

# Whole-exome sequencing identifies protein-coding variants associated with brain iron in 29,828 individuals

Received: 12 April 2024

Accepted: 16 June 2024

Published online: 02 July 2024

 Check for updates

Weikang Gong<sup>1,2,8</sup>✉, Yan Fu<sup>3,8</sup>, Bang-Sheng Wu<sup>1,8</sup>, Jingnan Du<sup>4,8</sup>, Liu Yang<sup>1</sup>, Ya-Ru Zhang<sup>1</sup>, Shi-Dong Chen<sup>1</sup>, JuJiao Kang<sup>5,6</sup>, Ying Mao<sup>1</sup>, Qiang Dong<sup>1</sup>, Lan Tan<sup>3</sup>, Jianfeng Feng<sup>5,6,7</sup>, Wei Cheng<sup>1,5,6</sup>✉ & Jin-Tai Yu<sup>1</sup>✉

Iron plays a fundamental role in multiple brain disorders. However, the genetic underpinnings of brain iron and its implications for these disorders are still lacking. Here, we conduct an exome-wide association analysis of brain iron, measured by quantitative susceptibility mapping technique, across 26 brain regions among 26,789 UK Biobank participants. We find 36 genes linked to brain iron, with 29 not being previously reported, and 16 of them can be replicated in an independent dataset with 3,039 subjects. Many of these genes are involved in iron transport and homeostasis, such as *FTH1* and *MLX*. Several genes, while not previously connected to brain iron, are associated with iron-related brain disorders like Parkinson's (*STAB1*, *KCNA10*), Alzheimer's (*SHANK1*), and depression (*GFAP*). Mendelian randomization analysis reveals six causal relationships from regional brain iron to brain disorders, such as from the hippocampus to depression and from the substantia nigra to Parkinson's. These insights advance our understanding of the genetic architecture of brain iron and offer potential therapeutic targets for brain disorders.

Iron is crucial for numerous physiological processes, including neurotransmitter synthesis, myelin formation, DNA synthesis and mitochondrial functions, and it profoundly influences neurodevelopment, cognition, and brain outcomes<sup>1,2</sup>. Our brain maintains a precise regulation of iron homeostasis. Any disturbance in this intricate balance, whether through iron overload or deficiency, may result in the emergence of brain disorders<sup>3</sup>. For example, iron accumulation might facilitate neuronal cell death in some neurodegenerative diseases, such as Alzheimer's disease (AD)<sup>4</sup>. In addition, the substantia nigra often has excess iron in Parkinson's disease (PD), which possibly

promotes oxidative stress and neuronal damage<sup>5</sup>. Moreover, cerebral iron deficiency, linked to alterations in hippocampal glucocorticoid receptor signaling, has been implicated in inducing depression<sup>6</sup>. Given iron's important role in brain development and its connection to multiple brain disorders, understanding the genetic architecture of brain iron accumulation can provide insights into brain development and the underlying mechanisms of these disorders, allowing for designing better diagnostic and therapeutic strategies.

Quantitative susceptibility mapping (QSM) is an emerging technique that enables the non-invasive measurement of brain iron

<sup>1</sup>School of Data Science, Department of Neurology and National Center for Neurological Disorders, Huashan Hospital, State Key Laboratory of Medical Neurobiology and MOE Frontiers Center for Brain Science, Fudan University, Shanghai, China. <sup>2</sup>Centre for Functional MRI of the Brain (fMRIB), Nuffield Department of Clinical Neurosciences, Wellcome Centre for Integrative Neuroimaging, University of Oxford, Oxford OX3 9DU, UK. <sup>3</sup>Department of Neurology, Qingdao Municipal Hospital, Qingdao University, 266071 Qingdao, China. <sup>4</sup>Department of Psychology, Center for Brain Science, Harvard University, Cambridge, MA 02138, USA. <sup>5</sup>Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, 200433 Shanghai, China. <sup>6</sup>Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence, Ministry of Education, Fudan University, 200433 Shanghai, China. <sup>7</sup>Department of Computer Science, University of Warwick, Coventry, UK. <sup>8</sup>These authors contributed equally: Weikang Gong, Yan Fu, Bang-Sheng Wu, Jingnan Du.

✉ e-mail: [weikanggong@fudan.edu.cn](mailto:weikanggong@fudan.edu.cn); [wcheng@fudan.edu.cn](mailto:wcheng@fudan.edu.cn); [jintai\\_yu@fudan.edu.cn](mailto:jintai_yu@fudan.edu.cn)

levels with high spatial resolution and sensitivity<sup>7</sup>. Built upon susceptibility-weighted MRI (swMRI), QSM has been demonstrated to be more sensitive to reflect tissue iron contents both phenotypically and genetically than other swMRI-derived measures, such as T2\*, and has shown higher robustness to acquisition noises and increased reproducibility<sup>8</sup>. A high positive correlation between brain iron level and QSM has been established from postmortem studies<sup>9</sup>. Previous studies also showed that brain iron has a high heritability in multiple regions, such as the putamen, substantia nigra, and pallidum<sup>8</sup>. While genome-wide association studies (GWAS) have found several loci associated with brain iron, these findings remain constrained to a few brain regions and common genetic variations (minor allele frequency > 1%)<sup>8</sup>. Additionally, many loci identified by GWAS map to noncoding regions of the genome, posing challenges in exploring the underlying mechanism. To overcome these limitations, a powerful technique, whole-exome sequencing (WES)<sup>10</sup>, can be used to identify protein-coding variants that are associated with brain iron. A large-scale exome-wide association analysis on multiple brain regions can uncover the intricate genetic architecture of brain iron accumulation and potentially highlight neural pathways crucial to iron-related brain disorders.

In this study, we conducted the most extensive exome-wide association study (EWAS) of brain iron accumulation to date. Leveraging genetic, brain imaging and phenotypic data from 26,789 subjects in the UK Biobank dataset, we systematically identified protein-coding variants associated with brain iron and studied the relationships between iron-related genes and brain disorders and phenotypes. Specifically, this study has four major goals. Firstly, we will identify rare and common genes that are associated with brain iron accumulation across multiple brain regions covering subcortical and cerebellar structures. Secondly, we aim to explore the biological functions of the identified genes, such as the biological pathways in which they are enriched in. Thirdly, we aim to explore the relationships between brain iron-related genes and disorders, including whether regional brain iron accumulation has causal relationships to multiple brain disorders. Finally, we used phenome-wide association study (PheWAS) to identify genetic associations of brain iron-related genes with a broad set of phenotypic variables.

## Results

### An overview of data and analysis pipeline

Our study primarily used brain imaging and phenotypic and genetic data from the UK Biobank, including 26 regional QSM features extracted from the swMRI data, exome-sequencing data, and diverse phenotypes for phenome-wide association studies. In the primary analysis of EWAS, we included a total of 29,828 individuals of white British ethnicity without illness conditions of brain cancer, stroke, or dementia, aged between 40 and 69, with ~52% of them being females. Among them, 26,789 of them are in the discovery set and the remaining 3039 are in the replication set (see the “Methods” section). A summary of the demographic information is provided in Table 1. A total of 18,800 rare genes and 41,790 common variants were analyzed in this study (see the “Methods” section).

A comprehensive visualization of our analysis pipeline is provided in Fig. 1. The study was initiated with the discovery of EWAS ( $N = 26,789$ ) to identify genetic variants correlated with brain iron as measured by QSM techniques. (Fig. 1a and b). Validation and replication studies ( $N = 3039$ ) were conducted to verify the robustness of our findings (Fig. 1c). Further, we examined the functionality and organization of the identified genes (Fig. 1d) and performed Mendelian randomization (MR) analysis to investigate the potential causal relationship between brain iron and multiple brain disorders (Fig. 1e). Lastly, a PheWAS was conducted to explore a broad set of brain iron-phenotype associations (Fig. 1e).

