Supplementary Information for The role of genetically predicted serum iron levels on neurodegenerative and cardiovascular traits

Supplementary Methods

GWAS summary statistics for meta-analysis of serum iron levels.

Baltimore Longitudinal Study on Aging (BLSA)

The Baltimore Longitudinal Study on Aging (BLSA) study is a population-based study in the Baltimore-Washington DC area aim to characterized important factors that contribute to healthy aging.[1] This is an open enrollment cohort that began in 1958, with approximately 1100 active participants. Blood samples were collected for DNA extraction, and genome-wide genotyping was completed for 1231 subjects using Illumina 550K. This analysis focused on a subset of the participants (n = 480) with European genetic ancestry and assessment of circulating iron. Serum iron was measured using a Fe slide method (VITROS 750, Johnson & Johnson). The BLSA is currently supported by the Intramural Research Program of the NIH, National Institute on Aging. The study protocol was approved by the Internal Review Board of the Intramural Research Program of the National Institutes of Health and all participants provided written informed consent.

Genotype and Imputation: The analysis was restricted to subjects with European ancestry and each analysis was further adjusted for the top two principal components derived from an EIGENSTRAT analysis utilizing ~10,000 randomly selected SNPs from the 550K SNP panel.[2] Genotyping was completed for 848 participants of European genetic ancestry (49.2%, n = 236 female) using a call rate of >98.5% without sex discrepancy based on X-chromosome heterozygosity. 501,704 autosomal SNPs passed quality control (completeness $\geq$ 99%, MAF $\geq$ 1%, HWE $\geq 10^{-4}$) were used to for imputation. Imputation to the Haplotype Reference Consortium (HRC) panel was conducted using the University of Michigan Imputation Server with HRC panel v1.1 reference panel.[3] The mean age was 74.7 ± 12.6 years.

Benyamin et al. 2014 meta-analysis

Datasets and analyses have been previously described.[4] In brief, 23,986 individuals of European genetic ancestry from a total of 11 cohorts were included in the discovery cohort for the meta-analysis or serum iron levels.

Moksnes et al. 2022 meta-analysis

Datasets and analyses have been previously described.[5] In brief, serum iron levels were meta-analyzed in three cohorts (Trøndelag Health Study – HUNT, Michigan Genomics Initiative, and SardiNIA), resulting in 236,612 individuals of European genetic ancestry included.

Montreal Heart Institute (MHI) Biobank

The MHI Biobank was initiated in 2008. It is an ongoing prospective hospital cohort with approximately 17,000 patients that have been recruited at the MHI. For all participants, data is collected via a questionnaire administered by a research nurse at study entry and

the patient questionnaire data is validated using available MHI electronic hospital records. Blood, DNA, and plasma are collected at baseline. Prospective follow-up is conducted by interview questionnaires every 4 years and using the MHI electronic hospital records. Recruitment of new participants into the MHI Biobank was put on pause in late 2019 but follow-up questionnaire and data curation activities have been maintained. Socio-demographic and lifestyle phenotypes were collected at the study baseline. Cardiovascular, non-vascular and drug-use phenotypes include all events that may have occurred prior to the baseline or during the follow-up period.

DNA samples from participants to the MHI Biobank (n = ~17,000) were genotyped on the Illumina GSAv3.0_MD bead chip at the Beaulieu Saucier Pharmacogenomics Centre. Standard quality control checks were performed, and then the genotypes were imputed with the TOPMed imputation reference panel. Autosomal and chromosome X genotyped variants and variants with an imputation score ≥ 0.6 with a minor allele frequency ≥ 0.01 (in each subgroup analyzed) were retained for analysis.

GWAS on iron levels was performed with SAIGE (version 0.44.5) using individuals of European genetic ancestry, including sex-specific analyses (including related individuals) with the following covariates: age, sex (excluded for sex-specific analyses) and 10 principal components of genetic ancestry. 9,125,531, 9,127,895 and 6,390,858 variants were tested for association in the sex-combined, male-only and female-only analyses. Additionally, hospital clinic and assay were included as fixed effects covariates using dummy coding. Iron values came from one entry for each participant, where the lab value that was the closest to the baseline date was selected. Iron levels from 1286 males and 803 females were tested for association. Mean iron average levels were 13.03 ± 6.91 umol/L (range 1.4 to 70.0) in females and 13.43 ± 6.91 umol/L (range 1.0 to 55.6) in males. The mean age was 63.27 ± 13.19 years in females and 66.25 ± 10.10 years in males.

Michigan Genomics Initiative (MGI) Biobank

The Michigan Genomic Initiative (MGI) biobank has been previously described.[6] Briefly, MGI participants are recruited primarily through surgical encounters at Michigan Medicine. This analysis used genetic and electronic health records for 70,439 individuals (freeze 5). Each participant is genotyped for ~570K genetic variants via genome-wide genotyping arrays and then imputed to the Trans-Omics for Precision Medicine (TOPMed) reference panel. Post-imputation quality control was performed where variants with a poor imputation score (< 0.3) and very low minor allele frequency (<0.01%) were excluded. We restricted the analysis to MGI participants of inferred European ancestry based on estimated European global ancestry fraction > 0.9 using ADMIXTURE[7] (n = 54,076).

In this European genetic ancestry cohort, 12,527 MGI participants had at least one iron measurement (LOINC code 2498-4) (7,287 females and 5,240 males). MGI participants included in this analysis had an average age of 59.94 ±15.18 years (range: 19-88).

GWAS were performed on rank-based inverse normalized transformation of the mean iron levels to overcome potential issues of reduced power and inflated type I error due to

violation of the normality assumption of the outcome. GWAS was performed using the SAIGE software with the following covariates: age, age-squared, sex, the age by sex interaction term, the first 20 principal components of genetic ancestry and the genotyping chip.[8] In sex-specific analysis, the same linear mixed model was specified in SAIGE with the same covariates other than sex and the age by sex interaction term.

