1	Ger	nome-wide meta-analysis of iron status biomarkers and the	
2	effect of iron on all-cause mortality in HUNT		
3	Marta R Moksnes, ^{1,†} Ailin Falkmo Hansen, ¹ Sarah E Graham, ² Sarah A Gagliano Taliun, ^{3,4} Kuan		
4		15.16 - 17.18 - 17.18 - 19. 19 19 19 10	
5	Maso	n, ""," Francesco Cucca, ""," David Schlessinger, "Gonçalo R. Abecasis, "Stephen	
6	Burge	ss, ^{15,20} Bjørn Olav Åsvold, ^{1,21,22} Jonas B Nielsen, ^{1,2,23,24} Kristian Hveem, ^{1,22,*} Cristen J	
7	Willer	, ^{1,2,5,25,*} Ben M Brumpton. ^{1,22,26,†*}	
8			
0	4	K.C. Johann Contex for Constin Enidemiclogy, Department of Public Health and Nursing, Foculty of Medicine	
10	1	R.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Norsing, Faculty of Medicine	
11	2	Division of Cardiovaccular Medicine, Department of Internal Medicine, Liniversity of Michigan, App Arber, MI	
12	2		
13	3	Faculty of Medicine, University of Montréal, Montréal, OC H3T 1.14, Canada	
14	4	Montréal Heart Institute Montréal, QC H1T 1C8, Canada	
15	5	Department of Computational Medicine and Bioinformatics. University of Michigan, Ann Arbor, MI, USA	
16	6	Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston,	
17		MA, USA	
18	7	Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA	
19	8	Department of Clinical Chemistry, St. Olavs hospital Trondheim University Hospital, Trondheim, Norway	
20	9	Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI, USA	
21	10	Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI, USA	
22	11	Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK	
23	12	Clinical Pharmacology and Therapeutics Section, Institute for Infection and Immunity, St George's, University	
24		of London, London, UK	
25	13	Clinical Pharmacology Group, Pharmacy and Medicines Directorate, St George's University Hospitals NHS	
26		Foundation Trust, London, UK	
27	14	Novo Nordisk Research Centre Oxford, Old Road Campus, Oxford, UK	
28	15	Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK	
29	16	National Institute for Health Research Cambridge Biomedical Research Centre, University of Cambridge and	
30		Cambridge University Hospitals, Cambridge, UK	
31	17	Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche (CNR), Cagliari, Italy	
32	18	Dipartimento di Scienze Biomediche, Università degli Studi di Sassari, Sassari, Italy	
33	19	Laboratory of Genetics, National Institute on Aging, US National Institutes of Health, Baltimore, Maryland	
34	20	Medical Research Council Biostatistics Unit, University of Cambridge, Cambridge, UK	
35	21	Department of Endocrinology, Clinic of Medicine, St. Olavs hospital Trondheim University Hospital,	
30		Trondheim, Norway	
31 20	22	HUNT Research Centre, Department of Public Health and Nursing, NTNU – Norwegian University of Science	
20 20		and Lechnology, Levanger, Norway	
39 40	23	Department of Epidemiology Research, Statens Serum Institute, Copenhagen, Denmark	
40 41	24	Department of Luman Canadian University Hospital, Copenhagen, Denmark	
41 12	20	Clinic of Medicine St. Oleve beenitel Trendheim University Heenitel Trendheim Nerwey	
43	20 ∗ T⊦	ese authors contributed equally (co-last authors)	
44	11	יישט מעוויטיס טרווושעוכע פעעמויץ (טריומסו מעוויטיס)	
45	†Corres	ponding authors: marta.r.moksnes@ntnu.no, ben brumpton@ntnu.no	
	100100	And a second management of the second s	

It is made available under a CC-BY 4.0 International license .

46 Abstract

47 Iron is essential for many biological processes, but iron levels must be tightly

48 regulated to avoid harmful effects of both iron deficiency and overload. Here, we

- 49 perform genome-wide association studies on four iron related biomarkers (serum iron,
- 50 serum ferritin, transferrin saturation, total iron binding capacity) in the Trøndelag
- 51 Health Study (HUNT), the Michigan Genomics Initiative (MGI) and the SardiNIA
- 52 study, followed by their meta-analysis with publicly available summary statistics,
- analyzing up to 257 953 individuals. We identify 127 genetic loci associated with iron
- 54 traits. Among 19 novel protein-altering variants, we observe a rare missense variant
- 55 (rs367731784) in HUNT, which suggests a role for *DNAJC13* in transferrin recycling.
- 56 We further validate the latest genetic risk scores for each biomarker in HUNT (6%
- 57 variance in serum iron explained) and present linear and non-linear Mendelian
- 58 randomization analyses of the traits on all-cause mortality. We find evidence of a
- 59 harmful effect of increased serum iron and transferrin saturation in linear analyses that
- 60 estimate population-averaged effects. However, there was weak evidence of a
- 61 protective effect of increasing serum iron at the very low end of its distribution. Our
- 62 findings contribute to our understanding of the genes affecting iron status and its
- 63 consequences on human health.
- 64

It is made available under a CC-BY 4.0 International license .

65 Introduction

Iron is essential for a variety of physiological processes in the human body, but excess iron is
toxic. Iron overload is associated with a wide range of health problems, including liver
damage, type 2 diabetes, cardiovascular disease and neurodegenerative diseases such as
Alzheimer's disease¹⁻³, while long-term iron deficiency causes anemia, which can disrupt
cognitive function and the immune system⁴⁻⁶. Because of the damaging effects of both
deficiency and overload, iron metabolism is tightly regulated⁷.
Iron is bound, transported and delivered around the body by the transferrin glycoprotein⁸,

while the main intracellular iron storage, ferritin, provides a long-term reserve of iron for formation of hemoglobin and other heme proteins^{9–11}. Serum iron, serum ferritin, transferrin saturation percentage (TSP) and the total iron binding capacity (TIBC) of transferrin are biochemical measurements that are commonly used together to assess an individual's iron status¹².

79

Mutations in various iron metabolism genes can cause both iron deficiency and overload^{13–15}. 80 81 Genetic variants in the transferrin gene, TF, and in the homeostatic iron regulator gene, HFE, have been estimated to account for about 40% of genetic variation in transferrin levels¹⁶. A 82 recent genome-wide association study (GWAS) meta-analysis¹⁷ of serum iron, ferritin, TSP 83 84 and TIBC from Iceland, UK and Denmark reported 46 novel loci associated with at least one 85 of these biomarkers, implicating proteins involved in iron homeostasis. Identifying additional 86 genetic loci associated with iron status could further increase our understanding of 87 pathomechanisms underlying dysregulated iron levels. Furthermore, genetic variants from the most recent study¹⁷ could improve existing genetic risk scores (GRS) that have been widely 88 89 used to assess the causal associations of iron status on a range of outcomes using Mendelian 90 Randomization^{18–23} (MR). However, the new GRSs have not vet been validated in an 91 independent study. Further, despite the observed damaging effects of both very high and very 92 low iron stores, no previous MR studies have investigated the shape of the associations 93 between genetically proxied iron status biomarkers and mortality. By validating the most 94 recent genetic risk scores and using MR in an independent study (HUNT), we provide robust 95 and novel insights into the causal associations between iron status biomarkers and all-cause 96 mortality, particularly regarding non-linear relationships.

97

It is made available under a CC-BY 4.0 International license .

98 To discover novel genetic variants associated with iron status, we combine three 99 approaches: (i) genome-wide association studies of variants deeply imputed from the TOPMed reference panel²⁴ in the Trøndelag Health Study (HUNT)²⁵ and the Michigan 100 101 Genomics Initiative (MGI), as well as variants imputed from a cohort specific reference panel in SardiNIA²⁶ (ii) association tests with genotyped coding variants 102 103 selected from low-coverage (5x) whole-genome sequencing, (iii) genome-wide meta-104 analyses of HUNT, MGI, SardiNIA and summary statistics from deCODE, Interval and the Danish Blood Donor Study (DBDS)¹⁷. The analyses included up to 257 953 105 106 individuals (57% females, 43% males) with measured iron status biomarkers. We 107 evaluate the variance explained by previously published variants for serum iron, serum 108 ferritin, TSP and TIBC in HUNT. Furthermore, we use the genetic variants for the 109 iron status biomarkers to estimate the average causal effect of a population shift in the 110 biomarker distributions on all-cause mortality (the population-averaged effect), and 111 for the first time, investigate the shape of the causal relationships using non-linear 112 Mendelian Randomization. 113 **Results** 114 115 Discovery of genetic loci associated with iron status We identified 127 genetic loci (82 novel for at least one trait) associated (p-value $< 5 \times 10^{-8}$) 116 117 with the four iron traits, serum iron, serum ferritin, TIBC and TSP, (Supplemental Table 1, 118 Supplemental Figures 1-4) in genome-wide association meta-analyses of the iron status 119 biomarkers in 6 cohorts: HUNT, MGI, SardiNIA, deCODE, Interval and DBDS 120 (Supplemental Table 2). Among the 77 unique index variants (the variants with the lowest p-121 value) in novel loci that had been imputed in more than one study, 60 had consistent 122 directions of effects across all the analyzed studies. We also identified three novel missense 123 variants associated with at least one iron trait among the variants selected for targeted 124 genotyping in HUNT (Supplemental Table 3). 125 126 Genes in several associated loci coded for proteins with established functions in iron 127 homeostasis (TF [transferrin]) SLC25A37 [mitoferrin-1], SLC25A28 [mitoferrin-2], SLC11A2 128 [divalent metal-transporter 1] and SLC40A1 [ferroportin-1], HFE [homeostatic iron

- regulator], *TFRC* [transferrin receptor], *TFR2* [transferrin receptor 2], *HAMP* [hepcidin],
- 130 ERFE [erythroferrone], HMOX1 [heme oxygenase], IREB2 [iron responsive element binding

It is made available under a CC-BY 4.0 International license .