**Table 1 | Demographic information of participants in this study**

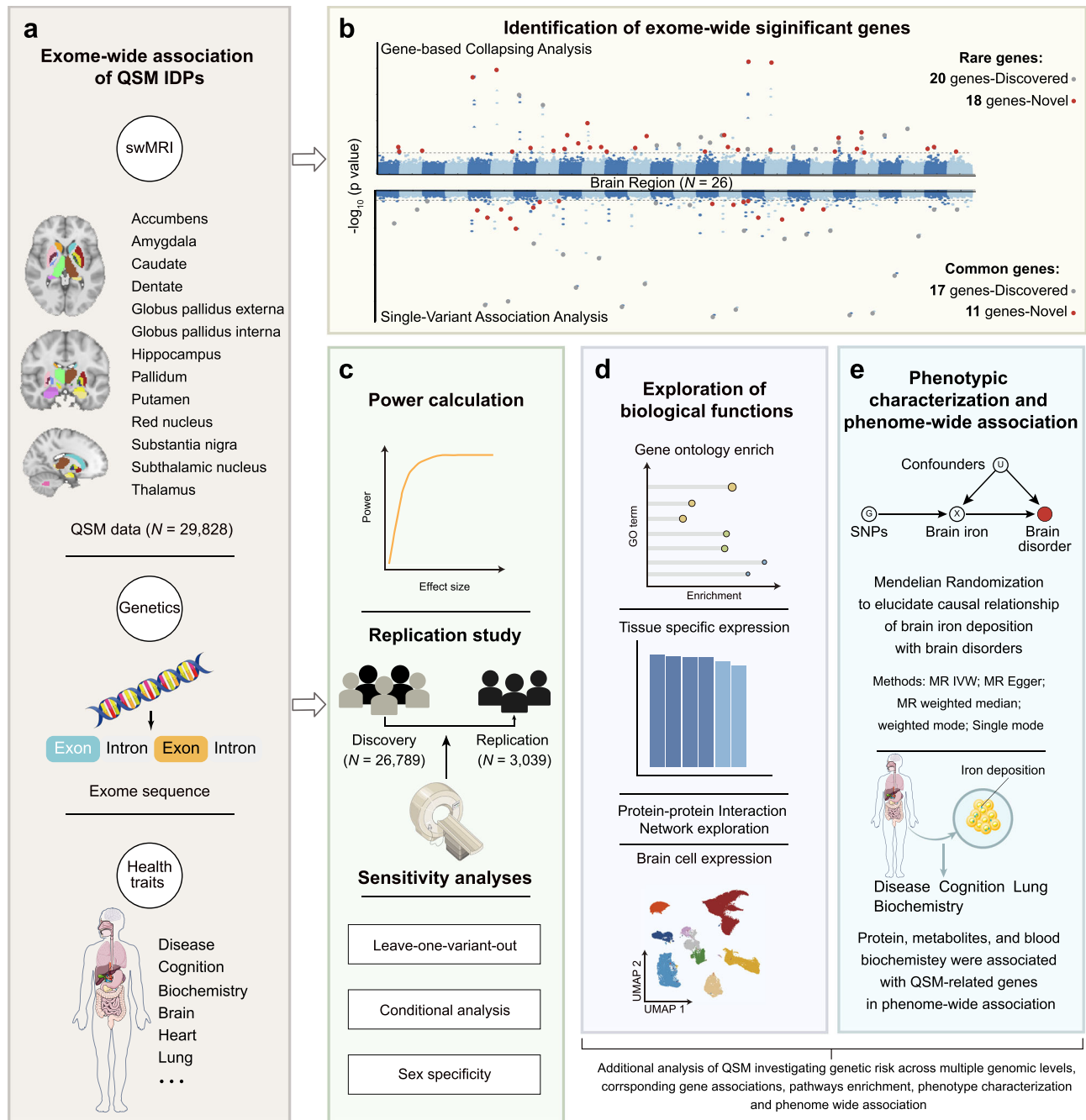
	Discovery set	Replication set
<i>N</i>	26,789	3039
Age (years; mean ± sd)	55.22 (7.43)	52.92(7.34)
Sex (female; percent)	14,041 (52.4)	1621 (53.3)
BMI	26.57 (4.17)	26.34 (4.14)
<b>Educational qualification (%)</b>		
A levels	1520 (5.7)	172 (5.7)
College	12,061 (45.2)	1404 (46.3)
CSE	726 (2.7)	88 (2.9)
None of above	1830 (6.9)	138 (4.6)
NVQ	4142 (15.5)	529 (17.5)
O levels	2966 (11.1)	332 (11.0)
Other	3451 (12.9)	368 (12.1)
<b>Smoking status (%)</b>		
Never	16,335 (61.1)	1938 (63.9)
Previous	8863 (33.1)	932 (30.7)
Current	1543 (5.8)	164 (5.4)
<b>Drinking status (%)</b>		
Never	519 (1.9)	58 (1.9)
Previous	552 (2.1)	46 (1.5)
Current	25,712 (96.0)	2935 (96.6)

### Rare protein-coding variants associated with brain iron

We conducted an exome-wide association study on 26,789 subjects to identify rare variants associated with brain iron levels across 26 brain regions. Out of the resulting associations, a total of 207 reached exome-wide significance (Bonferroni corrected  $p < 1.7 \times 10^{-8}$ , as detailed in the “Methods” section), covering 24 out of the 26 investigated brain regions (no associations were found for the left and right amygdala) (Figs. 2a and 3, Supplementary Data 1). These identified variants were mapped to 20 different protein-coding genes, with 18 of them not being previously reported in GWAS, while 2 overlapped with findings from a previous study<sup>8</sup>. The Manhattan plots and Q-Q plots for each brain region can be found in Supplementary Fig. S1. We further performed a replication study using 3039 subjects (see the “Methods” section). From the previously mentioned 20 genes, 4 are significant in both sets ( $p < 0.05$ ), encompassing 24 gene-based associations (Supplementary Data 2). This was notable considering the number of genes to replicate is just one under the null hypothesis of no associations.

We performed a power analysis of our WES analysis based on the method of a previous study<sup>11</sup>. The range of variants’ effect sizes is from 0.03 to 0.08 based on an analysis of our WES results (the regression coefficients and standard error)<sup>11</sup>. Therefore, given a sample size of  $N = 29,828$ , the estimated power is from 39.7% to 100% for the smallest effect to the largest to reach the significance level of uncorrected  $p < 1.7 \times 10^{-8}$  (Supplementary Data 3).

Furthermore, leave-one-variants out (LOVO) analysis was conducted for each of the significant associations to assess the sensitivity of our findings to analytical approaches (see the “Methods” section). The results showed that over 99.4% of the significant associations involving the above 20 significant genes remained significant in the LOVO analysis. This suggests that most significant associations arise from multiple contributing rare variants. There were a few exceptions. For instance, the association between the thalamus and *ULBP2* was influenced by single variants. The LOVO findings are cataloged in Supplementary Data 4. In addition, we conducted a conditional analysis to evaluate whether the identified rare variant signals were independent of nearby common variants (see the “Methods” section). As a result, all variants are significant, indicating that the rare variants



**Fig. 1** | **A summary of the analysis pipeline of the current study.** **a** The UK Biobank data used in the current study, including quantitative susceptibility mapping (QSM) features derived from susceptibility-weighted MRI, whole-exome sequencing data and phenotypic data. **b** Exome-wide association study (EWAS) of brain iron with rare and common variants. **c** Validation study and replication analysis of EWAS. **d** Post EWAS analysis, including gene set enrichment analysis, protein–protein interaction, tissue and single-cell expression, and **e** phenotypic-wide association study and Mendelian randomization analysis. swMRI susceptibility-weighted MRI, QSM quantitative susceptibility mapping, MR Mendelian randomization, IVW inverse variance weighted. This figure was partly generated using Servier Medical Art, provided by Servier, licensed under a Creative Commons Attribution 4.0 unported license.

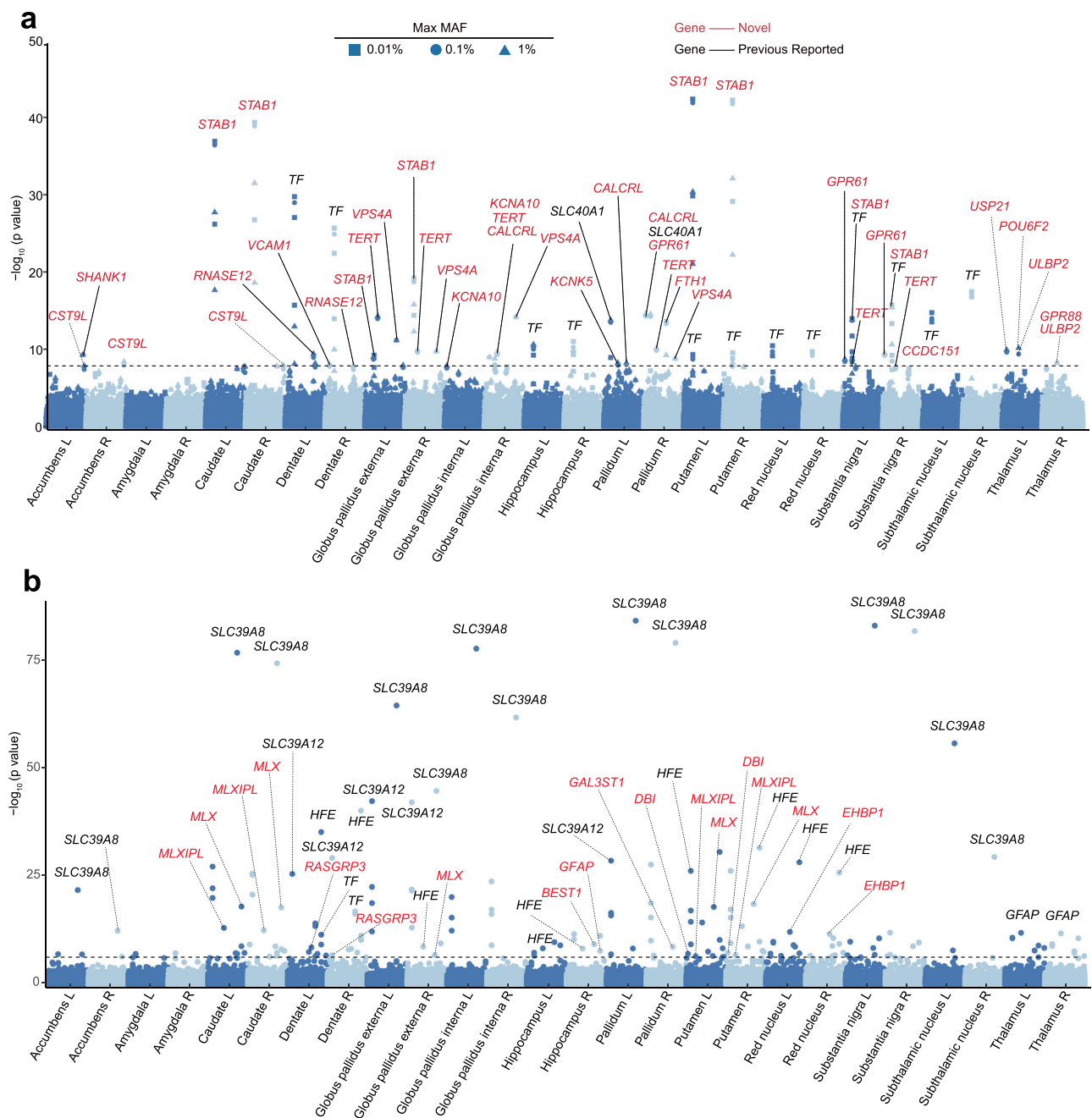
signals are independent of the nearby common variants (Supplementary Data 5). Additionally, sex-specific EWAS were executed. We found that 12 of the 20 rare genes are significant for males, and for females, 9 of the 20 rare genes are significant ( $p < 1.7 \times 10^{-8}$ ). (Supplementary Data 6).

