Polygenic architecture of a novel MRI endophenotype: the "hemochromatosis brain"

Robert Loughnan[1,2,3], Jonathan Ahern[1,2], Mary Boyle[1,4], Terry L. Jernigan[1,2,5,6] , Donald J. Hagler, Jr.[3,6], John R. Iversen[2,7], Oleksandr Frei[8], Diana M. Smith[2,3], Ole Andreassen[8], Leo Sugrue[10], Wesley K. Thompson[11], Anders Dale[1,3,6,9], Andrew J. Schork[12,13,14], Chun Chieh Fan[6,11]

1. Department of Cognitive Science, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
2. Center for Human Development, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92161, USA
3. Center for Multimodal Imaging and Genetics, University of California, San Diego School of Medicine, 9444 Medical Center Dr, La Jolla, CA 92037, USA
4. Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Baltimore, MD 21205, USA
5. Department of Psychiatry, University of California, San Diego School of Medicine, 9500 Gilman Drive, La Jolla, CA 92037, USA
6. Department of Radiology, University of California, San Diego School of Medicine, 9500 Gilman Drive, La Jolla, CA 92037, USA
7. Swartz Center for Computational Neuroscience, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92161, USA
8. NORMENT Centre, Division of Mental Health and Addiction, Oslo University Hospital & Institute of Clinical Medicine, University of Oslo, Oslo, Norway
9. Department of Neuroscience, University of California, San Diego School of Medicine, 9500 Gilman Drive, La Jolla, CA 92037, USA
10. Department of Radiology and Biomedical Imaging and Department of Psychiatry, University of California, San Francisco, 505 Parnassus Avenue, San Francisco, CA 94143, USA
11. Center for Population Neuroscience and Genetics, Laureate Institute for Brain Research, Tulsa, OK 74103, USA
12. Institute of Biological Psychiatry, Mental Health Center - Sct Hans, Copenhagen University Hospital, Copenhagen, Denmark
13. Neurogenomics Division, The Translational Genomics Research Institute (TGEN), Phoenix, AZ, USA
14. Section for Geogenetics, GLOBE Institute, Faculty of Health and Medical Sciences, Copenhagen University, Copenhagen, Denmark

Abstract

Iron is an important element required for the body to function properly; however, hemochromatosis, an excess of iron in healthy tissue, can lead to cell damage. C282Y is a genetic variant prevalent in European populations, which is known to increase iron levels in the body and is responsible for most hemochromatosis cases. Recent work from our group analyzing brain Magnetic Resonance (MR) images established, in a large sample of older adults, that individuals who were homozygous for C282Y showed evidence consistent with substantial iron deposition localized to brain regions related to the motor system. These individuals also had nearly two-fold increased risk for developing movement disorders, like Parkinson's Disease (PD). Here, we describe the broader polygenic architecture of this novel endophenotype - the "hemochromatosis brain." This multivariate MRI phenotype captures average brain-wide differences of C282Y homozygote individuals and may reflect consequences of iron deposition localized to motor circuits. To do this we trained a classifier in a sample of 960 individuals from the UKBiobank to distinguish C282Y homozygotes from other genotypes using phenotypes derived from MR images sensitive to iron. This classifier was then applied to an independent set of 35,283 individuals for which MRI were available in the UKBiobank. This generated, for each individual, a PolyVoxel Score (PVS) that captures variation in MRI phenotypes associated previously with C282Y homozygosity and may reflect the degree of brain-specific iron deposition or other disease mechanisms. In this independent set we showed the PVS significantly predicted risk for PD (OR=1.37, p=0.0089). Studying the genetics of the "hemochromatosis brain" we conducted a GWAS on the PVS to show it to be highly heritable (h^2_{snp}=0.381, SE=0.033) with 43 genome-wide significant loci. Many of these signals are proximal to genes known to be involved in iron homeostasis (TF, rs6794370, p=2.43×10^{-81}; HFE rs1800562, p=6.93×10^{-75}; TMPRSS6, rs2413450, p=2.52×10^{-50}). Discovered genomic variants overlapped with previous GWAS studies of iron/red blood cells, non-iron blood markers and brain and cognitive traits. Cell type enrichment implicated glial cells to be the principal cell class related to the PVS GWAS signal, in line with previous work finding that these cells to have the highest iron concentrations in the brain. Finally, tissue-specific analysis identified the substantia nigra and hippocampus to have enriched epigenetic markers to our GWAS findings. We have made available a python program to generate the PVS from MRI scans. In conclusion, we have shown that the multivariate MRI phenotype associated with C282Y homozygosity is, in fact, highly heritable, polygenic, and represents a novel brain endophenotype of PD. Our unique approach of combining classical forward and reverse genetics approaches to define and describe potentially disease mediating endophenotypes could be applied more broadly to other clinically relevant disorders.

Intro

Iron is required for the body to function properly; however, an excess of iron in healthy tissue can lead to cell damage. C282Y is a genetic variant with a homozygosity prevalence of approximately 0.6% in European populations which is known to increase iron levels in the body and is responsible for the majority of hemochromatosis cases[1,2]. Recent work from our group analyzing brain Magnetic Resonance Images (MRI) established, in a large sample of older adults, that individuals who were homozygous for C282Y showed evidence of substantial iron deposition localized to brain regions related to movement (basal ganglia and cerebellum)[3] – a novel endophenotype which we refer to in the current work as the "Hemochromatosis Brain". Male individuals with this genotype also had nearly two-fold increased risk for developing movement disorders, like Parkinson's Disease (PD). In this paper we seek to ask whether there are other genetic routes leading to iron buildup observed in the "Hemochromatosis Brain," and does this degree of brain related iron build up predict risk for movement disorders.