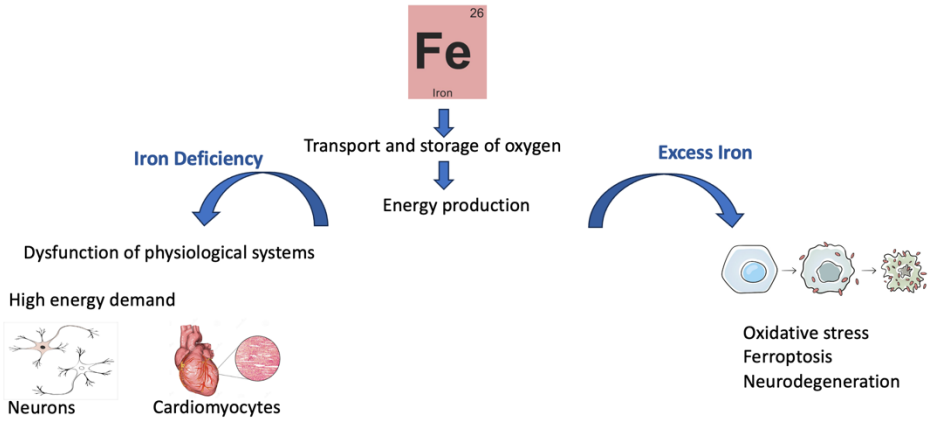
SardiNIA

The SardiNIA study has been previously described.[9] Briefly, it is a large population-based study which consists of 6,921 individuals, males and females, aged 14 to 102 years, characterized for several quantitative traits and medical conditions, and representing >60% of the adult population of four villages in the Lanusei Valley of Sardinia. Among them, 6,023 samples (2564 males and 3459 females, with mean age 43.66 ± 17.59 years (range 14 to 101.3)), were used in the serum iron association analysis.

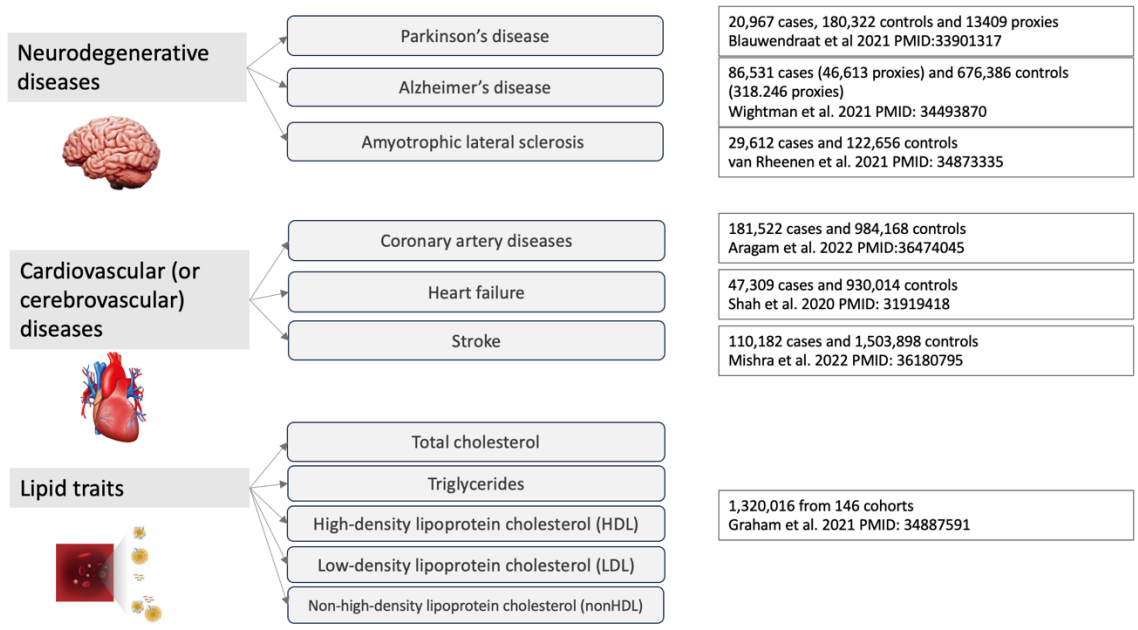
All samples have been genotyped with four Illumina arrays (OmniExpress, ImmunoChip, Cardio-MetaboChip and ExomeChip). Imputation was performed on a genome-wide scale using a Sardinian sequence-based reference panel of 3,514 individuals and the software Minimac5 on pre-phased genotypes, as previously described.[10,11] After imputation, variants with an imputation score > 0.3 were retained for those with a minor allele frequency ≥ 0.01. Variants with an imputation score > 0.6 were retained for those with a minor allele frequency < 0.01 as per the standard SardiNIA-cohort thresholds. About 22 million variants (20,143,392 SNPs and 1,688,858 indels) have been included in the analyses.[10]

Sex-specific GWAS was run on the autosomes using the q.emmax option in EPACTS-v3.2.6 (which conducts association testing adjusted for familiar relationships using a genetic variant-based kinship matrix), including related individuals.[12] Specifically, the inverse normal transformations of the residuals for serum iron were tested for association with the following covariates: age and age-squared.

