A genome-wide meta-analysis yields 46 new loci associating with biomarkers of iron homeostasis

Steven Bell1,2,35, Andreas S. Rigas3,35, Magnus K. Magnusson4,5,35, Egil Ferkingstad4,35, Elias Allara1,2,35, Gyda Bjornsdottr4, Anna Ramond1,2,6, Erik Sørensen3, Gisli H. Halldorsson6,4, Dirk S. Paul1,2, Kristoffer S. Burgdorf3, Hannes P. Eggertsson6,4, Joanna M. M. Howson2, Lise W. Thørner3, Snaedis Kristmundsdottir4, William J. Astle1,2,7,8, Christian Erikstrup9, Jon K. Sigurdsson4,

Dragana Vuckovic1,8, Khoa M. Dinh9, Vinicius Tragante4,10, Praveen Surendran2,11, Ole B. Pedersen1,2,12, Brynjar Vidarsson13, Tao Jiang1,2,8, Helene M. Paarup14, Pall T. Onundarson5,15, Parsa Akbari1,2,8, Kaspar R. Nielsen16, Sigrun H. Lund4,4, Kristinn Juleusson4, Magnus I. Magnusson4, Michael L. Frigge4,4, Asmundur Oddsson4,4, Isleifur Olafsson17, Stephen Kaptoge1,2, Henrik Hjalgrim18, Gudmundur Runarsson13, Angela M. Wood1,2, Ingileif Jonsdottir6,4, Thomas F. Hansen19,20,21, Olof Sigurdardottir22,

Hreinn Stefansson4,4, David Rye23, DBDS Genomic Consortium*, James E. Peters2, David Westergaard24, Hilma Holm4,4, Nicole Soranzo1,8,25, Karina Banasik24, Gudmar Thorleifsson4, Willem H. Ouwehand1,8,25,26, Unnur Thorsteinsdottir4,5, David J. Roberts1,27,28, Patrick Sulem4,4, Adam S. Butterworth1,2, Daniel F. Gudbjartsson4,4,29, John Danesh1,2,8, Søren Brunak24, Emanuele Di Angelantonio1,2,36,35, Henrik Ullum3,36,35 & Kari Stefansson4,5,36,4,5,36

Iron is essential for many biological functions and iron deficiency and overload have major health implications. We performed a meta-analysis of three genome-wide association studies from Iceland, the UK and Denmark of blood levels of ferritin (N = 246,139), total iron binding capacity (N = 135,430), iron (N = 163,511) and transferrin saturation (N = 131,471). We found 62 independent sequence variants associating with iron homeostasis parameters at 56 loci, including 46 novel loci. Variants at DUOX2, F5, SLC11A2 and TMPRSS6 associate with iron deficiency anemia, while variants at TF, HFE, TFR2 and TMPRSS6 associate with iron overload. A HBS1L-MYB intergenic region variant associates both with increased risk of iron overload and reduced risk of iron deficiency anemia. The DUOX2 missense variant is present in 14% of the population, associates with all iron homeostasis biomarkers, and increases the risk of iron deficiency anemia by 29%. The associations implicate proteins contributing to the main physiological processes involved in iron homeostasis: iron sensing and storage, inflammation, absorption of iron from the gut, iron recycling, erythropoiesis and bleeding/menstruation.

A full list of author affiliations appears at the end of the paper.
Iron is an essential element for a wide variety of metabolic processes such as oxygen transport, cellular respiration, and redox reactions in numerous metabolic pathways. For this reason, iron homeostasis is tightly regulated on cellular and systemic levels to ensure a balance between uptake, transport, storage, and utilization. Iron deficiency is one of the five leading causes of disability worldwide, especially among children and women of childbearing age. Similarly, iron overload is associated with an increased risk of several major chronic conditions, including diabetes and liver disease.

Four iron biomarkers are used for clinical assessment of iron status: serum ferritin, serum iron, and total iron-binding capacity (TIBC) are measured directly, while transferrin saturation (TSAT) is derived as serum iron divided by TIBC. While serum ferritin correlates well with body iron stores in non-inflamed individuals, TSAT measures the proportion of iron-binding sites of transferrin that are occupied by iron. TSAT indicates the availability of iron for erythropoiesis and is low in iron deficiency and high during iron overload. In some forms of anemia (e.g., anemia of inflammation) the iron is not transported efficiently to the bone marrow for erythropoiesis, despite adequate iron stores. Since in this situation there is adequate ferritin but low TSAT, it is useful to evaluate TSAT in addition to ferritin.

Genome-wide association studies (GWAS) have previously investigated the association between sequence variants and iron homeostasis biomarkers. The largest study to date yielded 11 loci: ABO, ARNTL, FADS2, HFE, NAT2, SLC40A1, TEX14, TF, TFR2, TFRC, and TMPRSS6 associating with one or more iron homeostasis biomarkers (ferritin, iron, TIBC or TSAT). To search for additional sequence variants associated with iron homeostasis, we performed a GWAS meta-analysis of ferritin, serum iron, TIBC, and TSAT in Iceland and blood donor studies from the UK (INTERVAL study) and Denmark (Danish Blood Donor Study). This was followed by cross-referencing of iron-associated loci with clinically relevant phenotypes (including iron deficiency anemia (IDA), iron overload, and red blood cell indices). We report associations with iron homeostasis biomarkers for 62 independent sequence variants at 56 loci, including 46 novel loci. Based on a literature review, we categorize 25 of these loci as associated with clinically relevant phenotypes (including iron deficiency anemia, iron overload, and red blood cell indices).