To address this question, we trained a classifier to predict C282Y homozygosity status from T2-weighted MRI scans, which are sensitive to iron accumulation, in a subsample of 960 individuals taken from UK Biobank. We then deploy this classifier on 35,283 independent MRI images, also from UK Biobank non of whom are C282Y homozygotes, to generate a PolyVoxel Score (PVS)[4] for each individual capturing the degree to which they resemble the archetypal "Hemochromatosis Brain". Next, we demonstrate the utility of this PVS by showing that it specifically predicts risk for movement disorders in this independent sample. Finally, we perform a Genome Wide Association Study (GWAS) on the PVS to uncover the heritability and genetic variants leading to variability of iron accumulation localized to motor circuits of the brain.

Results

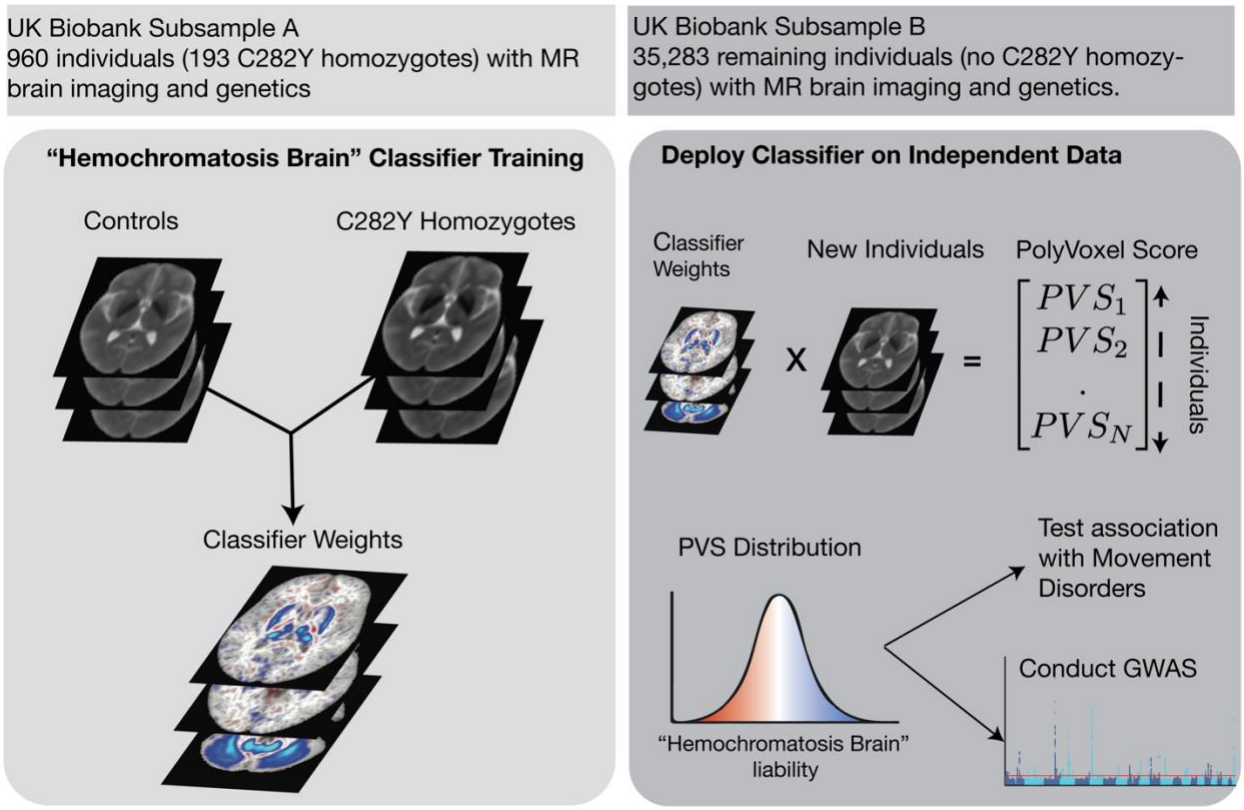

Figure 1 Overview of study design. Left panel: "Hemochromatosis Brain" Classifier training in Subsample A to differentiate controls from C282Y homozygotes from T2-weighted MRI scans. Right panel: deploy classifier in Subsample B to generate PolyVoxel Score (PVS) capturing "Hemochromatosis Brain" liability. We then test this PVS against movement (and other neurological) disorders and conduct GWAS to find variants associated with this liability.

	Subsample	
	A	**B**
n	960	35,283
Age, mean (SD)	64.3 (7.5)	64.1 (7.5)
Female, n (%)	575 (60.0)	18,378 (52.1)
C282Y Homozygote	193 (20.0)	0 (0.0)
C282Y Heterozygote	0 (0.0)	4,935 (14.0)

Table 1 Sample characteristics of two subsamples. Figure 1 displays how subsamples are used.

"Hemochromatosis Brain" Classifier Training

Figure 1 presents a graphical workflow of the analysis performed in this manuscript. Performing 5-fold cross validation in Subsample A, consisting of 193 C282Y homozygotes and 767 controls, established that the "Hemochromatosis Brain" classifier can predict C282Y status with high accuracy from T2-Weighted scans attaining an AUC=0.86, see Figure 2A. eFigure 1 shows results of hyperparameter tuning. Figure 2B displays classifier weights across the brain, as anticipated these show strong resemblance to regional association maps between T2 hypointensity and C282Y homozygosity shown in our previous work[33].