#### Common protein-coding variants associated with brain iron

Expanding upon the above analyses, we performed an EWAS to identify common protein-coding variants associated with brain iron (see the “Methods” section). In summary, 105 associations were discovered

that pass the genome-wide significance threshold (Bonferroni-corrected  $p < 4.6 \times 10^{-8}$ , as detailed in the “Methods” section) (Figs. 2b, 3, Supplementary Data 7). The identified variants map to 17 genes, with 11 of them not being previously reported<sup>8</sup>. It is worth noting that, as a comparison to ref. 8, our analyses specifically focused on variants on the exome regions and used a broader array of brain regions of interest. Manhattan plots and Q–Q plots for each brain region are shown in Supplementary Fig. S2.

As in the rare variants analysis, we performed a replication study. Among the above 17 identified genes, 13 of them can still be found



**Fig. 2 | Exome-wide association analysis of rare and common protein-coding genes with brain iron across 26 regions. a** Significance levels of genes that are mapped from rare variants. **b** Significance levels of genes that are mapped from common variants. The  $p$ -values reported are two-sided and unadjusted. The  $x$ -axis represents the brain regions analyzed in the current study (L and R represent left

and right brain regions), and the  $y$ -axis represents the  $-\log_{10} p$ -values of each gene. The gray dashed line is the exome-wide significance threshold, based on the Bonferroni-corrected  $p < 0.05$ . Genes in red are discovered by the current study, and in black are previously reported. MAF minor allele frequency, R right, L left.

significant (Supplementary Data 8, uncorrected  $p < 0.05$ ). In expectation, the number of genes that can be replicated is one.

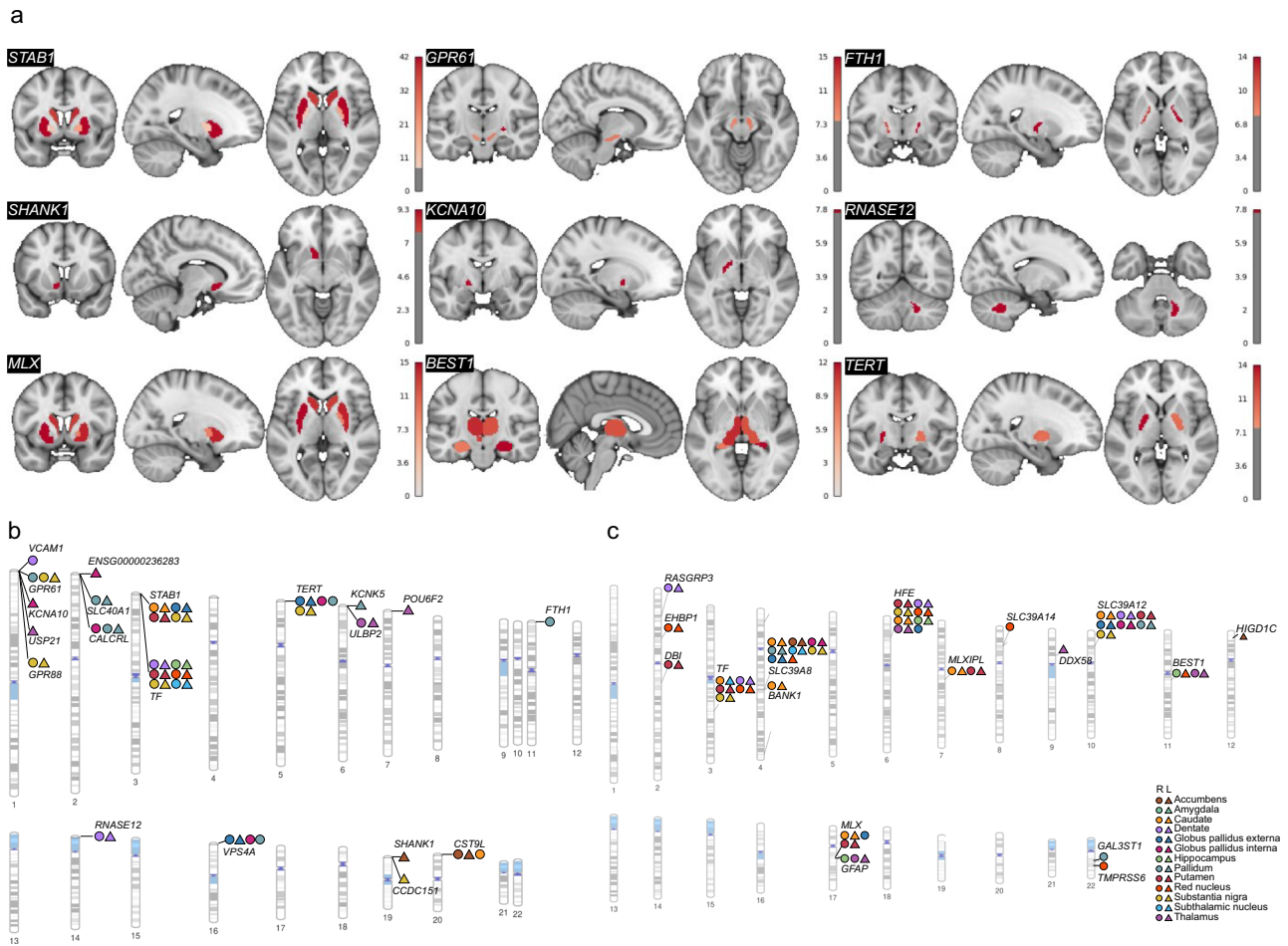
### Brain iron-associated genes are significant in existing GWAS of brain-related traits and diseases

We performed a literature search to explore how the 36 brain iron-associated genes in our analysis overlap with existing GWAS findings. We found that 3 genes are significant in AD GWAS, 4 genes are associated with cognitive-related traits, 2 genes are associated with depression-related traits, 1 gene is associated with PD, and 7 genes are reported in bipolar disorder and schizophrenia GWAS. The full list of our literature search results is shown in Supplementary Data 9.

### Functional enrichment and biological validation of brain iron-associated genes

To dive deeper into and verify the biological attributes of the identified genes, we conducted a functional enrichment analysis. Results demonstrated that brain iron-related genes are robustly enriched in iron-related functions. Notably, the pathway of intracellular iron ion homeostasis showed the highest statistical significance ( $p = 3.5 \times 10^{-12}$ ) followed by iron ion homeostasis ( $p = 3.4 \times 10^{-11}$ ) and transition metal ion transport ( $p = 7.4 \times 10^{-11}$ ) (Fig. 4a, Supplementary Data 10). Furthermore, the protein-protein map of these genes forms a dense network associated with iron, including key genes in the *SLC* families and *FTH1* (visible as brown clusters in Fig. 4b, Supplementary Data 11).





**Fig. 3 | Brain-wide and chromosome-wide mapping of brain iron-associated genes. a** Brain-wide association map of 16 example genes that are associated with brain iron. **b** Ideogram of rare genes that influence brain iron. **c** Ideogram of common genes that influence brain iron. R right, L left.

We then tested whether the identified genes were differentially expressed across various tissues, using the GTEx database (12). Notably, our genes exhibited significant differential expression in brain tissues compared to others (Fig. 4c). Notably, these included regions in our analysis, such as the substantia nigra, putamen, and cerebellum ( $p < 1.0 \times 10^{-4}$ ) (Fig. 4c, Supplementary Data 12). Leveraging single-cell RNA sequencing data of the human brain, we found that brain iron-associated genes showed higher expression levels in both excitatory and inhibitory neurons (Fig. 4d). Removing the top associated genes still kept most of the enriched pathways and differential expressed genes significantly (Supplementary Fig. S3, Supplementary Data 13 and 14). These analyses verify the biological relevance of our findings across both rare and common protein-coding variants.

### Mendelian randomization analysis of regional brain iron and brain disorders

Brain iron plays a crucial role in various brain disorders. To investigate whether regional brain iron has causal relationships with brain disorders, we conducted a two-sample Mendelian randomization (MR) analysis. Four brain disorders were selected for this study: depression, bipolar disorder, PD, and AD. Our primary focus was to understand the influence of brain iron on these disorders. Therefore, regional brain iron features were used as exposures, while the brain disorders were used as outcomes in the MR analysis (see the “Methods” section).

