Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson's disease

Eating Disorders Working Group of the Psychiatric Genomics Consortium; Lin, B.; Slagboom, P. Eline; op Landt Slof, M.C.T.; Boomsma, D.I.; Ligthart, Lannie; Posthuma, Danielle

published in
Nature Genetics
2020

DOI (link to publisher)
10.1038/s41588-020-0610-9

document version
Publisher's PDF, also known as Version of record

ARTICLES
https://doi.org/10.1038/s41588-020-0610-9
nature genetics

Check for updates

Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson's disease

Julien Bryois[1,209], Nathan G. Skene[2,3,4,5,209], Thomas Folkmann Hansen[6,7,8], Lisette J. A. Kogelman[6], Hunna J. Watson[9,10,11], Zijing Liu[4,5], Eating Disorders Working Group of the Psychiatric Genomics Consortium*, International Headache Genetics Consortium*, 23andMe Research Team*, Leo Brueggeman[12], Gerome Breen[13,14], Cynthia M. Bulik[1,9,15], Ernest Arenas[2], Jens Hjerling-Leffler[2]✉ and Patrick F. Sullivan[1,16]✉

Genome-wide association studies have discovered hundreds of loci associated with complex brain disorders, but it remains unclear in which cell types these loci are active. Here we integrate genome-wide association study results with single-cell transcriptomic data from the entire mouse nervous system to systematically identify cell types underlying brain complex traits. We show that psychiatric disorders are predominantly associated with projecting excitatory and inhibitory neurons. Neurological diseases were associated with different cell types, which is consistent with other lines of evidence. Notably, Parkinson's disease was genetically associated not only with cholinergic and monoaminergic neurons (which include dopaminergic neurons) but also with enteric neurons and oligodendrocytes. Using post-mortem brain transcriptomic data, we confirmed alterations in these cells, even at the earliest stages of disease progression. Our study provides an important framework for understanding the cellular basis of complex brain maladies, and reveals an unexpected role of oligodendrocytes in Parkinson's disease.

Understanding the genetic basis of complex brain disorders is critical for developing rational therapeutics. In the past decade, genome-wide association studies (GWASs) have identified thousands of highly significant loci[1–4]. However, interpretation of GWASs remains challenging. First, >90% of the identified variants are located in noncoding regions[5], complicating precise identification of risk genes. Second, extensive linkage disequilibrium present in the human genome confounds efforts to pinpoint causal variants. Finally, it remains unclear in which tissues and cell types these variants are active, and how they disrupt specific biological networks to impact disease risk.

Functional genomic studies of the brain are now seen as critical for interpretation of GWAS findings, as they can identify functional regions (for example, open chromatin, enhancers and transcription-factor-binding sites) and target genes (via chromatin interactions and expression quantitative trait loci)[6]. Gene regulation varies substantially across tissues and cell types[7,8], and hence it is critical to perform functional genomic studies in empirically identified cell types or tissues.

Multiple groups have developed strategies to identify tissues associated with complex traits[9–13], but few have focused on the identification of salient cell types within a tissue. Furthermore, previous studies used a small number of cell types derived from one or few different brain regions[3,11–17]. For example, we recently showed that, among 24 brain cell types, 4 types of neuron were consistently associated with schizophrenia[11]. We were explicit that this conclusion was limited by the relatively few brain regions studied; other cell types from unsampled regions could conceivably contribute to the disorder.

Here, we integrate a wider range of gene expression data—tissues across the human body and single-cell gene expression data from an entire nervous system—to identify tissues and cell types underlying a large number of complex traits (Fig. 1a,b). We find that psychiatric and cognitive traits are generally associated with similar cell types whereas neurological disorders are associated with different cell types. Notably, we show that Parkinson's disease is associated with cholinergic and monoaminergic neurons, enteric neurons and oligodendrocytes, providing new clues into its etiology.

[1]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [2]Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden. [3]UCL Institute of Neurology, Queen Square, London, UK. [4]Division of Brain Sciences, Department of Medicine, Imperial College, London, UK. [5]UK Dementia Research Institute at Imperial College, London, UK. [6]Danish Headache Center, Dept of Neurology, Copenhagen University Hospital, Glostrup, Denmark. [7]Institute of Biological Psychiatry, Copenhagen University Hospital MHC Sct Hans, Roskilde, Denmark. [8]Novo Nordic Foundations Center for Protein Research, Copenhagen University, Copenhagen, Denmark. [9]Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [10]School of Psychology, Curtin University, Perth, Western Australia, Australia. [11]Division of Paediatrics, School of Medicine, The University of Western Australia, Perth, Western Australia, Australia. [12]Department of Psychiatry, University of Iowa Carver College of Medicine, University of Iowa, Iowa City, IA, USA. [13]Institute of Psychiatry, MRC Social, Genetic and Developmental Psychiatry Centre, King's College London, London, UK. [14]National Institute for Health Research Biomedical Research Centre, South London and Maudsley National Health Service Trust, London, UK. [15]Department of Nutrition, University of North Carolina, Chapel Hill, NC, USA. [16]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA. [209]These authors contributed equally: Julien Bryois, Nathan G. Skene. *Lists of authors and their affiliations appear at the end of the paper. ✉e-mail: jens.hjerling-leffler@ki.se; patrick.sullivan@ki.se

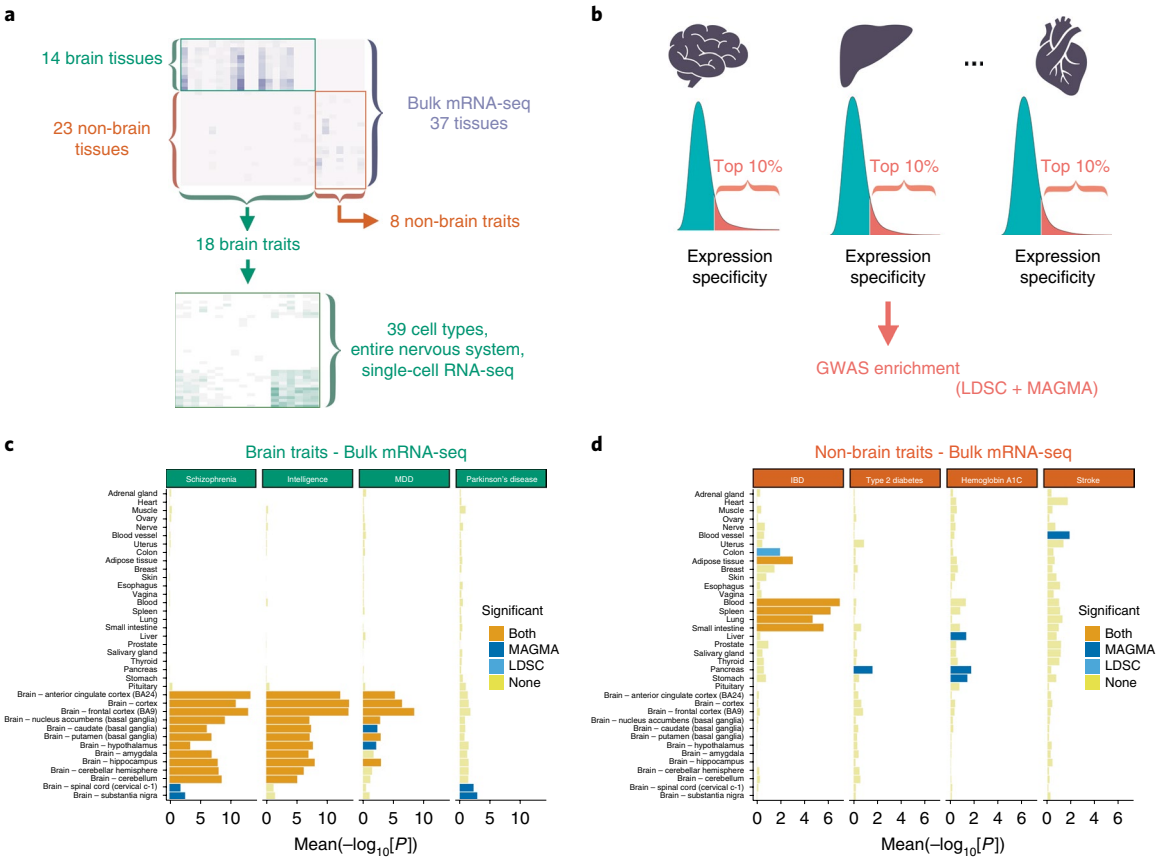


Fig. 1 | Study design and tissue-level associations. a, Heat map of associations between trait and tissue/cell type ($-\log_{10}[P]$) for the selected traits. **b**, Associations between trait and tissue/cell type were performed using MAGMA and LDSC (testing for enrichment in genetic association of the 10% most specific genes in each tissue/cell type). **c**, Tissue–trait associations for selected brain-related traits. **d**, Tissue–trait associations for selected non-brain-related traits. The mean strength of association ($-\log_{10}[P]$) of MAGMA and LDSC is shown, and the bar color indicates whether the tissue is significantly associated with both methods, one method or none (significance threshold: 5% false discovery rate). IBD, inflammatory bowel disease.

Results

Association of traits with tissues by using bulk RNA sequencing. Our goal was to use GWAS results to identify relevant tissues and cell types. Our primary focus was human phenotypes whose etiopathology is based in the central nervous system (CNS). We thus obtained 18 sets of GWAS summary statistics for brain-related complex traits. For comparison, we included GWAS summary statistics for eight diseases and traits with large sample sizes whose etiopathology is not rooted in the CNS (Methods).

We first aimed to identify human tissues showing enrichment for genetic associations using bulk-tissue RNA sequencing (RNA-seq; 37 tissues) from the Genotype-Tissue Expression (GTEx) project[7]. To robustly identify tissues implied by these 26 GWASs, we used 2 approaches (MAGMA[18] and LDSC[12,19]) that employ different assumptions (Methods). For both methods, we tested whether the 10% most specific genes in each tissue were enriched in genetic associations with the different traits (Fig. 1b).

Examination of non-brain-related traits found, as expected, associations with salient tissues. For example, as shown in Fig. 1d and Supplementary Table 1, inflammatory bowel disease was strongly associated with immune tissues (blood and spleen) and alimentary tissues impacted by the disease (small intestine and colon). Lung and adipose tissues were also significantly associated with inflammatory bowel disease, possibly because of the high specificity of immune genes in these two tissues (Extended Data Fig. 1). Type 2 diabetes was associated with the pancreas, while hemoglobin A1C, which is used to diagnose type 2 diabetes and monitor glycemic controls in individuals with diabetes, was associated with the pancreas, liver and stomach (Fig. 1d). Stroke and coronary artery disease were most associated with blood vessels and waist-to-hip ratio was most associated with adipose tissue (Fig. 1d and Supplementary Fig. 1).