**Results**

**Overview.** We performed a meta-analysis of four iron-related biomarkers: ferritin \((N = 246,139)\), serum iron \((N = 163,511)\), TIBC \((N = 135,430)\), and TSAT \((N = 131,471)\), combining GWAS results from Iceland, the UK, and Denmark (Fig. 1, Supplementary Data 1). We found associations with iron homeostasis biomarkers represented by 62 sequence variants at 56 loci, of which 46 have not been reported in the previous GWAS on iron homeostasis and are therefore considered novel (Table 1, Table 2, Fig. 2, and Supplementary Data 2). For each locus, we report the lead variant \((P < 0.01)\) and additional uncorrelated variants \((P < 3.0 \times 10^{-8})\) within the locus with genome-wide significance. Our criteria for statistical significance have been previously described (see "Methods"). A variant-to-gene mapping algorithm that takes into account gene location, variant effect (for coding variants), and effect on gene expression (eQTL) for each variant (lead variant and LD class) was used to choose a single candidate gene for each locus (see "Methods"). Twenty-five of the 62 iron homeostasis-associated sequence variants have a high-confidence predicted causal gene, 23 variants have multiple top-scoring genes, 36 variants have at least one coding variant or eQTL in the LD class, and 13 variants have more than one gene with coding variants and/or eQTL in the LD class (Supplementary Data 3). The LD class of a variant is defined as all variants having \(r^2 > 0.8\) with the variant. Linkage disequilibrium (LD) \((r^2)\) is estimated based on the Icelandic population. In cases where variants had more than one top-scoring gene, the gene closest to the lead variant was selected, except for two loci where likely candidate genes were present among the top-scoring genes (FTL (ferritin light chain) and HAMP (hepcidin)) (Supplementary Data 3). Fourteen of the variants associated with more than one biomarker, bringing the total number of observed associations to 87 (Supplementary Data 2). All our associations have \(P < 3.0 \times 10^{-8}\). We replicated the association of all 11 previously reported variants, 10 at genome-wide significance (Supplementary Data 2). In addition, we found six rare variants (minor allele frequency (MAF) < 1%), six low-frequency variants \((1% \leq MAF < 5%)\), and 37 common variants that have previously not been reported to associate with iron homeostasis biomarkers (Supplementary Data 2). Forty-six variants associate with a single iron biomarker (ferritin, 34;
| Gene | Position (chr: bp) | Marker | Minor allele | Major allele | MAF (%) | Minor/maj | MAF (%) | Phenotype | P value | The effect in SD (95% CI) |
|------|-------------------|--------|--------------|-------------|---------|----------|---------|-----------|---------|--------------------------|
| LEPR | rs469882 chr1:91064875 | C/A | 0.566 | 0.208 | 1.03 × 10^-4 | Iron | 0.026 (0.018, 0.034) | 8.42 × 10^-4 |
| ZNF644 | rs879870 chr1:169549811 | T/C | 0.037 | 0.039 | 1.00 × 10^-1 | Iron | -0.093 (-0.13, -0.059) | 2.15 × 10^-4 |
| IL6R | rs1051130 chr1:227044347 | T/C | 0.107 | 0.384 | 1.00 × 10^-1 | Iron | 0.026 (0.017, 0.035) | 1.02 × 10^-4 |
| ERFE | rs879870 chr1:169549811 | T/C | 0.097 | 0.384 | 1.00 × 10^-1 | Iron | 0.019 (0.011, 0.027) | 4.65 × 10^-5 |
| TSAT | rs1051130 chr1:227044347 | T/C | 0.924 | 0.684 | 1.00 × 10^-1 | Iron | 0.033 (0.024, 0.042) | 1.10 × 10^-4 |
| MYB | rs9399136 chr6:135081201 | C/T | 25.9 | 74.1 | 1.00 × 10^-1 | Iron | 0.033 (0.024, 0.042) | 1.10 × 10^-4 |
| HGFAC | rs9399136 chr6:135081201 | C/T | 0.483 | 0.517 | 1.00 × 10^-1 | Iron | 0.033 (0.024, 0.042) | 1.10 × 10^-4 |
| IHK1 | rs17580 chr14:94380925 | A/T | 0.0361 | 0.964 | 1.00 × 10^-1 | Iron | -0.023 (-0.043, -0.003) | 2.65 × 10^-4 |
| DUOX2 | rs57659670 chr15:45106240 | C/T | 0.130 | 0.870 | 1.00 × 10^-1 | Iron | 0.067 (0.056, 0.078) | 5.32 × 10^-4 |
| ABCA5 | rs77262773 chr17:69253570 | T/C | 0.126 | 0.874 | 1.00 × 10^-1 | Iron | -0.058 (-0.076, -0.040) | 5.73 × 10^-4 |
| ABCA5 | rs2005682 chr19:35456759 | T/A | 0.434 | 0.566 | 1.00 × 10^-1 | Iron | -0.032 (-0.042, -0.022) | 6.25 × 10^-4 |
| PRRG2 | rs112727702 chr19:49587947 | T/G | 0.130 | 0.874 | 1.00 × 10^-1 | Iron | -0.058 (-0.076, -0.040) | 5.73 × 10^-4 |
| PRGG2 | rs1132274 chr20:17615510 | A/C | 0.130 | 0.874 | 1.00 × 10^-1 | Iron | -0.032 (-0.042, -0.022) | 6.25 × 10^-4 |

**Iron homeostasis variants and protein quantitative loci (pQTL)**. To gain further insight into the biological pathways involved in iron homeostasis, we tested for association of the
The rs762752083[T] stop-gained variant at GCKR

62 iron homeostasis variants (including all variants with $r^2 \geq 0.8$
with any iron homeostasis variants) with an expression of 4792

The loci in the context of systemic iron homeostasis. Based on a
literature review, we placed 24 of the 56 candidate genes, as well as
the female-specific candidate gene VWF, into 6 categories representing
the main physiological processes involved in iron homeostasis: hepcidin regulation and iron storage (FTL, HAMP, HFE, TMPRSS6, TFR2, TFRC, TF, MTM4, and SERPIN1), inflammation (IL6R, NOD1, and IKZF1), gut absorption (SLC11A2, SLC40A1, EGLN3, and DUOX2), iron recycling (SLC11A2, SLC40A1, STAB1, TRIB1, and MAFF), erythropoiesis (ERFE, SLC25A37, MYB, and HK1) and bleeding/menstruation (F5 and VWF) (Fig. 4).

Hepcidin regulation and iron storage: Synthesis of the iron
homeostasis hormone hepcidin (HAMP) is under tight regulation
by the liver iron sensing and signaling cascade involving several
proteins, including those encoded by HFE, TMPRSS6, TF, TFR2, and
TFRC. Hepcidin as the major iron homeostasis hormone regulates iron transport from cells through inhibition (and degradation) of ferroportin in cells, such as intestinal epithelial
and liver cells and macrophages. HAMP, HFE, TMPRSS6, TF, TFR2, and TFRC along with the iron storage protein ferritin light chain (encoded by FTL) all have variants associated with iron biomarkers. Furthermore, the MTMR4 variant rs34523089 (MAF = 14.1%), associates with ferritin ($\beta = 0.069$ SD [0.059, 0.078], $P = 3.2 \times 10^{-48}$). MTMR4 has been shown to localize to early endosomes where it interacts with and dephosphorylates activated R-Smads, thus negatively regulating transforming growth factor $\beta$ (TGF$\beta$) signaling and TGF$\beta$1 has been shown to activate hepcidin mRNA expression.

The MTMR4 variant also associates with hepcidin protein levels in our pQTL study, similar to what was seen with variants in the known hepcidin regulators, TMPRSS6, and HFE (Supplementary Data 9). The SERPINA1 p.Glu288Val variant (rs17580[A], MAF = 3.79%) associates with increased TIBC ($\beta = 0.076$ SD [0.053, 0.099], $P = 1.2 \times 10^{-10}$). SERPINA1 encodes the protease inhibitor (PI) alpha-1-antitrypsin (A1AT) and the p.Glu288Val variant—also known as the PI S allele—is associated with A1AT-deficiency (A1ATD). Liver disease in A1ATD has been linked to liver iron overload, and recently A1AT was shown to increase hepcidin expression through proteolytic cleavage and inhibition of TMPRSS6.

**Inflammation:** IL6 and its receptor IL6R are important inflammatory mediators positively regulating liver hepcidin during inflammation. The IL6R p.Asp358Ala variant (rs2228145[C], MAF = 41%) that associates with decreased risk of rheumatoid arthritis associates with an increase in serum iron ($\beta = 0.026$ SD [0.018, 0.034], $P = 8.4 \times 10^{-11}$). Leptin and its receptor LEPR, in addition to its central role as an adipokine, have been shown to control cellular immune responses in several pathological situations including rheumatic diseases. The intergenic variant rs35945185[A] (MAF = 36.5%) linked to LEPR associates with iron ($\beta = 0.031$ SD [0.023, 0.039], $P = 1.54 \times 10^{-13}$). The IL6R and LEPR associated variants (rs2228145[C], rs35945185[A]) both are negatively associated.

**Fig. 2** Manhattan plots for iron homeostasis biomarker meta-analysis results for ferritin ($N = 246,139$), serum iron ($N = 163,511$), total iron-binding capacity (TIBC, $N = 135,430$), and transferrin saturation (TSAT, $N = 131,471$). Variants are plotted by chromosomal position ($x$-axis) and $-\log_{10}$ $P$ values ($y$-axis). A likelihood ratio test was used when testing for the association. Blue = novel loci (not reported in previous iron GWAS studies), red = previously reported loci.

