribution given to the Institute for Maternal and Child Health IRCCS Burlo Garofolo, Trieste, Italy—RC 01/21 to MPC, and D70-RESRICGIROTTO to GG. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank the inhabitants of the VB that made this study possible, the local administrations, the Tortona and Genova archdiocese, and the ASL-22, Novi Ligure (AL) for their support. We also thank Prof. Daniela Toniolo for the project supervision, Clara Camaschella for data collection supervision and organisation of the clinical data collection, Fiammetta Viganò for technical help, and Corrado Masciullo and Massimiliano Cocca for building the analysis platform. INTERVAL. The academic coordinating centre for INTERVAL was supported by core funding from NIHR Blood and Transplant Research Unit in Donor Health and Genomics (NIHR BTRU-2014-10024), UK Medical Research Council (MR/L003120/1), British Heart Foundation (SP/09/002; RG/13/13/30194; RG/18/13/33946) and the NIHR Cambridge BRC. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. The academic coordinating center would like to thank the blood donor center staff and blood donors for participating in the INTERVAL trial. KORA_F3. The study was initiated and financed by the Helmholtz Zentrum München—German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. NBS. The Nijmegen Biomedical Study is a population-based survey conducted at the Department for Health Evidence and the Department of Laboratory Medicine of the Radboud University Medical Center, Nijmegen, the Netherlands. Principal investigators of the Nijmegen Biomedical Study are Lambertus Kiemeney, André Verbeek, Dorine Swinkels, and Barbara Franke. We thank Doorlène van Tienoven and Anneke Geurts-Moespot for serum hepcidin measurements. PREVEND. PREVEND genetics was supported by the Dutch Kidney Foundation (Grant E033), the EU project grant GENECURE (FP-6 LSHM CT 2006 037697), the National Institutes of Health (grant 2R01LM010098), The Netherlands organisation for health research and development (NWO-Groot grant 175.010.2007.006, NWO VENI grant 916.761.70, ZonMw grant 90.700.441), and the Dutch Inter University Cardiology Institute Netherlands (ICIN). N. Verweij is supported by NWO VENI grant 016.186.125. We thank all individuals for participating in the PREVEND study. UK Biobank. This research has been conducted using the UK Biobank resource (Reference 88349).

Author contributions

Analysis: E.A., S.Be., R.S., S.J.K., L.G., D.M.G., F.W., V.T., A.M., S.K., J.Do., J.M., C.F., M.A., J.T., S.M., C.H., T.T., M.P.C., N.V., Y.J.V., T.E.G., E.F., M.M., A.G., K.I., A.Per. Data acquisition: W.O., D.J.R., J.Do., S.R.O., M.H.L., H.U., O.B.P., S.Br., K.B., C.E., DBDS, C.P., P.P.P., M.A., A.Pec., P.H., L.A.K., F.C.S., T.E.G., FinnGen, M.G., J.Da., A.S.B., E.D.A. Design: E.A., S.Be., D.Gil., L.G., R.T.L., M.H.L., H.U., L.F., S.Ba., G.G., P.H., L.A.K., S.Bu., B.B., K.S., M.G., A.M.W., A.S.B., E.D.A. Interpretation: E.A., S.Be., D.Gil., L.G., V.T., A.M., E.B., W.O., D.J.R., C.M.D., N.P., J.F.W., K.H., F.Q., P.S., D.Gu., L.D., K.C., M.I., S.Bu., B.B., K.O., D.S., K.S., M.M., K.I., J.Da., A.Per., A.M.W., A.S.B., E.D.A. Supervision: S.R.O., H.U., O.B.P., C.P., P.P.P., D.Gir., M.A., J.T., A.Pet., O.P., I.R., G.G., K.O., D.S., K.S., M.M., A.G., M.G., K.I., J.Da., A.Per., A.M.W., A.S.B., E.D.A. All authors read and approved the manuscript.