For brain-related traits (Fig. 1c, Supplementary Fig. 1 and Supplementary Table 1), 13 of 18 traits were significantly associated with 1 or more GTEx brain regions. For example, schizophrenia, intelligence, educational attainment, neuroticism, body mass index (BMI) and major depressive disorder (MDD) were most significantly associated with the brain cortex, frontal cortex or anterior cingulate cortex, while Parkinson's disease was most significantly associated with the substantia nigra (as expected) and spinal cord (Fig. 1c). Alzheimer's disease was associated with tissues with prominent roles in immunity (blood and spleen) consistent with other studies[16,20,21], but also with the substantia nigra and spinal cord, while stroke was associated with blood vessels (consistent with a role of arterial pathology in stroke)[22].

In conclusion, we show that tissue-level gene expression allows identification of relevant tissues for complex traits, indicating that our methodology is suitable to explore associations between trait and gene expression at the cell-type level.

Association of brain complex traits with cell types. We leveraged gene expression data from 39 broad categories of cell types from the mouse central and peripheral nervous system[23] to systematically map brain-related traits to cell types (Fig. 2a and Extended Data Fig. 2).

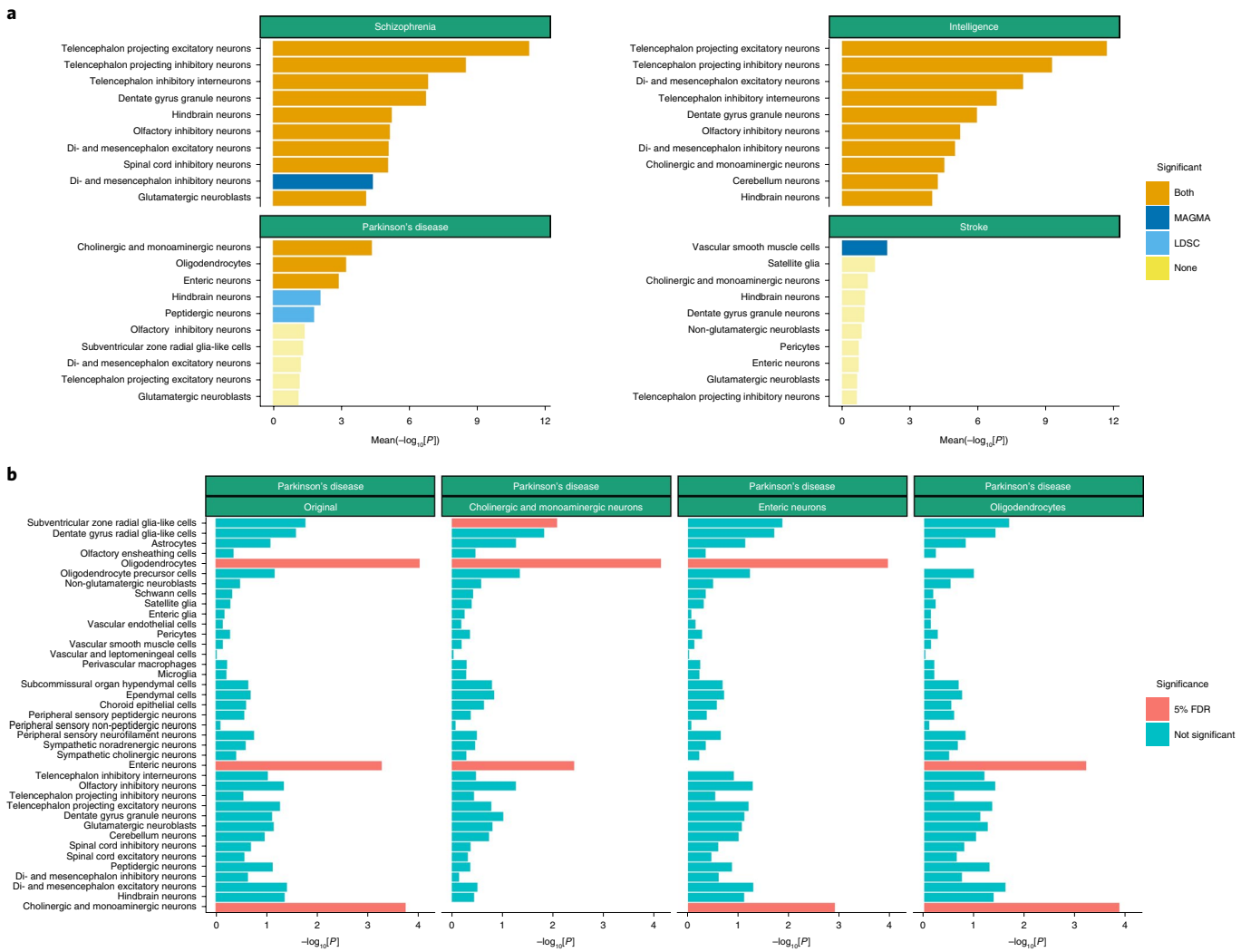


Fig. 2 | Association of selected brain-related traits with cell types from the entire nervous system. a, Associations of the 10 most associated cell types. **b**, Conditional analysis results for Parkinson's disease using MAGMA. The label indicates the cell type the association analysis is being conditioned on. The mean strength of association ($-\log_{10}[P]$) of MAGMA and LDSC is shown, and the bar color indicates whether the cell type is significantly associated with both methods, one method or none (significance threshold: 5% false discovery rate).

Our use of mouse data to inform human genetic findings was carefully considered (see Discussion).

As in our previous study of schizophrenia based on a small number of brain regions[11], we found the strongest signals for telencephalon projecting neurons (that is, excitatory neurons from the cortex, hippocampus and amygdala), telencephalon projecting inhibitory neurons (that is, medium spiny neurons from the striatum) and telencephalon inhibitory neurons (Fig. 2a and Supplementary Table 2). We also found that other types of neuron were associated with schizophrenia albeit less significantly (for example, dentate gyrus granule neurons). Other psychiatric and cognitive traits had similar cellular association patterns to schizophrenia (Extended Data Figs. 2 and 3 and Supplementary Table 2). We did not observe significant associations with immune or vascular cells for any psychiatric disorders or cognitive traits.

Neurological disorders generally implicated fewer cell types, possibly because the neurological GWAS had a lower signal than the GWASs of cognitive, anthropometric and psychiatric traits (Supplementary Fig. 2). Consistent with the genetic correlations (Supplementary Note), the pattern of associations for neurological disorders was distinct from that of psychiatric disorders (Extended Data Figs. 2 and 3), reflecting that neurological disorders have minimal functional overlap with psychiatric disorders[24].

Stroke was significantly associated with vascular smooth muscle cells (Fig. 2a), consistent with an important role of vascular processes for this trait. Alzheimer's disease had the strongest signal in microglia, as reported previously[10,16,25], but the association did not survive multiple testing correction.

We found that Parkinson's disease was significantly associated with cholinergic and monoaminergic neurons (Fig. 2a). This cluster consists of neurons (Supplementary Table 3) that are known to degenerate in Parkinson's disease[26–28], such as dopaminergic neurons from the substantia nigra (the hallmark of Parkinson's disease), but also serotonergic and glutamatergic neurons from the raphe nucleus[29], noradrenergic neurons[30], and neurons from afferent nuclei in the pons[31] and the medulla (the brain region associated with the earliest lesions in Parkinson's disease[26]). In addition, hindbrain neurons and peptidergic neurons were also significantly associated with Parkinson's disease (with LDSC alone). Interestingly, we also found that enteric neurons were significantly associated with Parkinson's disease (Fig. 2a), which is consistent with Braak's hypothesis, which postulates that Parkinson's disease could start in

the gut and travel to the brain via the vagus nerve[32,33]. Furthermore, we found that oligodendrocytes (mainly sampled in the midbrain, medulla, pons, spinal cord and thalamus; Supplementary Fig. 3) were significantly associated with Parkinson's disease, indicating a strong glial component to the disorder. This finding was unexpected but consistent with the strong association of the spinal cord at the tissue level (Fig. 1c), as the spinal cord contains the highest proportion of oligodendrocytes (71%) in the nervous system[23]. Together, these findings provide genetic evidence for a role of enteric neurons, cholinergic and monoaminergic neurons, and oligodendrocytes in Parkinson's disease etiology.

Neuronal prioritization in the mouse CNS. A key goal of this study was to prioritize specific cell types for follow-up experimental studies. As our metric of gene expression specificity was computed based on all cell types in the nervous system, it is possible that the most specific genes in a given cell type capture genes that are shared within a high-level category of cell types (for example, neurons). To rule out this possibility, we computed new specificity metrics based only on neurons from the CNS. We then tested whether the 10% most specific genes for each CNS neuron were enriched in genetic association for the brain-related traits that had a significant association with a CNS neuron (13/18) in our initial analysis.

Using the CNS neuron gene expression specificity metrics, we observed a reduction in the number of neuronal cell types associated with the different traits (Extended Data Fig. 4), suggesting that some of the signal was driven by core neuronal genes. However, we found that multiple neuronal cell types remained associated with a number of traits. For example, we found that telencephalon projecting excitatory and projecting inhibitory neurons were strongly associated with schizophrenia, bipolar disorder, educational attainment and intelligence using both LDSC and MAGMA. Similarly, telencephalon projecting excitatory neurons were significantly associated with BMI, neuroticism, MDD, autism and anorexia using one of the two methods, while hindbrain neurons and cholinergic and monoaminergic neurons remained significantly associated with Parkinson's disease.

Together, these results suggest that specific types of CNS neuron can be prioritized for follow-up experimental studies for multiple traits.

Trait and cell-type associations conditioning on other traits. As noted above, the patterns of associations of psychiatric and cognitive traits were highly correlated across the 39 different cell types tested (Extended Data Fig. 3). For example, the Spearman rank correlation of cell-type associations ($-\log_{10}[P]$) between schizophrenia and intelligence was 0.96 (0.94 for educational attainment) as both traits had the strongest signal in telencephalon projecting excitatory neurons and little signal in immune or vascular cells. In addition, we observed that genes driving the association signal in the top cell types of the two traits were enriched in relatively similar Gene Ontology (GO) terms involving neurogenesis and synaptic processes (Supplementary Note). We evaluated two possible explanations for these findings: schizophrenia and intelligence are both associated with the same genes that are specifically expressed in the same cell types; or schizophrenia and intelligence are associated with different sets of genes that are both specific to the same cell types. Given that these two traits have a significant negative genetic correlation ($r_g = -0.22$, from GWAS results alone) (Supplementary Table 4), we hypothesized that the strong overlap in cell-type associations for schizophrenia and intelligence was due to the second explanation.

To evaluate these hypotheses, we tested whether the 10% most specific genes for each cell type were enriched in genetic associations for schizophrenia controlling for the gene-level genetic association of intelligence using MAGMA (and vice versa) and found that the patterns of associations were largely unaffected. Similarly, we found that controlling for educational attainment had little effect on the schizophrenia associations and vice versa (Extended Data Fig. 5). In other words, genes driving the cell-type associations of schizophrenia appear to be distinct from genes driving the cell-type associations of cognitive traits.