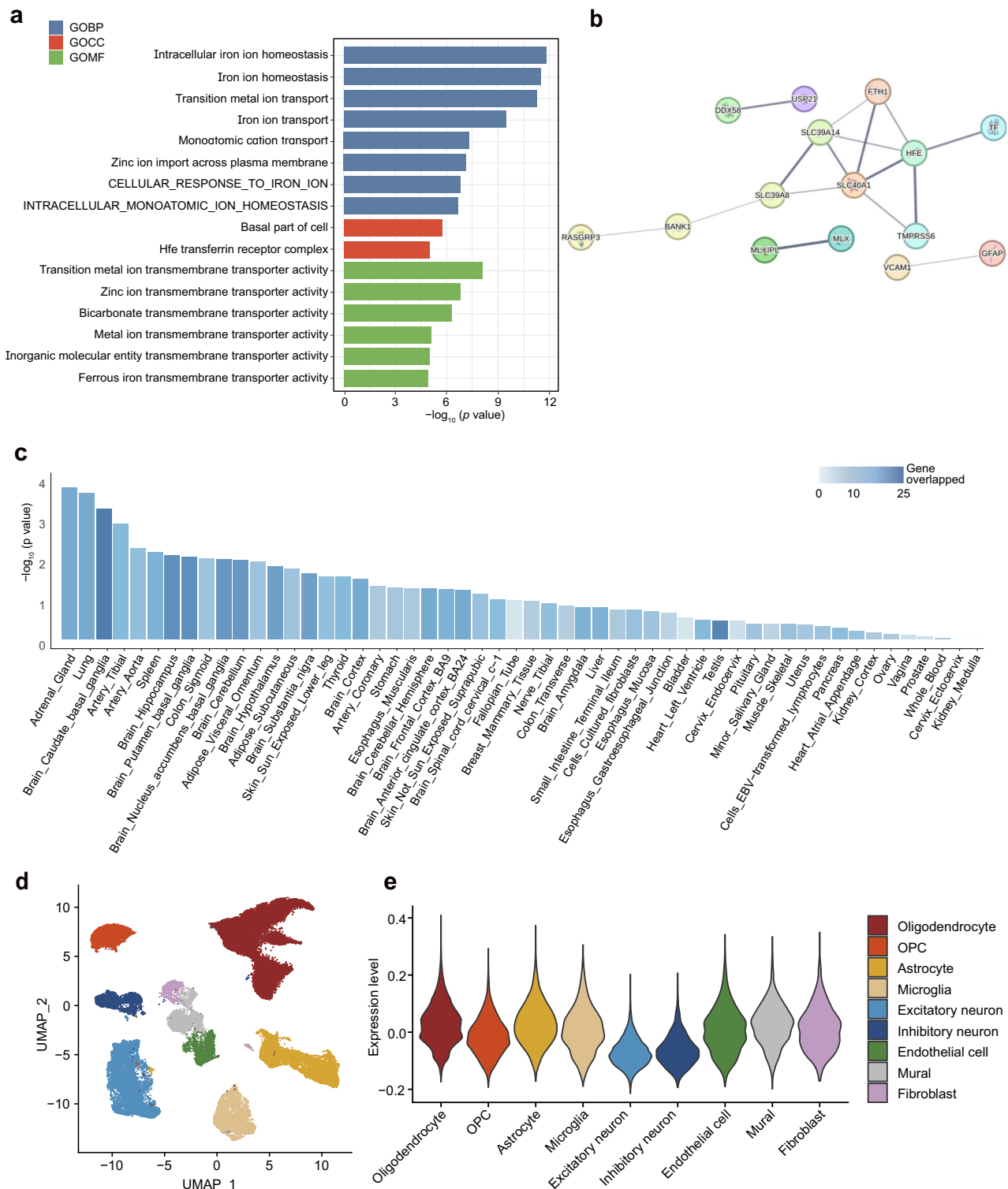
Our investigation revealed 11 significant causal relationships for the four candidate brain disorders (Table 2, Supplementary Data 15 and 16): (1) from the subthalamic nucleus, accumbens and

thalamus to bipolar disorder (top  $p = 3.6 \times 10^{-7}$ ,  $FDR = 2.2 \times 10^{-5}$ ); (2) from the caudate, substantia nigra, dentate and putamen to PD (top  $p = 1.0 \times 10^{-4}$ ,  $FDR = 0.0028$ ); (3) from the hippocampus and subthalamic nucleus to depression (top  $p = 7.6 \times 10^{-3}$ ,  $FDR = 0.046$ ). Additionally, we also conducted a reverse MR analysis (from disease to regional brain iron), revealing that the above significant associations are not significant in this analysis (Supplementary Data 17), indicating that the direction of causality was not biased by reverse causation. It is worth noting that while previous studies have documented statistical associations<sup>12–14</sup>, our results uncovered potential causal linkages from the brain to diseases.

Several sensitivity analyses were also conducted to assess the robustness of our results. MR results from alternative approaches are shown in Supplementary Data 16 and 18. The MR-Egger intercept was close to zero, and the pleiotropy test was not significant (Supplementary Data 15 and 17), suggesting that there was no directional pleiotropy in our analysis. In addition, we performed multivariate MR to assess the independence of causal effects (see the “Methods” section). The results revealed that the causal association between iron levels in 26 brain regions and brain disorders was not significant (Supplementary Data 19). This observation suggests that the causal relationship between brain iron levels and brain disorders may not be entirely independent across different brain regions.

### PheWAS of brain iron-associated genes

Brain iron plays a pivotal role in numerous fundamental biological processes. Conducting a PheWAS enables us to attain a comprehensive understanding of the diverse impacts of brain iron across the



**Fig. 4 | Biological functions of genes associated with brain iron. a** Gene set enrichment analysis of significant genes identified in the exome-wide association analysis, using Gene Ontology and KEGG Ontology database. The  $p$ -values reported are two-sided and unadjusted. **b** Protein-protein interaction network of significant brain iron-related genes. **c** Tissue-wide differential expression analysis of significant brain iron-related genes using GTEx dataset. The  $p$ -values reported are two-sided

and unadjusted. **d** Uniform manifold approximation and projection (UMAP) visualization of human brain single-cell sequencing data. **e** Expression levels of significant brain iron-related genes in human brain single-cell sequencing data. The colors indicate the tissue types. GOBP gene ontology biological process, GOCC gene ontology cellular component, GOMF gene ontology molecular function, OPC oligodendrocyte progenitor cell.

phenotypic landscape and enhances the validity of the findings of the current study<sup>15–17</sup>.

The phenotypes considered in this analysis encompass cognition metrics, neurological and psychiatric conditions, blood chemistry

measures, neuroimaging phenotypes, and plasma protein levels. Among the assessed phenotypes, the most robust associations were observed between brain iron-related genes and plasma proteins. *STAB1* is associated with multiple plasma protein, including, e.g. MME

**Table 2 | Causal relationships identified by Mendelian randomization analysis**

Brain iron (exposure)	Diseases (outcome)	Beta	p value	FDR
Subthalamic nucleus (right)	Bipolar	0.011	3.59E-07	2.15E-05
Accumbens (left)	Bipolar	0.023	3.11E-05	9.33E-04
Substantia nigra (right)	PD	0.010	1.45E-04	2.90E-03
Substantia nigra (left)	PD	0.011	1.33E-03	1.92E-02
Caudate (right)	PD	0.013	1.60E-03	1.92E-02
Dentate (right)	PD	0.006	4.25E-03	4.25E-02
Putamen (left)	PD	0.010	5.11E-03	4.34E-02
Thalamus (left)	Bipolar	0.024	5.79E-03	4.34E-02
Putamen (right)	PD	0.010	6.71E-03	4.47E-02
Subthalamic nucleus (right)	Depression	0.003	7.67E-03	4.60E-02
Hippocampus (right)	Depression	0.017	8.49E-03	4.63E-02

The iron in brain regions are the exposure variables and the brain disorders are the outcome variables.

( $p = 6.7 \times 10^{-13}$ ), *ECE1* ( $p = 4.0 \times 10^{-9}$ ), *OMD* ( $p = 9.2 \times 10^{-8}$ ), *LTBP2* ( $p = 5.6 \times 10^{-6}$ ). *TERT* is associated with proteins *EDA2R* ( $p = 1.4 \times 10^{-6}$ ) and *FOLR3* ( $p = 5.7 \times 10^{-6}$ ). *SHANK1* is associated with protein *NTRK2* ( $p = 8.9 \times 10^{-6}$ ) (Fig. 5a). Genes mapped to common variants have more significant levels of associations, as shown in Fig. 5b.

Our exploration of cognitive metrics and disease phenotypes revealed five significant gene-phenotype associations, all mapped to common variants (FDR-corrected  $p < 0.05$ ): *SLC39A8* is associated with fluid intelligence ( $p = 1.4 \times 10^{-10}$ ) and prospective memory ( $p = 1.5 \times 10^{-9}$ ); *HFE* is associated with multiple sclerosis ( $p = 2.2 \times 10^{-6}$ ), fibrosis liver ( $p = 5.3 \times 10^{-9}$ ) and broader liver disease ( $p = 6.5 \times 10^{-5}$ ); *EHBPI* is associated with the fluid intelligence ( $p = 9.7 \times 10^{-5}$ ); *MLX* is associated with migraine ( $p = 1.2 \times 10^{-3}$ ) (Fig. 5b). These associations were especially noteworthy as these genes ranked among the most significantly associated in our primary association analyses (Fig. 2b).

In other phenotypic variables that we analyzed, for genes mapped to rare variants, *RNASE12* is associated with the insulin-like growth factor 1 (IGF-1,  $p = 1.0 \times 10^{-4}$ ). Common variants show stronger phenotype-wide associations (Fig. 5b). For instance, the gene *BEST1* is associated with the ratio of omega-6 to omega-3 fatty acids ( $p = 1.4 \times 10^{-38}$ ), as determined by nuclear magnetic resonance (NMR) spectroscopy.

## Discussion

In our large-scale EWAS of brain iron, we uncovered the genetic underpinnings of brain iron accumulation using high-quality genetic data and QSM from swMRI scans of 29,828 UK Biobank participants. Our discovery of EWAS revealed 36 genes ( $N = 26,789$ ), mapped from either rare or common protein-coding variants, associated with brain iron accumulation. Remarkably, 29 of these genes were not previously reported. 16 of them can be replicated in an independent dataset with 3039 individuals. Our functional enrichment analysis revealed that the identified genes are enriched in biological pathways involving ion transport and homeostasis. MR analysis identified several interesting causal relationships between regional brain iron accumulation to brain disorders, such as from the substantia nigra to PD, and from the accumbens nucleus to AD, and from the hippocampus to depression. Furthermore, our phenotype-wide association study highlighted genes like *SLC39A8* and *EHBPI*'s associations with fluid intelligence and *HFE*'s linked to multiple sclerosis and liver-related diseases.