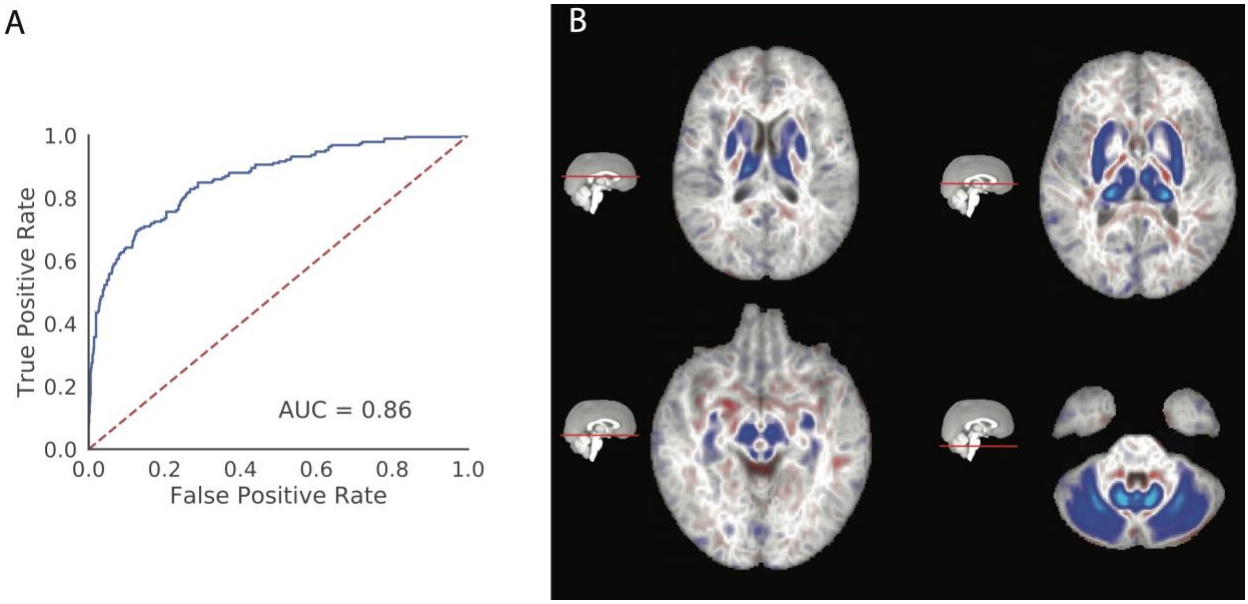

Figure 2 Panel A: Receiver Opererator Characteristic (ROC) curve of the "Hemochromatosis Brain" Classifier classifying controls vs C282Y homozygotes in 5-fold cross valdiation from Subsample A. Panel B: classifier weights across whole Subsample A (before regularization), units are arbitrary as PVS result is ultimately quantile transformed. Largest (negative) weights are found in basal ganglia and cerebelum; see[3] for more in depth discussion of anatomical distribution of association map between T2 hypointensity and C282Y homozygosity.

PolyVoxel Score Association with Neurological Disorders

Next, we deployed this classifier to unseen scans of 35,283 individuals from Subsample B (containing no C282Y homozygotes), to generate a PolyVoxel Score (PVS) for each individual quantifying their "Hemochromatosis Brain" liability as shown in right panel of Figure 1. Subsample B included no C282Y homozygotes, nevertheless, we found that this PVS significantly and specifically predicted risk for PD in this sample (OR=1.37, Z=2.62, p=0.0089) and showed trending associations for Abnormalities of Gait and Mobility (OR=1.13, Z=1.85, p=0.064) – see Figure 3A and eTable 1. Additionally, by segmenting the PVS into quartiles we observed that this measure showed a dosage dependent association with risk for PD, with individuals in the 4th quartile being over 3 times more likely to develop Parkinson's vs individuals in the 1st quartile (OR=3.28, Z=2.9, p=0.0035) – see Figure 3B and eTable 2. We did not find evidence of a PVS association with diagnoses of any other disorders tested.

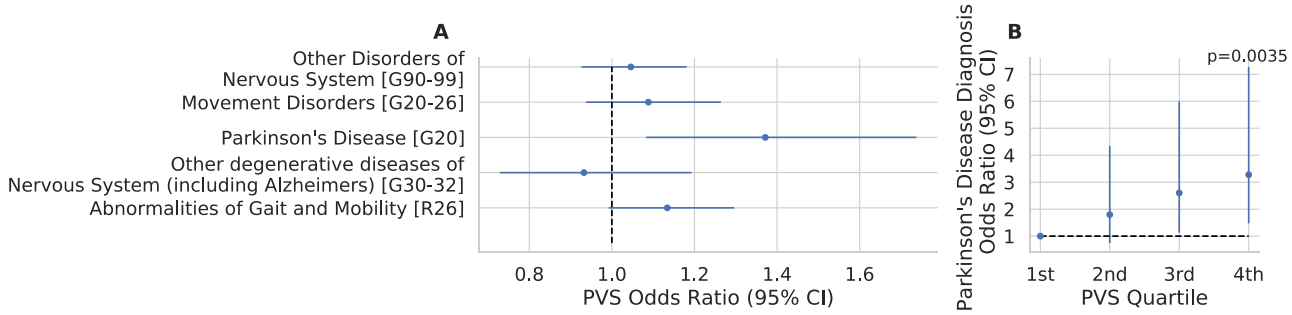

Figure 3 Association of PolyVoxel Score (PVS) with neurological diagnoses in Subsample B. Panel A: PVS odds ratio predicting each nuerological disorder. Panel B: quartile regression results associating each quartile of PVS with risk for PD (reference group is 1st quartile). eTable1 and eTable2 show regression results.

GWAS of "Hemochromatosis Brain"

We then sought to understand the polygenic signal associated with the "Hemochromatosis Brain" by conducting a Genome-Wide Association Study with the PVS in Subsample B as the phenotype of interest. We restricted this analysis to a single homogenous ancestry group, as is considered best practice for GWAS studies[5], reducing the sample size to 30,709 European ancestry individuals (containing no C282Y homozygotes). This analysis revealed 43 genome wide significant loci associated with the "Hemochromatosis Brain" phenotype and a high heritability: h^2_{ldsc}=0.381 (0.033) – see Figure 4A for Manhattan and QQ plots. Among the top hits we found many that are known regulators of iron homeostasis: for example, rs6794370 (TF; p=2.43$\times 10^{-81}$) transferrin is the principal glycoprotein responsible for mediating transport of iron through blood plasma; rs1800562 (HFE; p=6.93$\times 10^{-75}$) is the same C282Y mutation on which the classifier was trained (rediscovered with the heterozygotes); rs2413450 (TMPRSS6, p=2.52$\times 10^{-50}$), TMPRSS6 is part of production signaling pathway of hepcidin, a key hormonal regulator of iron absorption in humans; rs13107325 (p=3.66$\times 10^{-41}$) and rs12304921 (p=5.05$\times 10^{-21}$) are linked respectively to metal transporters ZIP8 and DMT1, the latter being a gene highly expressed in the gut and previously linked with microcytic anemia with iron overload[6,7].