Trait and cell-type associations conditioning on cell types. Many neuronal cell types passed our stringent significance threshold for multiple brain traits (Fig. 2a). This could be because gene expression profiles are highly correlated across cell types and/or because many cell types are independently associated with the different traits. To address this, we performed univariate conditional analysis using MAGMA, testing whether cell-type associations remained significant after controlling for the 10% most specific genes from other cell types (Supplementary Table 5). We observed that multiple cell types were independently associated with age at menarche, anorexia, autism, bipolar disorder, BMI, educational attainment, intelligence, MDD, neuroticism and schizophrenia (Supplementary Fig. 4). As in our previous study[11], we found that the association between schizophrenia and telencephalon projecting inhibitory neurons (that is, medium spiny neurons) was independent from telencephalon projecting excitatory neurons (that is, pyramidal neurons). For Parkinson's disease, enteric neurons, oligodendrocytes and cholinergic and monoaminergic neurons were independently associated with the disorder (Fig. 2b), suggesting that these three different cell types play an independent role in the etiology of the disorder.

Replication in other single-cell RNA-seq datasets. To assess the robustness of our results, we repeated these analyses in independent datasets. A key caveat is that these other datasets did not sample the entire nervous system as in the analyses above.

First, we used a single-cell RNA-seq dataset that identified 88 broad categories of cell types from 9 mouse brain regions[34]. We found similar patterns of association in this external dataset (Fig. 3a, Extended Data Fig. 6 and Supplementary Table 6). Notably, for schizophrenia, we strongly replicated associations with neurons from the cortex, hippocampus and striatum. We also observed similar cell-type associations for other psychiatric and cognitive traits (Fig. 3a and Extended Data Figs. 6 and 7). For neurological disorders, we found that stroke was significantly associated with mural cells while Alzheimer's disease was significantly associated with microglia (Extended Data Fig. 6). The associations of Parkinson's disease with neurons from the substantia nigra and oligodendrocytes were significant at a nominal level in this dataset ($P = 0.006$ for neurons from the substantia nigra; $P = 0.027$ for oligodendrocytes using LDSC). By computing gene expression specificity within neurons, we replicated our findings that neurons from the cortex can be prioritized for multiple traits (schizophrenia, bipolar disorder, educational attainment, intelligence, BMI, neuroticism, MDD and anorexia; Extended Data Fig. 8).

Second, we reanalyzed these GWAS datasets using our previous dataset[11] (24 cell types from 5 mouse brain regions; Fig. 3b, Extended Data Fig. 9 and Supplementary Table 7). We again found strong associations of pyramidal neurons from the somatosensory cortex, pyramidal neurons from region 1 of the cornu ammonis (CA1) of the hippocampus (both corresponding to telencephalon projecting excitatory neurons in our main dataset) and medium spiny neurons from the striatum (corresponding to telencephalon projecting inhibitory neurons) with psychiatric and cognitive traits. MDD and autism were most associated with neuroblasts, while intracranial volume was most associated with neural progenitors. The association of dopaminergic adult neurons with Parkinson's disease was significant at a nominal level using LDSC ($P = 0.01$), while oligodendrocytes did not replicate in this dataset, perhaps because they

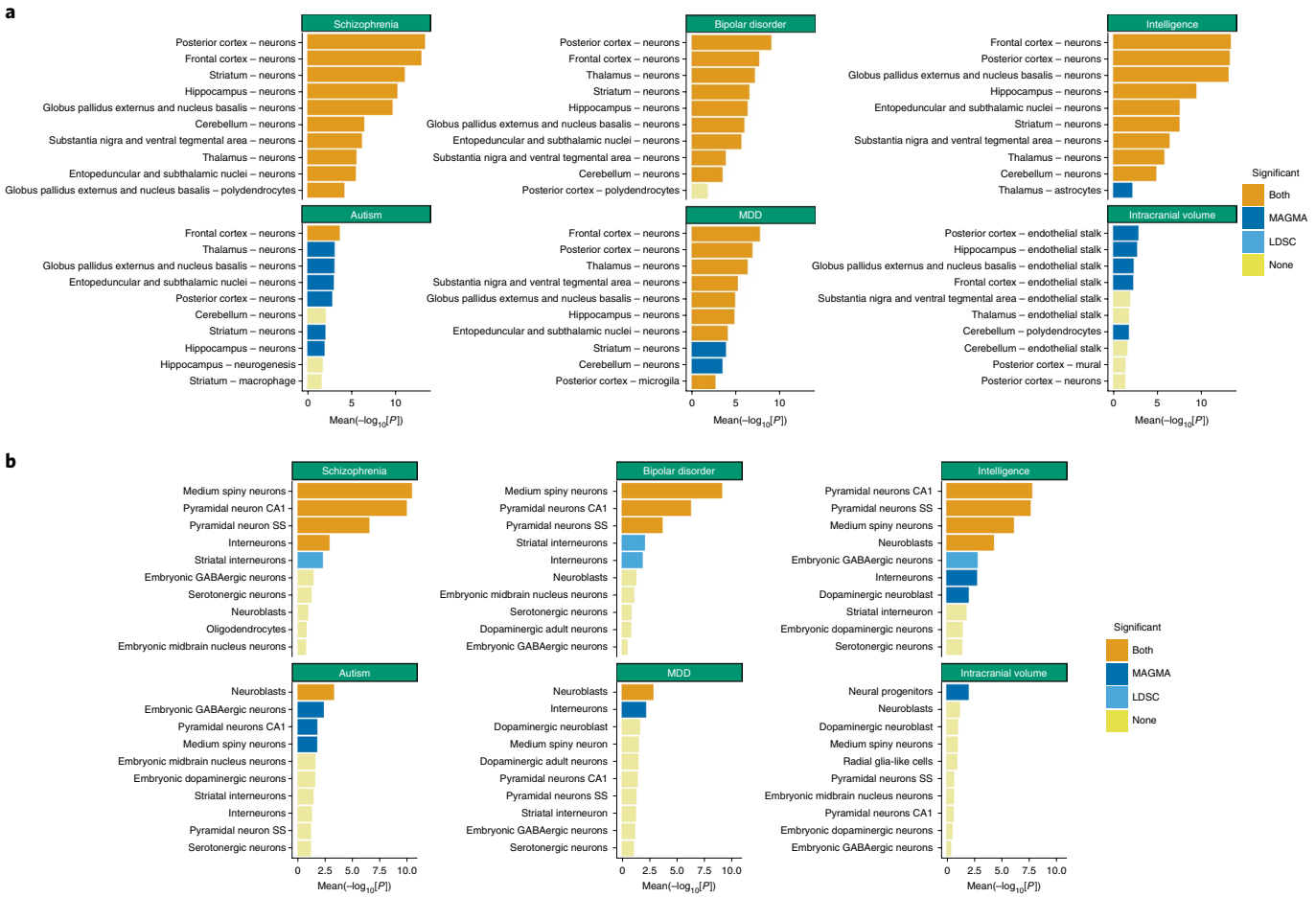


Fig. 3 | Replication of associations between cell type and trait in mouse datasets. a, Tissue–trait associations for the 10 most associated cell types among 88 cell types from 9 different brain regions. **b**, Tissue–trait associations for the 10 most associated cell types among 24 cell types from 5 different brain regions. The mean strength of association ($-\log_{10}[P]$) of MAGMA and LDSC is shown, and the bar color indicates whether the cell type is significantly associated with both methods, one method or none (significance threshold: 5% false discovery rate). SS, somatosensory cortex; CA1, cornu ammonis region 1.

were not sampled from the regions affected by the disorder (that is, spinal cord, pons, medulla or midbrain). A within-neuron analysis again found that projecting excitatory (that is, pyramidal CA1) and projecting inhibitory neurons (that is, medium spiny neurons) can be prioritized for multiple traits (schizophrenia, bipolar disorder, intelligence, educational attainment and BMI). In addition, neuroblasts could be prioritized for MDD and neural progenitors could be prioritized for intracranial volume (Extended Data Fig. 10).

Third, we evaluated a human dataset consisting of 15 different cell types from the cortex and hippocampus[35] (Fig. 4a and Supplementary Table 8). We replicated our findings with psychiatric and cognitive traits being associated with pyramidal neurons (excitatory) and interneurons (inhibitory) from the somatosensory cortex and hippocampus. We also replicated the association of Parkinson's disease with oligodendrocytes (enteric neurons and cholinergic and monoaminergic neurons were not sampled in this dataset). No cell types reached our significance threshold using specificity metrics computed within neurons, possibly because of similarities in the transcriptomes of neurons from the cortex and hippocampus.

Fourth, we evaluated a human dataset consisting of 35 different cell types from 3 different brain regions (visual cortex, frontal cortex and cerebellum) (Fig. 4b and Supplementary Table 9)[36]. We found that schizophrenia, educational attainment, neuroticism and BMI were associated with excitatory neurons, while bipolar disorder was associated with both excitatory and inhibitory neurons. As observed previously[10,16,25], Alzheimer's disease was significantly associated with microglia. Oligodendrocytes were not significantly associated with Parkinson's disease in this dataset, again possibly because the spinal cord, pons, medulla and midbrain were not sampled. No cell types reached our significance threshold using specificity metrics computed within neurons in this dataset.

Validation of oligodendrocyte pathology in Parkinson's disease. We investigated the role of oligodendrocytes in Parkinson's disease. First, we confirmed the association of oligodendrocytes with Parkinson's disease by combining evidence across all datasets (Fisher's combined probability test, $P=2.5\times10^{-7}$ using MAGMA and 6.3×10^{-3} using LDSC; Supplementary Table 2 and Supplementary Fig. 5). In addition, oligodendrocytes remained significantly associated with Parkinson's disease after conditioning on the top neuronal cell type in each dataset ($P=1.2\times10^{-7}$, Fisher's combined probability test).

Second, we tested whether genes with rare variants associated with parkinsonism (Supplementary Table 10) were specifically expressed in cell types from the mouse nervous system (Methods). As for the common variant, we found the strongest enrichment for cholinergic and monoaminergic neurons (Supplementary Table 11). However, we did not observe any significant enrichments for oligodendrocytes or enteric neurons for these genes.