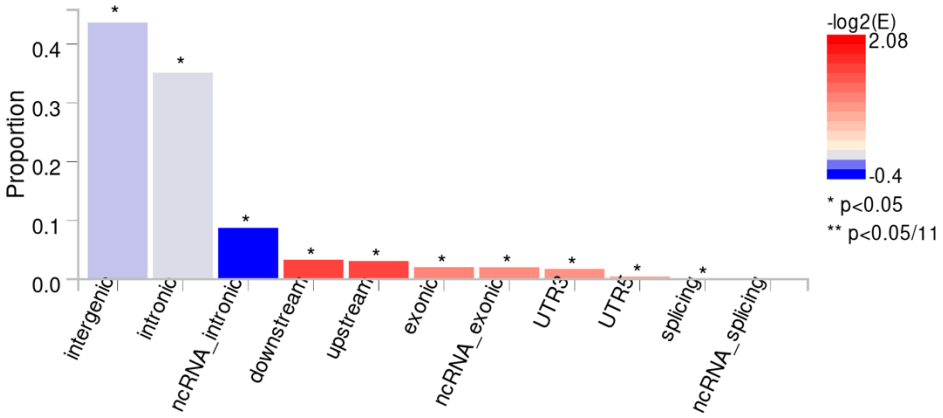
Supplementary Figures



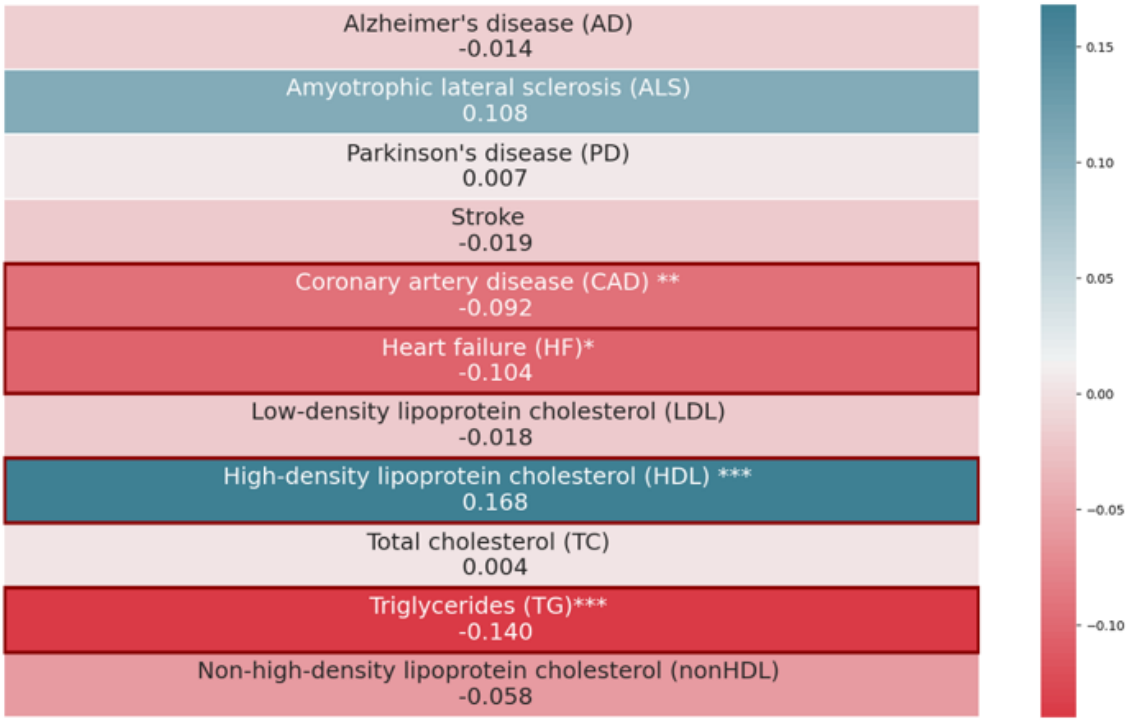
Supplementary Figure 1. Schematic overview of the impact of iron deficiency or overload on physiological systems.



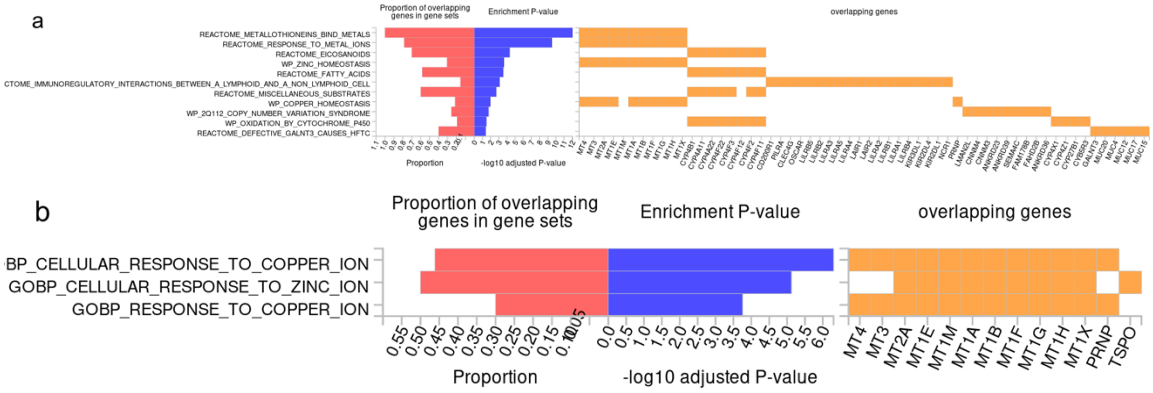
Supplementary Figure 2. Overview of eleven neurodegenerative, cardiovascular, cerebrovascular and lipid GWAS summary statistics datasets for analyses.



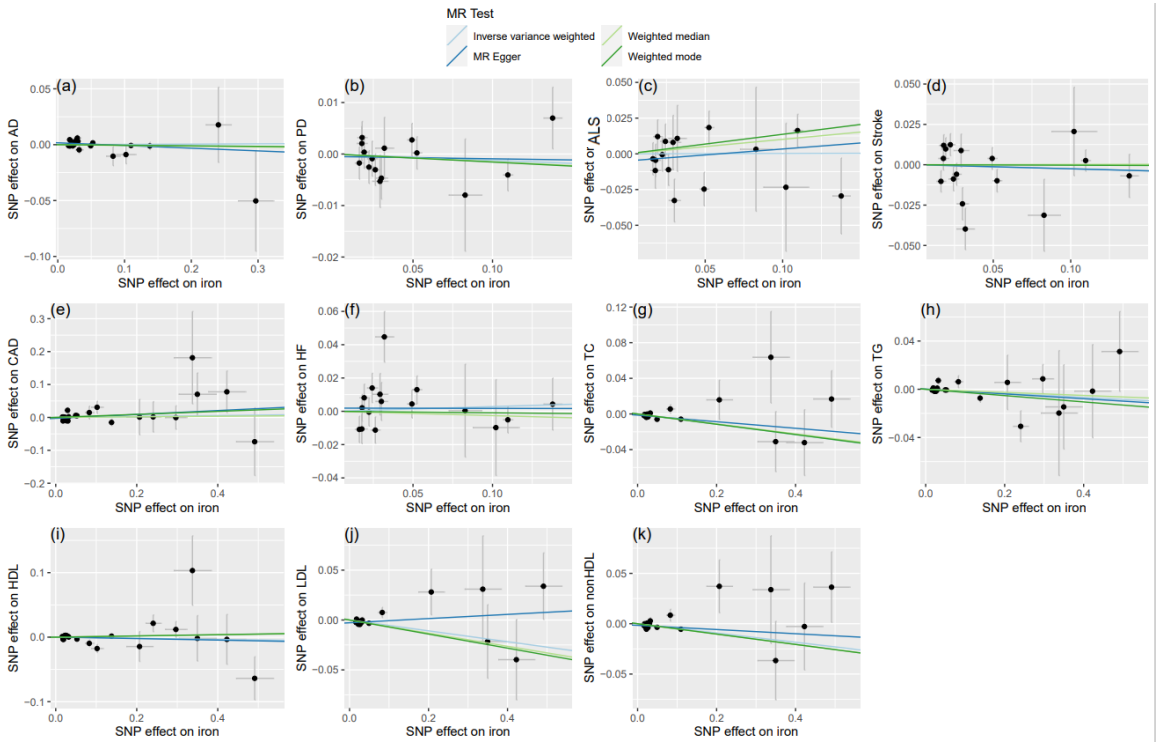
Supplementary Figure 3. Proportion of all variants in LD with independent significant ($p<5x10^{-8}$) variants with corresponding functional annotation in ANNOVAR as performed by FUMA. Colours represent log2(enrichment) relative to variants in the 1000 Genomes European super-population.