Competing interests

R.S. is currently employed at Astra Zeneca. N.V. is an employee and stockholder of Regeneron Pharmaceuticals. J.Da. serves on scientific advisory boards for AstraZeneca, Novartis, and UK Biobank, and has received multiple grants from academic, charitable, and industry sources outside of the submitted work. All other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-024-07115-3.

Correspondence and requests for materials should be addressed to Elias Allara or Emanuele Di Angelantonio.

Peer review information *Communications Biology* thanks Carla Cerami, Christopher D. Vulpe and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Joao Valente.

[1]BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [2]NIHR Blood and Transplant Research Unit in Donor Health and Behaviour, Cambridge, UK. [3]Victor Phillip Dahdaleh Heart and Lung Research Institute, University of Cambridge, Cambridge, UK. [4]Precision Breast Cancer Institute, Department of Oncology, University of Cambridge, Cambridge, UK. [5]Cancer Research UK Cambridge Centre, Li Ka Shing Centre, University of Cambridge, Cambridge, UK. [6]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK. [7]Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Boston, MA, USA. [8]Department of Medicine, Harvard Medical School, Boston, MA, USA. [9]Department of Statistics, Federal University of Parana, Curitiba, Brazil. [10]Division of Aging, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [11]Genetic Epidemiology Lab, Institute for Molecular Medicine Finland, Helsinki, Finland. [12]deCODE genetics/Amgen Inc., Reykjavik, Iceland. [13]UCL Institute of Health Informatics, London, UK. [14]BHF Centre of Research Excellence, School of Clinical Medicine, Addenbrooke's Hospital, Cambridge, UK. [15]Department of Haematology, University of Cambridge, Cambridge, UK. [16]NHS Blood and Transplant, Cambridge Biomedical Campus, Cambridge, UK. [17]Department of Haematology, Cambridge University Hospitals NHS Trust, Cambridge, UK. [18]Department of Haematology, University College London Hospitals NHS Trust, London, UK. [19]Radcliffe Department of Medicine, University of Oxford, John Radcliffe Hospital, Oxford, UK. [20]Department of Haematology, Churchill Hospital, Headington, Oxford, UK. [21]Department of Clinical Immunology, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark. [22]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [23]Statens Serum Institut, Copenhagen, Denmark. [24]Department of Immunology, Zealand University Hospital, Køge, Denmark. [25]Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [26]Department of Immunology, Aarhus University Hospital, Aarhus, Denmark. [27]Eurac Research, Institute for Biomedicine, Bolzano, Italy. [28]Department of Neurology, General Central Hospital, Bolzano, Italy. [29]Department of Medicine, Section of Internal Medicine, EuroBloodNet Referral Center, University Hospital of Verona, Verona, Italy. [30]Finnish Red Cross Blood Service, Helsinki, Finland. [31]msg life central europe gmbh, München, Germany. [32]Institute of Epidemiology, Helmholtz Munich, Neuherberg, Germany. [33]German Center for Diabetes Research (DZD), Neuherberg, Germany. [34]Institute for Medical Information Processing, Biometry and Epidemiology, Medical Faculty, Ludwig-Maximilians-Universität München, München, Germany. [35]Faculty of Medicine, University of Split, Split, Croatia. [36]Centre for Global Health Research, Usher Institute, University of Edinburgh, Edinburgh, Scotland Edinburgh, UK. [37]Medical Research Council Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK. [38]Centre for Cardiovascular Sciences, Queens Medical Research Institute, University of Edinburgh, Edinburgh, Scotland, UK. [39]Genomics Research Centre, Human Technopole, Milan, Italy. [40]Donor Studies, Department of Donor Medicine Research, Sanquin Research, Amsterdam, The Netherlands. [41]Department of Public and Occupational Health, Amsterdam Public Health Research Institute, Amsterdam UMC, Amsterdam, The Netherlands. [42]Longitudinal studies section, National Institute on Aging, Baltimore, MD, USA. [43]Geriatric Unit, Azienda Sanitaria Firenze (ASF), Florence, Italy. [44]Institute for Maternal and Child Health - IRCCS, Burlo Garofolo, Trieste, Italy. [45]Department of Medicine, Surgery and Health Sciences, University of Trieste, Trieste, Italy. [46]Department of Cardiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. [47]Regeneron Genetics Center, Tarrytown, NY, USA. [48]Department of Cardiology, University Medical Center Utrecht, Utrecht, The Netherlands. [49]IQ Health, Radboud University Medical Center, Nijmegen, The Netherlands. [50]Department of Urology, Radboud University Medical Center, Nijmegen, The Netherlands. [51]Department of Laboratory Medicine, Radboud University Medical Center, Nijmegen, The Netherlands. [52]School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland. [53]Department of Nutrition, Harvard T. H. Chan School of Public Health, Boston, MA, USA. [54]Cambridge Baker Systems Genomics Initiative, Baker Heart and Diabetes Institute, Melbourne, Australia. [55]Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge, UK. [56]Cambridge Baker Systems Genomics Initiative, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [57]Medical Research Council Biostatistics Unit, Cambridge, UK. [58]Australian Centre for Precision Health & Allied Health and Human Performance, University of South Australia, Adelaide, Australia. [59]South Australian Health and Medical Research Institute, Adelaide, Australia. [60]Neurogenetic Systems Analysis Group, Institute of Neurogenomics, Helmholtz Munich, Neuherberg, Germany. [61]Institute of Human Genetics, School of Medicine, Technical University of Munich, München, Germany. [62]Sanquin Blood Bank, Amsterdam, The Netherlands. [63]Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland. [64]Department of Human Genetics, Wellcome Sanger Institute, Hinxton, UK. [65]Cambridge Centre of Artificial Intelligence in Medicine, Cambridge, UK. [66]Health Data Science Centre, Human Technopole, Milan, Italy. [146]These authors contributed equally: Elias Allara, Steven Bell. [147]These authors jointly supervised this work: John Danesh, Alexandre Pereira, Angela M. Wood, Adam S. Butterworth, Emanuele Di Angelantonio. ✉e-mail: ea431@medschl.cam.ac.uk; ed303@medschl.cam.ac.uk