Third, we applied expression-weighted cell-type enrichment (EWCE)[10] to test whether genes that are upregulated/downregulated in post-mortem brains from humans with Parkinson's disease

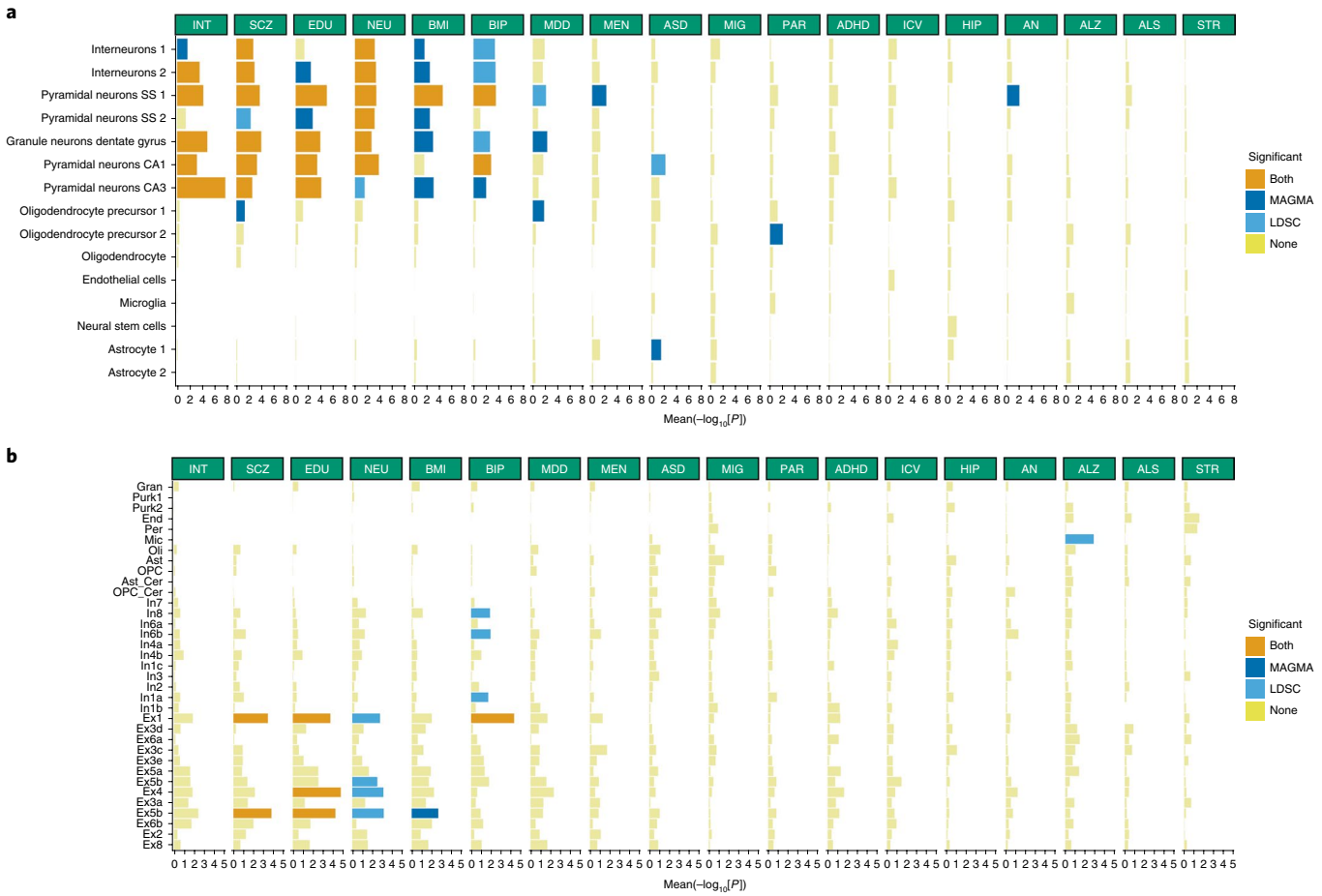


Fig. 4 | Human replication of associations between cell type and trait. a, Associations between cell type and trait for 15 cell types (derived from single-nuclei RNA-seq) from 2 different brain regions (cortex and hippocampus). **b**, Associations between cell type and trait for 35 cell types (derived from single-nuclei RNA-seq) from 3 different brain regions (frontal cortex, visual cortex and cerebellum). The mean strength of association ($-\log_{10}[P]$) of MAGMA and LDSC is shown, and the bar color indicates whether the cell type is significantly associated with both methods, one method or none (significance threshold: 5% false discovery rate). INT, intelligence; SCZ, schizophrenia; EDU, educational attainment; NEU, neuroticism; BMI, body mass index; BIP, bipolar disorder; MDD, major depressive disorder; MEN, age at menarche; ASD, autism spectrum disorder; MIG, migraine; PAR, Parkinson's disease; ADHD, attention deficit hyperactivity disorder; ICV, intracranial volume; HIP, hippocampal volume; AN, anorexia nervosa; ALZ, Alzheimer's disease; ALS, amyotrophic lateral sclerosis; STR, stroke. SS1, somatosensory cortex type 1; SS2, somatosensory cortex type 2; CA1, cornu ammonis region 1; CA3, cornu ammonis region 3.

(from six separate cohorts) were enriched in cell types located in the substantia nigra and ventral midbrain (Fig. 5). Three of the studies had a case–control design and measured gene expression in: the substantia nigra of 9 controls and 16 cases[37]; the medial substantia nigra of 8 controls and 15 cases[38]; and the lateral substantia nigra of 7 controls and 9 cases[38]. In all three studies, downregulated genes in Parkinson's disease were specifically enriched in dopaminergic neurons (consistent with the loss of this particular cell type in disease), while upregulated genes were significantly enriched in cells from the oligodendrocyte lineage. This suggests that an increased oligodendrocyte activity or proliferation could play a role in Parkinson's disease etiology. Surprisingly, no enrichment was observed for microglia, despite recent findings[39,40].

We also analyzed gene expression data from post-mortem human brains that had been scored by neuropathologists for their Braak stage[41]. Differential expression was calculated between brains with Braak scores of 0 (controls) and brains with Braak scores of 1–2, 3–4 and 5–6. At the later stages (Braak scores 3–4 and 5–6), downregulated genes were specifically expressed in dopaminergic neurons, while upregulated genes were specifically expressed in oligodendrocytes (Fig. 5), as observed in the case–control studies. Moreover, Braak stages 1 and 2 are characterized by little degeneration in the substantia nigra, and consistently, we found that downregulated genes were not enriched in dopaminergic neurons at this stage. Notably, upregulated genes were already strongly enriched in oligodendrocytes at Braak stages 1–2. These results not only support the genetic evidence indicating that oligodendrocytes may play a causal role in Parkinson's disease but also indicate that their involvement precedes the emergence of pathological changes in the substantia nigra.

Discussion

In this study, we used gene expression data from cells sampled from the entire nervous system to systematically map cell types to GWAS results from multiple psychiatric, cognitive and neurological complex phenotypes.

We note several limitations. First, we emphasize that we can implicate a particular cell type, but it is premature to exclude cell types for which we do not have data[11]. Second, we used gene expression data from mice to understand human phenotypes. We believe our approach is appropriate for several reasons. First, crucially, the key findings were replicated in human data. Second, single-cell RNA-seq is achievable in mouse but difficult in human neurons (where single-nuclei RNA-seq is typical[35,36,42,43]). In the brain, differences between single-cell and single-nuclei RNA-seq are important as transcripts

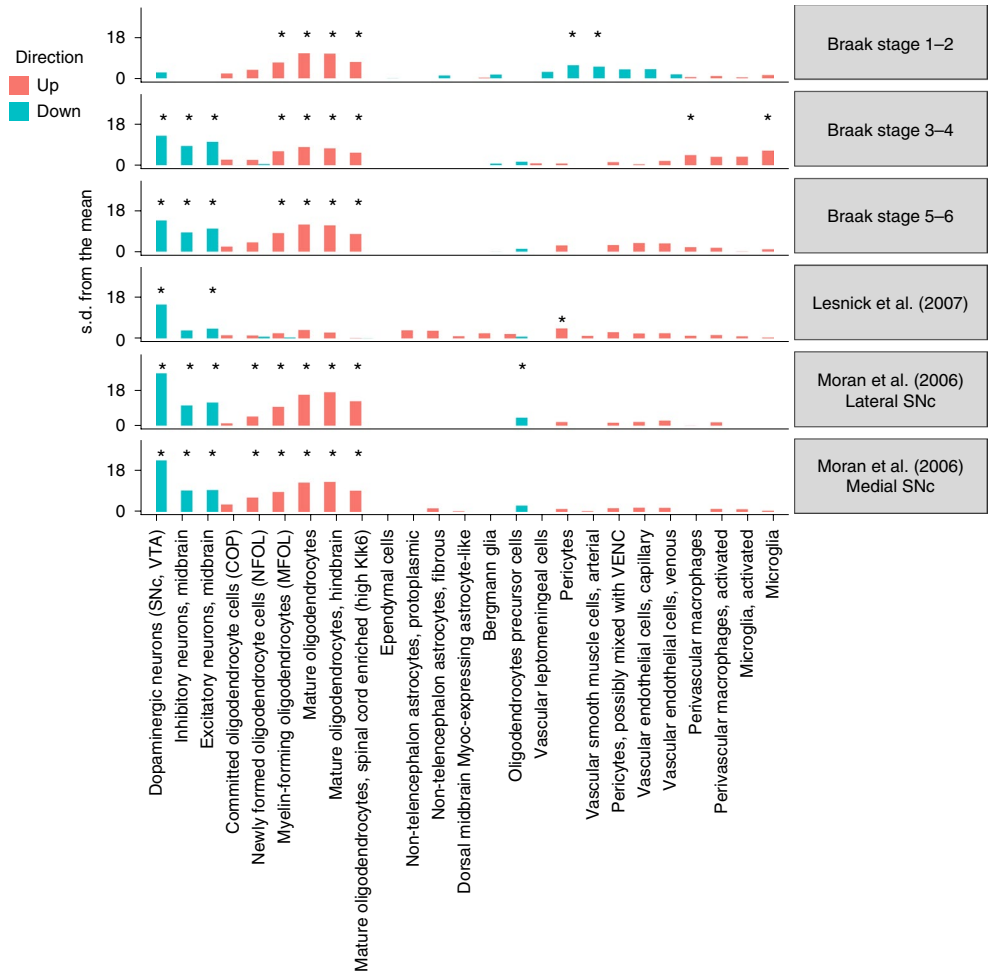


Fig. 5 | Enrichment of Parkinson's disease differentially expressed genes in cell types from the substantia nigra. Enrichment of the 500 most upregulated/downregulated genes (Braak stage 0 versus Braak stages 1–2, 3–4 and 5–6, as well as cases versus controls) in post-mortem human substantia nigra gene expression samples. The enrichments were obtained using EWCE[10]. An asterisk shows significant enrichments after multiple testing correction ($P < 0.05/(25 \times 6)$). SNc, substantia nigra par compacta; VTA, ventral tegmental area.

that are missed by sequencing nuclei are important for psychiatric disorders[11], and we previously showed that dendritically transported transcripts are specifically depleted from nuclei datasets[11] (confirmed in four additional datasets; Supplementary Fig. 6). Third, correlations in gene expression for cell type across species are high (median correlation 0.68; Supplementary Fig. 7), and as high as or higher than correlations across methods within cell type and species (single-cell versus single-nuclei RNA-seq, median correlation 0.6)[44]. Fourth, we evaluated only protein-coding genes with 1:1 orthologs between mice and humans, which are highly conserved. Fifth, we previously showed that gene expression data cluster by cell type and not by species[11], indicating broad conservation of core brain cellular functions across species. Sixth, we used a large number of genes to map cell types to traits (~1,500 genes for each cell type), minimizing potential bias due to individual genes differentially expressed across species. Seventh, if there were strong differences in cell-type gene expression between mice and humans, we would not expect that specific genes in mouse cell types would be enriched in genetic associations with human disorders. However, it remains possible that some cell types have different gene expression patterns between mice and humans, are present in only one species, have a different function or are involved in different brain circuits.