- 131 protein 2], *EPAS1* [endothelial PAS Domain Protein 1] and *TMPRSS6* [transmembrane serine
- 132 protease 6])^{7,27–29} Four of these loci (*SLC25A28*, *HMOX1*, *IREB2*, *EPAS1*) had not been
- 133 reported for iron status biomarkers in GWAS studies before, providing additional confidence
- 134 in the GWAS we report. With two exceptions (HAMP, TFR2), these genes were the nearest
- 135 gene to the index variant in the locus.
- 136

137 Protein-altering variants in meta-analysis loci

- 138
- 139 We identified 32 protein-altering single nucleotide polymorphisms (SNPs) in the meta-
- 140 analysis, which were either index variants or variants in strong linkage disequilibrium (LD)
- 141 (R^2 >0.8 or D'=1.0) with an index variant (Supplemental Table 1). In addition to SNPs known
- 142 to be related to diseases such as hemochromatosis, atransferrinemia and iron deficiency
- 143 anemia^{15,30–34}, and variants that had previously been reported for at least one of the iron
- 144 traits^{17,35}, we identified 11 protein-altering variants in novel iron status loci (Supplemental
- 145 Table 4): rs9427398 (FCGR2A), rs2437150 (SPRTN), rs1047891 (CPS1), rs41274050
- 146 (A1CF), rs1935 (JMJD1C), rs3742049 (COQ5), rs4149056 (SLCO1B1), rs2070863
- 147 (SERPINF2), rs883541 (WIPI1), rs1800961 (HNF4A) and rs738409 (PNPLA3).). In known
- 148 iron status loci, we further identified eight protein-altering variants not previously reported
- 149 for any iron traits: rs367731784 (DNAJC13), rs3812594 (SEC16A), rs34376913 (C9orf163),
- 150 rs445520 (SLC11A2), rs28929474 (SERPINA1), rs737700 (C16orf71), rs77542162 (ABCA6)
- 151 and rs34654230 (*RCN3*).
- 152

153 Custom genotyped variants in HUNT: Protein-altering variants in iron status

- 154 **loci**
- 155 Among the targeted candidate variants in HUNT identified by sequencing and clinical
- 156 studies, we identified three additional, novel protein-altering variants (Supplemental Table 3)
- 157 that were not included in the meta-analyses, and which were associated with iron status
- 158 biomarkers. These were located in NRM (rs374815811), HLA-DRB5 (rs701884) and TFR2
- 159 (chr7:100629337:A:T, GRCh38).
- 160
- 161 Heritability and genetic correlation of iron status markers
- 162 We estimated the respective narrow-sense SNP heritability (variance explained, $V_g/V_p \pm$
- 163 1SE) of serum iron (0.15 \pm 0.01), TIBC (0.43 \pm 0.01) and TSP (0.21 \pm 0.01) in HUNT using
- 164 genome-wide complex trait analysis (GCTA)³⁶. We found pair-wise genetic correlations

It is made available under a CC-BY 4.0 International license .

between 11% and 75% (Supplemental Table 5) for the four iron status biomarkers using LD

166 Score Regression (LDSC)³⁷ with the meta-analysis summary statistics. The TSP phenotype

167 was derived from the serum iron and TIBC measurements, giving rise to the two strongest

168 genetic correlations. The weakest correlation (iron vs TIBC) did not reach nominal

- 169 significance (p-value=0.35).
- 170

171 Functional mapping

172 We used Bayesian colocalization analysis to identify 94 unique pairs of GWAS loci and cis-

173 expression quantitative trait locus (cis-eQTL) signals that showed sufficient overlap in at

174 least one tissue to be consistent with a shared causal variant for the gene expression and the

- 175 iron status biomarker (Supplemental Table 6). We found associations in a range of tissues
- 176 which highlighted genes with established roles in iron metabolism (TF [posterior probability
- 177 of a common causal variant, PP4=0.96], TMPRSS6 [PP4=0.82], ERFE [PP4=0.97-0.98],
- 178 *IREB2* [PP4=0.80], *SLC40A1* [PP4=0.79-0.96])^{13,27}. Additionally, our results confirmed
- 179 previously reported genes (*DUOX2* [PP4=0.76], *HBS1L* [PP4=0.98], *IL6R* [PP4=0.81-0.82],
- 180 *SLC25A37* [PP4=0.85], *ABO* [PP4=0.97], *RNF43* [PP4=0.99])^{17,35}, and identified novel genes
- 181 interacting with previously reported genes, for example $DUOXA2^{38}$. Several iron status loci
- 182 were also colocalized with cis-eQTL signals for genes in the major histocompatibility
- 183 complex (MHC) other than HFE^{39} , as well as with transcription regulators^{40–42}, additional
- 184 transporter proteins 43,44 and transferases 45,46 .
- 185

186 Using Data-driven Expression Prioritized Integration for Complex Traits (DEPICT)⁴⁷ we

187 found an enriched (false discovery rate [FDR] <0.05) expression of ferritin associated genes

- 188 in the urogenital system, digestive system, and the hemic and immune system (Supplemental
- 189 Table 7). Serum iron, TSP and TIBC associated genes were not enriched in any tissue types
- 190 at FDR<0.05, however the strongest enrichment for genes in all three traits were found in
- 191 liver tissue, and particularly in hepatocytes (TSP, TIBC). The top ten genes per trait when
- 192 prioritized based on similarity between the associated (p-value $< 5 \times 10^{-8}$) loci, included
- 193 known iron regulatory genes (TFR2, HAMP, TFRC, SLC40A1), genes in which we had
- 194 identified protein-altering variants (IL6R, F5, GCKR, DUOX2, SERPINA1, ABCA6,
- 195 SLCO1B1), genes we found in the colocalization analysis (DUOXA2, IL1RN, SLC25A37), as
- 196 well as genes predicted to have iron ion binding and heme binding properties in gene
- 197 ontology analyses (*CYP3A43*, *CYP3A5*)⁴⁸ (Supplemental Table 8). Finally, we used DEPICT

It is made available under a CC-BY 4.0 International license .

198 and found gene sets enriched with iron status associated genes (Supplemental Table 9). Most 199 of the top ten gene sets at FDR<0.05 were from iron and TSP associated loci: Several were 200 related to the liver (including abnormal liver physiology and gene sets related to metabolic 201 processes), but also to inflammation (acute-phase response, decreased leukocyte cell 202 number), coagulation (coagulation factor protein-protein interaction networks) and 203 neurodevelopment (abnormal myelination). Most of the top ten gene sets enriched with 204 ferritin and TIBC associated loci did not reach FDR<0.05, but included decreased circulating 205 iron levels, decreased spleen iron level, gene sets related to red blood cells (decreased 206 hemoglobin, decreased hematocrit, erythrocyte homeostasis and differentiation), as well as 207 liver fibrosis and liver inflammation.

208

209 We identified the 1% top ranked genes per trait based on both physical distance to the

210 associated genetic variants and functional similarity to other associated genes (Supplemental

211 Tables 10-13) using Polygenic Priority Scores (PoPS)⁴⁹. The prioritized genes included the

212 main known iron regulatory genes, several genes that were nearest to the meta-analysis index

213 variants and novel genes in which we identified protein-altering variants (WIP11, SERPINF2,

- 214 *HNF4A*), further supporting a role for these genes in iron biology.
- 215

216 Phenome-wide association study (PheWAS) of biomarker loci

217 In total, 69 of the meta-analysis index variants were significantly associated (p-value <

 2.4×10^{-7}) with at least one additional phenotype ('phecode'⁵⁰), blood biomarker or continuous

trait in UK Biobank, and 97 phenotypes were significantly associated with at least one variant

220 (Supplemental Table 14). The associations spanned numerous biological domains, but most

221 associations were within the endocrine/metabolic (288 variant-trait associations), digestive

222 (139 variant-trait associations) and genitourinary (46 variant-trait associations) domains. The

strongest associations (p-values $< 1 \times 10^{-300}$) were within the hematopoietic, digestive and

endocrine/metabolic domains: The *HFE* variants rs1800562, rs144861591 and rs79220007

225 were associated with disorders of mineral metabolism, and in particular with disorders of iron

- 226 metabolism, the SLCO1B1 variant rs2900478 was associated with bilirubin, the
- 227 ASGR2;ASGR1 variant rs186021206 was associated with alkaline phosphatase, and the
- 228 GCKR variants rs1260326 and rs11336847 were associated with triglycerides. Overall, all the
- 229 GRSs for the four iron status biomarkers were associated with disorders of mineral
- 230 metabolism, in particular iron metabolism (Figure 1, Supplemental Table 15). Several GRSs

It is made available under a CC-BY 4.0 International license .