Running FUMA[8] on our GWAS summary statistics, we found 281 independent significant SNPs. These independent significant SNPs overlapped with hits from a total of 293 previous studies in GWAS catalog. Figure 4B shows the overlap of our discovered SNPs with broad categories of previously reported hits (see eTable 3 for terms used to define categories). We see "Iron/Red Blood Cells" is the largest category of previously reported results, followed by "Blood Markers (non Iron)" and "Brain/Cognition" traits. For this analysis we only found one independent significant SNP previously reported for PD rs13107325, discussed above, which was previously found to have opposing effects on PD and schizophrenia risk[9,10].

Figure 4C shows LDSC[11] (linkage disequilibrium score regression) derived genetic correlations of our PVS summary statistics results with a) results from us performing GWAS in the same UK Biobank Subsample B and b) relevant publicly available summary statistics. The only significant genetic correlation with the PVS was with intracranial volume (r_g=-0.21, SE=(0.04), p_{FDR}=1.7$\times 10^{-6}$). There were no significant genetic correlations between the "Hemochromatosis Brain" PVS and any other measures tested. We performed S-LDSC (stratified-LDSC) across 489 tissue-specific chromatin-based annotations across the body from peaks for six epigenetic marks[12]. Results revealed the hippocampus (H3K27ac, p_{FDR}=0.014 and H3K4me1, p_{FDR}=0.025) and substantia nigra (H3k27ac, p_{FDR}=0.036) displayed a significantly enriched signal. Performing cell type enrichment using FUMA[13] of "Hemochromatosis Brain" GWAS using cell types defined from PsychENCODE[14] displayed an enrichment for glial cells: oligodendrocytes (p_{FDR}=6.62$\times 10^{-5}$), astrocytes (p_{FDR}=1.7$\times 10^{-3}$) and microglia (p_{FDR}=1.8$\times 10^{-2}$) and no enrichment for neuronal cells– see Figure 4D. This pattern of enrichment exhibits a similar pattern followed by iron concentrations estimated from spectroscopy[15] – displayed in Figure 4E for comparison. See extended data tables for full summary of GWAS results.