A third limitation is that gene expression data were from adolescent mice. Although many psychiatric and neurological disorders have onsets in adolescence, some have onsets earlier (autism) or later (Alzheimer's and Parkinson's disease). It is thus possible that some cell types are vulnerable at specific developmental times. Data from studies mapping cell types across brain development and aging are required to resolve this issue.

We found that psychiatric traits implicated largely similar cell types. These biological findings are consistent with genetic and epidemiological evidence of a general psychopathy factor underlying diverse psychiatric disorders[24,45,46]. Although intelligence and educational attainment implicated similar cell types, conditional analyses showed that the same cell types were implicated for different reasons. This suggests that different sets of genes highly specific to the same cell types contribute independently to schizophrenia and cognitive traits.

Our findings for neurological disorders were strikingly different from those for psychiatric disorders. We found, in contrast to previous studies that either did not identify any cell-type associations with Parkinson's disease[47] or identified significant associations with cell types from the adaptive immune system[40], that cholinergic and monoaminergic neurons (which include dopaminergic neurons), enteric neurons and oligodendrocytes were significantly and independently associated with the disease. Our findings suggest that dopaminergic neuron loss in Parkinson's disease (the hallmark of the disease) is at least partly due to intrinsic biological mechanisms.

Interestingly, enteric neurons were also associated with Parkinson's disease. This result is in line with prior evidence

implicating the gut in Parkinson's disease. Notably, dopaminergic defects and Lewy bodies (that is, abnormal aggregates of proteins enriched in α-synuclein) are found in the enteric nervous system of individuals affected by Parkinson's disease[48,49]. In addition, Lewy bodies have been observed in individuals up to 20 years before their diagnosis[50], and sectioning of the vagus nerve (which connects the enteric nervous system to the CNS) was shown to reduce the risk of developing Parkinson's disease[51]. Therefore, our results linking enteric neurons with Parkinson's disease provide new genetic evidence for Braak's hypothesis[32].

The association of oligodendrocytes with Parkinson's disease was more unexpected. A possible explanation is that this association could be due to a related disorder (for example, multiple-system atrophy, characterized by parkinsonism and accumulation of α-synuclein in glial cytoplasmic inclusions[52]). However, this explanation is unlikely as multiple-system atrophy is a very rare disorder; hence, only a few individuals could have been included in the Parkinson's disease GWAS. In addition, misdiagnosis is unlikely to have led to the association of Parkinson's disease with oligodendrocytes. Indeed, we found a high genetic correlation between self-reported diagnosis from the 23andMe cohort and a previous GWAS of clinically ascertained Parkinson's disease[53].

We did not find an association of oligodendrocytes with parkinsonism for genes affected by rare variants. This result may reflect etiological differences between sporadic and familial forms of the disease or low statistical power. Previous evidence has suggested an involvement of oligodendrocytes in Parkinson's disease. For example, α-synuclein-containing inclusions have been reported in oligodendrocytes in the brains of individuals with Parkinson's disease[54]. These inclusions ('coiled bodies') are typically found throughout the brainstem nuclei and fiber tracts[55]. Although the presence of coiled bodies in oligodendrocytes is a common, specific and well-documented neuropathological feature of Parkinson's disease, the importance of this cell type and its early involvement in disease has not been fully recognized. Our findings suggest that alterations in oligodendrocytes occur at an early stage of disease, which precedes neurodegeneration in the substantia nigra, arguing for a key role of this cell type in Parkinson's disease etiology.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-020-0610-9.

Received: 23 July 2019; Accepted: 6 March 2020;
Published online: 27 April 2020

References

1. Pardiñas, A. F. et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018).
2. Lee, J. J., Wedow, R. & Okbay Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
3. Nagel, M. et al. Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nat. Genet.* **50**, 920–927 (2018).
4. Yengo, L. et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).
5. Maurano, M. T. et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
6. Akbarian, S. et al. The PsychENCODE project. *Nat. Neurosci.* **18**, 1707–1712 (2015).
7. Aguet, F. et al. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
8. Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–329 (2015).
9. Ongen, H. et al. Estimating the causal tissues for complex traits and diseases. *Nat. Genet.* **49**, 1676–1683 (2017).
10. Skene, N. G. & Grant, S. G. N. Identification of vulnerable cell types in major brain disorders using single cell transcriptomes and expression weighted cell type enrichment. *Front. Neurosci.* **10**, 1–11 (2016).
11. Skene, N. G. et al. Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.* **50**, 825–833 (2018).
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. Calderon, D. et al. Inferring relevant cell types for complex traits by using single-cell gene expression. *Am. J. Hum. Genet.* **101**, 686–699 (2017).
14. Savage, J. E. et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* **50**, 912–919 (2018).
15. Coleman, J. R. I. et al. Biological annotation of genetic loci associated with intelligence in a meta-analysis of 87,740 individuals. *Mol. Psychiatry* **24**, 182–197 (2019).
16. Jansen, I. E. et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).
17. 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).
18. 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).
19. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
20. Jevtic, S., Sengar, A. S., Salter, M. W. & McLaurin, J. A. The role of the immune system in Alzheimer disease: etiology and treatment. *Ageing Res. Rev.* **40**, 84–94 (2017).
21. Kunkle, B. W. et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).
22. O'Leary, D. H. et al. Carotid-artery intima and media thickness as a risk factor for myocardial infarction and stroke in older adults. *N. Engl. J. Med.* **340**, 14–22 (1999).
23. Zeisel, A. et al. Molecular architecture of the mouse nervous system. *Cell* **174**, 999–1014.e22 (2018).
24. Anttila, V. et al. Analysis of shared heritability in common disorders of the brain. *Science* **360**, (2018).
25. Keren-Shaul, H. et al. A unique microglia type associated with restricting development of Alzheimer's disease. *Cell* **169**, 1276–1290.e17 (2017).
26. Braak, H. et al. Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol. Aging* **24**, 197–211 (2003).
27. Sulzer, D. & Surmeier, D. J. Neuronal vulnerability, pathogenesis, and Parkinson's disease. *Mov. Disord.* **28**, 41–50 (2013).
28. Poewe, W. et al. Parkinson disease. *Nat. Rev. Dis. Primers* **3**, 17013 (2017).
29. Halliday, G. M. et al. Neuropathology of immunohistochemically identified brainstem neurons in Parkinson's disease. *Ann. Neurol.* **27**, 373–385 (1990).
30. Delaville, C., de Deurwaerdère, P. & Benazzouz, A. Noradrenaline and Parkinson's disease. *Front. Syst. Neurosci.* https://doi.org/10.3389/fnsys.2011.00031 (2011).
31. Rinne, J. O., Ma, S. Y., Lee, M. S., Collan, Y. & Röyttä, M. Loss of cholinergic neurons in the pedunculopontine nucleus in Parkinson's disease is related to disability of the patients. *Parkinsonism Relat. Disord.* **14**, 553–557 (2008).
32. Braak, H., Rüb, U., Gai, W. P. & Del Tredici, K. Idiopathic Parkinson's disease: possible routes by which vulnerable neuronal types may be subject to neuroinvasion by an unknown pathogen. *J. Neural Transm.* **110**, 517–536 (2003).
33. Liddle, R. A. Parkinson's disease from the gut. *Brain Res.* **1693**, 201–206 (2018).
34. Saunders, A. et al. Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* **174**, 1015–1030.e16 (2018).
35. Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955 (2017).
36. Lake, B. B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018).
37. Lesnick, T. G. et al. A genomic pathway approach to a complex disease: axon guidance and Parkinson disease. *PLoS Genet.* **3**, 0984–0995 (2007).
38. Moran, L. B. et al. Whole genome expression profiling of the medial and lateral substantia nigra in Parkinson's disease. *Neurogenetics* **7**, 1–11 (2006).
39. Kannarkat, G. T., Boss, J. M. & Tansey, M. G. The role of innate and adaptive immunity in Parkinson's disease. *J. Parkinsons Dis.* **3**, 493–514 (2013).

40. Gagliano, S. A. et al. Genomics implicates adaptive and innate immunity in Alzheimer's and Parkinson's diseases. *Ann. Clin. Transl. Neurol.* **3**, 924–933 (2016).
41. Dijkstra, A. A. et al. Evidence for immune response, axonal dysfunction and reduced endocytosis in the substantia nigra in early stage Parkinson's disease. *PLoS ONE* **10**, e0128651 (2015).
42. Lake, B. B. et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586–1590 (2016).
43. Sathyamurthy, A. et al. Massively parallel single nucleus transcriptional profiling defines spinal cord neurons and their activity during behavior. *Cell Rep.* **22**, 2216–2225 (2018).
44. Lake, B. B. et al. A comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA. *Sci. Rep.* **7**, 6031 (2017).
45. Caspi, A. et al. The p factor: one general psychopathology factor in the structure of psychiatric disorders? *Clin. Psychol. Sci.* **2**, 119–137 (2014).
46. Sullivan, P. F. & Geschwind, D. H. Defining the genetic, genomic, cellular, and diagnostic architectures of psychiatric disorders. *Cell* **177**, 162–183 (2019).
47. Reynolds, R. H. et al. Moving beyond neurons: the role of cell type-specific gene regulation in Parkinson's disease heritability. *NPJ Parkinsons Dis.* **5**, 6 (2019).
48. Singaram, C. et al. Dopaminergic defect of enteric nervous system in Parkinson's disease patients with chronic constipation. *Lancet* **346**, 861–864 (1995).
49. Wakabayashi, K., Takahashi, H., Takeda, S., Ohama, E. & Ikuta, F. Lewy bodies in the enteric nervous system in Parkinson's disease. *Arch. Histol. Cytol.* **52**, 191–194 (1989).
50. Stokholm, M. G., Danielsen, E. H., Hamilton-Dutoit, S. J. & Borghammer, P. Pathological α-synuclein in gastrointestinal tissues from prodromal Parkinson disease patients. *Ann. Neurol.* **79**, 940–949 (2016).
51. Svensson, E. et al. Vagotomy and subsequent risk of Parkinson's disease. *Ann. Neurol.* **78**, 522–529 (2015).
52. Gilman, S. et al. Second consensus statement on the diagnosis of multiple system atrophy. *Neurology* **71**, 670–676 (2008).
53. Nalls, M. A. et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet.* **46**, 989–993 (2014).
54. Wakabayashi, K., Hayashi, S., Yoshimoto, M., Kudo, H. & Takahashi, H. NACP/α-synuclein-positive filamentous inclusions in astrocytes and oligodendrocytes of Parkinson's disease brains. *Acta Neuropathol.* **99**, 14–20 (2000).
55. Seidel, K. et al. The brainstem pathologies of Parkinson's disease and dementia with Lewy bodies. *Brain Pathol.* **25**, 121–135 (2015).