DBDS Genomic Consortium

Karina Banasik[25], Jakob Bay[24], Jens Kjærgaard Boldsen[26], Thorsten Brodersen[24], Søren Brunak[25], Alfonso Buil Demur[67], Lea Arregui Nordahl Christoffersen[24], Maria Didriksen[21], Khoa Manh Dinh[21,26], Joseph Dowsett[21], Christian Erikstrup[26,68], Bjarke Feenstra[21,69], Frank Geller[21,69], Daniel Gudbjartsson[70], Thomas Folkmann Hansen[71], Dorte Helenius Mikkelsen[67], Lotte Hindhede[26], Henrik Hjalgrim[69,72], Jakob Hjorth von Stemann[21], Bitten Aagaard Jensen[73], Andrew Joseph Schork[67], Kathrine Kaspersen[26], Bertram Dalskov Kjerulff[26], Mette Kongstad[21], Susan Mikkelsen[26], Christina Mikkelsen[21], Janna Nissen[21], Mette Nyegaard[74], Sisse Rye Ostrowski[21,22], Ole Birger Pedersen[22,24], Liam James Elgaard Quinn[24], Þórunn Rafnar[70], Palle Duun Rohde[74], Klaus Rostgaard[69,72], Michael Schwinn[21], Erik Sørensen[21], Kari Stefansson[70], Hreinn Stefánsson[70], Lise Wegner Thørner[21], Unnur Þorsteinsdóttir[70], Mie Topholm Bruun[75], Henrik Ullum[23], Thomas Werge[22,67] & David Westergaard[25]