- 231 were associated with anemias (iron, ferritin, TSP) and with coagulation defects (ferritin,
- 232 TSP). Finally, the GRS for TIBC was also associated with liver cirrhosis without mention of
- alcohol, and the GRS for ferritin was associated with chronic non-alcoholic liver disease, and
- with phlebitis and thrombophlebitis.
- 235



236

237Figure 1: GRS-PheWAS: Phenome-wide associations between the GRS for each biomarker (serum iron [A], serum ferritin238[B], total iron binding capacity [C] and transferrin saturation percentage [D]) and 1 688 phecodes, blood biomarkers and239continuous traits in the UK Biobank. Triangles pointing upwards indicate a positive association between the phenotype and240the GRS (where a higher GRS score represents higher level of the iron status biomarker) and vice versa. Associations with241p-values < 10⁻³²² are plotted at 10⁻³²². The Bonferroni corrected p-value cut-off (2.4×10^7) is given as a red dotted line.

242

243 Linear Mendelian Randomization

244 The linear MR (ratio of coefficients method) indicated an increased mortality risk with

increased serum iron and TSP, with the point estimates suggesting that an increase of 1

standard deviation (SD) in both serum iron (1 SD = 6.3μ mol/L) and TSP (1 SD = 11.3

- 247 percentage points) would lead to an increased risk of mortality of 7% (Table 1). The
- 248 estimates for ferritin and TIBC were not statistically significant, however the point estimates
- for a 1 standard deviation increase in serum ferritin (1 SD = $46\mu g/L$) and TIBC (1 SD = 9.2

It is made available under a CC-BY 4.0 International license .

- 250 μmol/L) were a 4% increase and 3% decrease in mortality, respectively (Table 1). The
- estimate for ferritin was also very imprecise.
- 252
- .52
- Table 1: Linear Mendelian randomization ratio of coefficient estimates: Hazard ratios with 95% confidence intervals for all cause mortality are given per 1 standard deviation increase in the biomarker. The sample size is given as N.

Biomarker	Ν	Hazard ratio (95% CI)	P-value
Serum iron	56 654	1.07 (1.01 to 1.14)	0.03
Serum ferritin*	2 335	1.04 (0.26 to 4.22)	0.95
TSP	56 651	1.07 (1.02 to 1.12)	0.01
TIBC	56 654	0.97 (0.93 to 1.01)	0.13

- 255 Confidence interval (CI), Transferrin saturation percentage (TSP), Total iron binding capacity (TIBC)
- 256 *Non-pregnant women, 20-55 years old
- 257

258 Non-linear Mendelian Randomization

259 To investigate a potential non-linear causal association between iron status and all-cause

260 mortality, we used the GRSs as instruments for serum iron (F-statistic=3618, R^2 =0.06), TIBC

261 (F-statistic=8373, R^2 =0.129), TSP (F-statistic=6811, R^2 =0.107) and ferritin (F-

statistic=37.81, R²=0.015) in a non-linear Mendelian randomization analysis and estimated

the shape of the associations between the genetically predicted traits and all-cause mortality

- 264 (Figure 2). The median follow-up time was 23.6 years. After performing a statistical test for
- whether the best-fitting non-linear model of degree 1 fitted the data better than a linear model
- 266 (p-values: 0.60, 0.05, 0.23 and 1 for iron, ferritin, TIBC and TSP respectively), generally, we
- 267 did not find strong statistical evidence supporting a non-linear relationship over a linear one
- 268 for the associations between any of the genetically proxied iron traits and all-cause mortality.
- 269 However, the point estimates for serum iron do follow a J-shape, with a negative slope at
- 270 very low levels of serum iron and a constant positive slope above 10 µmol/L. The point
- 271 estimates were however imprecise at the tails of the distribution. The other analyses indicate
- a lower risk at higher TIBC and lower TSP and ferritin levels, with a weak indication (p-
- value=0.05) of a non-linear effect for ferritin. Post-hoc sensitivity analyses using genetic
- instruments that were consistent with systemic iron status (increased iron, ferritin, and TSP,
- and decreased TIBC) rather than just representing a single biomarker, gave similar results
- 276 (Supplemental Figure 5).
- 277

It is made available under a CC-BY 4.0 International license .



Figure 2: Non- linear Mendelian randomization: Dose-response curves (black) between iron traits and all-cause mortality in HUNT (gray lines give 95% confidence interval). The x-axis gives A: serum iron levels (μ mol/L), B: serum ferritin (μ g/L), C: transferrin saturation (%) and D: total iron binding capacity (TIBC) (μ mol/L). The y-axis gives the hazard ratios for allcause mortality with respect to the reference values (red dot), which represent the established target values (iron, TIBC, TSP)⁵¹ or median value (ferritin) for the traits. The curve gradients represent the localized average causal effect at each point.

278

It is made available under a CC-BY 4.0 International license .

285 Discussion

286 We performed the largest GWAS meta-analysis to-date of iron status biomarkers and 287 identified 127 genetic loci (82 novel) associated with iron status, including 19 novel protein-288 altering variants. Although 82 loci were classified as novel for at least one of the tested iron 289 traits, some had known associations with other tested traits or with related traits such as hemoglobin levels and red blood cell count and volume⁵². Because iron traits are biologically 290 291 linked, we might expect to find the same loci associated with several of the tested traits, as 292 seen for loci with established roles in iron metabolism (e.g. HFE, TF and TMPRSS6) and 293 others. We confirmed the genetic similarity between iron status biomarkers by observing a 294 genetic correlation of 75% between serum iron and TSP. However, most of the loci were 295 specifically associated with a single trait, which was in line with the low genetic correlation 296 between other pairs of biomarkers.

297

298 Overall, our findings are consistent with established knowledge about iron homeostasis and 299 the role of iron in various biological processes: The PheWAS analyses linked the meta-300 analysis loci to many different traits and phenotypes, particularly within the hematopoietic, 301 digestive, endocrine/metabolic and genitourinary domains. Further, in four novel loci, the 302 genes nearest to the index variant encoded proteins with established roles in iron regulation: 303 1) the mitochondrial iron transporter mitoferrin-2 (SLC25A28), 2) heme oxygenase 1 304 (*HMOX1*), which catalyzes heme degradation, 3) the iron-responsive element binding 2 305 (*IREB2*), and 4) *PAS1*, which regulates erythropoiesis according to cellular iron availability²⁸. 306 Further, the novel protein-altering variants were found in genes associated with various 307 biological traits and functions, which potentially highlight the many biological processes both 308 involved in and dependent on iron and iron regulation. These included, but were not limited 309 to, genes involved in or associated with: 1) Iron gut absorption, regulation and transport (*TFR2*, *SLC11A2*, *DNAJC13*)^{8,54,53}, where we found a rare (minor allele frequency) 310 (MAF)=0.0009) protein-altering variant with moderate effect size in DNAJC13, a gene 311 suggested to be involved in transferrin recycling⁵³. This variant was only imputed in HUNT, 312 where it was more than 100 times more common than in other non-Finnish Europeans 313 (https://gnomad.broadinstitute.org/variant/rs367731784)⁵⁵; 2) Concentrations of hemoglobin 314 (SPRTN, FCGR2A, CPS1, PNPLA3, ABCA6)^{49,56–58}, which holds more than two thirds of the 315 body's iron¹, and bilirubin $(SLCO1BI)^{59}$, which together with iron are products of heme 316 degradation; 3) Liver-related traits (ABCA6, HNF4A, PNPLA3)^{60,61}; 4) Iron-dependent 317

It is made available under a CC-BY 4.0 International license .

(putative) tumor suppression $(JMJD1C)^{62}$; 5) Iron accumulation in the brain (WIP1,

associated via its homolog $WIPI4^{63}$) and 6) Coagulation and immune response (SERPINA1,

320 *SERPINF2*)⁶⁴. A previous study also identified a different protein-altering SNP in the serine

321 protease inhibitor *SERPINA1*¹⁷. The transmembrane serine protease 6 (encoded by

322 *TMPRSS6*) is a negative regulator of hepcidin⁶⁵, a key hormone regulator of iron

323 homeostasis¹³. Given the role of the transmembrane serine protease 6 in iron regulation, both

324 the serine protease inhibitors could potentially affect iron regulation via this gene.