**Fig. 3** Venn diagram. Venn diagram showing loci (with predicted gene) harboring variants associated with ferritin, iron, TIBC, and/or TSAT.
with inflammatory markers serum amyloid A-1/A-2 proteins, and furthermore, the LEPR associated variant is also negatively associated with C-reactive protein (Supplementary Data 9). Also, the rs2529440[T] intron variant (MAF = 45%) at NOD1, encoding an intracellular innate immune pattern recognition sensor for bacterial cell components\(^\text{15}\), associates with a reduction in ferritin levels ($\beta = -0.035$, SD $[-0.014, -0.028]$, $P = 4.6 \times 10^{-23}$). Furthermore, the rs12718598[C] intron variant in IKZF1, encoding the lymphocyte specification and differentiation transcription factor Ikaros, shown to play a role in autoimmune diseases\(^\text{26}\), associates with increased serum iron levels ($\beta = 0.027$, SD $[0.019, 0.034]$, $P = 3.7 \times 10^{-11}$).

**Gut absorption**: Iron absorption is mediated by the two iron transporters DMT1 (encoded by SLC11A2) at the luminal and ferroportin (encoded by SLC40A1) basolaterally, both regulated by hepcidin signals and both harboring variants associated with iron homeostasis biomarkers\(^\text{13}\). Recently, hepcidin blocking of intestinal ferroportin was shown to inhibit HIF-2α expression, through increased intracellular iron and subsequent activation of iron-dependent prolyl hydroxylases, leading to reduced expression of iron absorptive proteins\(^\text{27}\). Mammalian HIF-α prolyl hydroxylases are encoded by the three genes EGLN1-3\(^\text{28}\). The rs996347[C] intron variant (MAF = 35%) at EGLN3 associates with increased ferritin ($\beta = 0.049$, SD $[0.042, 0.056]$, $P = 3.0 \times 10^{-41}$). EGLN3 is a likely candidate to mediate the inhibition of intestinal HIF2α expression, as it specifically inhibits HIF-2α rather than HIF-1α\(^\text{29,30}\). The DUOX2 p.His678Arg variant (rs75659670[C], MAF = 75%) associates with reduced ferritin ($\beta = -0.14$, SD $[-0.16, -0.13]$, $P = 1.1 \times 10^{-113}$), serum iron ($\beta = -0.042$, SD $[-0.056, -0.028]$, $P = 1.1 \times 10^{-8}$), and TSAT ($\beta = 0.058$, SD $[-0.074, -0.041]$, $P = 5.7 \times 10^{-12}$) and increased TIBC ($\beta = 0.077$, SD $[0.060, 0.094]$, $P = 3.7 \times 10^{-19}$). DUOX2 is expressed in the upper intestinal mucosa and may play a role in innate mucosal immunity\(^\text{10,31}\). Furthermore, in mouse models, DUOX1 and DUOX2 knockouts have a greater susceptibility to *Helicobacter felis* infection and inflammation\(^\text{32}\) and epidemiological studies have indicated that *H. pylori* infections in humans are associated with reduced iron stores\(^\text{33}\).

**Iron recycling**: Recycling of heme–iron takes place in the reticuloendothelial system in the spleen and liver, where old red cells are taken up and iron recycled back to the bone marrow, providing over 90% of the iron needed for the generation of heme in red cell precursors\(^\text{1}\). DMT1 and ferroportin also transport iron from endocytic vesicles and export iron out of the macrophage, respectively\(^\text{34}\). Furthermore, three uncorrelated rare variants (MAF < 1%) in STAB1 (p.Glu177Ter/rs762752083[T], p.Gly189Ser/rs750717575[A] and p.Glu527Lys/rs745795585[A]) and a variant in LD with a STAB1 variant (p.Ser451Thr/rs34216132[C], $r^2 > 0.99$ with the STAB1 variant p.Ser1089Gly/rs41292856[G]) (Supplementary Fig. 4) all associate with increased ferritin, with effects ranging from $0.17$ to $0.35$ SD ($P = 2.2 \times 10^{-8}$ to $2.6 \times 10^{-19}$). STAB1 is primarily expressed in M2-macrophages and sinusoidal endothelial cells\(^\text{35}\) and has been shown to affect phosphatidylserine-mediated uptake of aged red blood cells\(^\text{36,37}\). We also report associations of the intergenic variants rs2954029[T] (MAF = 48%) and rs6029148[A] (MAF = 7.1%) with reduced and increased ferritin (rs2954029[T]: $\beta = -0.024$, SD $[-0.031, -0.018]$, $P = 1.4 \times 10^{-12}$; rs6029148[A]: $\beta = 0.046$, SD $[0.033, 0.058]$, $P = 5.6 \times 10^{-12}$). Their closest protein-coding genes, TRIB1 (for rs2954029) and MAFB (for rs6029148), have both been shown to control the differentiation of macrophages\(^\text{38,39}\).

**Erythropoiesis**: The bone marrow relays signals inhibiting liver hepcidin synthesis under a state of stress erythropoiesis to make iron available to erythroid precursors\(^\text{40}\). Variants located close to two known iron regulators within the erythropoiesis compartment, the intergenic variant rs13253974[A] (MAF = 32%) near SLC25A37 (mitoferrin-1)\(^\text{41}\) and the intron variant rs13007705[T] at *ERFE* (erythroferrone)\(^\text{40,42}\) associate with increased ferritin...
IDA and iron overload. The two extremes of iron homeostasis, iron deficiency, and iron overload, are clinically important and associated with high disease burden. In iron deficiency, depletion of iron stores is followed by reduced iron availability for erythropoiesis, leading to IDA, presenting as hypochromic, microcytic anemia with low ferritin and/or low TSAT. Increased TSAT, most commonly defined as a saturation above 50%, is used as a screening marker for hemochromatosis and iron overload.

To understand how the 62 iron homeostasis variants affect either IDA or iron overload, we tested for association with IDA (defined as ever simultaneously having hemoglobin < 120 g/L for women, <130 g/L for men, MCV < 80 fl, MCH < 27 pg, and either ferritin < 10 mcg/L or TSAT < 16%; Ncases = 6476; Ncontrols = 362706) and iron overload (defined as TSAT ever >50%, Ncases = 4156, Ncontrols = 342647) (Fig. 5, Supplementary Data 12), correcting for 2 × 62 = 124 performed tests. The missense variants in DUOX2 (p.His678Arg; rs1800562[C]) and F5 (p.Arg534Gln, rs6025[T]) associate with IDA (DUOX2 p.His678Arg: OR = 1.29 [1.20–1.39], P = 2.0 × 10−11; F5 p.Arg534Gln: OR = 0.60 [0.49–0.73], P = 3.4 × 10−7). The variants showing sexual dimorphism for the effect on ferritin also showed similar trends with regard to IDA (Supplementary Fig. 6, Supplementary Data 13). In addition, a 3.55 kb deletion in the SLC11A2 3′ untranslated region (3′ UTR) and its downstream intron associate with IDA through a recessive mode of inheritance (OR = 32.5 [10.0–105]; P = 6.4 × 10−9) (Fig. 5, Supplementary Data 12, Supplementary Fig. 7). A rare frameshift mutation in TMPRSS6 (p.Asn473ThrfsTer63, rs773570300) only detected in the Icelandic cohort (MAF = 0.16%) also associated with IDA (OR = 3.0 [2.1–4.4]; P = 1.2 × 10−8). The rs9399136[C] variant in the intergenic HBS1L/MYB region is the only variant to associate with both IDA (OR = 0.84 [0.80–0.89], P = 4.7 × 10−11) and iron overload (OR = 1.13 [1.07–1.20], P = 1.4 × 10−3). This variant has not been associated with iron homeostasis but has been associated with hematological traits and variants in the same region have been associated with fetal hemoglobin expression.

Additionally, variants in the iron homeostasis regulatory genes HFE, TMPRSS6, TF, and TFR2 associate with iron overload (Fig. 5, Supplementary Data 12).