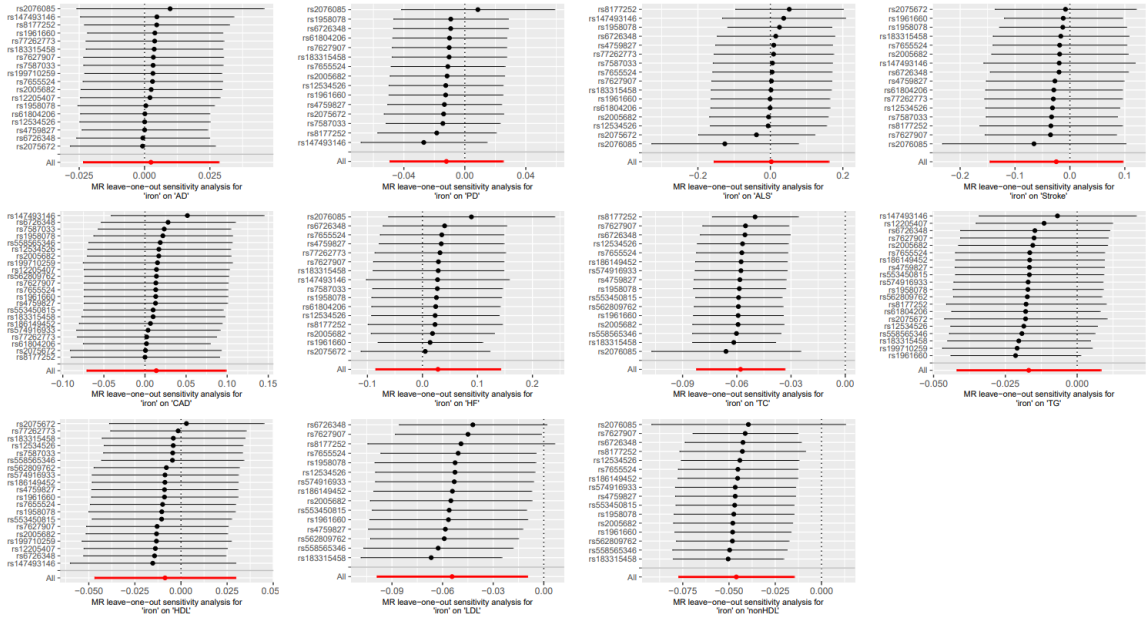
Supplementary Figure 4. Global genetic correlation results between iron and 11 neurodegenerative, cardiovascular, cerebrovascular or lipid traits. Red boxes mark the statistically significant results, * denotes p < 0.05; ** denotes p < 0.005; *** denotes p < 1e-5.



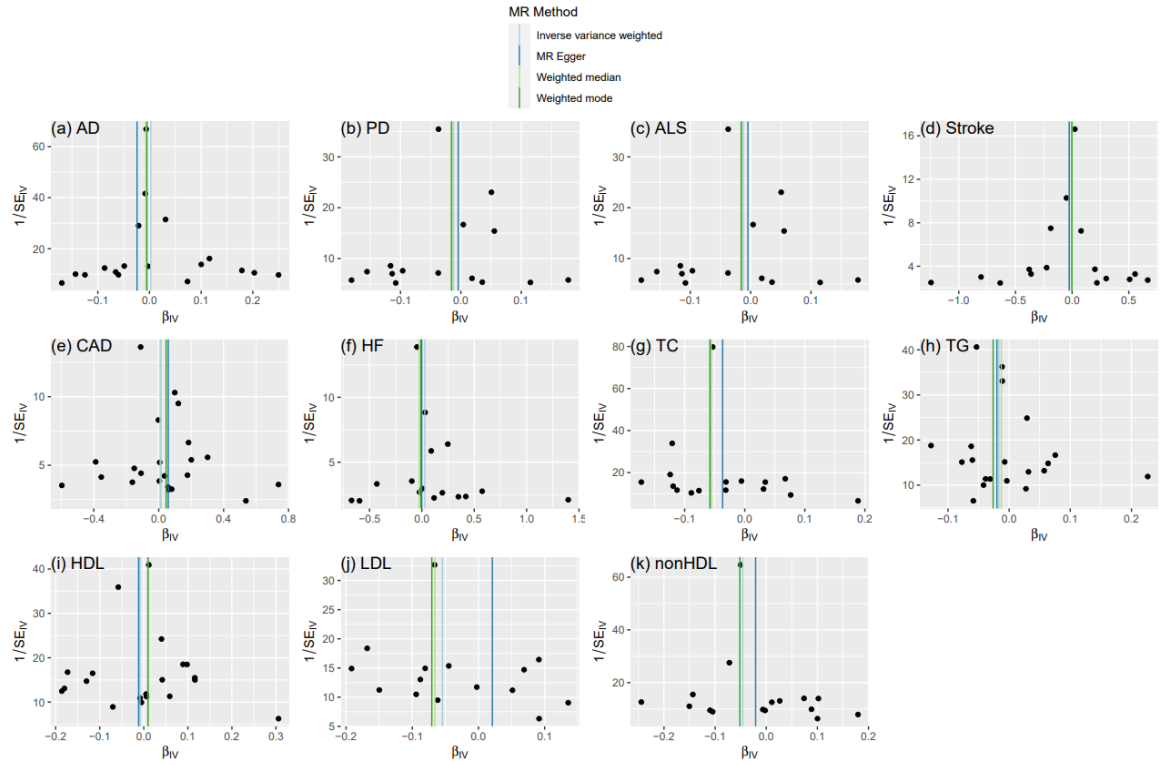
Supplementary Figure 5. **Gene-set enrichment analysis conducted in FUMA for protein-coding genes within the 106 loci exhibiting significant local genetic correlation between iron levels and Alzheimer's disease**. Results shown for all canonical pathways (MsigDB c2) and Gene Ontology (GO) biological processes (b) gene-sets. All protein coding genes were used as the background gene set, Bonferroni correction was applied, and a minimum of 5 genes overlapping with a gene-set was required.