325

326 Using DEPICT, we detected an enriched expression of genes in iron status loci in the liver.

327 This is in line with the important role of the liver in iron metabolism and storage 66 , including

328 hepcidin production. Consistent with previous studies, our analysis prioritized genes

329 encoding known hepcidin regulators such as HFE, TMPRSS6, TF, TFRC, TFR2, ERFE and

330 $IL6R^{67-70}$. Mutations in several of these genes have been demonstrated to cause diseases of

iron deficiency or overload^{1,15,33,71}. The associations with other genes related to inflammation,

both in the DEPICT and colocalization analyses, could possibly be related to the hepcidin

333 response to inflammation. The genes and gene sets prioritized by DEPICT pointed to several

different biological processes, which might reflect the numerous roles of iron in the body. A

335 limitation with using a similarity-based method for gene set and gene prioritization for iron

traits was that the software excluded the MHC region from the analysis, thereby also

excluding one of the most central genes in iron homeostasis, the *HFE* gene.

338

339 Colocalization analysis further linked the GWAS loci both to the liver and to iron overload:

340 iron status loci overlapped with cis-eQTLs for several of the hepcidin regulators, and genes

involved in other liver functions such as lipid and fatty acid metabolism (*ORMDL1*,

FADSI^{72,73}. The latter was also found in previous studies³⁵ and is in line with the results

343 from the PheWAS analysis. The colocalization of iron status loci and cis-eQTLs were

344 however found in several tissues, and not primarily in liver. A limitation to the analysis was

however that there were different sample sizes for each tissue, where liver had a relatively

346 small sample size and subsequently lower power than other tissues.

347

348 Because iron plays an essential role in so many biological processes, several Mendelian

randomization studies have explored the causal effect of iron status on a range of diseases $^{18-}$

²³. Despite the known harmful effects of both iron deficiency and overload, no previous MR

351 studies have investigated the shape of the exposure-outcome relationship. We therefore

It is made available under a CC-BY 4.0 International license .

352 assessed the causal effect of iron status biomarkers on all-cause mortality and investigated the 353 shape of these associations. We demonstrated that the GRSs based on the previous study 354 were good instruments for iron, TSP and TIBC in the independent HUNT study (variance 355 explained 6% (iron), 11% (TSP), 13% (TIBC), 1.6% (ferritin)). Using these, we found 356 evidence of a harmful effect of increased serum iron and TSP (derived from serum iron and 357 TIBC) in linear analyses that estimated population-averaged effects. The point estimates of 358 TIBC and ferritin were also suggestive of a harmful effect of increased iron status, although 359 the estimates were not statistically significant, and the ferritin estimate was very imprecise 360 due to the small sample size. In non-linear models, we did not find strong statistical evidence 361 supporting non-linear relationships over linear ones. However, there was weak evidence of a 362 protective effect of increasing serum iron at the very low end of its distribution, at serum iron levels below the normal range of 10-34 μ mol/L⁵¹. The results were supported by post-hoc 363 364 sensitivity analyses using only genetic variants consistent with systemic iron status. 365 366 The study had several clear limitations. First, in our GWAS analyses we did not adjust for 367 additional factors that could affect the biomarker concentrations, such as iron 368 supplementation, inflammatory status, alcohol consumption and menopausal status (except 369 for ferritin, where the full sample was pre-menopausal). These factors could therefore have 370 influenced the effect estimates, particularly for rare variants. Second, we would need a larger 371 sample size to confirm a non-linear shape of the exposure-outcome relationship at the 372 extremes of the biomarker distributions. The analysis of ferritin was particularly limited by 373 the small sample (N=2334) consisting of only relatively young, non-pregnant females, giving 374 a low number of strata and few deaths. Third, the association of the GRSs with all-cause 375 mortality could be attenuated because HUNT participants with suspected iron deficiency 376 anemia or phenotypic hemochromatosis were later contacted by the primary health care 377 services and offered treatment, and they could therefore have obtained a healthier iron status 378 than they would otherwise have had, causing the analysis to be less precise. Finally, although 379 the four biomarkers are commonly used to assess people's iron status, neither of them is a 380 very good individual predictor of iron stores, and the findings should therefore be interpreted 381 with caution. 382

383 In summary, we have increased the number of iron status associated loci through a

384 large GWAS meta-analysis and validated the latest genetic risk scores for four iron

385 status biomarkers. We find evidence of a harmful population-averaged effect of

It is made available under a CC-BY 4.0 International license .

- 386 genetically proxied serum iron and TSP, and weak evidence of a protective effect of
- 387 increasing serum iron in individuals at the very low end of its distribution. Our
- 388 findings contribute to our understanding of the genes affecting iron status and its
- 389 consequences on human health.
- 390

391 Methods

- 392 Cohort descriptions
- 393 Distributions of the age and sex of the HUNT, MGI and SardiNIA participants included in394 the current study are reported in Supplemental Table 2.
- 395

396 HUNT

- 397 The HUNT Study is a longitudinal population-based health study conducted in the county of
- Trøndelag, Norway since 1984^{25} . About 123 000 individuals (aged ≥ 20 years) have
- 399 participated in at least one of four surveys, and more than 70 000 of these participants have
- 400 been genotyped using one of three Illumina HumanCoreExome arrays: 12 v.1.0, 12 v.1.1 and
- 401 24 with custom content (UM HUNT Biobank v1.0). Sample and variant quality control (QC)
- 402 was performed using standard practices and has been reported in detail elsewhere⁷⁴. All
- 403 variants were imputed from the TOPMed reference panel (freeze 5)²⁴ using Minimac4 v1.0
- 404 (https://genome.sph.umich.edu/wiki/Minimac4). The reference panel is composed of 53 831
- 405 multi-ethnic samples and 410 323 831 SNP and indel variants at high depth (mean read depth
- 406 38.2X). Variants with a minor allele count (MAC) > 10 or imputation $R^2 \ge 0.3$ were included
- 407 in analysis. A subset of individuals was genotyped with additional custom content variants.
- 408

409 MGI

- 410 The Michigan Genomics Initiative (MGI) is a repository of genetic data and electronic
- 411 medical records from Michigan Medicine. Approximately 80 000 participants (aged
- ≥ 18 years) have predominantly been enrolled prior to surgical procedures with over 59 000
- 413 individuals genotyped using Illumina Infinium CoreExome-24. Following genotyping,
- 414 sample-level QC was performed to remove sex-mismatches, duplicates, samples with call
- 415 rate < 99%, or with estimated contamination > 2.5%. Variants with GenTrain score < 0.15,
- 416 Cluster Separation scores < 0.3, Hardy-Weinberg Equilibrium p-value among unrelated
- 417 European individuals $< 1 \times 10^{-4}$, or with evidence of batch effects (p-value $< 1 \times 10^{-3}$, Fisher's
- 418 exact test) were excluded. Imputation was performed using the TOPMed Imputation Server

It is made available under a CC-BY 4.0 International license .

419 (v1.2.7). Variants with MAF > 0.05% and imputation quality $R^2 \ge 0.3$ were included in

420 analysis.

421

422 SardiNIA

- 423 The SardiNIA study is a longitudinal population-based health study including 6 602
- 424 individuals from the Lanusei valley on Sardinia. The participants have been genotyped on
- 425 four different Illumina Infinium arrays, OmniExpress, Cardio-Metabochip⁷⁵, Immunochip⁷⁶
- 426 and Exome Chip), and then imputed from a SardiNIA specific sequencing panel (~4x
- 427 coverage) of 3 839 individuals, using Minimac3⁷⁷. Markers with imputation quality $R^2 > 0.3$
- 428 (or > 0.6 in variants with MAF < 1%) were retained, resulting in a total of ~ 19 million
- 429 genetic variants. Samples, genotyping, sequencing and variant calling have previously been

430 described elsewhere⁷⁸.

431

432 Iron Status Biomarkers

433 Distributions of the biomarker levels in the HUNT, MGI and SardiNIA participants included434 in the current study are reported in Supplemental Table 2.

435

436 HUNT

437 Non-fasting serum samples were drawn in 1995-1997 (HUNT2). Serum iron concentration

438 (µmol/L) was determined using a FerroZine method using a Hitachi 911 Autoanalyzer

- 439 (reagents from Boehringer, Germany). The serum transferrin concentration (µmol/L) was
- 440 analyzed by an immunoturbidimetric method using the Hitachi 911 Autoanalyzer (reagents
- 441 from DAKO A/S, Denmark), and calculated for a molecular weight of 79 570 Da. TIBC was
- 442 calculated as 2 x the serum transferrin concentration. The TSP was calculated as 100 x [serum
- 443 iron concentration/TIBC]. Serum ferritin was measured from serum samples using an Abbott
- 444 AxSYM analyzer (reagents from Abbott Laboratories, USA). In total, 56 667 HUNT
- 445 participants had measurements of serum iron and TIBC, 56 664 had measurements of TSP,
- 446 while ferritin was only measured in 2 334 women (fertile, non-pregnant, aged 20-55 years).