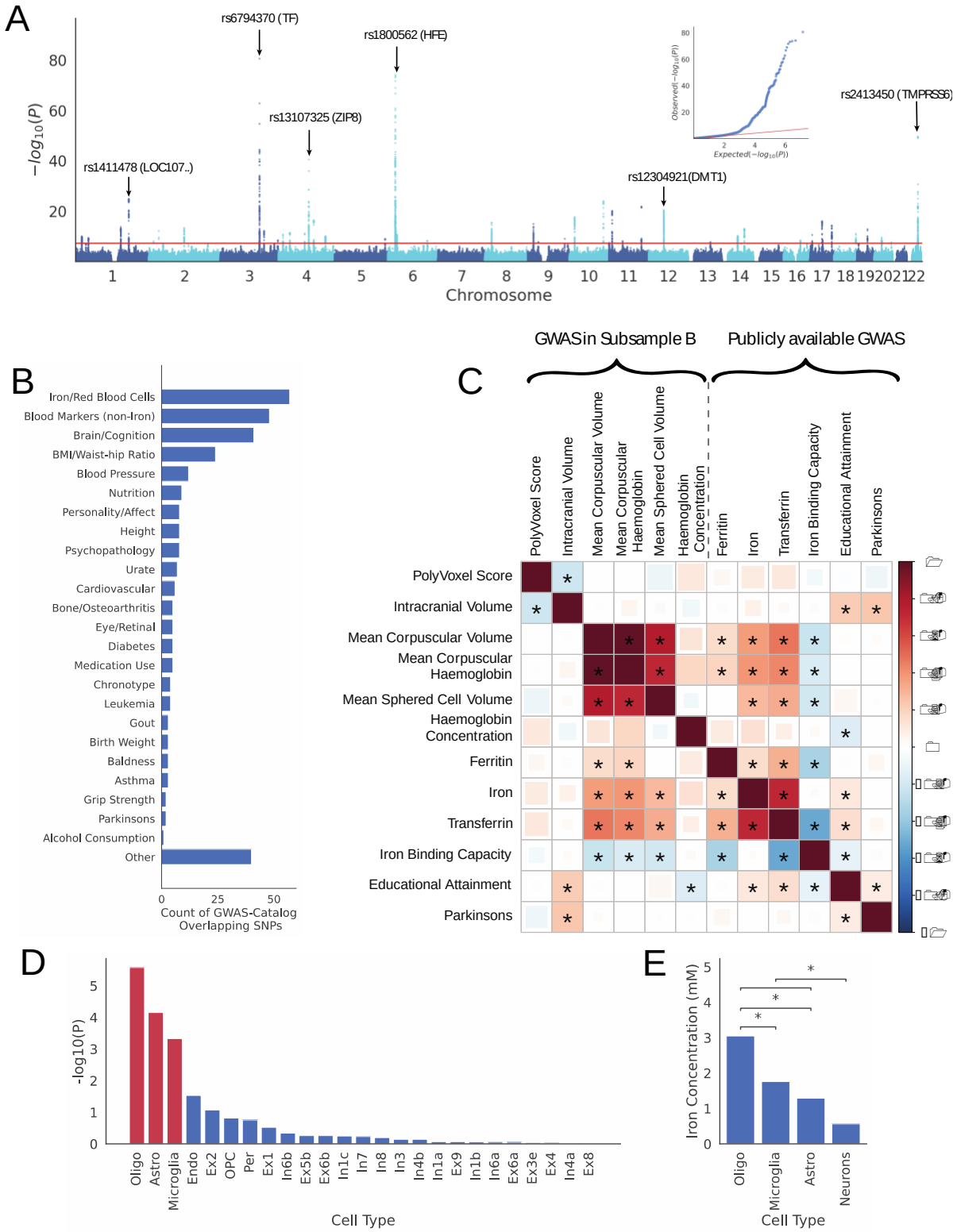

Figure 4 Results of GWAS on PVS of "Hemochromatosis Brain" in Subsample B. Panel A: Manhattan plot and QQ plot of GWAS result with peaks annotated – a total of 43 loci are discovered, see supplementary data tables for full list. Panel B: overlap of independent significant SNPs with previously reported SNPs in GWAS-Catalog. Panel C: LDSC genetic correlations of PVS with iron, brain, and cognitive measures, summary statistics were either generated on the same subsample B or publicly available.

Stars indicate FDR significant associations. Panel D: Cell type enrichment using FUMA, red bars indicate FDR significantly enriched cell types. Panel E: Iron concentration (estimated uing spectroscopy) of cell types with data from[15].

Discussion

In this manuscript we have presented a novel imaging endophenotype: the "Hemochromatosis Brain". This measure quantifies the degree to which individuals exhibit the archetypal pattern of T2-weighed signal suggest of iron deposition in motor circuits observed in MRI scans of C282Y homozygotes. We show that this measure predicts risk for Parkinson's disease and that this association displays a dose dependent response. Furthermore, we perform a GWAS on this "Hemochromatosis Brain" phenotype to show that it displays a large SNP-heritability (h^2_{ldsc}=0.381) given our modest sample size by GWAS standards. The top loci from this analysis mapped to genes known to regulate iron homeostasis. Furthermore, analysis revealed that the implicated genes were specifically enriched in glial cells – a cell group known to have higher concentrations of iron[15].

The results from this analysis validate an established relationship between iron and movement disorders, while providing a novel biomarker for PD risk. Previous post mortem and in vivo imaging studies have found higher iron deposition in substantia nigra, putamen and red nucleus[16], all regions with large weightings from the "Hemochromatosis Brain" classifier (Figure 2B and previous work[3]). Additionally, deficits in iron have also been associated with movement disorders, with anemic patients and male blood donors being more likely to suffer from PD[17,18]. Furthermore, Neurodegeneration with Brain Iron Accumulation describes a class of rare genetic mutations that result in iron dysregulation and a stereotyped pattern of iron accumulation in the basal ganglia consistent with the pattern observed in the "Hemochromatosis Brain"[19], with a clinical presentation of dystonia and other Parkinsonism symptoms.

Although epidemiological, imaging and postmortem studies suggest a link between PD and iron, GWAS of the disorder have not uncovered a major contribution from iron[20]. Indeed, in our study we only find one independent significant SNP (rs13107325, p=3.66$\times 10^{-41}$) overlapping with variants previously associated with PD. We also find that the "Hemochromatosis Brain" does not display any genetic correlation with the largest previous GWAS of PD. Our interpretation of this result is that iron dysregulation may represent one potential contributor to PD risk and without this subtype being identified in current PD GWAS these unmodelled heterogeneous genetic effects lead to null findings for iron related variants. This idea of genetic heterogeneity is thought to explain the modest GWAS results of Major Depression[21], and we aim to assess this explanation of our results in future work. Much of the current PD results highlight lysosome dysfunction as a principal mechanism underlying disease pathpysiology[22–24]. The contribution of iron dysregulation to PD risk may still converge on lysosome dysfunction. For example iron overload leading to oxidative damage and cell death could lead to deficits in clearing the damaged tissue, and ultimately disease presentation; this process would be exacerbated in patients with pre-existing lysosomal dysfunction.

Single cell enrichment analysis implicated glial, and not neuronal, cells as the principal cell class associated with our "Hemochromatosis Brain" GWAS results. This is consistent with findings that glial cells, particularly oligodendrocytes, showing larger concentrations of iron than neuronal cell types[15,25] and is also consistent with role of glial cells' in maintaining brain homeostasis[26]. Analysis of tissues specific enrichment (across the human body) found the

substantia nigra and hippocampus to be associated our GWAS discoveries. The former being the primary site of PD pathology[24] and the latter known to show accelerated atrophy in PD[27]. These provide further support that "Hemochromatosis Brain" represents an iron specific endophenotype of relevance to PD.

An important limitation of the current study is its observational nature. Although, T2-weighted MRI signal is known to be related to iron concentrations[28], and the GWAS signal we detect appears to implicate iron related genes we cannot conclude that iron specifically is imparting a causal impact on disease risk for PD. It is possible that iron may be serving as a marker for another biological process that is responsible for generating the differential risk for PD. In addition, T2-weighted signal is not solely reduced by higher iron concentrations; processes of edema and gliosis also impact T2-weighted intensity[29]. It is possible that some of the effects captured by the PVS are in fact related to these non-iron processes.

Here we have showcased a novel approach combining both forward and reverse genetics to elucidate and define an endophenotype of PD that could provide a marker for overall PD risk and an iron dysregulation subtype of the disorder. Beyond suggesting potential new directions for PD research, this approach could be applied more broadly to other genetic disorders that can be associated with a multivariate phenotype.

Methods

UK Biobank Sample

Genotypes, MRI scans, demographic and clinical data were obtained from the UK Biobank under accession number 27412, excluding 206 participants who withdrew their consent. All participants provided electronic signed informed consent and the study was approved by the UK Biobank Ethics and Governance Council. This resulted in a total sample of 36,243 individuals (18,953 females) with a mean age of 64.1 years (standard deviation 7.5 years), with genetics and qualified imaging. We split the sample into two non-overlapping groups: Subsample A for C282Y homozygote classifier training and Subsample B for deploying this classifier. Table 1 shows summarizes demographics of these two subsamples. The recruitment period for participants was from 2006 to 2010. Genotype, health record and neuroimaging data was collected from January 2006 to May 2021. Data analysis was conducted from January 2022 to October 2022. This study follows the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline for cross-sectional studies[30].