We tested the 62 iron homeostasis variants for association with the following eleven clinical manifestations of iron overload and/or iron deficiency based on various meta-analyses performed in Iceland using data from Iceland, UK, Denmark, and the USA: hemochromatosis, liver fibrosis/cirrhosis, liver cancer, type 2 diabetes, impotence, cardiomyopathy, osteoporosis, osteoarthritis, hyperpigmentation, amenorrhea, and restless leg syndrome (Supplementary Data 14). Taking all 62 × 11 = 682 tests into account using Bonferroni correction, the TMPRSS6 p.Val749Ala variant (rs855791[A]) associates with less risk of hemochromatosis (Ncases = 719, Ncontrols = 497001; OR = 0.80 [0.72–0.89], P = 6.1 × 10−5). The HFE p.Cys282Tyr variant (rs1800562[A]), the main variant associating with recessive hereditary hemochromatosis (type 1) associates with a higher risk of hemochromatosis (additive model: OR = 25.7 [21.6–30.5], P < 10−300; recessive model: OR = 218.9 [164.6–291.0], P < 10−300), liver fibrosis/cirrhosis (Ncases = 1043, Ncontrols = 705646; additive model:
Through a GWAS meta-analysis of the iron homeostasis biomarkers ferritin, serum iron, iron-binding capacity, and TSAT in Iceland, Denmark, and the UK, we have identified 56 loci harboring variants associated with one or more of these biomarkers, 46 of which are novel (including six rare variants, six low-frequency variants, and 37 common variants). Among the novel loci, variants in DUOX2 and SLC11A2 associate with increased risk of IDA, while the F5 rs6025[T] variant protects against IDA. Furthermore, the rs9399136[C] variant at the HBBS1/MYB locus is protective against IDA while increasing the risk of iron overload.

While most of these iron homeostasis variants show similar effects in Iceland, UK, and Denmark, the observed heterogeneity for a subset of the variants may reflect demographic, clinical, and environmental differences. In clinical populations, iron homeostasis markers are more frequently measured in individuals with...
and tissue damage (e.g., liver injury)\textsuperscript{66}, and also that we have more ferritin measurements (~246 K vs. ~131–163 K).

Iron deficiency is a major global health problem, especially for children and women\textsuperscript{2}. A worldwide survey in 2010 showed that one-third of the world population is anemic with iron deficiency being responsible for approximately half of that cases\textsuperscript{69}. In addition to the nonspecific symptoms of IDA, it also may contribute globally to reduced cognitive performance in children\textsuperscript{67}, adverse outcomes of pregnancies\textsuperscript{66}, and decline in cognition in the elderly\textsuperscript{68,69}. Despite the importance of iron deficiency and IDA, no systematic genetic studies looking at iron deficiency or IDA have been performed. Sequence variants that are common (at DUOX2 and the HBBS1-MYB intergenic region), low-frequency (at\textsuperscript{FS}), and rare (at TMPRSS6 and SLC11A2) associate with IDA (Fig. 5, Table 1). The association of the missense DUOX2 variant with all iron homeostasis markers, as well as with IDA is striking. That this association was seen in all three populations studied but not observed in previous GWAS of iron homeostasis is intriguing, however, it should be noted that Benyamin et al.\textsuperscript{6} reported a genome-wide significant association with ferritin near this locus (rs16976620). Our study is significantly larger and also benefits from more comprehensive imputation panels made available since then, which likely enabled us to not only detect an association at genome-wide significance but also map this to the likely causal gene with high confidence.

The phenotype of recessive IDA with low iron stores that we report with the rare 3.5 kb deletion within SLC11A2 is different from the previously reported recessive hypochromic anemia with iron overload associated with this gene\textsuperscript{55–57}. Further studies to define the pathways mediating the effects of the variants associating with IDA could help shed light on the pathophysiology of iron deficiency. Notably, neither any individual iron homeostasis variants nor the PRS for ferritin or TSAT associate with the risk of restless leg syndrome, a neurological disorder suggested being exacerbated by iron deficiency\textsuperscript{70}. Although this argues against a simple causal relationship between the two, a more complex relationship, e.g., through brain iron concentrations\textsuperscript{71} cannot be ruled out. Even though hereditary hemochromatosis is most often associated with HFE p.Cys282Tyr homozygosity, the penetrance is only 28% in males and much lower in females\textsuperscript{72}. The common missense variant in TMPRSS6 (rs855791[A], MAF = 43.1%) protects against hereditary hemochromatosis (OR = 0.80 [0.72–0.89], \(P = 6.1 \times 10^{-5}\)) and could thus be a modifying gene in this disease.

In summary, we have identified 46 novel loci affecting iron homeostasis. Many of the novel candidate genes have roles in homeostasis through mechanisms, such as absorption, iron recycling, erythropoiesis, and hepcidin regulation. Furthermore, we show an association of five of these loci with IDA, a major clinical entity that hitherto has not been studied thoroughly from a genetic point of view. This study reveals a substantial catalog of possible iron regulatory genes, awaiting further inquiry to fully elucidate their functional role.

Methods

Study subjects from Iceland. The Icelandic data (where around one-half of all individuals had repeated measurements) include the vast majority of all clinical laboratory results in Iceland from 1990 to 2017. Serum iron and TIBC were measured with colorimetric methods and serum ferritin was measured with an electrochemiluminescence immunoassay using reagents and calibrators and Cobas 6800 and 8000 modular instruments from Roche Diagnostics, Mannheim, Germany. Hemoglobin concentration measurements, as well as other basic hematology parameters used, were measured on EDTA anticoagulated blood using the Sysmex XN-1000 hematology analyzer.

All participants who donated samples gave informed consent and the National Bioethics Committee of Iceland approved the study (VSN-15-198) which was conducted in agreement with conditions issued by the Data Protection Authority of Iceland. Personal identities of the participant’s data and biological samples were
Association testing and meta-analysis. The four iron homeostasis biomarkers (ferritin, serum iron, TIBC, and TSAT) were each rank-based inverse normal transformed to a standard normal distribution (separately for each sex) and adjusted for age using a generalized additive model. In addition, for the UK cohort, the variants were adjusted for principal components. Phased SNPs by mass index, smoking levels, alcohol levels, and iron supplementation status. For each sequence variant, the mixed model implemented in the software BOLT-LMM v2.3,30, using the genotype as an additive covariate and the transformed quantitative trait as a response, was used to test for association with quantitative traits. Logistic regression was used for association between traits and case-control phenotypes, using software developed at deCODE genetics79.

We used LD score regression to account for distribution inflation in the dataset due to cryptic relatedness and population stratification84. LD score regression intercepts were as follows: ferritin: 1.032 (SE = 0.011), iron: 1.016 (SE = 0.025), TSAT: 1.029 (SE = 0.039), TIBC: 1.005 (SE = 0.030). This method is robust to test for association between sequence variants and binary traits, repressing trait status against expected genotype count. In the Icelandic data, we adjusted for sex, age, and county of birth by including these variables in the logistic regression model. In the UK and Danish data we adjusted for sex and age, as well as principal components in order to adjust for population stratification.

Results from the Icelandic, UK, and Danish datasets were combined using a fixed-effect inverse-variance weighted meta-analysis, where different datasets were allowed to have different population frequencies for alleles and genotypes but assumed to have a common effect. Heterogeneity in effect estimates was assessed for all associations using Cochran Q (P < 0.05). Nominal P-values were corrected for multiple testing using the Bonferroni correction than the commonly used threshold of 5 × 10−8 (which would not control FWER at 0.05) given that 40 million markers were tested while being more powerful than simply correcting for 40 million tests using a fixed threshold of 0.05%

10−8.