447

448 SardiNIA

- 449 Serum iron (µmol/L) and serum transferrin concentrations (µmol/L) were measured in fasting
- 450 blood-samples from individuals with genotype and imputation data from the SardiNIA
- 451 cohort. TIBC was calculated as 2 x the serum transferrin concentration. In total, 5 930 and

It is made available under a CC-BY 4.0 International license .

452 5 926 genotyped SardiNIA participants had measurements on serum iron and TIBC

453 respectively.

454

455 MGI

456 Serum iron concentration was measured using the Ferrozine Colorimetric assay, and serum

- 457 ferritin was measured using a Chemiluminescent Immunoassay. Serum transferrin
- 458 concentrations were measured using an Immunoturbidimetric assay, and TIBC was calculated
- 459 as 2 x the serum transferrin concentration. The TSP was calculated as 100 x [serum iron
- 460 concentration/TIBC]. For individuals with multiple measurements, the initial measurement
- 461 was used in the analyses. In total, 10 403, 9 480, 10 399 and 10 381 participants from MGI
- 462 had measurements of serum iron, serum ferritin, TIBC and TSP respectively.
- 463

464 Association analyses

465 Association analyses of all iron traits (iron, ferritin, TIBC, TSP) in HUNT were performed

466 using a linear mixed model regression under an additive genetic model for each variant as

- 467 implemented in BOLT-LMM v2.3.4⁷⁹, which also controls for relatedness between the
- 468 samples. Association analyses of all iron traits in MGI were performed using a linear
- 469 regression model in unrelated individuals using rvtests⁸⁰. In both HUNT and MGI, we
- 470 applied rank-based inverse normal transformation on the iron variables after adjusting for age
- 471 and sex using linear regression, and included age, sex, genotyping batch and the first 10
- 472 principal components (PCs) of ancestry as covariates. Association analyses of serum iron and
- 473 TIBC in SardiNIA was performed using age, age² and sex-adjusted inverse-normalized
- 474 residuals of TIBC or iron as input to the Efficient Mixed Model Association eXpedited
- 475 (EMMAX)⁸¹ single variant test as implemented in EPACTS
- 476 (<u>https://github.com/statgen/EPACTS</u>).
- 477
- 478 Additionally, we performed association analyses of serum iron, TIBC and TSP in HUNT with
- 479 361 965 additional custom content variants genotyped in 44 248 (serum iron, TIBC) or 44
- 480 246 individuals (TSP) using BOLT-LMM v2.3.4⁷⁹, including the same covariates and rank-
- 481 based normal transformation of the variables as was done in the main analyses.
- 482

It is made available under a CC-BY 4.0 International license .

483 Meta-analyses

- 484 We performed fixed-effect inverse-variance weighted meta-analysis of summary statistics for 485 iron (sample size N=236 612), ferritin (N=257 953), TIBC (N=208 422) and TSP (N=198 516) using METAL⁸². Serum iron and TIBC were meta-analyzed in all studies (HUNT, MGI, 486 487 SardiNIA and summary statistics from deCODE and Interval). SardiNIA did not have data on 488 serum ferritin and TSP and was therefore excluded from these meta-analyses, while the 489 available summary statistics for ferritin also included the DBDS study. To harmonize genomic positions from each study, we used LiftOver from UCSC⁸³ to map the data from 490 SardiNIA from Human Genome Build GRCh37 to GRCh38. Because standard errors were 491 492 not given in the available summary statistics from deCODE, Interval and DBDS, we 493 calculated them as the absolute value of the (effect size/qnorm(p-value/2)), where qnorm 494 represents the inverse standard normal distribution. In HUNT, MGI and SardiNIA we 495 performed genomic control correction of any analyses with an inflation factor $\lambda > 1$ prior to meta-analysis. We considered genetic loci reaching a p-value $< 5 \times 10^{-8}$ for follow-up analyses. 496
- 497

498 Definition of independent loci and locus novelty

- 499 Genetic loci were defined around variants with a genome-wide significant association with a
- 500 trait (p-value $< 5 \times 10^{-8}$). The locus borders were set 500 kb to each side of the highest and
- 501 lowest genetic positions reaching genome-wide significance in each region. Overlapping
- 502 genetic loci were merged if the index (lowest p-value) variants were in LD ($R^2 \ge 0.2$ and/or
- 503 D' \geq 0.8), or if one of the index variants was too rare to calculate LD with the other from our
- 504 reference panel of 5000 unrelated individuals in HUNT. A locus was classified as novel for a
- 505 given trait if it had not been reported previously for the trait. Previously published variants
- 506 were identified through a literature search and a look-up in the GWAS catalog⁸⁴.
- 507

508 Annotation of genetic variants

- 509 We used plink $v1.9^{85}$ with a reference panel of 5000 unrelated individuals in HUNT to
- 510 identify genetic variants in strong LD ($\mathbb{R}^2 \ge 0.8$) with the index variants, and annotated the
- 511 functional consequences and rsIDs of the index variants and LD proxies using ANNOVAR
- 512 $(v.2019Oct24)^{86}$ and the UCSC human genome browser⁸³.
- 513

It is made available under a CC-BY 4.0 International license .

514 Functional mapping of genetic variants

515 We used three different bioinformatic approaches to perform functional mapping and gene

516 prioritization of the meta-analysis summary level data: Bayesian colocalization analysis^{87,88},

- 517 DEPICT⁴⁷ and Polygenic Priority Scores⁸⁹.
- 518

519 To assess if any iron status loci were overlapping with significant cis-eQTL signals

520 and consistent with shared causal variants for iron status markers and gene expression

521 levels in specific tissue types, we used Bayesian colocalization analysis ('coloc') as

522 implemented in the R package coloc. We used cis-eQTL data from 27 general tissue

523 types (49 subtypes) in the individuals of European ancestry from the Genotype-Tissue

524 Expression (GTEx) portal, data set v8 (<u>https://www.gtexportal.org</u>) and the GWAS

525 meta-analysis results for each iron trait as input. For each tissue type, we analyzed all

526 genes whose expression were associated (p-value $< 5 \times 10^{-8}$) with at least one iron

527 status associated variant (p-value $< 5 \times 10^{-8}$), using effect sizes and standard errors for

528 each variant-trait association as input. The coloc software estimated the variance in

529 each trait (iron trait or gene expression level) from the sample sizes and minor allele

530 frequencies. We set the prior probability of a genetic variant being associated with

only iron traits, only gene expression or both traits to be 10^{-4} , 10^{-4} and 10^{-6}

respectively. We considered posterior probabilities (PP4) above 75% to give support

for a common causal variant for the iron trait and expression of the gene in the giventissue.

535

We performed gene set enrichment, gene prioritization and tissue/cell type enrichment tests on the iron trait loci (p-value $< 5 \times 10^{-8}$) using Data-driven Expression Prioritized Integration for Complex Traits (DEPICT) (v1.1, release 194)⁴⁷. Prior to the analysis we used LiftOver from the UCSC⁸³ to convert the genomic positions of the genetic variants from GRCh38 to GRCh37. Enrichment results with an FDR < 5% were considered significant.

541

542 Finally, we prioritized genes by computing Polygenic Priority Scores⁸⁹ from summary-level

543 data from each iron status biomarker. The method uses Multi-marker Analysis of GenoMic

544 Annotation (MAGMA)⁹⁰ to compute gene-level associations and gene-gene correlations from

545 the meta-analysis p-values and sample sizes and LD information from individuals of

546 European ancestry from the 1000 Genomes reference panel⁹¹. MAGMA is applied a second

It is made available under a CC-BY 4.0 International license .

547 time to perform enrichment analysis for genetic features. Genes are finally prioritized based

548 on a combination of physical distance to associated genetic variants and functional similarity

549 with other associated genes. We considered the 1% top ranked genes per biomarker to be

- 550 prioritized genes for the respective traits.
- 551

552 Heritability estimation

553 We estimated the narrow-sense additive SNP heritability of serum iron, TIBC and TSP in

554 HUNT using GCTA³⁶. Ferritin heritability was not estimated because of the low sample size

555 in HUNT. We created genetic relationship matrices (GRMs) based on 358 956 genotyped

autosomal variants in 56 667 individuals with serum iron and TIBC data, and 56 664

557 individuals with TSP data. We used the respective GRMs with GCTA-GREML (genomic-

558 relatedness-based restricted maximum-likelihood) to estimate the variance in each variable

that was explained by the genetic variants. We used age, sex, and genotyping batch as

560 covariates in the analyses, and we transformed the iron trait variables to normality with rank

561 based inverse normal transformation after regression on age and sex prior to the analyses.