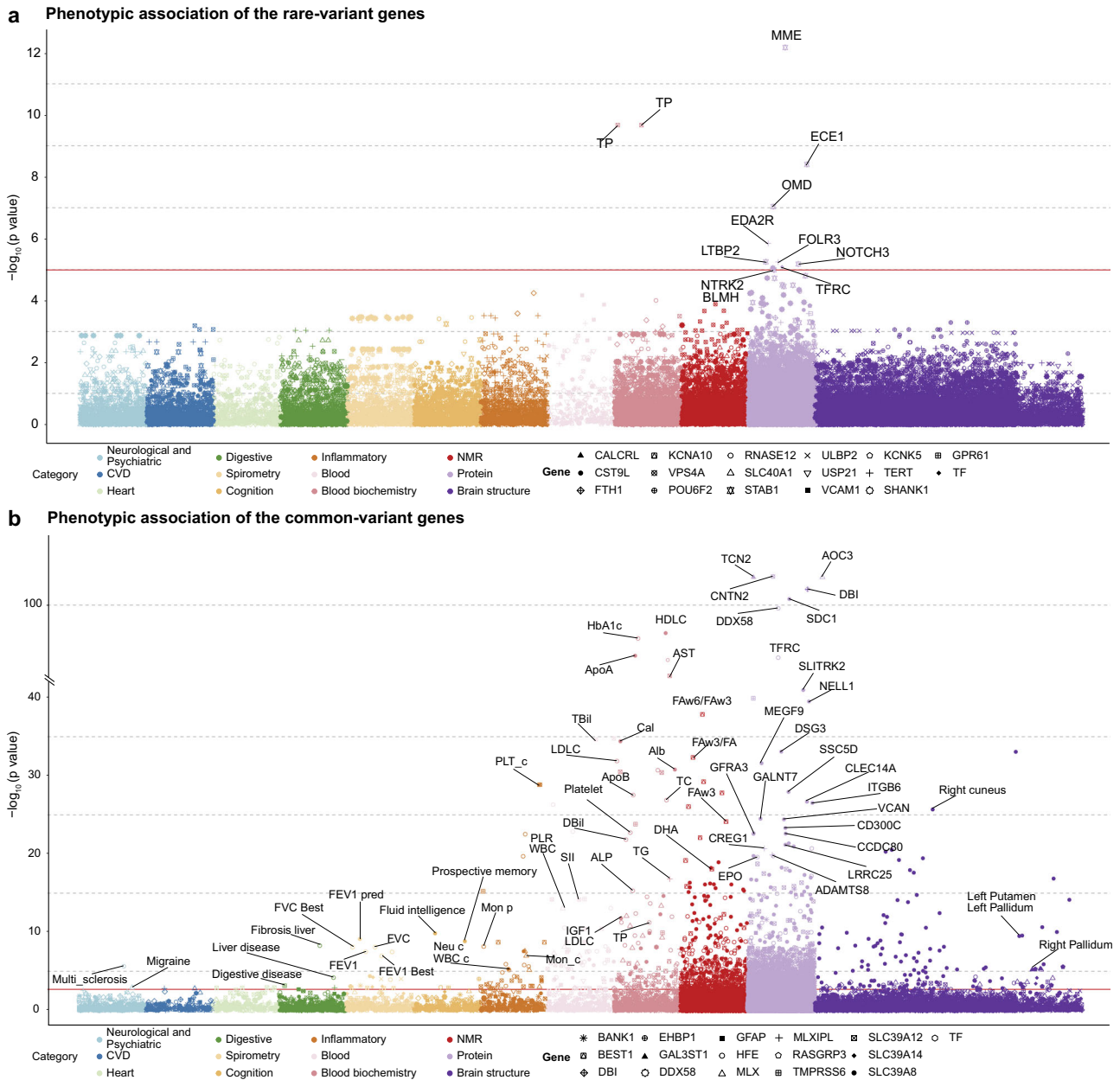
Our analyses revealed multiple rare genes that have been previously proposed as important contributors to the development of brain disorders or related to iron transport and ferritin (Fig. 3a, b). For example, the gene *STAB1* has the overall highest significant level in our

EWAS and displayed replicated associations with multiple brain regions, including those of the putamen, caudate, globus pallidus externa and substantia nigra. It has previously been reported as a candidate gene for bipolar disorder<sup>18</sup> and cerebrovascular diseases<sup>19</sup>. *FTH1* is associated with the pallidum in our EWAS. This gene was closely linked to various brain disorders, as it is responsible for encoding the heavy chain of ferritin<sup>20</sup>. Our rare variants analysis also identified the transferrin gene (*TF*) that is associated with the hippocampus. This gene was known to participate in iron transport. It exhibited an interesting convergence of rare and common variant evidence. *KCNA10* is associated with globus pallidus interna, whose changes in expression level were observed in PD patients<sup>21</sup>.

For common genes, owing to the involvement of a larger number of candidate brain regions than the previous study<sup>8</sup>, we identified several gene-brain associations (Fig. 3a, c). For example, *HFE* was associated with the dentate nucleus and red nucleus, which has been previously reported to lead to an increased risk of developing movement disorders<sup>22</sup>. The gene *MLX* is associated with putamen. This gene controls the transport and storage of ferrous iron<sup>23</sup>.

Importantly, our EWAS identified eight genes linked to the brain iron levels of substantia nigra (SN): *STAB1*, *TF*, *GPR61*, *TERT*, *SLC39A8*, *SLC39A12* and *HFE*. These genes are linked to PD's underlying mechanisms. For instance, previous research has suggested that dysregulated *STAB1* expression in microglia might play a role in the pathogenesis of PD<sup>24</sup>. The gene *SLC39A8*, encoding a metal ion transporter, has been linked to various conditions, including PD<sup>25</sup>. Interestingly, our MR analyses further revealed significant causal relationships between brain iron accumulation in the SN to PD. PD is marked by motor impairments, stemming largely from the loss of dopamine-producing neurons in the SN<sup>26</sup>. Dopamine, a pivotal neurotransmitter, governs our motor functions and coordination. A deficit in dopamine results in multiple movement challenges in PD patients, such as tremors, stiffness, bradykinesia, and balance issues. Notably, increased iron accumulation within the substantia nigra has been observed in PD patients, complementing our MR results and indicating a potential causal link between the SN and PD. Diving deeper into these biological pathways could enrich our understanding of PD, potentially leading to more targeted and effective therapeutic interventions<sup>27</sup>.

We identified four genes associated with the hippocampal iron level: *BEST1*, *HFE*, *TF*, and *GFAP*. These genes may provide insights into the connection between the hippocampus and depression. For example, Astrocytes, characterized by their expression of *GFAP*, play an important role in the central nervous system. They are abundant in the hippocampus, a central component of the limbic system that has long been theorized to play a role in depression's neuropathology<sup>28</sup>. Post-mortem studies have highlighted the involvement of cerebral astrocytes immunoreactive to *GFAP* in the pathogenesis of depression<sup>29,30</sup>. In addition, the *HFE* gene is primarily known for its role in hereditary hemochromatosis, a genetic disorder that causes the body to absorb too much iron. Mutations in the *HFE* gene can lead to excessive iron accumulation. Notably, the brain's uptake of blood iron is important for the optimal synthesis of neurotransmitters such as serotonin, dopamine, and noradrenaline. These neurotransmitters, involved in regulating emotional behaviors, rely on neuron aromatic hydroxylase, with iron acting as an important cofactor. Notably, noradrenaline affects neuroplasticity through the brain-derived neurotrophic factor, which is important for the functioning of prefrontal and hippocampal neurons implicated in depression<sup>31</sup>. Moreover, research has also established a correlation between severe depression symptoms and high body iron levels<sup>32,33</sup> as well as increased brain iron levels<sup>14</sup>. Our MR analysis revealed a significant causal linkage between the hippocampus and depression. In studies using mouse models, cerebral iron deficiency, leading to the suppression of the hippocampal glucocorticoid receptor signaling pathway, has been implicated in inducing depression<sup>6</sup>. A consistent observation among



**Fig. 5 | Phenome-wide association analysis of brain iron-associated genes.**

**a** Phenome-wide associations of genes mapped from rare variants. **b** Phenome-wide associations of genes mapped from common variants. Scatterplot showing associations between brain iron-related genes and a wide range of phenotypes, including 12 categories, listed at the bottom left of each figure. The y-axis indicates the  $-\log_{10}$  of the  $p$ -value for each association, and the x-axis represents different phenotype categories. The  $p$ -values shown are two-sided and unadjusted for multiple testing. Linear regression models and SKAT-O tests were used for gene-based

analysis, and the model adjusted for age, gender, and top 10 ancestral principal components. Red line in each figure is the FDR 0.05 correction threshold. CVD cardiovascular disease, FEV1 forced expiratory volume in 1 s, FEV1\_Best best measure of FEV1, FEV1\_predperc predicted percentage of FEV1/TP, total protein, HDLC high-density lipoprotein cholesterol, LDLC low-density lipoprotein cholesterol, WBC white blood cell count, Neu\_c neutrophil count, PLT platelet count, HBA1C glycated hemoglobin, PLR platelet-to-lymphocyte ratio, SII systemic immune-inflammation index.

depressed individuals is reduced hippocampal volume, correlating with the length and recurrence of depressive episodes<sup>34,35</sup>. Furthermore, the hippocampus closely interacts with other brain regions responsible for emotional and mood regulation, such as the amygdala and prefrontal cortex. Any malfunction within these networks might hamper emotional processing, thus amplifying the mood-related symptoms synonymous with depression. Iron potentially holds essential significance in the survival of hippocampus neurons, thereby influencing the progression of depression.

Our findings also reveal several insights into the relationships between brain iron accumulation and AD. We identified five genes

associated with the brain iron levels of the accumbens nucleus: *SHANK1*, *CST9L*, *HIGD1C* and *SLC39A8*. Notably, the cognitive deficits typical of AD often coincide with synaptic loss attributed to disruptions in the postsynaptic density. Such disruptions are evidenced by a marked reduction in *SHANK1* protein levels<sup>36</sup>. Given the accumbens nucleus's role in dopamine regulation<sup>37</sup>, the research highlighted the potential consequence of perturbed dopaminergic signaling in the context of AD<sup>38</sup>. Furthermore, excess iron levels in neural tissues can induce oxidative stress, known to adversely affect neural integrity and functionality<sup>39</sup>. Such oxidative stress has been linked to disturbing neurotransmission, particularly of dopamine, which in turn influences



the function of the accumbens nucleus<sup>40</sup>. In PheWAS, we also find that *SLC39A8* is associated with the cognition test scores (i.e., fluid intelligence and prospective memory), which aligns with the known characteristics of AD as a progressive neurodegenerative disorder characterized by memory loss and cognitive deficits.