Image acquisition

T1 weighted and diffusion weighted scans were collected from three scanning sites throughout the United Kingdom, all on identically configured Siemens Skyra 3T scanners, with 32-channel receiver head coils. For diffusion scans, multiple scans with no diffusion gradient were collected (b=0 s/mm^2) to fit diffusion models. The average of these b=0 scans was used as voxel-wise measures of T2-weighted intensities. Diffusion-weighted scans were collected using

a SE-EPI sequence at 2mm isotropic resolution. T1-weighted scans were collected using a 3D MPRAGE sequence at 1mm isotropic resolution. If a person had multiple (longitudinal) scans we used the first scan. Further details of image acquisition can be found here[31].

Image Preprocessing

Scans were corrected for nonlinear transformations provided by MRI scanner manufacturers[32,33], and T2-weighted and T1-weighted images were co-registered using mutual information[34]. Intensity inhomogeneity correction was performed by applying smoothly varying, estimated B1-bias fields[35]. Images were rigidly registered and resampled into alignment with a pre-existing, in-house, averaged, reference brain with 1.0 mm isotropic resolution[35].

C282Y Homozygote Classifier and PolyVoxel Score Generation

We aimed to train a classifier to predict C282Y homozygosity status from MRI scans. For training we generated a sample derived from the 193 C282Y homozygotes with qualified imaging and found covariate matched controls at a ratio of 4:1 (cases to controls)[36]. Controls were matched for sex, age, scanner and top ten principal components of genetic ancestry. This was performed as described elsewhere[3] with the exception that controls were selected as those with no C282Y mutations (i.e. C282Y heterozygotes were excluded from being cases or controls). This resulted in a subsample of 193 C282Y "cases" (112 female) and 767 "controls" (463 female), i.e. no C282Y mutations, for training the classifier – we refer to this sample as Subsample A. For fitting, hyperparameter tunning and evaluating the classifier we employed a 5 fold cross-validation scheme in Subsample A.

The C282Y homozygote classifier was used fit using a PolyVoxel Score (PVS) framework[4] as follows. Let $\boldsymbol{X^{train}}$ represent the pre-residualized imaging matrix with *N* rows of individuals and *M* columns of voxels in the training sample. Each element of this matrix, $x_{i,j}^{train}$ represents the voxel intensity for the i-th individual at the j-th voxel for a T2-weighted scan. Here each $\boldsymbol{X^{train}}$ was quantile transformed to a normal distribution and pre-residualized for covariates of age, sex, MRI scanner and top 10 principal components of genetics. Let $\mathbf{y}^{\boldsymbol{train}}$ represent the 1-dimensional vector such that:

$$\mathbf{y}^{\mathbf{train}} = \begin{cases} 1, & \text{if C282Y homozygote} \\ 0, & \text{otherwise} \end{cases}$$

We then regress each column of $\boldsymbol{X^{train}}$ with $\mathbf{y}^{\boldsymbol{train}}$ to generate a vector of univariate z-statistics, $\mathbf{z}$, of dimension *M* – where each element describes the univariate association of that voxel with C282Y homozygosity status. We then threshold this $\mathbf{z}$ to restrict to nominally significant voxels in which p<0.01. Next, we could use the vector $\mathbf{z}$ to generate out of sample PVS's from a test sample's imaging data, $\mathbf{X}^{\boldsymbol{test}}$, to estimate liability of individuals along the "Hemochroamtosis Brain" liability scale as follows: $PVS = \mathbf{z}\boldsymbol{X}^{\mathbf{test}}$ (these $\mathbf{z}$ statistics are displayed in Figure 2B). However, due to the correlation structure across voxels this gives suboptimal out of sample prediction. As such we use the correlation matrix of $\boldsymbol{X}^{train}$, $\boldsymbol{R}$, to re-

weight $\mathbf{z}$. $\boldsymbol{R}$ can be decomposed using singular value decomposition $\boldsymbol{R} = \boldsymbol{USU}^T$ ($\boldsymbol{U}$ – unitary matrix, $\boldsymbol{S}$ – diagonal matrix with singular values on its diagonal). We consider the regularized form of $\boldsymbol{R}$ as: $\boldsymbol{R_r} = \boldsymbol{US_rU}^T$, where $\boldsymbol{S}_\mathbf{r}$ is obtained from $\boldsymbol{S}$ by keeping r largest singular values and replacing the remaining with r^{th} largest. We then decorrelate association statistics to generate PVS weights as $\boldsymbol{w_r} = \boldsymbol{R_r}\mathbf{z}$, which were in turn used to generate PVS scores for each individual as $PVS_r = \boldsymbol{w_r}\ \boldsymbol{X}^{\mathbf{test}}$. During cross-validation in Subsample A nine PVS_r were generated with different values of r (1, 5, 10, 20, 50, 100, 200, 500, 1000 and M (i.e. no regularization)). The r value that maximized the association between PVS_r and $\mathbf{y}^{\boldsymbol{test}}$ was selected for displaying receiver operator characteristic (ROC) curves. For downstream analysis in Subsample B we used the mean optimal value, r=13, across folds of subsample A

Disease Enrichment of "Hemochromatosis Brain"