562

563 Genetic correlation between iron traits

We used the LDSC software³⁷ with the iron trait meta-analysis summary statistics and precomputed LD Scores for Europeans from the 1000 Genomes reference panel⁹¹, and estimated the pair-wise genetic correlations of the four iron traits. Prior to the analysis, we changed all p-values $< 1 \times 10^{-300}$ to the exact value 1×10^{-300} to make sure the software was able to read the smallest values and did not discard these SNPs. To ensure that only well imputed SNPs were included in the analysis and thereby avoid bias due to variable imputation quality, we filtered the input files to the HapMap3 reference panel prior to the

- 571 analysis, as recommended by the software developers (<u>https://github.com/bulik/ldsc/</u>).
- 572

573 Phenome-wide association tests (PheWAS)

We constructed GRSs for the iron status biomarkers by summing the product of the effect

575 size and the estimated allele count (dosage) for the index variants in genome-wide significant

576 loci (p-value $< \times 10^{-8}$). We used TOPMed imputed estimated allele counts and effect sizes

577 from the meta-analysis and calculated the GRS for participants of white British ancestry in

- 578 the UK Biobank. We tested the association of the GRSs (GRS-PheWAS) and the individual
- 579 index variants (single variant PheWAS) with 1 688 phecodes, continuous traits and blood

It is made available under a CC-BY 4.0 International license .

- 580 biomarkers. We used a logistic or linear regression model respectively to assess the 581 association of the single variant estimated allele counts ('dosage') or inverse normalized GRS 582 and each phecode or continuous trait/biomarker. For the GRS-PheWAS we included as 583 covariates sex and birth year for binary traits, and sex and age at measurement for continuous 584 traits. For the single variant PheWAS we used GWAS summary statistics generated with SAIGE v.29.4.2⁹², with sex and the first four PCs as covariates in addition to age at initial 585 assessment for quantitative traits and birth year for binary traits. To correct for multiple 586 testing, we used a Bonferroni corrected p-value significance cut-off of 2.4×10^{-7} , correcting 587 588 for the number of tested variants, phecodes, biomarkers and continuous traits. In total, 29 589 variants were excluded from the single variant PheWAS and GRS-PheWAS because they
- 590 were not imputed in UK Biobank (Supplemental Information, Section II)
- 591

592 Validation of genetic risk scores in HUNT

593 To validate the previously published results from Iceland, Denmark and UK, we created

594 weighted GRSs for each trait based on the published index variants and effect sizes¹⁷ using

595 the same approach as described in the previous section. We tested the predictive ability of

596 each GRS by regressing each trait on the respective GRS in the independent cohort, HUNT

597 $(N_{iron}=N_{TIBC}=56\ 667, N_{TSP}=56\ 664, N_{ferritin}=2334)$ and report the trait variance explained by

the GRS. In total, ten variants were excluded from the GRSs because they were not imputed

- 599 in HUNT (Supplemental Information, section III).
- 600

601 Mendelian randomization of iron status on all-cause mortality

602 To assess the causal association of iron status on all-cause mortality, we performed linear MR

analyses using the ratio of coefficients method⁹³, using GRSs as genetic instruments for the

604 four iron related traits. The GRSs were constructed as described above, using index variants

and external effect sizes from the previous independent meta-analysis¹⁷. We used linear

regression to estimate the associations between the iron related traits and the GRS, and a Cox

607 proportional hazards regression to estimate the association of the GRS with mortality. The

608 MR estimate was obtained as the ratio of the outcome-instrument and exposure-instrument

609 association estimates. The standard error was estimated as the standard error of the GRS-

610 mortality association divided by the GRS-biomarker association estimate.

611 To further assess the shape of the association, we performed a non-linear Mendelian

612 randomization with the fractional polynomial method^{94,95} in HUNT, using the same GRS as

It is made available under a CC-BY 4.0 International license .

genetic instrument for the iron traits. The method has been described in detail elsewhere^{95–98}. 613 614 In brief, each iron related trait was regressed on its respective GRS, and the population was 615 divided into 100 (iron, TIBC, TSP) or 20 (ferritin) strata based on the residual traits. The 616 number of strata was reduced for ferritin because of the lower sample size for this biomarker. 617 In each stratum we used linear regression to estimate the association of the GRS with the iron 618 trait, and Cox proportional hazards regression to estimate the association of the GRS with 619 mortality. We calculated the localized average causal effect (LACE) of the respective trait on 620 all-cause mortality in each stratum as the ratio of the GRS-outcome and GRS-exposure 621 associations. We performed meta-regression of the LACE against the mean of the exposure 622 in each stratum and tested whether the best-fitting fractional polynomial of degree 1 fitted the LACE estimates better than a linear model using the fractional polynomial method⁹⁴. 623 624

625 To further validate the selection of SNPs representing each biomarker in the non-linear MR,

626 we performed post-hoc sensitivity analyses rerunning the non-linear MR method with new

627 instruments that had stricter criteria for SNP inclusion. Here, we restricted the GRSs to index

628 variants from the previous study¹⁷ that were not only GWAS significant for at least one trait,

but also nominally significant (p-value<0.05) for the remaining traits. Further, we excluded

630 SNPs that had directions of effect that were not consistent with systemic iron status

631 (increasing serum iron, ferritin and TSP, and decreasing TIBC)⁹⁹. We used the remaining 14

632 SNPs (Supplemental Information, Section IV) to construct each of the four GRSs in the

633 analysis as described for the main analysis.

634

635 Ethics

636 All study participants have given informed consent. The analyses in HUNT has approval

637 from the Norwegian Data Protection Authority and the Regional Ethics Committee for

638 Medical and Health Research Ethics in Central Norway (REK Reference Number: 2014/144),

639 the analyses in MGI are approved by the Institutional Review Board of the University of

640 Michigan Medical School (IRB Reference Number: HUM00094409), and the analyses in UK

641 Biobank are covered by the ethics approval for UK Biobank studies (application 24460) from

the NHS National Research Ethics Service on 17th June 2011 (Ref 11/NW/0382) and

643 extended on 10th May 2016 (Ref 16/NW/0274).

It is made available under a CC-BY 4.0 International license .

644

645 Data Availability

646 Data generated or analyzed during this study are available from the corresponding authors

- 647 upon reasonable request.
- 648

649 Acknowledgements

650 The Trøndelag Health Study (The HUNT Study) is a collaboration between HUNT Research 651 Center (Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science 652 and Technology), Trøndelag County Council, Central Norway Regional Health Authority, 653 and the Norwegian Institute of Public Health. The genotyping in HUNT was financed by the 654 National Institutes of Health; University of Michigan; the Research Council of Norway; the 655 Liaison Committee for Education, Research, and Innovation in Central Norway; and the Joint 656 Research Committee between St Olav's hospital and the Faculty of Medicine and Health 657 Sciences, NTNU. The authors acknowledge the Michigan Genomics Initiative participants, 658 Precision Health at the University of Michigan, the University of Michigan Medical School 659 Central Biorepository, and the University of Michigan Advanced Genomics Core for 660 providing data and specimen storage, management, processing, and distribution services, and 661 the Center for Statistical Genetics in the Department of Biostatistics at the School of Public 662 Health for genotype data curation, imputation, and management in support of the research 663 reported in this publication.

664

665 Author Contributions

- 666 M.R.M analyzed the data and wrote the first draft of the manuscript. A.F.H., S.E.G, S.A.G.T,
- 667 K-H.W and W.Z. contributed to analyses. K.T. gave advice on phenotype definitions and
- hemochromatosis. S.B., A.M. and D.G. gave advice on non-linear Mendelian randomization.
- 669 B.M.B, J.B.N, K.H. and C.J.W conceived and designed the study. All authors interpreted
- 670 results and revised the paper.

671

672 Competing Interests

673 D.G is employed part-time by Novo Nordisk outside the submitted work. G.R.A. works for

- 674 Regeneron Pharmaceuticals. C.J.W.'s spouse works for Regeneron Pharmaceuticals. The
- 675 remaining authors have no competing interests to declare.

It is made available under a CC-BY 4.0 International license .