Eating Disorders Working Group of the Psychiatric Genomics Consortium

Roger Adan[17,18,19], Lars Alfredsson[20], Tetsuya Ando[21], Ole Andreassen[22], Jessica Baker[9], Andrew Bergen[23,24], Wade Berrettini[25], Andreas Birgegård[26,27], Joseph Boden[28], Ilka Boehm[29], Claudette Boni[30], Vesna Boraska Perica[31,32], Harry Brandt[33], Gerome Breen[13,14], Julien Bryois[1], Katharina Buehren[34], Cynthia Bulik[1,9,15], Roland Burghardt[35], Matteo Cassina[36], Sven Cichon[37], Maurizio Clementi[36], Jonathan Coleman[13,14], Roger Cone[38], Philippe Courtet[39], Steven Crawford[33], Scott Crow[40], James Crowley[16,26], Unna Danner[18], Oliver Davis[41,42], Martina de Zwaan[43], George Dedoussis[44], Daniela Degortes[45], Janiece DeSocio[46], Danielle Dick[47], Dimitris Dikeos[48], Christian Dina[49,50], Monika Dmitrzak-Weglarz[51], Elisa Docampo Martinez[52,53,54], Laramie Duncan[55], Karin Egberts[56], Stefan Ehrlich[29], Geòrgia Escaramís[52,53,54], Tõnu Esko[57,58], Xavier Estivill[52,53,54,59], Anne Farmer[13], Angela Favaro[45], Fernando Fernández-Aranda[60,61], Manfred Fichter[62,63], Krista Fischer[57], Manuel Föcker[64], Lenka Foretova[65], Andreas Forstner[37,66,67,68,69], Monica Forzan[36], Christopher Franklin[31], Steven Gallinger[70], Héléna Gaspar[13,14], Ina Giegling[71], Johanna Giuranna[64], Paola Giusti-Rodríquez[16], Fragiskos Gonidakis[72], Scott Gordon[73], Philip Gorwood[30,74], Monica Gratacos Mayora[52,53,54], Jakob Grove[75,76,77,78], Sébastien Guillaume[39], Yiran Guo[79], Hakon Hakonarson[79,80], Katherine Halmi[81], Ken Hanscombe[82], Konstantinos Hatzikotoulas[31], Joanna Hauser[83], Johannes Hebebrand[64], Sietske Helder[13,84], Anjali Henders[85], Stefan Herms[37,69], Beate Herpertz-Dahlmann[34], Wolfgang Herzog[86], Anke Hinney[64], L. John Horwood[28], Christopher Hübel[1,13], Laura Huckins[31,87], James Hudson[88], Hartmut Imgart[89], Hidetoshi Inoko[90], Vladimir Janout[91], Susana Jiménez-Murcia[60,61], Craig Johnson[92], Jennifer Jordan[93,94], Antonio Julià[95], Anders Juréus[1], Gursharan Kalsi[13], Deborah Kaminská[96], Allan Kaplan[97], Jaakko Kaprio[98,99], Leila Karhunen[100], Andreas Karwautz[101], Martien Kas[17,102], Walter Kaye[103], James Kennedy[97], Martin Kennedy[104], Anna Keski-Rahkonen[98], Kirsty Kiezebrink[105], Youl-Ri Kim[106], Katherine Kirk[73], Lars Klareskog[107], Kelly Klump[108], Gun Peggy Knudsen[109], Maria La Via[9], Mikael Landén[1,19], Janne Larsen[76,110,111], Stephanie Le Hellard[112,113,114], Virpi Leppä[1], Robert Levitan[115], Dong Li[79], Paul Lichtenstein[1], Lisa Lilenfeld[116], Bochao Danae Lin[17], Jolanta Lissowska[117], Jurjen Luykx[17], Pierre Magistretti[118,119], Mario Maj[120], Katrin Mannik[57,121], Sara Marsal[95], Christian Marshall[122],

Nicholas Martin[73], Manuel Mattheisen[26,27,75,123], Morten Mattingsdal[22], Sara McDevitt[124,125], Peter McGuffin[13], Sarah Medland[73], Andres Metspalu[57,126], Ingrid Meulenbelt[127], Nadia Micali[128,129], James Mitchell[130], Karen Mitchell[131], Palmiero Monteleone[132], Alessio Maria Monteleone[120], Grant Montgomery[73,85,133], Preben Bo Mortensen[76,110,111], Melissa Munn-Chernoff[9], Benedetta Nacmias[134], Marie Navratilova[65], Claes Norring[26,27], Ioanna Ntalla[44], Catherine Olsen[73], Roel Ophoff[17,135], Julie O'Toole[136], Leonid Padyukov[107], Aarno Palotie[58,99,137], Jacques Pantel[30], Hana Papezova[96], Richard Parker[73], John Pearson[138], Nancy Pedersen[1], Liselotte Petersen[76,110,111], Dalila Pinto[87], Kirstin Purves[13], Raquel Rabionet[139,140,141], Anu Raevuori[98], Nicolas Ramoz[30], Ted Reichborn-Kjennerud[109,142], Valdo Ricca[134,143], Samuli Ripatti[144], Stephan Ripke[145,146,147], Franziska Ritschel[29,148], Marion Roberts[13], Alessandro Rotondo[149], Dan Rujescu[62,71], Filip Rybakowski[150], Paolo Santonastaso[151], André Scherag[152], Stephen Scherer[153], Ulrike Schmidt[13], Nicholas Schork[154], Alexandra Schosser[155], Jochen Seitz[34], Lenka Slachtova[156], P. Eline Slagboom[127], Margarita Slof-Op 't Landt[157,158], Agnieszka Slopien[159], Sandro Sorbi[134,160], Michael Strober[161,162], Garret Stuber[9,163], Patrick Sullivan[1,16], Beata Świątkowska[164], Jin Szatkiewicz[16], Ioanna Tachmazidou[31], Elena Tenconi[45], Laura Thornton[9], Alfonso Tortorella[165,166], Federica Tozzi[167], Janet Treasure[13], Artemis Tsitsika[168], Marta Tyszkiewicz-Nwafor[150], Konstantinos Tziouvas[169], Annemarie van Elburg[18,170], Eric van Furth[157,158], Tracey Wade[171], Gudrun Wagner[101], Esther Walton[29], Hunna Watson[9,10,11], Thomas Werge[172], David Whiteman[73], Elisabeth Widen[99], D. Blake Woodside[173,174], Shuyang Yao[1], Zeynep Yilmaz[9,16], Eleftheria Zeggini[31,175], Stephanie Zerwas[9] and Stephan Zipfel[176]

[17]Brain Center Rudolf Magnus, Department of Translational Neuroscience, University Medical Center Utrecht, Utrecht, the Netherlands. [18]Center for Eating Disorders Rintveld, Altrecht Mental Health Institute, Zeist, the Netherlands. [19]Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. [20]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. [21]Department of Behavioral Medicine, National Institute of Mental Health, National Center of Neurology and Psychiatry, Tokyo, Japan. [22]NORMENT KG Jebsen Centre, Division of Mental Health and Addiction, University of Oslo, Oslo University Hospital, Oslo, Norway. [23]BioRealm, LLC, Walnut, CA, USA. [24]Oregon Research Institute, Eugene, OR, USA. [25]Department of Psychiatry, Center for Neurobiology and Behavior, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA. [26]Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden. [27]Center for Psychiatry Research, Stockholm Health Care Services, Stockholm City Council, Stockholm, Sweden. [28]Christchurch Health and Development Study, University of Otago, Christchurch, New Zealand. [29]Division of Psychological and Social Medicine and Developmental Neurosciences, Faculty of Medicine, Technische Universität Dresden, Dresden, Germany. [30]INSERM U894, Centre of Psychiatry and Neuroscience, Paris, France. [31]Wellcome Sanger Institute, Hinxton, Cambridge, UK. [32]Department of Medical Biology, School of Medicine, University of Split, Split, Croatia. [33]The Center for Eating Disorders at Sheppard Pratt, Baltimore, MD, USA. [34]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, RWTH Aachen University, Aachen, Germany. [35]Klinikum Frankfurt/Oder, Frankfurt, Germany. [36]Clinical Genetics Unit, Department of Woman and Child Health, University of Padova, Padua, Italy. [37]Institute of Medical Genetics and Pathology, University Hospital Basel, Basel, Switzerland. [38]Life Sciences Institute and Department of Molecular and Integrative Physiology, University of Michigan, Ann Arbor, MI, USA. [39]Department of Emergency Psychiatry and Post-Acute Care, CHRU Montpellier, University of Montpellier, Montpellier, France. [40]Department of Psychiatry, University of Minnesota, Minneapolis, MN, USA. [41]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. [42]School of Social and Community Medicine, University of Bristol, Bristol, UK. [43]Department of Psychosomatic Medicine and Psychotherapy, Hannover Medical School, Hannover, Germany. [44]Department of Nutrition and Dietetics, Harokopio University, Athens, Greece. [45]Department of Neurosciences, University of Padova, Padua, Italy. [46]College of Nursing, Seattle University, Seattle, WA, USA. [47]Department of Psychology, Virginia Commonwealth University, Richmond, VA, USA. [48]Department of Psychiatry, Athens University Medical School, Athens University, Athens, Greece. [49]L'institut du thorax, INSERM, CNRS, UNIV Nantes, Nantes, France. [50]L'institut du thorax, CHU Nantes, Nantes, France. [51]Department of Psychiatric Genetics, Poznań University of Medical Sciences, Poznań, Poland. [52]Barcelona Institute of Science and Technology, Barcelona, Spain. [53]Universitat Pompeu Fabra, Barcelona, Spain. [54]Centro de Investigación Biomédica en Red en Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain. [55]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA. [56]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital of Würzburg, Centre for Mental Health, Würzburg, Germany. [57]Estonian Genome Center, University of Tartu, Tartu, Estonia. [58]Program in Medical and Population Genetics, Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA. [59]Genomics and Disease, Bioinformatics and Genomics Programme, Centre for Genomic Regulation, Barcelona, Spain. [60]Department of Psychiatry, University Hospital of Bellvitge –IDIBELL and CIBERobn, Barcelona, Spain. [61]Department of Clinical Sciences, School of Medicine, University of Barcelona, Barcelona, Spain. [62]Department of Psychiatry and Psychotherapy, Ludwig-Maximilians-University (LMU), Munich, Germany. [63]Schön Klinik Roseneck affiliated with the Medical Faculty of the University of Munich (LMU), Munich, Germany. [64]Department of Child and Adolescent Psychiatry, University Hospital Essen, University of Duisburg-Essen, Essen, Germany. [65]Department of Cancer, Epidemiology and Genetics, Masaryk Memorial Cancer Institute, Brno, Czech Republic. [66]Institute of Human Genetics, University of Bonn School of Medicine & University Hospital Bonn, Bonn, Germany. [67]Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany. [68]Department of Psychiatry (UPK), University of Basel, Basel, Switzerland. [69]Department of Biomedicine, University of Basel, Basel, Switzerland. [70]Department of Surgery, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada. [71]Department of Psychiatry, Psychotherapy and Psychosomatics, Martin Luther University of Halle-Wittenberg, Halle, Germany. [72]1st Psychiatric Department, National and Kapodistrian University of Athens, Medical School, Eginition Hospital, Athens, Greece. [73]QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [74]CMME (Groupe Hospitalier Sainte-Anne), Paris Descartes University, Paris, France. [75]Department of Biomedicine, Aarhus University, Aarhus, Denmark.