[67]Institute of Biological Psychiatry, Mental Health Centre, Sct. Hans, Copenhagen University Hospital, Roskilde, Denmark. [68]Department of Clinical Medicine, Health, Aarhus University, Aarhus, Denmark. [69]Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark. [70]deCODE Genetics, Reykjavik, Iceland. [71]Danish Headache Center, Department of Neurology, Copenhagen University Hospital, Rigshospitalet-Glostrup, Copenhagen, Denmark. [72]Danish Cancer Society Research Center, Copenhagen, Denmark. [73]Department of Clinical Immunology, Aalborg University Hospital, Aalborg, Denmark. [74]Department of Health Science and Technology, Faculty of Medicine, Aalborg Univeristy, Aalborg, Denmark. [75]Department of Clinical Immunology, Odense University Hospital, Odense, Denmark.

FinnGen Consortium

Aarno Palotie[76,77,78], Mark Daly[76,77,78], Bridget Riley-Gills[79], Howard Jacob[79], Dirk Paul[80], Slavé Petrovski[80], Chia-Yen Chen[81], Sally John[81], George Okafo[82], Robert Plenge[83], Joseph Maranville[83], Mark McCarthy[84], Rion Pendergrass[84], Margaret G. Ehm[85], Kirsi Auro[86], Simonne Longerich[87], Anders Mälarstig[88], Anna Vlahiotis[88], Katherine Klinger[89], Clement Chatelain[89], Matthias Gossel[89], Karol Estrada[90], Robert Graham[90], Dawn Waterworth[91], Chris O'Donnell[92], Nicole Renaud[92], Tomi P. Mäkelä[93], Jaakko Kaprio[76], Minna Ruddock[94], Petri Virolainen[95], Antti Hakanen[95], Terhi Kilpi[96], Markus Perola[96], Jukka Partanen[97], Taneli Raivio[98], Jani Tikkanen[99], Raisa Serpi[99], Tarja Laitinen[100,101], Veli-Matti Kosma[102], Jari Laukkanen[103,104], Marco Hautalahti[105], Outi Tuovila[106], Raimo Pakkanen[106], Jeffrey Waring[79], Fedik Rahimov[79], Ioanna Tachmazidou[80], Zhihao Ding[7], Marc Jung[82], Hanati Tuoken[82], Shameek Biswas[83], David Pulford[107], Neha Raghavan[87], Adriana Huertas-Vazquez[87], Jae-Hoon Sul[87], Xinli Hu[88], Åsa Hedman[88], Ma´en Obeidat[92], Jonathan Chung[92], Jonas Zierer[92], Mari Niemi[92], Samuli Ripatti[76], Johanna Schleutker[95,115], Mikko Arvas[97], Olli Carpén[98,117], Reetta Hinttala[99], Johannes Kettunen[99,111], Arto Mannermaa[102], Katriina Aalto-Setälä[112], Mika Kähönen[100,101], Johanna Mäkelä[105], Reetta Kälviäinen[120], Valtteri Julkunen[120], Hilkka Soininen[120], Anne Remes[111], Mikko Hiltunen[121], Jukka Peltola[101], Minna Raivio[117], Pentti Tienari[117], Juha Rinne[115], Roosa Kallionpää[115], Juulia Partanen[116], Adam Ziemann[79], Nizar Smaoui[79], Anne Lehtonen[79], Susan