A major strength of our research lies in its pioneering use of the most extensive WES analyses for brain iron accumulation to date. This approach has led us to identify multiple genes not previously reported, further enriching our understanding of the genetic architecture behind brain iron accumulation. The current study also has some potential limitations. Our analysis predominantly encompasses participants of European genetic ancestries and is constrained by the available sample size for brain imaging data. Incorporating cortical brain regions in our study might have enhanced our findings, offering a more comprehensive understanding of the genetic structure and associations concerning regional brain iron levels. In addition, a large external validation dataset would be beneficial to verify the solidity of our findings.

In conclusion, our investigation identified 36 genes associated with brain iron accumulation across multiple brain regions. Many of these genes are enriched in pathways related to iron transport and homeostasis and are linked to iron-related brain disorders. Furthermore, our study also revealed several causal pathways from regional brain iron accumulation to disorders such as PD, AD and depression. These findings provided insights into the genetic architecture of brain iron accumulation and uncovered the important relationships of brain iron with several brain disorders and behavioral traits. We anticipate that our findings will serve as a groundwork for future research, aiding in the elucidation of how these genes impact brain iron levels and contribute to the onset and progression of brain disorders.

## Methods

### Study population

The UK Biobank (UKB) (<https://www.ukbiobank.ac.uk/>) is a population-based prospective cohort of ~500,000 participants aged from 40 to 69 years old at enrollment between 2006 and 2010<sup>41</sup>, among which whole-exome sequencing data are available for 454,787 participants. UKB received ethical approval from the National Health Service National Research Ethics Service (reference: 11/NW/0382) and all participants provided written informed consent. This study was conducted under application number 19542.

### Brain imaging-derived phenotypes

We used an automated quantitative susceptibility mapping (QSM) pipeline<sup>8</sup>, based on the susceptibility-weighted MRI (swMRI) data, to measure the brain iron. We used swMRI data from 37,213 subjects in the UK Biobank (release 2022). The detailed MRI data acquisition, quality control, and QSM processing pipelines have been reported in the original paper<sup>8</sup>. The individual space voxel-wise QSM maps are downloaded and warped to the MNI152 standard space based on the warp field maps. We extracted the median QSM values from 26 sub-cortical and cerebellum structures as shown in Fig. 1. Among them, 16 of the 26 subcortical structures were defined based on the original paper (the accumbens, amygdala, caudate, hippocampus, pallidum, putamen, substantia nigra and thalamus, both left and right)<sup>8</sup>. The corresponding field IDs for these imaging-derived phenotypes are listed in Supplementary Data 20. We further selected 10 additional regions of interest (the red nucleus, subthalamic nucleus, globus pallidus externa, globus pallidus interna and dentate, both left and right), based on the segmentation masks of the multi-contrast PD25 atlas<sup>42</sup> and the deep cerebellar nuclei probabilistic atlas<sup>43,44</sup>.

### Exome sequencing and quality control

Whole-exome sequencing was conducted on 454,787 participants in the Regeneron Genetics Center (RGC) and protocols were described in

detail elsewhere<sup>10</sup>. The OQFE WES pVCF files in GRCh38 human reference genome build<sup>45</sup> were utilized in this study, and we performed additional quality control similar to the previous study<sup>46</sup>. First, multi-allelic sites were split into bi-allelic sites and all calls that have a low genotype quality and extremely low/high genotype depth were set to no-call. Variants with call rate <90%, Hardy-Weinberg  $P$ -value <  $10^{-15}$ , and in Ensembl low-complexity regions were excluded. Samples withdrawn from the study, duplicates, with discordance between self-reported and genetically inferred sex, samples whose Ti/Tv, Het/Hom, SNV/indel, and number of singletons exceed 8 standard deviations from the mean were removed. We used King software to calculate the kinship coefficient using the high-quality variants (MAF > 0.1%, missingness < 1%, HWE  $P$  >  $10^{-6}$  and two rounds of pruning using --indep-pairwise 200 100 0.1 and --indep-pairwise 200 100 0.05). Unrelated samples were defined using the kinship coefficient threshold at 0.0884, indicating the 2nd relatedness. To maximize the sample size, participants related to multiple other individuals were first iteratively removed until none remained. Then, one of the remaining kinship pairs was removed at random. In this study, samples were mainly restricted to White British (filed 22006) and ancestry-specific principal components were calculated, which were used in the following analysis.

### Variant annotation

SnPEff<sup>47</sup> was used to annotate rare variants (MAF < 1%), and for those annotated with multiple consequences, the most severe consequence was kept for each gene transcript. Loss of function (LOF) was defined for variants annotated as stop gained, start lost, splice donor, splice acceptor, stop lost, or frameshift. Likely deleterious missense was determined if variants were consistently predicted as deleteriousness in SIFT<sup>48</sup>, PolyPhen2 HDIV and PolyPhen2 HVAR<sup>49</sup>; LRT<sup>50</sup>; and MutationTaster<sup>51</sup>. As for common variants (MAF ≥ 1%), ANNOVAR<sup>52</sup> was utilized to annotate variants using refGene as a reference panel, and those annotated as exonic, UTR3, or UTR5 were kept in the following single-variant EWAS analysis.

### Data partition

In this study, we partitioned the entire dataset into two subsets: a discovery set and a replication set. This partitioning was based on the availability of longitudinal swMRI scans for individual subjects. Two independent association tests were conducted as follows: Discovery set: This subset comprised baseline imaging data from 26,789 subjects who did not have longitudinal swMRI scans available. Statistical significance in the discovery set was determined using the genome-wide significance threshold. Replication set: This subset consisted of repeated scan brain imaging data from 3039 subjects. Statistical significance in the replication set was set at an uncorrected  $p$ -value threshold of <0.05, following the approach utilized in a previous study<sup>53</sup>.

### Exome-wide association analysis

EWAS was performed in the discovery and replication sets using unrelated British with both whole exome sequence (WES) and swMRI brain imaging data available using a generalized mixed model implemented in SAIGE-GENE+<sup>54</sup>. A total of 18,800 variants were analyzed. Rare pLOF and likely deleterious missense variants with MAF < 0.01 were collapsed in each gene, and the EWAS was calculated using the Burden test, SKAT, and SKAT-O, adjusting for age, sex, and the first ten principal components. We constructed six frequency-function collapsing masks for each gene in the gene-based collapsing test: for the frequency of variants, including MAF < 0.01, <0.001, and <0.0001; for the function of variants, pLOF and pLOF+ likely deleterious missense variants. For common variants, a single-variant association analysis was performed, adjusting for the same covariates as in the gene-based collapsing test. A sparse genetic relationship matrix was constructed using the high-quality variants with the recommended relative coefficient cutoff of 0.05. Bonferroni correction was applied and the significance threshold was set to

$p < 1.7 \times 10^{-8}$ . This is computed as  $0.05 / (18,800 \times 3 \times 2 \times 26)$ , where 18,800 is the number of candidate rare genes, 3 is the number of minor allele frequency cutoffs, 2 is the number of variants annotation groups, and 26 is the number of brain regions.

### Robustness of EWAS

To assess the robustness of the results and detect the variants driving the association in the collapsing test, we further performed leave-one-variant-out analysis. The variant maximized the  $p$ -value after removing the collapsing test was identified as the driving variant. For the significant gene–phenotype associations found in the gene-based EWAS, conditional analyses were further performed to evaluate the influence of the nearby common loci (defined as independent index variants after clumping ( $-\text{clump-p1 } 1 \times 10^{-5} \text{ --clump-r2 } 0.01$ ),  $\pm 500$  kb of the identified gene,  $\text{MAF} > 0.5\%$  in the UKB imputed data). For the significant gene–phenotype associations, gene-based collapsing tests were re-performed to condition on nearby common variants signals. For single variant EWAS, independent significant SNPs were further identified ( $p < 1.0 \times 10^{-5}$  and  $r^2 \leq 0.6$  within a 1 Mb window). Independent significant SNPs were then clumped to obtain lead SNPs ( $r^2 \leq 0.1$  within a 1 Mb window). Genomic loci were defined by merging lead SNPs within 250 kb. In addition, we performed subgroup analysis based on sex with the same covariates adjusted except for the subgroup factor.

### Common variants association analysis

For common exonic variants (MAF 1%), a single variant association analysis was conducted among the unrelated Caucasian cohort using SAIGE-GENE+<sup>54</sup>, adjusting age, gender, and ten principal components. Lead SNPs were identified as independent significant SNPs which meet significant thresholds and are independent of other significant SNPs with  $r^2 < 0.1$  within a 1 Mb window. The significance threshold was set to  $4.60 \times 10^{-8}$  ( $0.05 / (41,790 \times 26)$ , Bonferroni correction for 41,790 exonic coding SNPs in 26 brain regions).

### Tissue and pathway enrichment analysis

Tissue enrichment analysis was performed by the GENE2FUNC function in FUMA<sup>55</sup> with all mapped genes in EWAS as input. Briefly, Differentially Expressed Gene (DEG) sets were pre-calculated by performing a two-sided  $t$ -test for any one type of tissue against all other tissues of 54 tissue types based on data from the GTEx database<sup>56</sup>. Then hypergeometric tests were used to test brain iron-associated genes against each of the DEG sets. For pathway enrichment analysis, hypergeometric tests were also performed to test if brain iron-associated genes are overrepresented in any of the pre-defined gene sets, covering Gene Ontology, Reactome, GWAScatalog, and Immunologic signatures.

### Single-nucleus RNA sequencing data source and analyses

We used single-nucleus RNA sequencing (snRNA-seq) data of human brain vasculature obtained from a recent study conducted by Garcia et al. on the Gene Expression Omnibus database with the accession ID: GSE173731<sup>57</sup>. The clustering and annotation of the cell types were conducted via the metadata file provided by the authors. We also computed the gene set score of the brain iron-associated genes using results from both single and gene-based EWAS, using *AddModuleScore* function. The primary analysis and subsequent visualization were conducted using the R package Seurat.