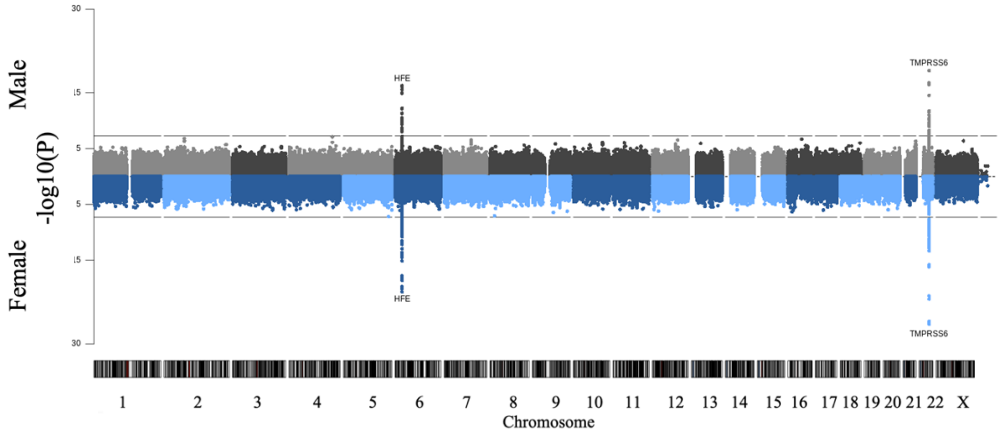
Supplementary Figure 6. Scatterplots displaying Mendelian randomization results with iron as the exposure and the 11 tested outcomes: Alzheimer's disease (a), Parkinson's disease (b), amyotrophic lateral sclerosis (ALS) (c), stroke (d), coronary artery disease (e), heart failure (f), total cholesterol (g), triglycerides (h) high density lipoprotein cholesterol (i), low density lipoprotein cholesterol (j) and non-high density lipoprotein cholesterol (k).



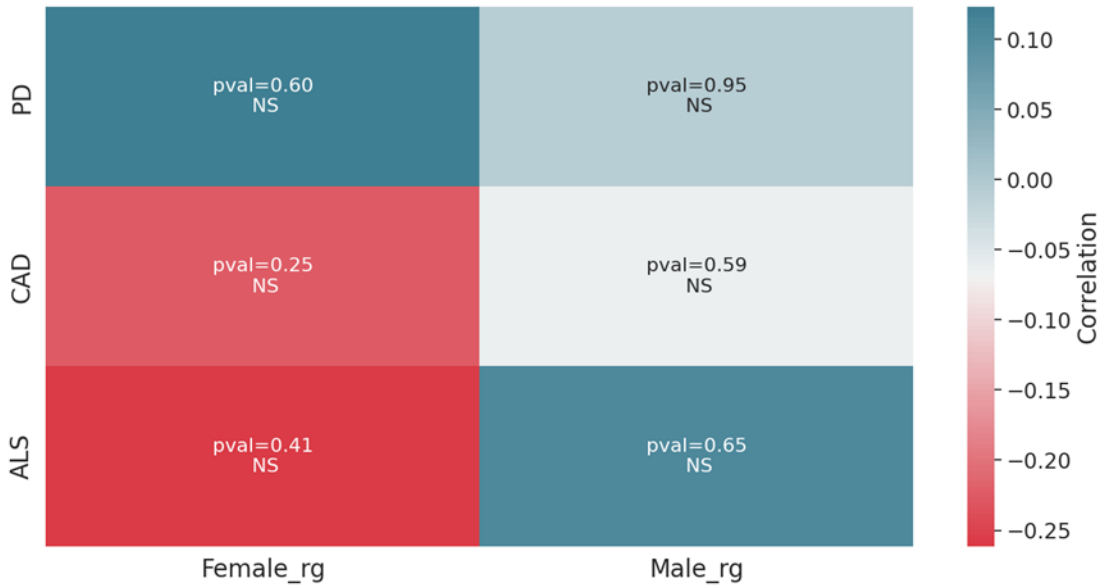
Supplementary Figure 7. Leave-one-out plots displaying Mendelian randomization results with iron as the exposure and the 11 tested outcomes: Alzheimer's disease (a), Parkinson's disease (b), amyotrophic lateral sclerosis (ALS) (c), stroke (d), coronary artery disease (e), heart failure (f), total cholesterol (g), triglycerides (h), high density lipoprotein cholesterol (i), low density lipoprotein cholesterol (j) and non-high density lipoprotein cholesterol (k).



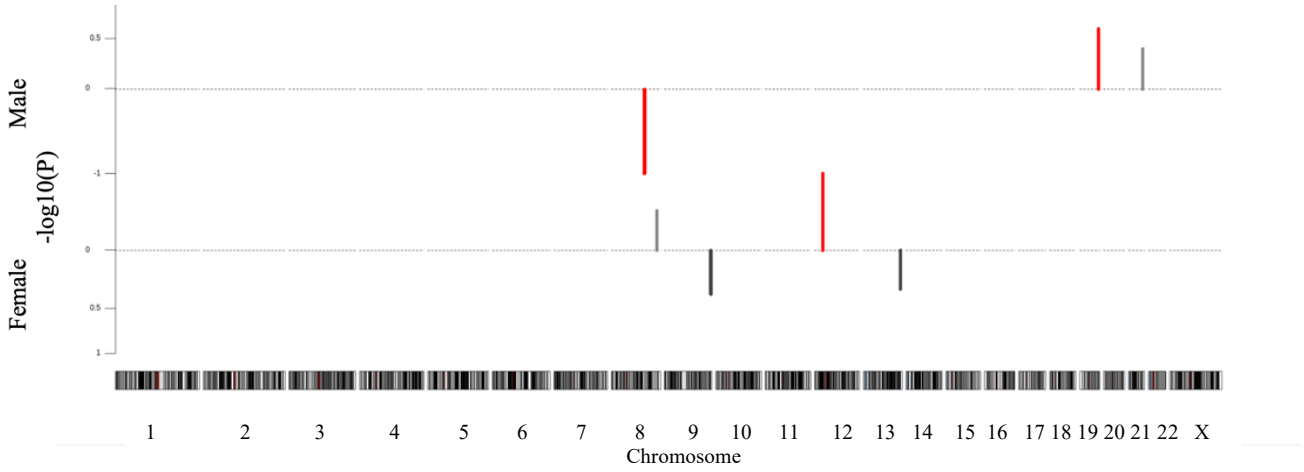
Supplementary Figure 8. Funnel plots displaying Mendelian randomization results with iron as the exposure and the 11 tested outcomes: Alzheimer's disease (a), Parkinson's disease (b), amyotrophic lateral sclerosis (ALS) (c), stroke (d), coronary artery disease (e), heart failure (f), total cholesterol (g), triglycerides (h), high density lipoprotein cholesterol (i), low density lipoprotein cholesterol (j) and non-high density lipoprotein cholesterol (k).