Eaton[81], Heiko Runz[81], Sanni Lahdenperä[81], Natalie Bowers[84], Edmond Teng[84], Fanli Xu[123], Laura Addis[123], John Eicher[123], Qingqin S. Li[126], Karen He[91], Ekaterina Khramtsova[91], Martti Färkkilä[117], Jukka Koskela[117], Sampsa Pikkarainen[117], Airi Jussila[101], Katri Kaukinen[101], Timo Blomster[111], Mikko Kiviniemi[120], Markku Voutilainen[115], Tim Lu[84], Linda McCarthy[123], Amy Hart[91], Meijian Guan[91], Jason Miller[87], Kirsi Kalpala[88], Melissa Miller[88], Xinli Hu[88], Kari Eklund[117], Antti Palomäki[115], Pia Isomäki[101], Laura Pirilä[115], Oili Kaipiainen-Seppänen[120], Johanna Huhtakangas[111], Nina Mars[1], Apinya Lertratanakul[79], Coralie Viollet[80], Marla Hochfeld[83], Jorge Esparza Gordillo[123], Fabiana Farias[87], Nan Bing[88], Xinli Hu[88], Margit Pelkonen[120], Paula Kauppi[117], Hannu Kankaanranta[108,109,125], Terttu Harju[111], Riitta Lahesmaa[115], Susan Eaton[81], Hubert Chen[84], Joanna Betts[123], Rajashree Mishra[123], Majd Mouded[129], Debby Ngo[129], Teemu Niiranen[110], Felix Vaura[110], Veikko Salomaa[110], Kaj Metsärinne[115], Jenni Aittokallio[115], Jussi Hernesniemi[101], Daniel Gordin[117], Juha Sinisalo[117], Marja-Riitta Taskinen[117], Tiinamaija Tuomi[117], Timo Hiltunen[117], Amanda Elliott[76,78,127], Mary Pat Reeve[76], Sanni Ruotsalainen[76], Audrey Chu[123], Dermot Reilly[128], Mike Mendelson[122], Jaakko Parkkinen[88], Tuomo Meretoja[130], Heikki Joensuu[113], Johanna Mattson[117], Eveliina Salminen[117], Annika Auranen[101], Peeter Karihtala[113], Päivi Auvinen[120], Klaus Elenius[115], Esa Pitkänen[76], Nina Mars[76], Mark Daly[76,77,78], Relja Popovic[79], Margarete Fabre[80], Jennifer Schutzman[84], Diptee Kulkarni[123], Alessandro Porello[91], Andrey Loboda[87], Heli Lehtonen[88], Stefan McDonough[88], Sauli Vuoti[131], Kai Kaarniranta[118,120], Joni A. Turunen[119,134], Terhi Ollila[117], Hannu Uusitalo[101], Juha Karjalainen[76], Mengzhen Liu[79], Stephanie Loomis[81], Erich Strauss[84], Hao Chen[84], Kaisa Tasanen[111], Laura Huilaja[111], Katariina Hannula-Jouppi[117], Teea Salmi[101], Sirkku Peltonen[115], Leena Koulu[115], David Choy[84], Ying Wu[88], Pirkko Pussinen[117], Aino Salminen[117], Tuula Salo[117], David Rice[117], Pekka Nieminen[117], Ulla Palotie[117], Maria Siponen[120], Liisa Suominen[120], Päivi Mäntylä[120], Ulvi Gursoy[115], Vuokko Anttonen[111], Kirsi Sipilä[114,135], Hannele Laivuori[76], Venla Kurra[101], Laura Kotaniemi-Talonen[101], Oskari Heikinheimo[117], Ilkka