Several studies have used gene-gene interaction analyses to explore the genetic basis of iron homeostasis. Despite the high prevalence of iron deficiency and iron overload disorders, the genetic basis of iron homeostasis is not fully understood. The identification of genetic variants associated with iron homeostasis could provide insights into the pathophysiology of iron deficiency and iron overload diseases and potentially identify new therapeutic targets.

### Variant-to-gene mapping

To predict the most likely causal gene for each variant we used an algorithm taking into account the gene location with regard to LD class (defined as all variants with r2 > 0.8 with the lead variant), the variant effect for coding variants, and the effect on gene expression (eQTL, restricting to the top cis-eQTL). The algorithm, called variant-to-gene mapping, considers all genes within the LD class ± 250 kb and outputs a score for each gene. The GWAS variant is not causal itself but in LD with the causal variant. To identify the likely causal gene, we defined all variants in linkage disequilibrium (r2 > 0.8) with the GWAS variant as the LD class. We assumed local effects, where genes overlapping the LD class interval receive a distance score of 5, while genes within 250 kb on each side of the LD class interval receive a distance score of 1. The score for each gene was calculated based on the gene density into account. In cases where more than one gene share the maximum score (for example, if the LD class contains four genes and they all have probability ≥ 0.25), we chose the gene with the most

**Study subjects from the UK.** The INTERVAL study is a prospective cohort study of approximately 45,000 blood donors, representative of the wider donor population, recruited from 1993 to 2000. Participants, aged 18 years and older, were recruited before 2012 and 2014 from 25 National Health Service Blood and Transplant static donor centers in England. All participants provided written, informed consent, and the study was approved by the Cambridge (East) Research Ethics Committee (ref: 11/EE/0155).

Ferritin measurement was based on the immunological agglutination principle with the enhancement of the reaction by latex. Particles coated with anti-ferritin antibodies agglutinated with ferritin and the precipitate was determined turbidimetrically at 570/800 nm. Serum iron was measured using a colorimetric method (FerroZine) without deproteinization. Under acidic conditions, iron was liberated from transferrin. A cobalt reduction with Ferrozine to form a colored complex. The color intensity is directly proportional to the iron concentration and was measured photometrically. TIBC was calculated by summing up serum iron and unsaturated iron-binding capacity, which was also measured photometrically. TSAT was calculated by dividing serum iron by TIBC concentration. All data points lying more than 4.5 interquartile range from the median were considered outliers and removed (591 for ferritin, 7 for transferrin, 65 for TSAT, and 37 for serum iron).

The genotyping protocol and quality control procedures for INTERVAL study samples have been described in detail previously73. Briefly, DNA extracted from buffy coat was used to assay approximately 820,000 variants and short insertions/deletions on the Affymetrix Axiom Biobank genotyping array (Affymetrix, Santa Clara, California, US). Genotyping was performed in multiple batches of approximately 4800 samples each. Sample QC was performed including exclusions for genotyping failure, low call rate, duplicate samples, extreme heterozygosity, and non-European descent. We used our high-resolution multiple imputations using a joint UK10K and 1 and 1000 Genomes Phase 3 (May 2013 release) reference panel and retained variants with a MAF ≥ 0.1% and/or INFO score ≥ 40 for analysis.

The meta-analyses of hemorrhagic stroke, liver fibrosis/cirrhosis, liver cancer, type 2 diabetes, osteoarthritis, impotence, cardiomyopathy, osteoporosis, hyperpigmentation, and anemia (Supplementary Data 4) include data from the UK Biobank, accessed under Application Number 56270.

**Study subjects from Denmark.** The Danish Blood Donor Study (DBDS), initiated in 2010 as a collaborative blood donor-oriented and generic research platform and is an on-going nation-wide prospective cohort with inclusion sites at all Danish blood collection facilities. Currently, more than 110,000 blood donors are participating, and more than 95% of invited blood donors are willing to participate2. Due to the step-wise roll-out of DBDS, an enrichment of individuals from the greater Copenhagen region (the capital) and the central region of Jutland (the second largest city) are present in this study. DBDS has secured necessary permissions and approval from the Danish Data Protection Agency (2007-58-0015), the Scientific Ethical Committee system (M-20090237), and the Ethical Committee, Copenhagen region (the capital) and the central region of Jutland (the second largest city) are present in this study. DBDS has secured necessary permissions and approval from the Danish Data Protection Agency (2007-58-0015) and the Scientific Ethical Committee system (M-20090237). Briefly, regarding the DBDS genotyping, DNA extraction was performed by Origen (Genetech Solutions, Rochester, NY, USA), and for 28,075 individuals using Abbott Architect i2000SR (Abbott Laboratories, Abbott Park, IL, USA), including 27 individuals that had genotypes called using joint calling with Graphtyper78. In total, 155,250 Icelanders, including 27 individuals that had genotypes called using joint calling with Graphtyper78. In total, 155,250 Icelanders, as well as the subsequent imputation, has been described in recent publications20,21. In summary, we sequenced the whole genomes of 28,075 Icelanders using whole-genome technology to a mean depth of at least 10X (median 32X). Single-nucleotide polymorphism (SNPs) and indels were identified and their genotypes called using joint calling with Graphyper28. In total, 155,250 Icelanders were genotyped using Illumina SNP chips and their genotypes were phased using long-range phasing22. All sequenced individuals were also chip-typed and long-range phasing performed to validate the information about genotypes that had been previously used to improve genotype calls. Genotypes of the 32 million high-quality sequence variants were imputed into all chip-typed Icelanders. Variants in the Icelandic and Danish cohorts were imputed based on the IMPUTE HHMM model29 as previously described30.

Variants in INTERVAL were imputed using the Sanger Imputation Server (https://imputation.sanger.ac.uk) which implements the Burrows–Wheeler transform imputation algorithm PBWT on whole chromosome. A combined UK10K and the 1000 Genomes Phase 3 reference panel was used32. Using genealogic information, the sequence variants were also imputed into relatives of the chip-typed further increasing the sample size for association analysis, allowing us to assess associations. The imputed variants had imputation information over 0.8. The GWAS from Denmark was performed using 19 million markers identified through whole-genome sequencing of 2816 Danes that were subsequently imputed into 84,386 chip-typed individuals. The GWAS from the UK was performed with 19 million markers from the UK10K and 1000 Genomes Phase 3 reference panel, imputed into 43,059 chip-typed individuals participating in the INTERVAL study. In total, 40 million markers were tested in the meta-analysis.

**10 COMMUNICATIONS BIOLOGY | https://doi.org/10.1038/s42003-020-01575-z | www.nature.com/commsbio**
significant eQTL if such information existed, otherwise the gene closest to the lead variant was selected. Relative values for the scoring for high- and moderate-impact values were based on enrichment analysis, as previously described, while the score of 50 for eQTL was determined in order to make coding and eQTL equally informative overall. Values for proximity were set to have some degree of preference for closeby genes, given otherwise equal evidence, while at the same time giving stronger weight to coding and eQTL than to proximity alone. Data sources for eQTL data are listed in Supplementary Data 18.

Genetic overlap with other traits. We calculated genetic correlations between pairs of traits using the cross-trait LD score regression method, in our meta-analysis using summary statistics from traits in the Icelandic and UK datasets. We used results for about 1.2 million variants, well imputed in both datasets and for LD information we used precomputed LD scores for European populations (available from the Broad Institute).

Heritability estimation. Heritability was estimated in the following two ways: (1) 2 × parent–offspring correlation, (2) 2 × full sibling correlation, using the Icelandic data (where all family relationships are known).