### Protein–protein interaction network

Protein–protein interaction based on the significant gene set derived from the single variant and gene-based EWAS results were investigated using the human STRING database<sup>58</sup>. Interactions with a confidence score of at least medium confidence were extracted and subsequently visualized in Cytoscape<sup>59</sup> (version 3.9.0). Proteins were further

clustered using the Markov clustering (MCL) algorithm to investigate the functional clusters using the default settings<sup>60</sup>.

### Phenome-wide association study

To explore broader phenome-wide associations and underlining mechanisms for the brain iron-related genes or variants derived from single and gene-based EWAS results, we investigated their associations with additional phenotypes. We mainly focused on brain structures ( $N = 220$ : thickness, surface area, volume for 68 cortical regions and volume for 16 subcortical structures), biochemistry ( $N = 30$ ), inflammatory ( $N = 11$ ) markers, metabolomics ( $N = 249$ ), and proteomics ( $N = 1463$ ). As for rare variants, gene-based linear mixed models in SAIGE-GENE+ were employed. And for common variants, single-variant association analysis was performed using linear regression by PLINK v2. Both models were adjusted for age, sex, and the first ten genetic principal components.

Standard Siemens Skyra 3T running VD13A SP4 with a 32-channel head coil was used to acquire the T1-weighted neuroimaging data with a resolution of  $1 \times 1 \times 1$  mm (Field 20252) (detailed acquisition protocol can be found at [https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/brain\\_mri.pdf](https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/brain_mri.pdf)). The cortical surface areas, volumes, and mean thickness for 68 cortical regions were extracted based on FreeSurfer's surface templates using *aparc atlas*<sup>61</sup>. The volume for 16 subcortical regions was estimated via FreeSurfer's *aseg tool*<sup>62</sup>.

Biochemistry or inflammatory markers were obtained from blood count data (Category 100081) and blood biochemistry data (Category 17518) based on UK Biobank blood samples (detailed protocol can be found at <https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=100080>). Four blood cell count ratios were additionally calculated for downstream analysis, including the neutrophils to lymphocytes ratio (NLR), platelet-to-lymphocyte ratio (PLR), lymphocyte-to-monocyte ratio (LMR), and the systemic immune-inflammation index (SII).

Nuclear magnetic resonance (NMR) metabolomics data (Category 220) were acquired from randomly selected EDTA plasma samples using a high-throughput NMR-based metabolic biomarker profiling platform. This platform covers 249 metabolic spanning multiple metabolic pathways, including lipoprotein lipids, fatty acids, fatty acid compositions, and various low-molecular-weight metabolites (detailed protocol can be found at <https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=220>).

Proteomics data (Category 1838) of 1463 proteins in plasma were measured by Olink Explore platform, using Proximity Extension Assay (PEA) (detailed information on sample collection, processing, normalization, and quality control procedures can be found at <https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=1839>).

### Causal relationship between regional brain iron and brain disorders

To investigate the causal relationships between QSM and multiple brain disorders, we first performed a Genome-Wide Association Study (GWAS) using imputed SNP (to the Haplotype Reference Consortium) genotype data obtained from the UK Biobank resource<sup>63</sup>. A total of 8,445,740 SNPs were included in the GWAS. Samples were mainly restricted to white British and those who were used in computing the principal components. Individuals with a missing genotype rate  $> 0.05$ , with mismatch self-reported (Data field 31) and genetic sex (Data field 22001), with abnormal sex chromosome aneuploidy, and have more than 10 putative third-degree relatives been further removed. We also excluded variants with call rate  $< 0.95$ ,  $\text{MAF} < 0.01$ , Hardy–Weinberg  $P$ -value  $< 10^{-6}$ , or imputation quality score  $< 0.5$ . GWAS were independently performed for each phenotype using linear regression models implemented in PLINK2<sup>64</sup>. Covariates included age, sex, and the first ten genetic principal components. A total of 26,776 to 28,129

participants with phenotype and covariates available for 26 subcortical and cerebellum structures were included in the final GWAS.

Our primary two-sample MR analyses were conducted using regional brain iron as exposure, and diseases as outcomes. For diseases, we leveraged GWAS summary data of iron-related brain disorders, including Parkinson's disease (Ncase = 33,674, Ncontrol = 449,056)<sup>65</sup>, Depression (excluding 23andMe and UK Biobank: Ncase = 45,591, Ncontrol = 97,674)<sup>66</sup>, Bipolar (Ncase = 20,352, Ncontrol = 31,358)<sup>67</sup>, and Alzheimer's disease (Ncase = 21,982, Ncontrol = 41,944)<sup>68</sup>. Instrumental variables (IV) were selected based on a significant level ( $p < 5 \times 10^{-8}$ ) and followed by LD clumping ( $R^2 > 0.001$ ). Then IV from the exposure and outcome data were harmonized to the same effect alleles. *F*-statistics were computed to assess the strength of the instruments. When only a single SNP was available, the Wald ratio was used to estimate the causality of exposure to outcome. When more than one SNP was available, the inverse-variance weighted (IVW) with multiplicative random effects method was employed<sup>69</sup>. MR-PRESSO test<sup>70</sup> was used to detect outliers, and if an outlier was detected, the original *p*-value was replaced by the outlier-corrected *p*-value. The *q*-value FDR approach was used to correct for multiple comparisons across brain regions and diseases<sup>71</sup>.

To assess the robustness of our analysis, several sensitivity analyses were conducted. This involved using different MR methods, including MR Egger<sup>72</sup>, Wald ratio<sup>73</sup>, and Weighted median<sup>74</sup>. The intercept of MR Egger was used to identify the presence of directional pleiotropy. Considering the similarity of genetic architecture between different brain regions, LASSO feature selection (`mv_lasso.feature_selection()` function in "TwoSampleMR") followed by a Multivariable MR (MVMR) was performed to further assess whether the causal effects were independent<sup>75</sup>. In addition, we also performed a reverse MR (from neurologic and psychiatric disorders to regional brain iron) to infer the direction of causality. The MR analysis was performed using the "TwoSampleMR" version 0.5.6 in R version 4.2.

### Power analysis

We simulated 1000 datasets for each combination of the estimated effect sizes (i.e., based on the regression coefficient and its estimated errors) and the cMAC per gene of our analysis, based on the method in ref. 11. Carrier status was randomized across  $N = 29,828$  (sample size for our collapsing tests) participants. A linear regression model was used to test for the association with a threshold of  $p < 1.7 \times 10^{-8}$  (corresponding to the significance threshold in our collapsing tests).

### Statistics and reproducibility

The code used in the paper is made publicly available for reproducibility purposes. Statistical analyses are given as well. There is no randomness for all results presented in this study.

### Reporting summary

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

### Data availability

The main data, including the individual-level phenotypic and genetic data used in this study, were accessed from the UK Biobank under application number 19542 and were available through UKB. The EWAS summary statistics are available at <https://doi.org/10.5281/zenodo.11170064><sup>76</sup>.

### Code availability

All software and R packages used to perform the analyses in this work are freely available online: SnpEff, <https://pcingola.github.io/SnpEff/>; SAIGE-GENE+, <https://github.com/saigegit/SAIGE>; R, <https://www.r-project.org/>; PLINK2, <https://www.cog-genomics.org/plink/2.0/>; FUMA, <https://fuma.ctglab.nl/>; MAGMA, <https://ctg.cncr.nl/software/magma/>; STRING, <https://www.string-db.org/>. The scripts<sup>76</sup> used to conduct the main analyses are available at <https://github.com/weikangong/BrainIronWES>.

## References

- Rouault, T. A. Iron metabolism in the CNS: implications for neurodegenerative diseases. *Nat. Rev. Neurosci.* **14**, 551–564 (2013).
- Hare D, Ayton S, Bush A, et al. A delicate balance: Iron metabolism and diseases of the brain[J]. *Frontiers in aging neuroscience*, 2013, 5: 34.
- Ndayisaba, A., Kaindlstorfer, C. & Wenning, G. K. Iron in neurodegeneration—cause or consequence? *Front. Neurosci.* **13**, 180 (2019).
- Lei, P. et al. Tau deficiency induces parkinsonism with dementia by impairing APP-mediated iron export. *Nat. Med.* **18**, 291–295 (2012).
- Liu, Z., Zhou, T., Ziegler, A. C., Dimitrion, P. & Zuo, L. Oxidative stress in neurodegenerative diseases: from molecular mechanisms to clinical applications. *Oxid. Med. Cell. Longev.* **2017**, 2525967 (2017).
- Zhang, H. et al. Cerebral iron deficiency may induce depression through downregulation of the hippocampal glucocorticoid-glucocorticoid receptor signaling pathway. *J. Affect. Disord.* **332**, 125–135 (2023).
- van der Weijden, C. W. et al. Quantitative myelin imaging with MRI and PET: an overview of techniques and their validation status. *Brain* **146**, 1243–1266 (2023).
- Wang, C. et al. Phenotypic and genetic associations of quantitative magnetic susceptibility in UK Biobank brain imaging. *Nat. Neurosci.* **25**, 818–831 (2022).
- Langkammer, C. et al. Quantitative susceptibility mapping (QSM) as a means to measure brain iron? A post mortem validation study. *Neuroimage* **62**, 1593–1599 (2012).
- Van Hout, C. V. et al. Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature* **586**, 749–756 (2020).
- Holstege, H. et al. Exome sequencing identifies rare damaging variants in ATP8B4 and ABCA1 as risk factors for Alzheimer's disease. *Nat. Genet.* **54**, 1786–1794 (2022).
- Du, G. et al. Dynamics of nigral iron accumulation in Parkinson's disease: from diagnosis to late stage. *Mov. Disord.* **37**, 1654–1662 (2022).
- Chen, Q. et al. Iron deposition in Parkinson's disease by quantitative susceptibility mapping. *BMC Neurosci.* **20**, 1–8 (2019).
- Duan, X. et al. Quantitative susceptibility mapping of brain iron deposition in patients with recurrent depression. *Psychiatry Investig.* **19**, 668 (2022).
- Diogo, D. et al. Phenome-wide association studies across large population cohorts support drug target validation. *Nat. Commun.* **9**, 4285 (2018).
- Hu, J. X., Thomas, C. E. & Brunak, S. Network biology concepts in complex disease comorbidities. *Nat. Rev. Genet.* **17**, 615–629 (2016).
- Shen, X. et al. A phenome-wide association and Mendelian Randomisation study of polygenic risk for depression in UK Biobank. *Nat. Commun.* **11**, 2301 (2020).
- Witt, S. H. et al. Investigation of manic and euthymic episodes identifies state-and trait-specific gene expression and STAB1 as a new candidate gene for bipolar disorder. *Transl. Psychiatry* **4**, e426–e426 (2014).
- Pan, Z. -L. & Chen, C. -Y. Analysis of multi-tissue transcriptomes reveals candidate genes and pathways influenced by cerebrovascular diseases. Preprint at *bioRxiv* <https://doi.org/10.1101/8068933> (2019).
- Shieh, J. T. et al. Heterozygous nonsense variants in the ferritin heavy chain gene FTH1 cause a novel pediatric neuroferritinopathy. Preprint at *medRxiv* <https://doi.org/10.1101/2023.01.30.23285099> (2023).
- Simunovic, F. et al. Gene expression profiling of substantia nigra dopamine neurons: further insights into Parkinson's disease pathology. *Brain* **132**, 1795–1809 (2009).
- Loughnan, R. et al. Association of Genetic Variant Linked to hemochromatosis with brain magnetic resonance imaging measures of iron and movement disorders. *JAMA Neurol.* **79**, 919–928 (2022).