Kalliala[117], Lauri Aaltonen[117], Varpu Jokimaa[115], Marja Vääräsmäki[111], Outi Uimari[111], Laure Morin-Papunen[111], Maarit Niinimäki[111], Terhi Piltonen[111], Katja Kivinen[76], Elisabeth Widen[76], Taru Tukiainen[76], Niko Välimäki[137], Eija Laakkonen[138], Jaakko Tyrmi[139,140], Heidi Silven[140], Eeva Sliz[140], Riikka Arffman[140], Susanna Savukoski[140], Triin Laisk[133], Natalia Pujol[133], Mengzhen Liu[79], Janet Kumar[85], Iiris Hovatta[137], Erkki Isometsä[117], Hanna Ollila[76], Jaana Suvisaari[110], Antti Mäkitie[132], Argyro Bizaki-Vallaskangas[101], Sanna Toppila-Salmi[136,141], Tytti Willberg[115], Elmo Saarentaus[76], Antti Aarnisalo[117], Elisa Rahikkala[111], Kristiina Aittomäki[124], Fredrik Åberg[142], Mitja Kurki[76,127], Aki Havulinna[76,110], Juha Mehtonen[76], Priit Palta[76], Shabbeer Hassan[76], Pietro Della Briotta Parolo[76], Wei Zhou[127], Mutaamba Maasha[127], Susanna Lemmelä[76], Manuel Rivas[143], Aoxing Liu[76], Arto Lehisto[76], Andrea Ganna[76], Vincent Llorens[76], Henrike Heyne[76], Nina Mars[76], Joel Rämö[76], Rodos Rodosthenous[76], Satu Strausz[76], Tuula Palotie[144], Kimmo Palin[137], Javier Garcia-Tabuenca[139], Harri Siirtola[139], Tuomo Kiiskinen[1], Jiwoo Lee[76,127], Kristin Tsuo[76,127], Kati Kristiansson[96], Kati Hyvärinen[30], Jarmo Ritari[30], Katri Pylkäs[140], Minna Karjalainen[140], Tuomo Mantere[99], Eeva Kangasniemi[100], Sami Heikkinen[121], Nina Pitkänen[95], Samuel Lessard[89], Lila Kallio[95], Tiina Wahlfors[96], Eero Punkka[98], Sanna Siltanen[100], Teijo Kuopio[103], Anu Jalanko[76], Huei-Yi Shen[76], Risto Kajanne[76], Mervi Aavikko[76], Helen Cooper[76], Denise Öller[76], Rasko Leinonen[76,145], Henna Palin[100], Malla-Maria Linna[98], Masahiro Kanai[127], Zhili Zheng[127], L. Elisa Lahtela[76], Mari Kaunisto[76], Elina Kilpeläinen[76], Timo P. Sipilä[76], Oluwaseun Alexander Dada[76], Awaisa Ghazal[76], Anastasia Kytölä[76], Rigbe Weldatsadik[76], Kati Donner[76], Anu Loukola[98], Päivi Laiho[96], Tuuli Sistonen[96], Essi Kaiharju[96], Markku Laukkanen[96], Elina Järvensivu[96], Sini Lähteenmäki[96], Lotta Männikkö[96], Regis Wong[96], Auli Toivola[96], Minna Brunfeldt[96], Hannele Mattsson[96], Sami Koskelainen[96], Tero Hiekkalinna[96], Teemu Paajanen[96], Shuang Luo[76], Shanmukha Sampath Padmanabhuni[76], Marianna Niemi[139], Javier Gracia-Tabuenca[139], Mika Helminen[139], Tiina Luukkaala[139], Iida Vähätalo[139], Jyrki Tammerluoto[76], Sarah Smith[105], Tom Southerington[105] & Petri Lehto[105]