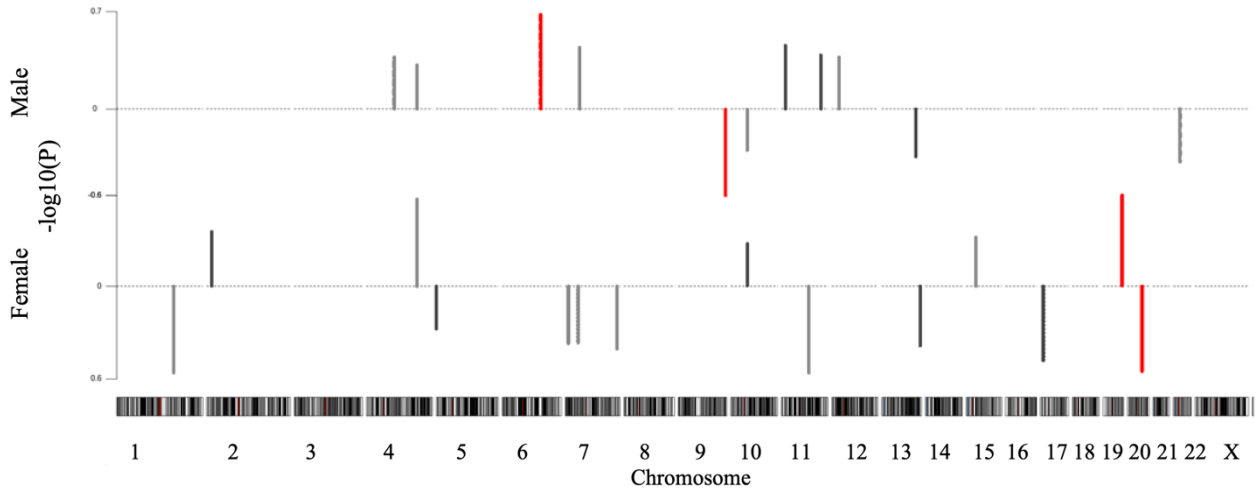
Supplementary Figure 9. Miami plot of sex-stratified GWAS meta-analysis of serum iron using a sample-size weighted z-score method resulting in a total sample size of 9 090 males (top) and 11 547 females (bottom) of European-like genetic ancestry. The dashed line represents the threshold to denote genome-wide significance ($5x10^{-8}$).



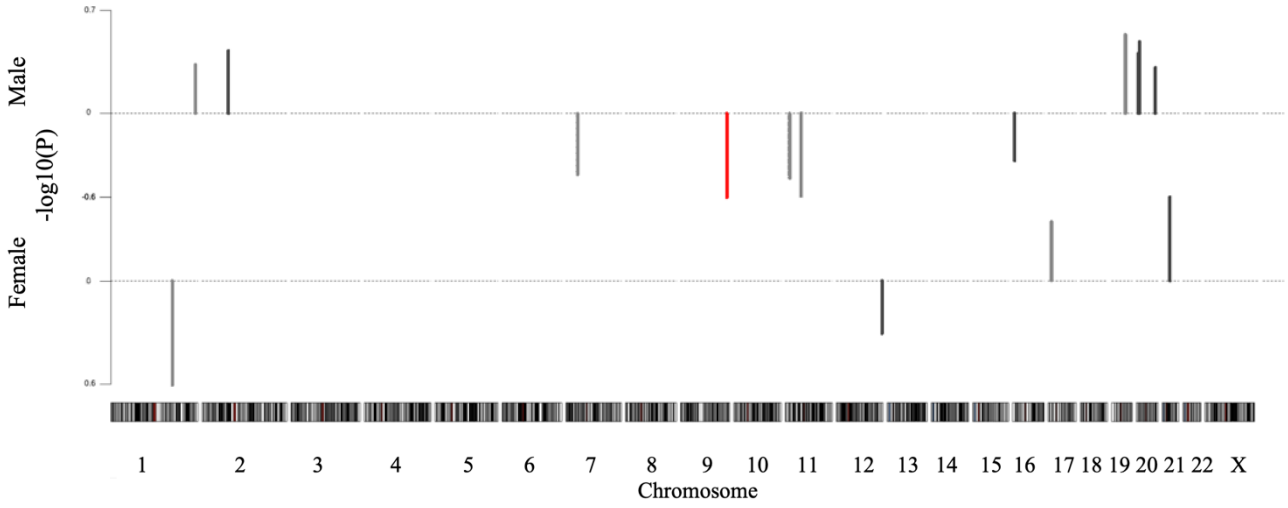
Supplementary Figure 10. Sex-stratified global genetic correlation results between iron levels and Parkinson's disease (PD), coronary artery disease (CAD) and amyotrophic lateral sclerosis (ALS).



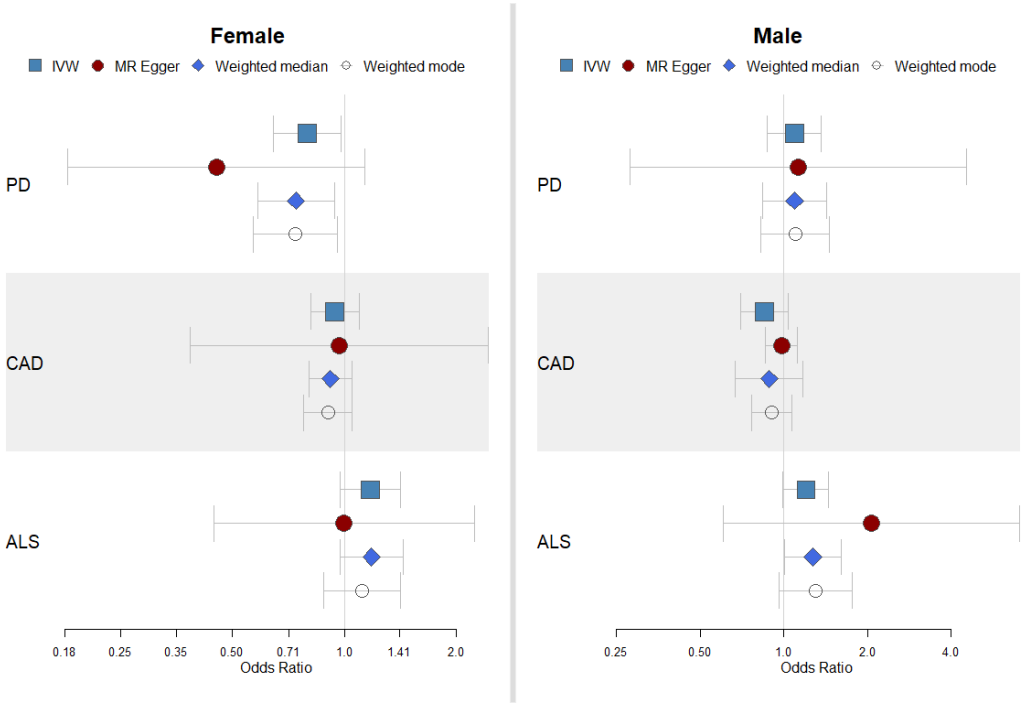
Supplementary Figure 11. Sex-stratified local genetic correlation results between iron levels and Parkinson's disease. Statistically significant regional correlations after Bonferroni correction are presented in red. Positive and negative scales are used to differentiate between positive and negative correlations, respectively.