Polygenic risk scores. We generated PRS for ferritin and TSAT and regressed the scores, along with sex, year of birth, and 20 principal components as covariates in logistic regression models against 11 clinical manifestations of iron deficiency on iron overload (restless legs, hemochromatosis, liver fibrosis/cirrhosis, liver cancer, type 2 diabetes, osteoarthritis, impotence, cardiomyopathy, osteoporosis, hyperpigmentation, and amenorrhea). Scores are based on a framework of 620,000 high-quality SNPs covering the whole genome, adjusted for LD using LDpred. The methods used to generate the PRS have been previously described in detail. For restless legs, the phenotype data is from Iceland while for the other ten phenotypes, data is from UK Biobank (restless legs were not available in UK Biobank). To minimize bias and/or overfitting, the geographical population with the phenotype data is not included when generating the scores. Thus, for ferritin, the PRS for restless legs is based on a Denmark + UK GWAS meta-analysis, while the PRS for the other phenotypes is based on Iceland + Denmark GWAS meta-analyses. For TSAT, the PRS for restless legs is based on the UK GWAS, while the PRS for the other phenotypes is based on the Icelandic GWAS.

Protein measurements (pQTL). During 2000–2019, we collected plasma samples from 40,004 Icelanders. Fifty-two percent of the samples were collected as part of the Icelandic Cancer Project, while the remaining samples (48%) were collected as part of the Icelandic population WGS data have been deposited at the European Variant Archive under accession code PRJEB15197. The authors declare that the data supporting the findings of this study are available within the article, its Supplementary Data files, and upon request. Overall meta-analysis summary statistics have been shared at https://www.decode.com/summarydata/. The UK Biobank data can be obtained upon application (ukbiobank.ac.uk). For this study, UK Biobank data was under project number 56270.

Code availability

We used publicly available software (URLs listed below) in conjunction with algorithms described in the Methods section: IBA (https://github.com/ibis/ibis); GenomeAnalysisTK (http://www.broadinstitute.org/gatk/); Picard tools 1.117 (https://broadinstitute.github.io/picard/); SAMtools 1.3 (http://samtools.github.io/); Bedtools v2.25.0-76-gc7e696e (https://github.com/arq5x/bedtools2/); Variant Effect Predictor (https://github.com/Ensembl/ensembl-vep/).

Received: 19 May 2020; Accepted: 7 December 2020; Published online: 03 February 2021

References

1. Muckenthaler, M. U., Rivella, S., Hentze, M. W. & Galy, B. A red carpet for iron metabolism. Cell 168, 344–361 (2017).
2. Vos, T. et al. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 390, 1211–1259 (2017).
3. Pilling, L. C. et al. Common conditions associated with hereditary haemochromatosis genetic variants: cohort study in UK Biobank. Br. Med. J. 364, k2229 (2019).
4. Adams, P. C. Epidemiology and diagnostic testing for haemochromatosis and iron overload. Int. J. Lab. Hematol. 37, 25–30 (2015).
5. Camaschella, C. Iron-deficiency anemia. N. Engl. J. Med. 372, 1832–1843 (2015).
6. Benyamin, B. et al. Novel loci affecting iron homeostasis and their effects in individuals at risk for haemochromatosis. Nat. Commun. 5, 4926 (2014).
7. Liao, M. et al. Genome-wide association study identifies variants in PMI2 associated with serum ferritin in a Chinese population. PLoS ONE 9, e105844 (2014).
8. Rafiuddin, M. L. et al. Genome-wide association study of iron traits and relation to diabetes in the Hispanic Community Health Study/Study of Latinos (HCHS/SOL): potential genomic intersection of iron and glucose regulation? Hum. Mol. Genet. 26, 1966–1978 (2017).
9. Stenvinkel, B. et al. Weighting sequence variants based on their annotation increases power of whole-genome association studies. Nat. Genet. 48, 314–317 (2016).
10. Sirotkmany, G., Donko, A. & Geiszt, M. Nox/Duox family of NADPH oxidases: lessons from knockout mouse models. Trends Pharmacol. Sci. 37, 318–327 (2016).
11. Rushton, D. H. & Barth, J. H. What is the evidence for gender differences in ferritin and haemoglobin? Crit. Rev. Oncol. Hematol. 73, 1–9 (2010).
12. Freitas, S. D. S. et al. Genetic variants of VWF gene in type 2 von Willebrand disease. Haemophilia 25, e78–e85 (2019).
13. Wang, C. Y. & Babitt, J. L. Liver iron sensing and body iron homeostasis. Hepatology 133, 18–29 (2019).
14. Yu, J. et al. TMTMR4 attenuates transforming growth factor beta (TGFbeta) signaling by dephosphorylating R-Smads in endosomes. J. Biol. Chem. 285, 8454–8462 (2010).
15. Chen, S. et al. Transforming growth factor beta (TGF-beta1) activates hepcidin mRNA expression in hepatocytes. J. Biol. Chem. 291, 13160–13174 (2016).
16. Stoller, J. K. & Aboussouan, L. S. A review of alpha(1)-antitrypsin deficiency. Am. J. Respir. Crit. Care Med. 185, 246–259 (2012).
17. Elzouki, A. N., Hultcrantz, R., Stal, P., Befrits, R. & Eriksson, S. Increased PIG gene frequency for alpha 1 antitrypsin in patients with genetic haemochromatosis. Gut 36, 922–926 (1995).
18. Schaefer, B. et al. Impaired hepcidin expression in alpha-1-antitrypsin deficiency associated with iron overload and progressive liver disease. Hum. Mol. Genet. 24, 6254–6263 (2015).
19. Nemeth, E. et al. IL-6 mediates hypoferremia of inflammation by inducing the synthesis of the iron regulatory hormone hepcidin. J. Clin. Investig. 113, 1271–1276 (2004).
20. Verga Falzacappa, M. V. et al. STA3 mediates hepatic hepcidin expression and its inflammatory stimulation. Blood 109, 353–358 (2007).
21. Weiss, G., Ganz, T. & Goodnough, L. T. Anemia of inflammation. Blood 133, 40–50 (2019).
22. Kim, K. et al. High-density genotyping of immune loci in Koreans and Europeans identifies eight new rheumatoid arthritis risk loci. Ann. Rheum. Dis. 74, e13 (2015).
23. Okada, Y. et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 506, 376 (2014).
24. Abella, V. et al. Leptin in the interplay of inflammation, metabolism and immune system disorders. Nat. Rev. Rheumatol. 13, 100–109 (2017).
25. Keesstra-Gounder, A. M. & Tools, R. M. NOD1 and NOD2: beyond peptidoglycan sensing. Trends Immunol. 38, 758–767 (2017).
26. Chen, Q. et al. Multiple functions of Il10ra in hematological malignancies, solid tumor and autoimmune diseases. Gene 684, 47–52 (2019).
27. Schwartz, A. J. et al. Hepatic hepcidin/intestinal HIF-2alpha axis maintains iron absorption during iron deficiency and overload. J. Clin. Investig. 129, 336–348 (2019).
28. Place, T. L. & Domann, F. E. Prolyl-hydroxylase 3: evolving roles for an ancient signaling protein. Hypoxia 2013, 13–17 (2013).
29. Appelhoff, R. J. et al. Differential function of the prolyl hydroxylases PHD1, PHD2, and PHD3 in the regulation of hypoxia-inducible factor. J. Biol. Chem. 279, 38458–38465 (2004).
30. Taniguchi, C. M. et al. Cross-talk between hypoxia and insulin signaling through Phd3 regulates hepatic glucose and lipid metabolism and ameliorates diabetes. Nat. Med. 19, 1325 (2013).
31. Ha, E. M., Oh, C. T., Bae, Y. S. & Lee, W. J. A direct role for dual oxidase in Drosophila gut immunity. Science 310, 847–850 (2005).
32. Creaser, B., El-Zaatari, M., Dang, D. T. & Merchant, J. L. Dual oxidase control release of hydrogen peroxide by the gastric epithelium to prevent Helicobacter felis infection and inflammation in mice. Gastroenterology 145, 1045–1054 (2013).
33. Hudak, L., Jaraisy, A., Haj, S. & Muhsen, A. An updated systematic review and meta-analysis on the association between Helicobacter pylori infection and iron deficiency anemia. *Helicobacter* **22**, https://doi.org/10.1111/hel.12330 (2017).