[76]The Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH), Aarhus, Denmark. [77]Centre for Integrative Sequencing, iSEQ, Aarhus University, Aarhus, Denmark. [78]Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark. [79]Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, USA. [80]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [81]Department of Psychiatry, Weill Cornell Medical College, New York, NY, USA. [82]Department of Medical and Molecular Genetics, King's College London, Guy's Hospital, London, UK. [83]Department of Adult Psychiatry, Poznań University of Medical Sciences, Poznań, Poland. [84]Zorg op Orde, Leidschendam, the Netherlands. [85]Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia. [86]Department of General Internal Medicine and Psychosomatics, Heidelberg University Hospital, Heidelberg University, Heidelberg, Germany. [87]Department of Psychiatry, and Genetics and Genomics Sciences, Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [88]Biological Psychiatry Laboratory, McLean Hospital/Harvard Medical School, Boston, MA, USA. [89]Eating Disorders Unit, Parklandklinik, Bad Wildungen, Germany. [90]Department of Molecular Life Science, Division of Basic Medical Science and Molecular Medicine, School of Medicine, Tokai University, Isehara, Japan. [91]Faculty of Health Sciences, Palacky University, Olomouc, Czech Republic. [92]Eating Recovery Center, Denver, CO, USA. [93]Department of Psychological Medicine, University of Otago, Christchurch, New Zealand. [94]Canterbury District Health Board, Christchurch, New Zealand. [95]Rheumatology Research Group, Vall d'Hebron Research Institute, Barcelona, Spain. [96]Department of Psychiatry, First Faculty of Medicine, Charles University, Prague, Czech Republic. [97]Center for Addiction and Mental Health, Department of Psychiatry, Institute of Medical Science, University of Toronto, Toronto, Ontario, Canada. [98]Department of Public Health, University of Helsinki, Helsinki, Finland. [99]Institute for Molecular Medicine Finland, Helsinki Institute of Life Science, University of Helsinki, Helsinki, Finland. [100]Institute of Public Health and Clinical Nutrition, Department of Clinical Nutrition, University of Eastern Finland, Kuopio, Finland. [101]Eating Disorders Unit, Department of Child and Adolescent Psychiatry, Medical University of Vienna, Vienna, Austria. [102]Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, the Netherlands. [103]Department of Psychiatry, University of California San Diego, San Diego, CA, USA. [104]Department of Pathology and Biomedical Science, University of Otago, Christchurch, New Zealand. [105]Health Services Research Unit, University of Aberdeen, Aberdeen, UK. [106]Department of Psychiatry, Seoul Paik Hospital, Inje University, Seoul, Korea. [107]Rheumatology Unit, Department of Medicine, Center for Molecular Medicine, Karolinska Institutet and Karolinska University Hospital, Stockholm, Sweden. [108]Department of Psychology, Michigan State University, East Lansing, MI, USA. [109]Department of Mental Disorders, Norwegian Institute of Public Health, Oslo, Norway. [110]National Centre for Register-Based Research, Aarhus BSS, Aarhus University, Aarhus, Denmark. [111]Centre for Integrated Register-based Research (CIRRAU), Aarhus University, Aarhus, Denmark. [112]Department of Clinical Science, K.G. Jebsen Centre for Psychosis Research, Norwegian Centre for Mental Disorders Research (NORMENT), University of Bergen, Bergen, Norway. [113]Dr. Einar Martens Research Group for Biological Psychiatry, Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway. [114]Department of Clinical Medicine, Laboratory Building, Haukeland University Hospital, Bergen, Norway. [115]Institute of Medical Science, University of Toronto, Toronto, Ontario, Canada. [116]American School of Professional Psychology, Argosy University, Northern Virginia, Arlington, VA, USA. [117]Department of Cancer Epidemiology and Prevention, M Skłodowska-Curie Cancer Center - Oncology Center, Warsaw, Poland. [118]BESE Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia. [119]Department of Psychiatry, University of Lausanne-University Hospital of Lausanne (UNIL-CHUV), Lausanne, Switzerland. [120]Department of Psychiatry, University of Campania 'Luigi Vanvitelli', Naples, Italy. [121]Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland. [122]Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada. [123]Department of Psychiatry, Psychosomatics and Psychotherapy, University of Würzburg, Würzburg, Germany. [124]Department of Psychiatry, University College Cork, Cork, Ireland. [125]Eist Linn Adolescent Unit, Bessborough, Health Service Executive South, Cork, Ireland. [126]Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia. [127]Molecular Epidemiology Section (Department of Medical Statistics), Leiden University Medical Centre, Leiden, the Netherlands. [128]Department of Psychiatry, Faculty of Medicine, University of Geneva, Geneva, Switzerland. [129]Division of Child and Adolescent Psychiatry, Geneva University Hospital, Geneva, Switzerland. [130]Department of Psychiatry and Behavioral Science, University of North Dakota School of Medicine and Health Sciences, Fargo, ND, USA. [131]National Center for PTSD, VA Boston Healthcare System, Department of Psychiatry, Boston University School of Medicine, Boston, MA, USA. [132]Department of Medicine, Surgery and Dentistry 'Scuola Medica Salernitana', University of Salerno, Salerno, Italy. [133]Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia. [134]Department of Neuroscience, Psychology, Drug Research and Child Health (NEUROFARBA), University of Florence, Florence, Italy. [135]Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, CA, USA. [136]Kartini Clinic, Portland, OR, USA. [137]Center for Human Genome Research at the Massachusetts General Hospital, Boston, MA, USA. [138]Biostatistics and Computational Biology Unit, University of Otago, Christchurch, New Zealand. [139]Saint Joan de Déu Research Institute, Saint Joan de Déu Barcelona Children's Hospital, Barcelona, Spain. [140]Institute of Biomedicine (IBUB), University of Barcelona, Barcelona, Spain. [141]Department of Genetics, Microbiology and Statistics, University of Barcelona, Barcelona, Spain. [142]Institute of Clinical Medicine, University of Oslo, Oslo, Norway. [143]Department of Health Science, University of Florence, Florence, Italy. [144]Department of Biometry, University of Helsinki, Helsinki, Finland. [145]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. [146]Stanley Center for Psychiatric Research, Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA. [147]Department of Psychiatry and Psychotherapy, Charité - Universitätsmedizin, Berlin, Germany. [148]Eating Disorders Research and Treatment Center, Department of Child and Adolescent Psychiatry, Faculty of Medicine, Technische Universität Dresden, Dresden, Germany. [149]Department of Psychiatry, Neurobiology, Pharmacology, and Biotechnologies, University of Pisa, Pisa, Italy. [150]Department of Psychiatry, Poznań University of Medical Sciences, Poznań, Poland. [151]Department of Neurosciences, Padua Neuroscience Center, University of Padova, Padua, Italy. [152]Institute of Medical Statistics, Computer and Data Sciences, Jena University Hospital, Jena, Germany. [153]Department of Genetics and Genomic Biology, The Hospital for Sick Children, Toronto, Ontario, Canada. [154]J. Craig Venter Institute (JCVI), La Jolla, CA, USA. [155]Department of Psychiatry and Psychotherapy, Medical University of Vienna, Vienna, Austria. [156]Department of Pediatrics and Center of Applied Genomics, First Faculty of Medicine, Charles University, Prague, Czech Republic. [157]Center for Eating Disorders Ursula, Rivierduinen, Leiden, the Netherlands. [158]Department of Psychiatry, Leiden University Medical Centre, Leiden, the Netherlands. [159]Department of Child and Adolescent Psychiatry, Poznań University of Medical Sciences, Poznań, Poland. [160]IRCCS Fondazione Don Carlo Gnocchi, Florence, Italy. [161]Department of Psychiatry and Biobehavioral Science, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, CA, USA. [162]David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. [163]Department of Cell Biology and Physiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [164]Department of Environmental Epidemiology, Nofer Institute of Occupational Medicine, Lodz, Poland. [165]Department of Psychiatry, University of Naples SUN, Naples, Italy. [166]Department of Psychiatry, University of Perugia, Perugia, Italy. [167]Brain Sciences Department, Stremble Ventures, Limassol, Cyprus. [168]Adolescent Health Unit, Second Department of Pediatrics, 'P. & A. Kyriakou' Children's Hospital, University of Athens, Athens, Greece. [169]Pediatric Intensive Care Unit, 'P. & A. Kyriakou' Children's Hospital, University of Athens, Athens, Greece. [170]Faculty of Social and Behavioral Sciences, Utrecht University, Utrecht, the Netherlands. [171]School of Psychology, Flinders University, Adelaide, South Australia, Australia. [172]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark. [173]Department of Psychiatry, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada. [174]Toronto General Hospital, Toronto, Ontario, Canada. [175]Institute of Translational Genomics, Helmholtz Zentrum München, Neuherberg, Germany. [176]Department of Internal Medicine VI, Psychosomatic Medicine and Psychotherapy, University Medical Hospital Tübingen, Tübingen, Germany.