We wanted to test the utility and specificity of the "Hemochromatosis Brain" PVS in distinguishing cases and controls of PD and other disorders of the brain. To do this we defined Subsample B in UK Biobank who were the remainder of individuals with neuroimaging that were not in Subsample A. This resulted in 35,283 individuals (18,278 females, 0 C282Y homozygotes). For this sample we generated a PVS for each individual from their imaging data using the C282Y homozygote classifier, as described above, trained in the full Subsample A (regularization parameter set to r=13 as the mean optimal value across cross-validation). This PVS is a single score per individual that in essence quantifies the degree of iron accumulation in motor circuits observed in C282Y homozygotes. We then associated this "Hemochromatosis Brain" PVS with 5 disorders/classes of disorders of the brain using logistic models with covariates of age, sex and top ten components of genetic ancestry. These 5 different disorders/categories were: Other disorders of the Nervous System (ICD10: G90-99), Movement Disorders (ICD10: G20-G26), Parkinson's Disease (ICD10: G20), Other degenerative diseases of the Nervous System (including Alzheimer's) (ICD10: G30-32) and Abnormalities of Gait and Mobility (ICD10: R26). These diagnoses were extracted from UK Biobank field 41270. We additionally performed quartile regression, in which individuals were assigned to quartiles 1-4 according to their PVS with 1st being the lowest 25% of PVS, 2nd being scores between 25th and 50th percentile and so on. A model was then fit to predict PD from these quartiles as a categorical variables covarying for the same covariates described above.

GWAS of "Hemochromatosis Brain"

In Subsample B we then performed a GWAS with the "Hemochromatosis Brain" PVS as the phenotype of interest, covarying for age, sex and top ten components of genetic ancestry. We restricted this analysis to individuals declared as self-identified "White British" and similar genetic ancestry (using Data Field: 22006), this resulted in 30,709 individuals. Additionally, to enable the calculation of genetic correlations of related traits in UK Biobank, we performed GWAS for intracranial volume, as a measure of brain size, and the closest available proxies of blood iron[37]: mean corpuscular volume, mean corpuscular hemoglobin, mean sphere cell volume and hemoglobin concentration. If there were multiple instances of a variable we took their mean value for each subject. For these measures, as well as the PVS, we quantile transformed them to a normal distribution to enforce normality and reduce leverage of

outliers. We used sample and variant QC's of: --geno 0.05, --hwe 1e-12, --maf 0.005 and --mind 0.1 using PLINK (v2.00a3.6LM)[38]. This resulted in 7,131,446 remaining variants and no individuals removed. PLINK was used for performing GWAS across these 6 traits (using –glm).

Locus Definition and Discovery Overlap GWAS Catalog

Summary statistics were uploaded to FUMA (FUMA: v1.4.1, MAMGA v1.08)[8] for locus definition and gene mapping. FUMA provides an output of independent significant SNPs (genome-wide significant SNPs that are independent of each other at r^2<0.6) and finds which of these overlap with previous studies reported in GWAS catalog (e104_r2021-09-15). We use this to define broad categories of studies that overlapped with our discoveries, see eTable3 for search terms used to define categories. We enforced that each independent significant SNP could be counted for multiple categories (if it appeared in multiple studies across categories), but it could not be counted more than once within a category (in a case where a SNP occurred in multiple studies within the same category). The results of this analysis are shown in Figure 4B and extended data tables.

Genetic Correlation

LDSC[11,39] was used to estimate h^2_{snp} from summary statistics using 1k genomes European ancestry reference panel. LDSC was also used to calculate genetic correlations with GWAS of intracranial volume and four blood iron proxies that we conducted in Subsample B, as well as publicly available summary statistics of four blood iron traits[37], Parkinson's disease[20] and educational attainment[40].

Single Cell Enrichment

We performed cell type enrichment using the FUMA platform[8]. Association statistics at the gene level were calculated using MAGMA (v1.08)[41]. This gene-based distribution was then associated with expression patterns of 25 cell types from PsychENCODE single-cell RNA-seq data collected from the adult brain[14] by performing a gene-property analysis[13].

Tissue-Specific Enrichment

We used stratified LDSC (S-LDSC)[42] to examine tissue-type specific enrichment of the "Hemochromatosis Brain" GWAS results. We analyzed annotations derived from 489 tissue-specific chromatin assays of bulk tissue from peaks for six epigenetic marks. These data were collected as part of the Roadmap Epigenomics and ENCODE projects[43,44], annotations were downloaded from previous analysis[12] (see URLs). We controlled for baseline annotations as recommended by S-LDSC[42]. We reported the signed enrichment Z statistics, as well as the corresponding FDR corrected multiple comparisons adjusted p-values.

Data Availability: UK Biobank data was accessed under accession number 27412. Summary statistics from GWAS of "Hemochromatosis Brain" PVS are available for download. Researchers can apply for access to UK Biobank data at https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access.

Code Availability: We have created a python command line program to generate PVS's for T2-Weighted images registered in Montreal Neurological Institute space using classification weights presented in this analysis - https://github.com/robloughnan/pvs.

UK biobank Statement: UK Biobank's research resource is a major contributor in the advancement of modern medicine and treatment, enabling better understanding of the prevention, diagnosis and treatment of a wide range of serious and life-threatening illnesses – including cancer, heart diseases and stroke. UK Biobank is generously supported by its founding funders the Wellcome Trust and UK Medical Research Council, as well as the Department of Health, Scottish Government, the Northwest Regional Development Agency, British Heart Foundation and Cancer Research UK. The organisation has over 150 dedicated members of staff, based in multiple locations across the UK.

URLS
PLINK (https://www.cog-genomics.org/plink/2.0/)
FUMA (https://fuma.ctglab.nl)
LDSC and sLDSC (https://github.com/bulik/ldsc)
Annotations for sLDSC tissue-specific analysis (https://alkesgroup.broadinstitute.org/LDSCORE/)

Funding: This work was supported by grants R01MH122688 and RF1MH120025 funded by the National Institute for Mental Health (NIMH) in addition to Lundbeck Foundation fellowship (R335-2019-2318).

Conflict of Interest Statement: Dr. Andreassen has received speaker's honorarium from Lundbeck, and is a consultant to HealthLytix. Dr. Dale is a Founder of and holds equity in CorTechs Labs, Inc, and serves on its Scientific Advisory Board. He is a member of the Scientific Advisory Board of Human Longevity, Inc. and receives funding through research agreements with General Electric Healthcare and Medtronic, Inc. The terms of these arrangements have been reviewed and approved by UCSD in accordance with its conflict of interest policies. The other authors declare no competing interests.