34. Drakesmith, H., Nemeth, E. & Ganz, T. Iron out ferroportin. *Cell Metab.* **22**, 777–787 (2015).

35. Kzhzhikowska, J., Gratchev, A. & Goerdert, S. Stabilin-1, a homeostatic scavenger receptor with multiple functions. *J. Cell. Mol. Med.* **10**, 635–649 (2006).

36. Lee, S. J., Park, S. Y., Jung, M. Y., Bae, S. M. & Kim, I. L. System mechanism for phosphatidylserine-dependent erythrocytosis in mouse liver. *Blood* **117**, 5213–5223 (2011).

37. Park, S. Y. et al. Stabilin-1 mediates phosphatidylserine-dependent clearance of cell corpses in alternatively activated macrophages. *J. Cell Sci.* **122**, 3365–3373 (2009).

38. Satoh, T. et al. Critical role of Trib1 in differentiation of tissue-resident M2-like macrophages. *Nature* **495**, 524 (2013).

39. Aziz, A., Soucie, E., Sarrazin, S. & Sieweke, M. H. MafB/c-Maf deficiency enables self-renewal of differentiated functional macrophages. *Science* **326**, 867–871 (2009).

40. Kautz, L. et al. Identification of erythроferrone as an erythroid regulator of iron metabolism. *Nat. Genet.* **46**, 678–684 (2014).

41. Troedec, M. B. et al. Targeted deletion of the mouse Mitoferrin 1 gene: anemia from protoporphyrin. *Blood* **117**, 5494–5502 (2011).

42. Annes, J., Sridhar, S. & Gooden, J. H. Erythроferrone inhibits the induction of hepcidin by BMP6. *Blood* **132**, 1473–1477 (2018).

43. Danjou, F. et al. Genome-wide association analyses based on whole-genome sequencing in Sardinia provide insights into regulation of hemoglobin levels. *Nat. Genet.* **47**, 1264 (2015).

44. Thein, S. L. et al. Intergenic variants of HBS1L-MYB are responsible for a significant proportion of Hb F-elevations in the baboon by prolonged erythropoietic stress. *Blood* **116**, 428–430 (2010).

45. Zappa, L., Malpica, A., Zecca, L. & Iadecola, C. Iron deficiency increases brain ferritin content in human and mouse brain. *J. Neurosci.* **36**, 13610–13620 (2016).

46. Heeney, M. M. & Finberg, K. E. Iron-refractory iron deficiency anemia (IRIDA). *Hematol. Oncol. Clin. North Am.* **28**, 637–652 (2014).

47. Rigas, A., Pedersen, O. B., Erikstrup, C. & Ullum, H. Iron deficiency among blood donors: experience from the Danish Blood Donor Study and from the Copenhagen ferritin monitoring scheme. *Transfus. Med.* **29**, 23–27 (2019).

48. Kell, D. B. & Pretorius, E. Serum ferritin is an important inflammatory disease marker, as it is mainly a leakage product from damaged cells. *Metallomics* **6**, 748–773 (2014).

49. McCann, J. C. & Ames, B. N. An overview of evidence for a causal relation between iron deficiency during development and deficits in cognitive or behavioral function. *Am. J. Clin. Nutr.* **85**, 931–945 (2007).

50. Andreo, M., Le Squere, P., Estivin, S. & Gentric, A. Anemia and cognitive performances in the elderly: a systematic review. *Eur. J. Neurol.* **20**, 1234–1240 (2013).

51. Guo, S. Y. et al. Restless legs syndrome: from pathophysiology to clinical diagnosis and management. *Pront. Aging Neurosci.* **9**, 171 (2017).

52. Ward, R. J., Ziucka, F. A., Dynn, J. H., Crichton, R. R. & Zecca, L. The role of iron in brain ageing and neurodegenerative disorders. *Lancet Neurol.* **13**, 1045–1060 (2014).

53. Allen, K. J. et al. Iron-overload-related disease in HFE hereditary hemochromatosis. *N. Engl. J. Med.* **358**, 221–230 (2008).

54. Kell, D. B. & Pretorius, E. Serum ferritin is an important inflammatory disease marker, as it is mainly a leakage product from damaged cells. *Metallomics* **6**, 748–773 (2014).

55. Menzel, S. et al. A QTL in chromosome 6q23 influencing fetal hemoglobin levels in adults. *Proc. Natl Acad. Sci. USA* **104**, 11346–11351 (2007).

56. Desimone, J., Biel, M. & Heller, P. Maintenance of fetal hemoglobin (HbF) elevations in the baboon by prolonged erythropoietic stress. *Blood* **60**, 519–523 (1982).

57. Lavelle, D., Desimone, J., Heller, P., Zwiers, D. & Hall, L. On the mechanism of Hb F-elevations in the baboon by erythropoietic stress and pharmacological manipulation. *Blood* **67**, 1083–1089 (1986).

58. Bianchi, M. & Magnani, M. Hexokinase mutations that produce nonphosphoryc hemolytic-anemia. *Blood Cells Mol. Dis.* **21**, 2–8 (1995).

59. Camaschella, C. Iron deficiency. *Blood* **133**, 30–39 (2014).

60. Van Cott, E. M., Khor, B. & Zehnder, J. L. Factor V Leiden. *Am. J. Hematol.* **91**, 46–49 (2016).

61. Kassebaum, N. J. et al. A systematic analysis of global anemia burden from 1990 to 2010. *Blood* **123**, 615–624 (2014).

62. McLaren, C. E. et al. Distribution of transferrin saturation in an Australian population: relevance to the early diagnosis of haemochromatosis. *Gastroenterology* **114**, 543–549 (1998).

63. Kanai, M. et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–398 (2018).

64. Menzel, S. et al. A QTL influencing F-cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. *Nat. Genet.* **39**, 1197–1199 (2007).

65. Barton, J. C. & Adams, P. C. Clinical guidelines: HFE hemochromatosis-screening, diagnosis and management. *Nat. Rev. Gastroenterol. Hepatol.* **7**, 482–484 (2010).

66. O’Keefe, S. T., Gavin, K. & Lavan, N. Iron status and restless legs syndrome in the elderly. *Ageing Age Mech.* **23**, 200–203 (1994).

67. Sun, E. R., Chen, C. A., Ho, G., Earley, C. J. & Allen, R. P. Iron and the restless legs syndrome. *Sleep* **21**, 371–377 (1998).

68. Allen, R. P. et al. Evidence-based and consensus clinical practice guidelines for the iron treatment of restless legs syndrome/Wills-Ekbo disease in adults and children: an IRLSSG task force report. *Sleep Med.* **41**, 27–44 (2018).

69. Bonferroni, C. & Duca, P. Two new human DMT1 gene mutations in a patient with microcytic anemia, low ferritinemia, and liver iron overload. *Blood* **107**, 4168–4170 (2006).

70. Iolascon, A. et al. Microcytic anemia and hepatic iron overload in a child with compound heterozygous mutations in DMT1 (SLC11A2). *Blood* **107**, 349–354 (2006).

71. Mifsud, M. P. et al. Identification of a human mutation of DMT1 in a patient with microcytic anemia and iron overload. *Blood* **105**, 1337–1342 (2005).

72. Yanatori, I. & Kishi, F. DMT1 and iron transport. *Free Radic. Biol. Med.* **133**, 53–63 (2019).

73. Hubert, N. & Hentze, M. W. Previously uncharacterized isoforms of divalent metal transporter (DMT)-1: implications for regulation and cellular function. *Proc. Natl Acad. Sci. USA* **99**, 12345–12350 (2002).