23. Guo, W. et al. Super-enhancer-driven MLX mediates redox balance maintenance via SLC7A11 in osteosarcoma. *Cell Death Dis.* **14**, 439 (2023).
24. Corces, M. R. et al. Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat. Genet.* **52**, 1158–1168 (2020).
25. Nebert, D. W. & Liu, Z. SLC39A8 gene encoding a metal ion transporter: discovery and bench to bedside. *Hum. Genom.* **13**, 1–21 (2019).
26. Hirtz, D. et al. How common are the “common” neurologic disorders? *Neurology* **68**, 326–337 (2007).
27. Armstrong, M. J. & Okun, M. S. Diagnosis and treatment of Parkinson disease: a review. *JAMA* **323**, 548–560 (2020).
28. Campbell, S., Marriott, M., Nahmias, C. & MacQueen, G. M. Lower hippocampal volume in patients suffering from depression: a meta-analysis. *Am. J. Psychiatry* **161**, 598–607 (2004).
29. O'Leary, L. A. et al. Widespread decrease of cerebral vimentin-immunoreactive astrocytes in depressed suicides. *Front. Psychiatry* **12**, 640963 (2021).
30. Kim, R., Healey, K. L., Sepulveda-Orengo, M. T. & Reissner, K. J. Astroglial correlates of neuropsychiatric disease: from astrocytopathy to astrogliosis. *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* **87**, 126–146 (2018).
31. Berthou, C., Iliou, J. P. & Barba, D. Iron, neuro-bioavailability and depression. *EJHaem* **3**, 263–275 (2022).
32. Richardson, A. C. et al. Higher body iron is associated with greater depression symptoms among young adult men but not women: observational data from the daily life study. *Nutrients* **7**, 6055–6072 (2015).
33. Hidese, S., Saito, K., Asano, S. & Kunugi, H. Association between iron-deficiency anemia and depression: a web-based Japanese investigation. *Psychiatry Clin. Neurosci.* **72**, 513–521 (2018).
34. Frodl, T. et al. Hippocampal changes in patients with a first episode of major depression. *Am. J. Psychiatry* **159**, 1112–1118 (2002).
35. Roddy, D. W. et al. The hippocampus in depression: more than the sum of its parts? Advanced hippocampal substructure segmentation in depression. *Biol. Psychiatry* **85**, 487–497 (2019).
36. Grabrucker, A. M. et al. Amyloid beta protein-induced zinc sequestration leads to synaptic loss via dysregulation of the ProSAP2/Shank3 scaffold. *Mol. Neurodegener.* **6**, 1–20 (2011).
37. Jackson, M. E. & Moghaddam, B. Amygdala regulation of nucleus accumbens dopamine output is governed by the prefrontal cortex. *J. Neurosci.* **21**, 676–681 (2001).
38. Pan, X. et al. Dopamine and dopamine receptors in Alzheimer's disease: a systematic review and network meta-analysis. *Front. Aging Neurosci.* **11**, 175 (2019).
39. Li, L.-B. et al. Iron exposure and the cellular mechanisms linked to neuron degeneration in adult mice. *Cells* **8**, 198 (2019).
40. Ferreira, A., Neves, P. & Gozzelino, R. Multilevel impacts of iron in the brain: the cross talk between neurophysiological mechanisms, cognition, and social behavior. *Pharmaceuticals* **12**, 126 (2019).
41. Sudlow, C. et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
42. Xiao, Y. et al. A dataset of multi-contrast population-averaged brain MRI atlases of a Parkinson's disease cohort. *Data Brief* **12**, 370–379 (2017).
43. Diedrichsen, J., Balsters, J. H., Flavell, J., Cussans, E. & Ramnani, N. A probabilistic MR atlas of the human cerebellum. *Neuroimage* **46**, 39–46 (2009).
44. Diedrichsen, J. et al. Imaging the deep cerebellar nuclei: a probabilistic atlas and normalization procedure. *Neuroimage* **54**, 1786–1794 (2011).
45. Szustakowski, J. D. et al. Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. *Nat. Genet.* **53**, 942–948 (2021).
46. Jurgens, S. J. et al. Analysis of rare genetic variation underlying cardiometabolic diseases and traits among 200,000 individuals in the UK Biobank. *Nat. Genet.* **54**, 240–250 (2022).
47. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).
48. Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for genomes. *Nat. Protoc.* **11**, 1–9 (2016).
49. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **Chapter 7**, Unit 7.20 (2013).
50. Chun, S. & Fay, J. C. Identification of deleterious mutations within three human genomes. *Genome Res.* **19**, 1553–1561 (2009).
51. Schwarz, J. M., Rödelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **7**, 575–576 (2010).
52. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
53. Elliott, L. T. et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210–216 (2018).
54. Zhou, W. et al. SAIGE-GENE+ improves the efficiency and accuracy of set-based rare variant association tests. *Nat. Genet.* **54**, 1466–1469 (2022).
55. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
56. Battle, A., Brown, C. D., Engelhardt, B. E. & Montgomery, S. B. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
57. Garcia, F. J. et al. Single-cell dissection of the human brain vasculature. *Nature* **603**, 893–899 (2022).
58. Szklarczyk, D. et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–d613 (2019).
59. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
60. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584 (2002).
61. Desikan, R. S. et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–980 (2006).
62. Fischl, B. et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341–355 (2002).
63. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
64. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015).
65. Nalls, M. A. et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* **18**, 1091–1102 (2019).
66. Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
67. Stahl, E. A. et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* **51**, 793–803 (2019).
68. Kunkle, B. W. et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates A $\beta$ , tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).



69. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665 (2013).
70. Verbanck, M., Chen, C. Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 693–698 (2018).
71. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA* **100**, 9440–9445 (2003).
72. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
73. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**, e34408 (2018).
74. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
75. Meeks, K. A. et al. Mendelian randomization analyses suggest a causal role for circulating GIP and IL-1RA levels in homeostatic model assessment-derived measures of  $\beta$ -cell function and insulin sensitivity in Africans without type 2 diabetes. *Genome Med.* **15**, 108 (2023).
76. Gong, W. weikanggong/BrainIronWES: code for brain iron WES study. *Zenodo* <https://doi.org/10.5281/zenodo.11170064> (2024).

## Acknowledgements

We gratefully thank all UK Biobank participants for their time and UK Biobank team members for collating the data. J.-T. Yu was supported by grants from the Science and Technology Innovation 2030 Major Projects (2022ZD0211600), National Natural Science Foundation of China (82071201, 81971032, 92249305), Shanghai Municipal Science and Technology Major Project (No.2018SHZDZX01), Research Start-up Fund of Huashan Hospital (2022QD002), Excellence 2025 Talent Cultivation Program at Fudan University (3030277001), Shanghai Talent Development Funding for The Project (2019074), and ZHANGJIANG LAB, Tianqiao and Chrissy Chen Institute, and the State Key Laboratory of Neurobiology and Frontiers Center for Brain Science of Ministry of Education, Fudan University. W. Cheng was supported by grants from the National Natural Sciences Foundation of China (no. 82071997) and the Shanghai Rising-Star Program (no. 21QA1408700). J.F. Feng was supported by the National Key R&D Program of China (No.2018YFC1312904 and No. 2019YFA0709502), the Shanghai Municipal Science and Technology Major Project (No. 2018SHZDZX01), the 111 Project (No. B18015), Shanghai Center for Brain Science and Brain-Inspired Technology and Zhangjiang Lab. Figure 1 was partly generated using Servier Medical Art, provided by Servier, licensed under a Creative Commons Attribution 4.0 unported license.

## Author contributions

W.K. Gong, W. Cheng, J.-T. Yu and J.F. Feng designed the study. W.K. Gong, Y. Fu, B.-S. Wu, and J.N. Du conducted the main analyses and drafted the manuscript. W.K. Gong, Y. Fu, B.-S. Wu, J.N. Du, L. Yang, Y.-R. Zhang, S.-D. Chen, J.J. Kang contributed to data collection and analyses. W.K. Gong, J.-T. Yu, W. Cheng, Y. Mao, Q. Dong, L. Tan, and J.F. Feng critically revised the manuscript. All authors reviewed and approved the final version.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-49702-2>.

**Correspondence** and requests for materials should be addressed to Weikang Gong, Wei Cheng or Jin-Tai Yu.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions 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.

**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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024