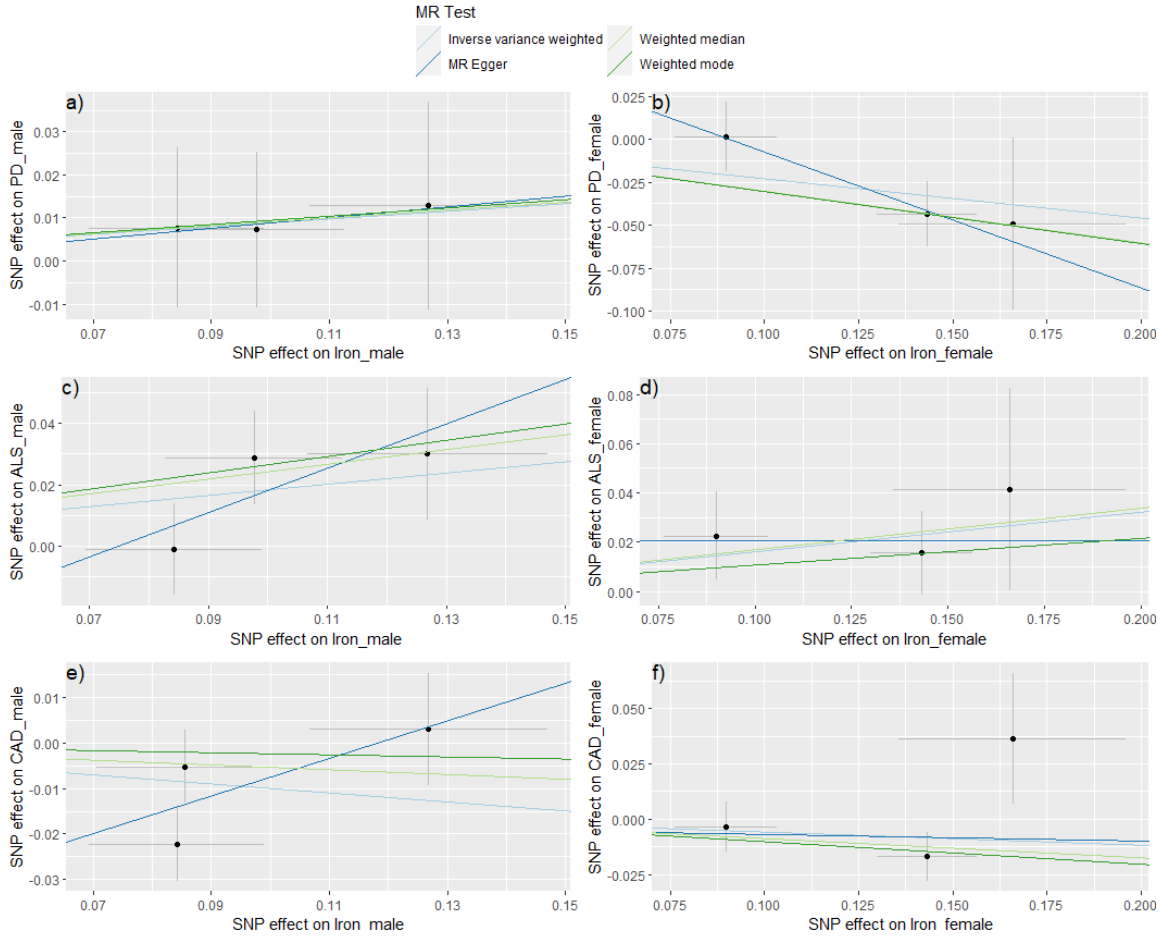
Supplementary Figure 12. Sex-stratified local genetic correlation results between iron levels and coronary artery disease. Statistically significant regional correlations after Bonferroni correction are presented in red. Positive and negative scales are used to differentiate between positive and negative correlations, respectively.



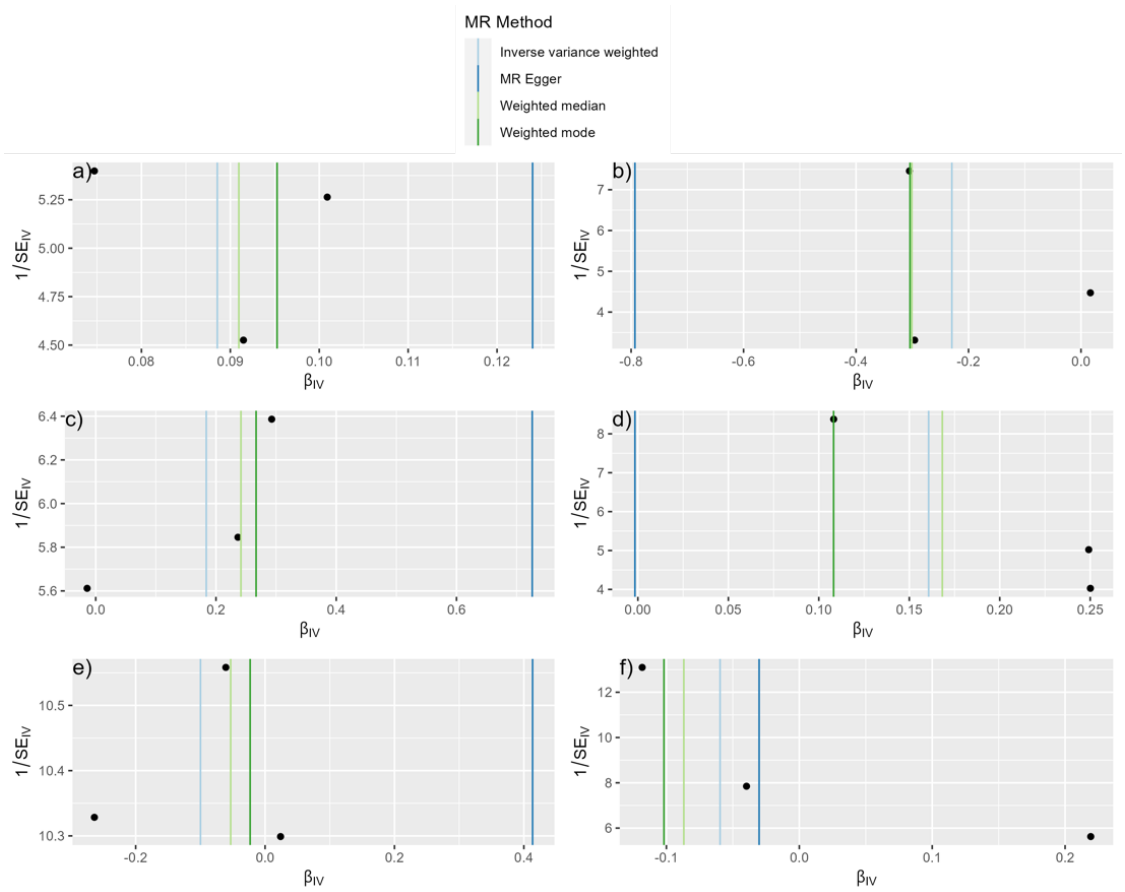
Supplementary Figure 13. Sex-stratified local genetic correlation results between iron levels and amyotrophic lateral sclerosis. Statistically significant regional correlations after Bonferroni correction are presented in red. Positive and negative scales are used to differentiate between positive and negative correlations, respectively.