1. Bomford, A. Genetics of haemochromatosis. *Lancet* **360**, 1673–1681 (2002).
2. Powell, L. W., Seckington, R. C. & Deugnier, Y. Haemochromatosis. *Lancet* **388**, 706–716 (2016).
3. 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).
4. Loughnan, R. J. *et al.* Generalization of cortical MOSTest genome-wide associations within and across samples. *Neuroimage* **263**, 119632 (2022).
5. Marees, A. T. *et al.* A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int. J. Methods Psychiatr. Res.* **27**, 1–10 (2018).
6. Beaumont, C. *et al.* Two new human DMT1 gene mutations in a patient with microcytic anemia, low ferritinemia, and liver iron overload. *Blood* **107**, 4168–4170 (2006).
7. Bardou-Jacquet, E. *et al.* A novel N491S mutation in the human SLC11A2 gene impairs protein trafficking and in association with the G212V mutation leads to microcytic anemia and liver iron overload. *Blood Cells. Mol. Dis.* **47**, 243–248 (2011).
8. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
9. Smeland, O. B. *et al.* Genome-wide Association Analysis of Parkinson's Disease and Schizophrenia Reveals Shared Genetic Architecture and Identifies Novel Risk Loci. *Biol. Psychiatry* **89**, 227–235 (2021).
10. Pickrell, J. K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48**, 709–717 (2016).
11. Bulik-Sullivan, B. *et al.* LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
12. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
13. Watanabe, K., Umićević Mirkov, M., de Leeuw, C. A., van den Heuvel, M. P. & Posthuma, D. Genetic mapping of cell type specificity for complex traits. *Nat. Commun.* **10**, (2019).
14. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science (80-.).* **8464**, (2018).
15. Reinert, A., Morawski, M., Seeger, J., Arendt, T. & Reinert, T. Iron concentrations in neurons and glial cells with estimates on ferritin concentrations. *BMC Neurosci.* **20**, 1–14 (2019).
16. Wang, J. Y. *et al.* Meta-analysis of brain iron levels of Parkinson's disease patients determined by postmortem and MRI measurements. *Sci. Rep.* **6**, 1–13 (2016).
17. Savica, R. *et al.* Anemia or low hemoglobin levels preceding Parkinson disease. *Neurology* **73**, 1381 LP – 1387 (2009).
18. Logroscino, G., Chen, H., Wing, A. & Ascherio, A. Blood donations, iron stores, and risk of Parkinson's disease. *Mov. Disord.* **21**, 835–838 (2006).
19. Hayflick, S. J., Kurian, M. A. & Hogarth, P. Chapter 19 - Neurodegeneration with brain iron accumulation. in *Neurogenetics, Part I* (eds. Geschwind, D. H., Paulson, H. L. & Klein, C. B. T.-H. of C. N.) **147**, 293–305 (Elsevier, 2018).
20. 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).
21. Flint, J. & Kendler, K. S. The Genetics of Major Depression. *Neuron* **81**, 484–503 (2014).
22. Wallings, R. L., Humble, S. W., Ward, M. E. & Wade-Martins, R. Lysosomal Dysfunction at the Centre of Parkinson's Disease and Frontotemporal Dementia/Amyotrophic Lateral Sclerosis. *Trends Neurosci.* **42**, 899–912 (2019).
23. Navarro-Romero, A., Montpeyó, M. & Martinez-Vicente, M. The Emerging Role of the Lysosome in Parkinson's Disease. *Cells* **9**, 1–25 (2020).
24. Poewe, W. *et al.* Parkinson disease. *Nat. Rev. Dis. Prim.* **3**, 1–21 (2017).
25. Connor, J. R. & Menzies, S. L. Relationship of iron to oligodendrocytes and myelination. *Glia* **17**, 83–93 (1996).
26. Jäkel, S. & Dimou, L. Glial cells and their function in the adult brain: A journey through the history of their ablation. *Front. Cell. Neurosci.* **11**, 1–17 (2017).
27. Camicioli, R. *et al.* Parkinson's disease is associated with hippocampal atrophy. *Mov. Disord.* **18**, 784–790 (2003).
28. Ordidge, R. J., Gorell, J. M., Deniau, J. C., Knight, R. A. & Helpern, J. A. Assessment of relative brain iron concentrations using T2-weighted and T2*-weighted MRI at 3 Tesla. *Magn. Reson. Med.* **32**, 335–341 (1994).
29. Prayson, R. *Neuropathology*. (2011).
30. Von Elm, E. *et al.* The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Epidemiology* **18**, 800–804 (2007).
31. Alfaro-Almagro, F. *et al.* Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage* **166**, 400–424 (2018).
32. Jovicich, J. *et al.* Reliability in multi-site structural MRI studies: Effects of gradient non-linearity correction on phantom and human data. *Neuroimage* **30**, 436–443 (2006).
33. Wald, L., Schmitt, F. & Dale, A. Systematic spatial distortion in MRI due to gradient non-linearities. *Neuroimage* **13**, 50 (2001).
34. Wells, W. Multi-modal volume registration by maximization of mutual information. *Med. Image Anal.* **1**, 35–51 (1996).
35. Hagler, D. J. *et al.* Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. *Neuroimage* **202**, (2019).
36. Rothman, K., Lash, T., VanderWeele, T. & Haneuse, S. *Modern Epidemiology*. *Postgraduate Medical Journal* **63**, (Wolters Kluwer, 2021).
37. Bell, S. *et al.* A genome-wide meta-analysis yields 46 new loci associating with biomarkers of iron homeostasis. *Commun. Biol.* **4**, (2021).
38. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, (2015).
39. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
40. Okbay, A. *et al.* Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nat. Genet.* **54**, (2022).
41. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput. Biol.* **11**, 1–19 (2015).

42. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
43. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–329 (2015).
44. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).