[76]Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland. [77]Broad Institute of MIT and Harvard, Boston, MA, USA. [78]Massachusetts General Hospital, New Boston, NY, USA. [79]Abbvie, Chicago, IL, USA. [80]Astra Zeneca, Cambridge, United Kingdom. [81]Biogen, Cambridge, MA, USA. [82]Boehringer Ingelheim, Ingelheim am Rhein, Germany. [83]Bristol Myers Squibb, New York, NY, USA. [84]Genentech, San Francisco, CA, USA. [85]GlaxoSmithKline,

Collegeville, PA, USA. [86]GlaxoSmithKline, Espoo, Finland. [87]Merck, Kenilworth, NJ, USA. [88]Pfizer, New York, NY, USA. [89]Translational Sciences, Sanofi R&D, Framingham, MA, USA. [90]Maze Therapeutics, San Francisco, CA, USA. [91]Janssen Research & Development, LLC, Spring House, PA, USA. [92]Novartis Institutes for BioMedical Research, Cambridge, MA, USA. [93]HiLIFE, University of Helsinki, Helsinki, Finland. [94]Arctic biobank/University of Oulu, Oulu, Finland. [95]Auria Biobank/University of Turku/Hospital District of Southwest Finland, Turku, Finland. [96]THL Biobank/Finnish Institute for Health and Welfare (THL), Helsinki, Finland. [97]Finnish Red Cross Blood Service/Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland. [98]Helsinki Biobank/Helsinki University and Hospital District of Helsinki and Uusimaa, Helsinki, Finland. [99]Northern Finland Biobank Borealis/University of Oulu/Northern Ostrobothnia Hospital District, Oulu, Finland. [100]Finnish Clinical Biobank Tampere/University of Tampere/Pirkanmaa Hospital District, Tampere, Finland. [101]Pirkanmaa Hospital District, Tampere, Finland. [102]Biobank of Eastern Finland/University of Eastern Finland/Northern Savo Hospital District, Kuopio, Finland. [103]Central Finland Biobank/University of Jyväskylä/Central Finland Health Care District, Jyväskylä, Finland. [104]Central Finland Health Care District, Jyväskylä, Finland. [105]FINBB - Finnish biobank cooperative, Kuopio, Finland. [106]Business Finland, Helsinki, Finland. [107]GlaxoSmithKline, Stevenage, United Kingdom. [108]University of Gothenburg, Gothenburg, Sweden. [109]Seinäjoki Central Hospital, Seinäjoki, Finland. [110]Finnish Institute for Health and Welfare (THL), Helsinki, Finland. [111]Northern Ostrobothnia Hospital District, Oulu, Finland. [112]Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. [113]Department of Oncology, Helsinki University Hospital Comprehensive Cancer Center and University of Helsinki, Helsinki, Finland. [114]Research Unit of Oral Health Sciences Faculty of Medicine, University of Oulu, Oulu, Finland. [115]Hospital District of Southwest Finland, Turku, Finland. [116]Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Helsinki, Finland. [117]Hospital District of Helsinki and Uusimaa, Helsinki, Finland. [118]Department of Molecular Genetics, University of Lodz, Lodz, Poland. [119]Helsinki University Hospital and University of Helsinki, Helsinki, Finland. [120]Northern Savo Hospital District, Kuopio, Finland. [121]University of Eastern Finland, Kuopio, Finland. [122]Novartis, Boston, MA, USA. [123]GlaxoSmithKline, Brentford, United Kingdom. [124]Department of Medical Genetics, Helsinki University Central Hospital, Helsinki, Finland. [125]Tampere University, Tampere, Finland. [126]Janssen Research & Development, LLC, Titusville, NJ 08560, USA. [127]Broad Institute, Cambridge, MA, USA. [128]Janssen Research & Development, LLC, Boston, MA, USA. [129]Novartis, Basel, Switzerland. [130]Department of Breast Surgery, Helsinki University Hospital Comprehensive Cancer Center and University of Helsinki, Helsinki, Finland. [131]Janssen-Cilag Oy, Espoo, Finland. [132]Department of Otorhinolaryngology - Head and Neck Surgery, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. [133]Estonian biobank, Tartu, Estonia. [134]Eye Genetics Group, Folkhälsan Research Center, Helsinki, Finland. [135]Medical Research Center, Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland. [136]University of Eastern Finland and Kuopio University Hospital, Department of Otorhinolaryngology, Kuopio, Finland. [137]University of Helsinki, Helsinki, Finland. [138]University of Jyväskylä, Jyväskylä, Finland. [139]University of Tampere, Tampere, Finland. [140]University of Oulu, Oulu, Finland. [141]Department of Allergy, Helsinki University Hospital and University of Helsinki, Helsinki, Finland. [142]Transplantation and Liver Surgery Clinic, Helsinki University Hospital, Helsinki University, Helsinki, Finland. [143]University of Stanford, Stanford, CA, USA. [144]University of Helsinki and Hospital District of Helsinki and Uusimaa, Helsinki, Finland. [145]European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, UK.