67	6
----	---

677 References

678	1.	Andrews, N. C. Disorders of iron metabolism. N. Engl. J. Med. 341, 1986–1995
679		(1999).
680	2.	Rajpathak, S. N. et al. The role of iron in type 2 diabetes in humans. Biochim. Biophys.
681		Acta - Gen. Subj. 1790, 671–681 (2009).
682	3.	Belaidi, A. A. & Bush, A. I. Iron neurochemistry in Alzheimer's disease and
683		Parkinson's disease: targets for therapeutics. J. Neurochem. 139, 179–197 (2016).
684	4.	Jáuregui-Lobera, I. Iron deficiency and cognitive functions. Neuropsychiatr. Dis.
685		Treat. 10, 2087–2095 (2014).
686	5.	Das, I. et al. Impact of iron deficiency anemia on cell-mediated and humoral immunity
687		in children: A case control study. J. Nat. Sci. Biol. Med. 5, 158-163 (2014).
688	6.	Ahluwalia, N., Sun, J., Krause, D., Mastro, A. & Handte, G. Immune function is
689		impaired in iron-deficient, homebound, older women. Am. J. Clin. Nutr. 79, 516-521
690		(2004).
691	7.	Silva, B. & Faustino, P. An overview of molecular basis of iron metabolism regulation
692		and the associated pathologies. Biochim. Biophys. Acta - Mol. Basis Dis. 1852, 1347-
693		1359 (2015).
694	8.	Gomme, P. T. & McCann, K. B. Transferrin: Structure, function and potential
695		therapeutic actions. Drug Discov. Today 10, 267–273 (2005).
696	9.	Wang, W. et al. Ferritin H is a novel marker of early erythroid precursors and
697		macrophages. Histopathology 62, 931–940 (2013).
698	10.	Koury, M. J. & Ponka, P. New insights into erythropoiesis: The roles of folate, vitamin
699		B 12, and iron. Annu. Rev. Nutr. 24, 105-131 (2004).
700	11.	Philpott, C. C. The flux of iron through ferritin in erythrocyte development. Curr.
701		<i>Opin. Hematol.</i> 25 , 183–188 (2018).
702	12.	Pfeiffer, C. M. & Looker, A. C. Laboratory methodologies for indicators of iron
703		status : strengths, limitations, and analytical challenges. 106, 1606–1614 (2017).
704	13.	Lal, A. Iron in Health and Disease: An Update. Indian J. Pediatr. 87, 58-65 (2020).
705	14.	Abbaspour, N., Hurrell, R. & Kelishadi, R. Review on iron and its importance for
706		human health. J. Res. Med. Sci. 19, 3–11 (2014).
707	15.	Finberg, K. E. et al. Mutations in TMPRSS6 cause iron-refractory iron deficiency
708		anemia (IRIDA). Nat. Genet. 40, 569–571 (2008).

709	16.	Benyamin, B. et al. Variants in TF and HFE Explain ~40% of Genetic Variation in
710		Serum-Transferrin Levels. Am. J. Hum. Genet. 84, 60-65 (2009).
711	17.	Bell, S. et al. A genome-wide meta-analysis yields 46 new loci associating with
712		biomarkers of iron homeostasis. Commun. Biol. 4, (2021).
713	18.	Gill, D. et al. Associations of genetically determined iron status across the phenome: A
714		mendelian randomization study. PLoS Med. 16, 1-16 (2019).
715	19.	Gill, D. et al. Effects of Genetically Determined Iron Status on Risk of Venous
716		Thromboembolism and Carotid Atherosclerotic Disease: A Mendelian Randomization
717		Study. J. Am. Heart Assoc. 8, (2019).
718	20.	Gill, D. et al. The Effect of Iron Status on Risk of Coronary Artery Disease: A
719		Mendelian Randomization Study - Brief Report. Arterioscler. Thromb. Vasc. Biol. 37,
720		1788–1792 (2017).
721	21.	Del Greco, F. M. et al. Serum iron level and kidney function: A Mendelian
722		randomization study. Nephrol. Dial. Transplant. 32, 273-278 (2017).
723	22.	Pichler, I. et al. Serum Iron Levels and the Risk of Parkinson Disease: A Mendelian
724		Randomization Study. PLoS Med. 10, (2013).
725	23.	Gill, D., Monori, G., Tzoulaki, I. & Dehghan, A. Iron status and risk of stroke: A
726		Mendelian randomization study. Stroke 49, 2815–2821 (2018).
727	24.	Taliun, D. et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed
728		Program. <i>Biorxiv</i> 1–46 (2019).
729	25.	Krokstad, S. et al. Cohort profile: The HUNT study, Norway. Int. J. Epidemiol. 42,
730		968–977 (2013).
731	26.	Pilia, G. et al. Heritability of cardiovascular and personality traits in 6,148 Sardinians.
732		PLoS Genet. 2, 1207–1223 (2006).
733	27.	Hentze, M. W., Muckenthaler, M. U., Galy, B. & Camaschella, C. Two to Tango:
734		Regulation of Mammalian Iron Metabolism. Cell 142, 24-38 (2010).
735	28.	Sanchez, M., Galy, B., Muckenthaler, M. U. & Hentze, M. W. Iron-regulatory proteins
736		limit hypoxia-inducible factor-2a expression in iron deficiency. Nat. Struct. Mol. Biol.
737		14 , 420–426 (2007).
738	29.	Paradkar, P. N., Zumbrennen, K. B., Paw, B. H., Ward, D. M. & Kaplan, J. Regulation
739		of Mitochondrial Iron Import through Differential Turnover of Mitoferrin 1 and
740		Mitoferrin 2. Mol. Cell. Biol. 29, 1007-1016 (2009).
741	30.	Whitfield, J. B. et al. Effects of HFE C282Y and H63D polymorphisms and polygenic
742		background on iron stores in a large community sample of twins. Am. J. Hum. Genet.

743		66 , 1246–1258 (2000).
744	31.	De Tayrac, M. et al. Genome-wide association study identifies TF as a significant
745		modifier gene of iron metabolism in HFE hemochromatosis. J. Hepatol. 62, 664–672
746		(2015).
747	32.	Montosi, G. et al. Autosomal-dominant hemochromatosis is associated with a mutation
748		in the ferroportin (SLC11A3) gene. J. Clin. Invest. 108, 619-623 (2001).
749	33.	Roetto, A. et al. Hemochromatosis due to mutations in transferrin receptor 2. Blood
750		Cells. Mol. Dis. 29, 465–470 (2002).
751	34.	Bogdan, A. R., Miyazawa, M., Hashimoto, K. & Tsuji, Y. Regulators of Iron
752		Homeostasis: New Players in Metabolism, Cell Death, and Disease. Trends Biochem.
753		<i>Sci.</i> 41 , 274–286 (2016).
754	35.	Benyamin, B. et al. Novel loci affecting iron homeostasis and their effects in
755		individuals at risk for hemochromatosis. Nat. Commun. 5, 5926 (2014).
756	36.	Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-
757		wide complex trait analysis. Am. J. Hum. Genet. 88, 76-82 (2011).
758	37.	Bulik-Sullivan, B. et al. LD score regression distinguishes confounding from
759		polygenicity in genome-wide association studies. Nat. Genet. 47, 291-295 (2015).
760	38.	Poncelet, L., Dumont, J. E., Miot, F. & De Deken, X. The Dual Oxidase Duox2
761		stabilized with DuoxA2 in an enzymatic complex at the surface of the cell produces
762		extracellular H2O2 able to induce DNA damage in an inducible cellular model. Exp.
763		<i>Cell Res.</i> 384 , 111620 (2019).
764	39.	Horton, R. et al. Gene map of the extended human MHC. Nat. Rev. Genet. 5, 889-899
765		(2004).
766	40.	Gao, X. et al. HES6 acts as a transcriptional repressor in myoblasts and can induce the
767		myogenic differentiation program. J. Cell Biol. 154, 1161-1171 (2001).
768	41.	Li, Y. et al. ZNF322, a novel human C 2H 2 Krüppel-like zinc-finger protein,
769		regulates transcriptional activation in MAPK signaling pathways. Biochem. Biophys.
770		Res. Commun. 325, 1383–1392 (2004).
771	42.	Napolitano, G. & Ballabio, A. TFEB at a glance. J. Cell Sci. 129, 2475–2481 (2016).
772	43.	Jutabha, P. et al. Human sodium phosphate transporter 4 (hNPT4/SLC17A3) as a
773		common renal secretory pathway for drugs and urate. J. Biol. Chem. 285, 35123-
774		35132 (2010).
775	44.	Fleming, J. C. et al. Characterization of a murine high-affinity thiamine transporter,
776		Slc19a2. Mol. Genet. Metab. 74, 273–280 (2001).

777	45.	Saidu, Y. Physicochemical features of rhodanese: A review. African J. Biotechnol. 3,
778		370–374 (2004).
779	46.	Thompson, M. A. et al. Human indolethylamine N-methyltransferase: cDNA cloning
780		and expression, gene cloning, and chromosomal localization. Genomics 61, 285-297
781		(1999).
782	47.	Pers, T. H. et al. Biological interpretation of genome-wide association studies using
783		predicted gene functions. Nat. Commun. 6, 5890 (2015).
784	48.	Yu, T. et al. The prognostic value of differentially expressed CYP3A subfamily
785		members for hepatocellular carcinoma. Cancer Manag. Res. 10, 1713–1726 (2018).
786	49.	Kichaev, G. et al. Leveraging Polygenic Functional Enrichment to Improve GWAS
787		Power. Am. J. Hum. Genet. 104, 65–75 (2019).
788	50.	Wei, W. Q. et al. Evaluating phecodes, clinical classification software, and ICD-9-CM
789		codes for phenome-wide association studies in the electronic health record. PLoS One
790		12 , 1–16 (2017).
791	51.	Rustad, P. et al. The Nordic Reference Interval Project 2000: Recommended reference
792		intervals for 25 common biochemical properties. Scand. J. Clin. Lab. Invest. 64, 271-
793		284 (2004).
794	52.	Astle, W. J. et al. The Allelic Landscape of Human Blood Cell Trait Variation and
795		Links to Common Complex Disease. Cell 167, 1415-1429.e19 (2016).
796	53.	Fujibayashi, A. et al. Human RME-8 is involved in membrane trafficking through
797		early endosomes. Cell Struct. Funct. 33, 35-50 (2008).
798	54.	Gunshin, H. et al. Slc11a2 is required for intestinal iron absorption and erythropoiesis
799		but dispensable in placenta and liver. J. Clin. Invest. 115, 1258–1266 (2005).
800	55.	Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation
801		in 141,456 humans. Nature 581, 434–443 (2020).
802	56.	Vuckovic, D. et al. The Polygenic and Monogenic Basis of Blood Traits and Diseases.
803		<i>Cell</i> 182 , 1214-1231.e11 (2020).
804	57.	Chen, M. H. et al. Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667
805		Individuals from 5 Global Populations. Cell vol. 182 (2020).
806	58.	Oskarsson, G. R. et al. Predicted loss and gain of function mutations in ACO1 are
807		associated with erythropoiesis. Commun. Biol. 3, 1-10 (2020).
808	59.	Johnson, A. D. et al. Genome-wide association meta-analysis for total serum bilirubin
809		levels. Hum. Mol. Genet. 18, 2700-2710 (2009).
810	60.	Nielsen, J. B. et al. Loss-of-function genomic variants highlight potential therapeutic