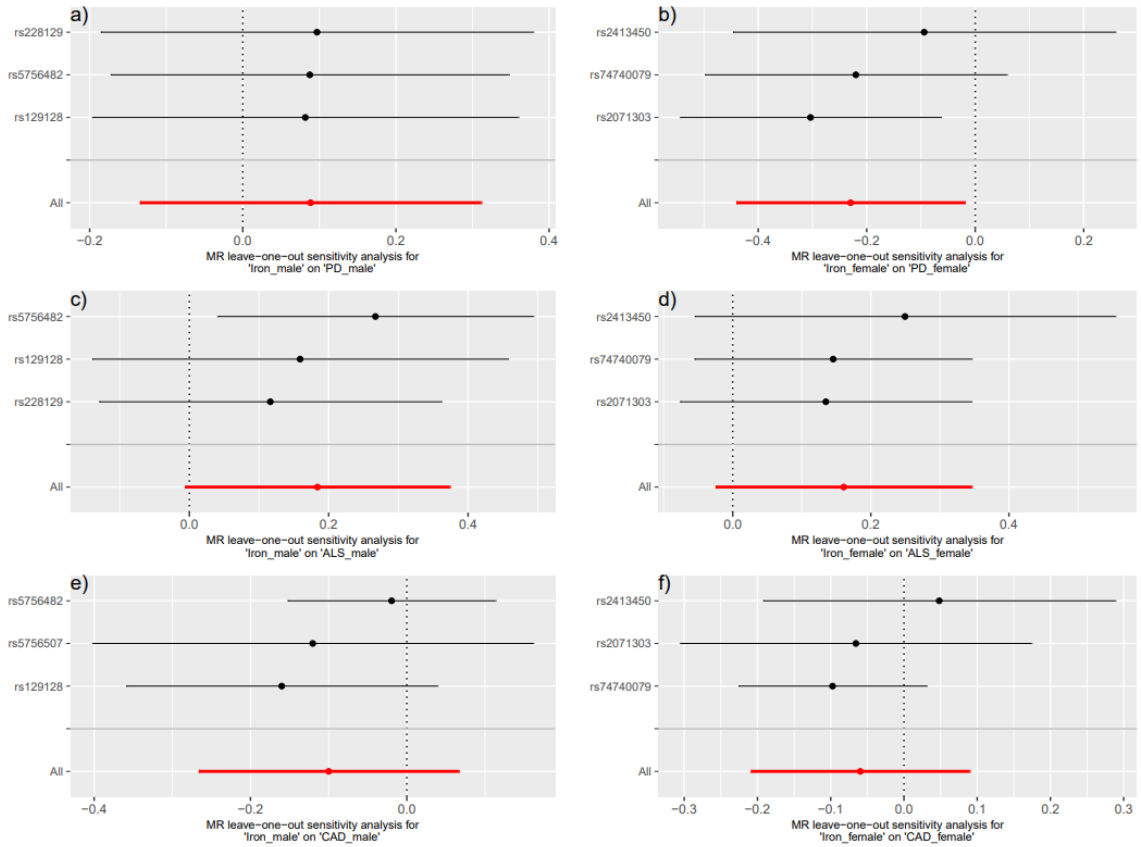
Supplementary Figure 14. Sex-specific Mendelian randomization results. Analyses conducted in males (a) and females (b) with iron as the exposure and either Parkinson's disease (PD), coronary artery disease (CAD) or amyotrophic lateral sclerosis (ALS) as the outcome.



Supplementary Figure 15. Scatter plots displaying sex-specific Mendelian randomization results with iron as exposure and the 11 tested outcome: (a) Parkinson's disease for males (b) Parkinson's disease for females (PD), (c) coronary artery disease for males (CAD), (d) CAD for females, (e) amyotrophic lateral sclerosis (ALS), for males (f) ALS for females.



Supplementary Figure 16. Funnel plots displaying sex-specific Mendelian randomization results with iron as exposure and the tested outcomes: (a) Parkinson's disease for males (b) Parkinson's disease for females (PD), (c) coronary artery disease for males (CAD), (d) CAD for females, (e) amyotrophic lateral sclerosis (ALS), for males (f) ALS for females.



Supplementary Figure 17. Leave-one-out plots displaying sex-specific Mendelian randomization results with iron as exposure and the tested outcomes: (a) Parkinson's disease for males (b) Parkinson's disease for females (PD), (c) coronary artery disease for males (CAD), (d) CAD for females, (e) amyotrophic lateral sclerosis (ALS), for males (f) ALS for females.

Supplementary References

1. Shock, N., Greulick, R. & Andres, R. Normal Human Aging: The Baltimore Study of Aging. *NIH Publication* 84–2450 (1984).
2. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904–909 (2006).
3. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284–1287 (2016).
4. Benyamin, B. *et al.* Novel loci affecting iron homeostasis and their effects in individuals at risk for hemochromatosis. *Nat Commun* **5**, 4926 (2014).
5. Moksnes, M. R. *et al.* Genome-wide meta-analysis of iron status biomarkers and the effect of iron on all-cause mortality in HUNT. *Commun Biol* **5**, 591 (2022).
6. Zawistowski, M. *et al.* The Michigan Genomics Initiative: A biobank linking genotypes and electronic clinical records in Michigan Medicine patients. *Cell genomics* **3**, 100257 (2023).
7. Alexander, D. H. & Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**, (2011).
8. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet* **50**, 1335–1341 (2018).
9. Pilia, G. *et al.* Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet* **2**, (2006).
10. Orrù, V. *et al.* Complex genetic signatures in immune cells underlie autoimmunity and inform therapy. *Nat Genet* **52**, 1036–1045 (2020).
11. Pistis, G. *et al.* Rare variant genotype imputation with thousands of study-specific whole-genome sequences: implications for cost-effective study designs. *Eur J Hum Genet* **23**, 975–83 (2015).
12. Kang, H. M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* **42**, 348–54 (2010).