74. Hentze, M. W., Muckenthaler, M. U. & Andrews, N. C. Balancing acts: molecular control of mammalian iron metabolism. *Cell* **117**, 285–297 (2004).
Danish Administrative Regions

The Danish Administrative Regions; The Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome. The Health and Social Care Research and Development Division (Welsh Government), 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. The Novo Nordisk Foundation (NNF14CC0001 and NNF17OC0027594). The Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No. 115881 (RHAP-SODY).

The authors thank all the blood banks in Denmark for both collecting and contributing data to this study. Danish Blood Donor Research Fund. Aarhus University, Copenhagen University Hospital banks in Denmark for both collecting and contributing data to this study. Medical Research Council Biostatistics Unit, Cambridge Institute of Public Health, Cambridge, UK. Department of Communications Biology, University of Cambridge, Cambridge, UK. 2British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. 3Department of Clinical Immunology, Copenhagen University Hospital, Copenhagen, Denmark. 4deCODE genetics/Amgen Inc., Reykjavik, Iceland. 5Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland. 6Department of Infectious Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK. 7Medical Research Council Biostatistics Unit, Cambridge Institute of Public Health, Cambridge, UK. 8Department of Human Genetics, Wellcome Sanger Institute, Hinxton, Cambridge, UK. 9Department of Clinical Immunology, Aarhus University Hospital, Aarhus, Denmark. 10Department of Cardiology, Division Heart & Lungs, University Medical Center Utrecht, Utrecht, The Netherlands. 11Rutherford Fund Fellow, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. 12Department of Clinical Immunology, Odense University Hospital, Odense, Denmark. 13The Laboratory in Mjöd, RAM, Reykjavik, Iceland. 14Department of Clinical Immunology, Aarhus University Hospital, Aarhus, Denmark. 15Department of Cardiology, Division Heart & Lungs, University Medical Center Utrecht, Utrecht, The Netherlands. 16Department of Laboratory Hematology, Landspitali, the National University Hospital of Iceland, Reykjavik, Iceland. 17Department of Clinical Immunology, Aalborg University Hospital, Aalborg, Denmark. 18Department of Clinical Biochemistry, Landspitali, the National University Hospital of Iceland, Reykjavik, Iceland. 19Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark. 20Danish Headache Center, Department of Neurology, Rigshospitalet-Glostrup, Glostrup, Denmark. 21Institute of Biological Psychiatry, Copenhagen University Hospital MHC Scit. Hans, Roskilde, Denmark. 22Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark. 23Department of Clinical Biochemistry, Akureyri Hospital, Akureyri, Iceland. 24Translational Disease Systems Biology, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. 25Department of Haematology, University of Cambridge, Cambridge, UK. 26UK National Health Service Blood and Transplant, Cambridge Biomedical Campus, Cambridge, UK. 27Radcliffe Department of Medicine and National Health Service Blood and Transplant, John Radcliffe Hospital, Oxford, UK. 28UK National Health Service Blood and Transplant, John Radcliffe Hospital, Oxford OX3 9QX, UK. 29School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland. 30These authors contributed equally: Steven Bell, Andreas S. Rigas, Magnus K. Magnusson, Egil Feringstad, Elisa Allara. 31These authors jointly supervised this work: Emanuele Di Angelantonio, Henrik Ullum, Kari Stefansson. 32A list of authors and their affiliations appears at the end of the paper. 33email: magnus.magnusson@decode.is; ed303@medschl.cam.ac.uk; Henrik.Ullum@regionh.dk; kstefans@decode.is

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s42003-020-01575-z.

Correspondence and requests for materials should be addressed to M.K.M., E.F.A, H.U. or K.S.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2021

1The National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics at the University of Cambridge, University of Cambridge, Cambridge, UK. 2British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. 3Department of Clinical Immunology, Copenhagen University Hospital, Copenhagen, Denmark. 4deCODE genetics/Amgen Inc., Reykjavik, Iceland. 5Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland. 6Department of Infectious Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK. 7Medical Research Council Biostatistics Unit, Cambridge Institute of Public Health, Cambridge, UK. 8Department of Human Genetics, Wellcome Sanger Institute, Hinxton, Cambridge, UK. 9Department of Clinical Immunology, Aarhus University Hospital, Aarhus, Denmark. 10Department of Cardiology, Division Heart & Lungs, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands. 11Rutherford Fund Fellow, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. 12Department of Clinical Immunology, Odense University Hospital, Odense, Denmark. 13The Laboratory in Mjöd, RAM, Reykjavik, Iceland. 14Department of Clinical Immunology, Næstved Hospital, Næstved, Denmark. 15The Laboratory in Mjöd, RAM, Reykjavik, Iceland. 16Department of Clinical Immunology, Aalborg University Hospital, Aalborg, Denmark. 17Department of Clinical Biochemistry, Landspitali, the National University Hospital of Iceland, Reykjavik, Iceland. 18Department of Clinical Immunology, Aalborg University Hospital, Aalborg, Denmark. 19Department of Clinical Biochemistry, Landspitali, the National University Hospital of Iceland, Reykjavik, Iceland. 20Danish Headache Center, Department of Neurology, Rigshospitalet-Glostrup, Glostrup, Denmark. 21Institute of Biological Psychiatry, Copenhagen University Hospital MHC Scit. Hans, Roskilde, Denmark. 22Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark. 23Department of Clinical Biochemistry, Akureyri Hospital, Akureyri, Iceland. 24Translational Disease Systems Biology, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. 25Department of Haematology, University of Cambridge, Cambridge, UK. 26UK National Health Service Blood and Transplant, Cambridge Biomedical Campus, Cambridge, UK. 27Radcliffe Department of Medicine and National Health Service Blood and Transplant, John Radcliffe Hospital, Oxford, UK. 28UK National Health Service Blood and Transplant, John Radcliffe Hospital, Oxford OX3 9QX, UK. 29School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland. 30These authors contributed equally: Steven Bell, Andreas S. Rigas, Magnus K. Magnusson, Egil Feringstad, Elisa Allara. 31These authors jointly supervised this work: Emanuele Di Angelantonio, Henrik Ullum, Kari Stefansson. 32A list of authors and their affiliations appears at the end of the paper. 33email: magnus.magnusson@decode.is; ed303@medschl.cam.ac.uk; Henrik.Ullum@regionh.dk; kstefans@decode.is

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

© The Author(s) 2021
DBDS Genomic Consortium

**Denmark** Steffen Andersen30, Karina Banasik24, Søren Brunak24, Kristoffer Burgdorf3, Christian Erikstrup9, Thomas F. Hansen19,20,21, Henrik Hjalgrim18, Gregor Jemec31, Poul Jennum32, Pär Johansson3, Kasper R. Nielsen16, Mette Nyegaard33, Helene M. Paarup14, Ole B. Pedersen12, Mikkel Petersen9, Erik Sørensen3, Henrik Ullum3,36,38 & Thomas Werge34

**Iceland** Daniel F. Gudbjartsson4,29, Kari Stefansson4,5,36, Hreinn Stefánsson4 & Unnur Thorsteinsdóttir4,5

30Department of Finance, Copenhagen Business School, Copenhagen, Denmark. 31Department of Clinical Medicine, Sealand University Hospital, Roskilde, Denmark. 32Department of Clinical Neurophysiology at University of Copenhagen, Copenhagen, Denmark. 33Department of Biomedicine, Aarhus University, Aarhus, Denmark. 34Institute of Biological Psychiatry, Mental Health Centre Sct. Hans, Copenhagen University Hospital, Roskilde, Denmark.