811		targets for cardiovascular disease. Nat. Commun. 11, (2020).
812	61.	Sookoian, S. & Pirola, C. J. Liver enzymes, metabolomics and genome-wide
813		association studies: From systems biology to the personalized medicine. World J.
814		<i>Gastroenterol.</i> 21 , 711–725 (2015).
815	62.	Johansson, C. et al. The roles of Jumonji-type oxygenases in human disease.
816		<i>Epigenomics</i> 6 , 89–120 (2014).
817	63.	Khandia, R. et al. A Comprehensive Review of Autophagy and Its Various Roles in
818		Infectious, Non-Infectious, and Lifestyle Diseases: Current Knowledge and Prospects
819		for Disease Prevention, Novel Drug Design, and Therapy. Cells 8, 674 (2019).
820	64.	Law, R. H. P. et al. An overview of the serpin superfamily. Genome Biol. 7, 1-11
821		(2006).
822	65.	Wang, C. Y., Meynard, D. & Lin, H. Y. The role of TMPRSS6/matriptase-2 in iron
823		regulation and anemia. Front. Pharmacol. 5 MAY, 1-6 (2014).
824	66.	Rishi, G. & Subramaniam, V. N. The liver in regulation of iron homeostasis. Am. J.
825		Physiol Gastrointest. Liver Physiol. 313, G157–G165 (2017).
826	67.	Darshan, D. & Anderson, G. J. Interacting signals in the control of hepcidin
827		expression. BioMetals 22, 77-87 (2009).
828	68.	Arezes, J. et al. Erythroferrone inhibits the induction of hepcidin by BMP6. Blood 132,
829		1473–1477 (2018).
830	69.	Armitage, A. E. et al. Hepcidin regulation by innate immune and infectious stimuli.
831		<i>Blood</i> 118 , 4129–4139 (2011).
832	70.	Lin, L. et al. Iron transferrin regulates hepcidin synthesis in primary hepatocyte culture
833		through hemojuvelin and BMP2/4. <i>Blood</i> 110 , 2182–2189 (2007).
834	71.	Pietrangelo, A. Hereditary Hemochromatosis — A New Look at an Old Disease. N.
835		Engl. J. Med. 350, 2383–2397 (2004).
836	72.	Siow, D., Sunkara, M., Morris, A. & Wattenberg, B. Regulation of de novo
837		sphingolipid biosynthesis by the ORMDL proteins and sphingosine kinase-1. Adv.
838		<i>Biol. Regul.</i> 57 , 42–54 (2015).
839	73.	Cho, H. P., Nakamura, M. & Clarke, S. D. Cloning, expression, and fatty acid
840		regulation of the human Δ -5 desaturase. J. Biol. Chem. 274 , 37335–37339 (1999).
841	74.	Ferreira, M. A. et al. Shared genetic origin of asthma, hay fever and eczema elucidates
842		allergic disease biology. Nat. Genet. 49, 1752-1757 (2017).
843	75.	Voight, B. F. et al. The Metabochip, a Custom Genotyping Array for Genetic Studies
844		of Metabolic, Cardiovascular, and Anthropometric Traits. PLoS Genet. 8, 1–12 (2012).

It is made available under a CC-BY 4.0 International license .

845	76.	Parkes, M., Cortes, A., Van Heel, D. A. & Brown, M. A. Genetic insights into
846		common pathways and complex relationships among immune-mediated diseases. Nat.
847		Rev. Genet. 14, 661–673 (2013).
848	77.	Das, S. et al. Next-generation genotype imputation service and methods. Nat Genet 48,
849		1284–1287 (2016).
850	78.	Sidore, C. et al. Genome sequencing elucidates Sardinian genetic architecture and
851		augments association analyses for lipid and blood inflammatory markers. Nat. Genet.
852		47 , 1272–1281 (2015).
853	79.	Loh, P. R. et al. Efficient Bayesian mixed-model analysis increases association power
854		in large cohorts. Nat. Genet. 47, 284–290 (2015).
855	80.	Zhan, X., Hu, Y., Li, B., Abecasis, G. R. & Liu, D. J. RVTESTS: An efficient and
856		comprehensive tool for rare variant association analysis using sequence data.
857		Bioinformatics 32, 1423–1426 (2016).
858	81.	Kang, H. M. et al. Variance component model to account for sample structure in
859		genome-wide association studies. Nat. Genet. 42, 348-354 (2010).
860	82.	Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of
861		genomewide association scans. Bioinformatics 26, 2190-2191 (2010).
862	83.	Kent, W. J. et al. The Human Genome Browser at UCSC. Genome Res. 12, 996–1006
863		(2002).
864	84.	Buniello, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide
865		association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res.
866		47 , D1005–D1012 (2019).
867	85.	Chang, C. C. et al. Second-generation PLINK: Rising to the challenge of larger and
868		richer datasets. Gigascience 4, 1–16 (2015).
869	86.	Wang, K., Li, M. & Hakonarson, H. ANNOVAR: Functional annotation of genetic
870		variants from high-throughput sequencing data. Nucleic Acids Res. 38, 1-7 (2010).
871	87.	Giambartolomei, C. et al. Bayesian Test for Colocalisation between Pairs of Genetic
872		Association Studies Using Summary Statistics. PLoS Genet. 10, (2014).
873	88.	Wallace, C. Statistical testing of shared genetic control for potentially related traits.
874		Genet. Epidemiol. 37, 802–813 (2013).
875	89.	Weeks, E. M. et al. Leveraging polygenic enrichments of gene features to predict
876		genes underlying complex traits and diseases. medRxiv (2020)
877		doi:10.1101/2020.09.08.20190561.

878 90. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized

It is made available under a CC-BY 4.0 International license .

879		Gene-Set Analysis of GWAS Data. PLoS Comput. Biol. 11, 1–19 (2015).
880	91.	Altshuler, D. L. et al. A map of human genome variation from population-scale
881		sequencing. Nature 467, 1061–1073 (2010).
882	92.	Zhou, W. et al. Efficiently controlling for case-control imbalance and sample
883		relatedness in large-scale genetic association studies. Nat. Genet. 50, 1335-1341
884		(2018).
885	93.	Palmer, T. M. et al. Instrumental variable estimation of causal risk ratios and causal
886		odds ratios in mendelian randomization analyses. Am. J. Epidemiol. 173, 1392–1403
887		(2011).
888	94.	Staley, J. R. & Burgess, S. Semiparametric methods for estimation of a nonlinear
889		exposure-outcome relationship using instrumental variables with application to
890		Mendelian randomization. Genet. Epidemiol. 41, 341–352 (2017).
891	95.	Burgess, S., Davies, N. M. & Thompson, S. G. Instrumental variable analysis with a
892		nonlinear exposure-outcome relationship. Epidemiology 25, 877-885 (2014).
893	96.	Burgess, S. et al. Dose-response relationship between genetically proxied average
894		blood glucose levels and incident coronary heart disease in individuals without
895		diabetes mellitus. Diabetologia 64, 845-849 (2021).
896	97.	Malik, R. et al. Relationship between Blood Pressure and Incident Cardiovascular
897		Disease: Linear and Nonlinear Mendelian Randomization Analyses. Hypertension
898		2004–2013 (2021) doi:10.1161/HYPERTENSIONAHA.120.16534.
899	98.	Sun, Y. Q. et al. Body mass index and all cause mortality in HUNT and UK Biobank
900		studies: Linear and non-linear mendelian randomisation analyses. BMJ 364, 1-10
901		(2019).
902	99.	Wish, J. B. Assessing iron status: beyond serum ferritin and transferrin saturation.
903		<i>Clin. J. Am. Soc. Nephrol.</i> 1 Suppl 1 , 4–8 (2006).

904