International Headache Genetics Consortium

Verneri Anttila[177], Ville Artto[178], Andrea Carmine Belin[179], Irene de Boer[180], Dorret I. Boomsma[181], Sigrid Børte[182], Daniel I. Chasman[183], Lynn Cherkas[184], Anne Francke Christensen[185], Bru Cormand[186], Ester Cuenca-Leon[177], George Davey-Smith[187], Martin Dichgans[188], Cornelia van Duijn[189], Tonu Esko[57], Ann Louise Esserlind[190], Michel Ferrari[180], Rune R. Frants[180], Tobias Freilinger[191], Nick Furlotte[192], Padhraig Gormley[177], Lyn Griffiths[193], Eija Hamalainen[194], Thomas Folkmann Hansen[6], Marjo Hiekkala[195], M. Arfan Ikram[189], Andres Ingason[196], Marjo-Riitta Järvelin[197], Risto Kajanne[194], Mikko Kallela[178], Jaakko Kaprio[98,99], Mari Kaunisto[195], Lisette J. A. Kogelman[6], Christian Kubisch[198], Mitja Kurki[177], Tobias Kurth[199], Lenore Launer[200], Terho Lehtimaki[201], Davor Lessel[198], Lannie Ligthart[181], Nadia Litterman[192], Arn van den Maagdenberg[180], Alfons Macaya[202], Rainer Malik[188], Massimo Mangino[184], George McMahon[187], Bertram Muller-Myhsok[203], Benjamin M. Neale[177], Carrie Northover[192], Dale R. Nyholt[193], Jes Olesen[190], Aarno Palotie[58,99,137], Priit Palta[194], Linda Pedersen[182], Nancy Pedersen[1], Danielle Posthuma[181], Patricia Pozo-Rosich[204], Alice Pressman[205], Olli Raitakari[206], Markus Schürks[199], Celia Sintas[186], Kari Stefansson[196], Hreinn Stefansson[196], Stacy Steinberg[196], David Strachan[207], Gisela Terwindt[180], Marta Vila-Pueyo[202], Maija Wessman[195], Bendik S. Winsvold[182], Huiying Zhao[193] and John Anker Zwart[182]

[177]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [178]Department of Neurology, Helsinki University Central Hospital, Helsinki, Finland. [179]Karolinska Institutet, Stockholm, Sweden. [180]Leiden University Medical Centre, Leiden, the Netherlands. [181]VU University, Amsterdam, the Netherlands. [182]Oslo University Hospital and University of Oslo, Oslo, Norway. [183]Harvard Medical School, Cambridge, MA, USA. [184]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [185]Danish Headache Center, Copenhagen University Hospital, Copenhagen, Denmark. [186]University of Barcelona, Barcelona, Spain. [187]Medical Research Council (MRC) Integrative Epidemiology Unit, University of Bristol, Bristol, UK. [188]Institute for Stroke and Dementia Research, Munich, Germany. [189]Erasmus University Medical Centre, Rotterdam, the Netherlands. [190]Danish Headache Center, Department of Neurology, Rigshospitalet, Glostrup, Denmark. [191]University of Tübingen, Tübingen, Germany. [192]23&Me Inc., Mountain View, CA, USA. [193]Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia. [194]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. [195]Folkhälsan Institute of Genetics, Helsinki, Finland. [196]Decode genetics Inc., Reykjavik, Iceland. [197]University of Oulu, Biocenter Oulu, Finland. [198]University Medical Center Hamburg-Eppendorf, Hamburg, Germany. [199]Harvard Medical School, Boston, MA, USA. [200]National Institute on Aging, Bethesda, MD, USA. [201]School of Medicine, University of Tampere, Tampere, Finland. [202]Vall d'Hebron Research Institute, Barcelona, Spain. [203]Max Planck Institute of Psychiatry, Munich, Germany. [204]Headache Research Group, Universitat Autònoma de Barcelona, Barcelona, Spain. [205]Sutter Health, Sacramento, CA, USA. [206]Department of Medicine, University of Turku, Turku, Finland. [207]Population Health Research Institute, St George's University of London, London, UK.

23andMe Research Team

Michelle Agee[208], Babak Alipanahi[208], Adam Auton[208], Robert Bell[208], Katarzyna Bryc[208], Sarah Elson[208], Pierre Fontanillas[208], Nicholas Furlotte[208], Karl Heilbron[208], David Hinds[208], Karen Huber[208], Aaron Kleinman[208], Nadia Litterman[208], Jennifer McCreight[208], Matthew McIntyre[208], Joanna Mountain[208], Elizabeth Noblin[208], Carrie Northover[208], Steven Pitts[208], J. Sathirapongsasuti[208], Olga Sazonova[208], Janie Shelton[208], Suyash Shringarpure[208], Chao Tian[208], Joyce Tung[208], Vladimir Vacic[208] and Catherine Wilson[208]

[208]23andMe, Inc., Mountain View, CA, USA.

Methods

GWAS results. Our goal was to use GWAS results to identify relevant tissues and cell types. Our primary focus was human phenotypes whose etiopathology is based in the CNS. We thus obtained 18 sets of GWAS summary statistics from European samples for brain-related complex traits. These were selected because they had at least one genome-wide significant association (as of 2018; for example, Parkinson's disease, schizophrenia and IQ (intelligence quotient)). For comparison, we included GWAS summary statistics for eight diseases and traits with large sample sizes whose etiopathology is not rooted in the CNS (for example, type 2 diabetes). The selection of these conditions allowed contrasts of tissues and cells highlighted by our primary interest in brain phenotypes with non-brain-related traits.

The phenotypes were: schizophrenia[1], educational attainment[2], intelligence[14], BMI[4], bipolar disorder[56], neuroticism[3], MDD[57], age at menarche[58], autism[59], migraine[60], amyotrophic lateral sclerosis[61], ADHD[62], Alzheimer's disease[16], age at menopause[63], coronary artery disease[64], height[4], hemoglobin A1c[65], hippocampal volume[66], inflammatory bowel disease[67], intracranial volume[68], stroke[69], type 2 diabetes mellitus[70], type 2 diabetes adjusted for BMI[70], waist–hip ratio adjusted for BMI[71] and anorexia nervosa[72].

For Parkinson's disease, we performed an inverse-variance-weighted meta-analysis[73] using summary statistics from Nalls et al.[53] (9,581 cases, 33,245 controls) and summary statistics from 23andMe (12,657 cases, 941,588 controls). We found a very high genetic correlation (r_g)[74] between the results from these cohorts ($r_g = 0.87$, s.e. = 0.068) with little evidence of sample overlap (LDSC bivariate intercept = 0.0288, s.e. = 0.0066). The *P* values from the meta-analysis strongly deviated from the expected (Supplementary Fig. 8) but the trend was consistent with polygenicity (LDSC intercept = 1.0048, s.e. = 0.008) rather than uncontrolled inflation[74]. In this new meta-analysis, we identified 61 independent loci associated with Parkinson's disease (49 reported previously[17] and 12 novel; Supplementary Fig. 9). The 10,000 most associated SNPs from the 23andMe cohort are available in Supplementary Table 12.

Gene expression data. We collected publicly available single-cell RNA-seq data from different studies. The core dataset of our analysis is a study that sampled more than 500,000 single cells from the entire mouse nervous system (19 regions) and identified 39 broad categories (level 4) and 265 refined cell types (level 5)[23]. The 39 cell types expressed a median of 16,417 genes, had a median unique molecular identifier (UMI) total count of ~8.6 million and summed the expression of a median of 1,501 single cells (Supplementary Table 13). The replication datasets were: a mouse study that sampled 690,000 single cells from 9 brain regions (frontal cortex, striatum, globus pallidus externus/nucleus basalis, thalamus, hippocampus, posterior cortex, entopeduncular nucleus/subthalamic nucleus, substantia nigra/ventral tegmental area and cerebellum) and identified 565 cell types[34] (note that we averaged the UMI counts by broad categories of cell type in each brain region, resulting in 88 different cell types); our prior mouse study of ~10,000 cells from 5 different brain regions (and samples enriched for oligodendrocytes, dopaminergic neurons, serotonergic neurons and cortical parvalbuminergic interneurons) that identified 24 broad categories and 149 refined cell types[11]; a study that sampled 19,550 nuclei from frozen adult human post-mortem hippocampus and prefrontal cortex and identified 16 cell types[35]; a study that generated 36,166 single-nuclei expression measurements (after quality control) from the human visual cortex, frontal cortex and cerebellum[36]. We also obtained bulk-tissue RNA-seq gene expression data from 53 tissues from the GTEx consortium[7] (v8, median across samples).

Gene expression data processing. All datasets were processed uniformly. First we computed the mean expression for each gene in each cell type from the single-cell expression data (if this statistic was not provided by the authors). We used the pre-computed median expression across individuals for the GTEx dataset and excluded tissues that were not sampled in at least 100 individuals, non-natural tissues (for example, Epstein–Barr virus–transformed lymphocytes) and testis tissues (outlier using hierarchical clustering). We then averaged the expression of tissues by organ (with the exception of brain tissues) resulting in gene expression profiles of a total of 37 tissues. For all datasets, we filtered out any genes with non-unique names, genes not expressed in any cell types, non-protein-coding genes and, for mouse datasets, genes that had no expert-curated 1:1 orthologs between mice and humans (Mouse Genome Informatics, The Jackson Laboratory, version 11/22/2016). Gene expression was then scaled to a total of 1 million UMIs (or transcripts per million (TPM)) for each cell type/tissue. We then calculated a metric of gene expression specificity by dividing the expression of each gene in each cell type by the total expression of that gene in all cell types, leading to values ranging from 0 to 1 for each gene (0: meaning that the gene is not expressed in that cell type; 0.6: that 60% of the total expression of that gene is performed in that cell type; 1: that 100% of the expression of that gene is performed in that cell type). The 10% most specific genes (Supplementary Tables 14 and 15) in each tissue/cell type partially overlapped for related tissues/cell types, did not overlap for unrelated tissue/cell types and allowed us to cluster related tissues/cell types as expected (Supplementary Figs. 10 and 11).

MAGMA primary and conditional analyses. MAGMA (v1.06b)[18] is a software for gene-set enrichment analysis using GWAS summary statistics. Briefly, MAGMA computes a gene-level association statistic by averaging *P* values of SNPs located around a gene (taking into account LD structure). The gene-level association statistic is then transformed to a *z* score. MAGMA can then be used to test whether a gene set is a predictor of the gene-level association statistic of the trait (*z* score) in a linear regression framework. MAGMA accounts for a number of important covariates such as gene size, gene density, mean sample size for tested SNPs per gene, the inverse of the minor allele counts per gene and the log of these metrics.

For each GWAS summary statistic, we excluded any SNPs with INFO score <0.6, with minor allele frequency <1% or with estimated odds ratio >25 or smaller than 1/25, the major histocompatibility complex region (chr6: 25–34 Mb) for all GWAS and the *APOE* region (chr19: 45020859–45844508) for the Alzheimer's GWAS. We set a window of 35 kilobases (kb) upstream to 10 kb downstream of the gene coordinates to compute gene-level association statistics and used the European reference panel from phase 3 of the 1000 Genomes Project[75] as the reference population. For each trait, we then used MAGMA to test whether the 10% most specific gene in each tissue/cell type was associated with gene-level genetic association with the trait. Only genes with at least 1 TPM or 1 UMI per million in the tested cell type were used for this analysis. The significance level of the different cell types was highly correlated with the effect size of the cell type (Supplementary Fig. 12), with values ranging between 0.999 and 1 across the 18 brain-related traits in the Zeisel et al. dataset[23]. The significance threshold was set to a 5% false discovery rate across all tissues/cell types and traits within each dataset.

MAGMA can also perform conditional analyses given its linear regression framework. We used MAGMA to test whether cell types were associated with a specific trait conditioning on the gene-level genetic association of another trait (*z* score from MAGMA.out file) or to look for associations of cell types conditioning on the 10% most specific genes from other cell types by adding these variables as covariates in the model.

To test whether MAGMA was well calibrated, we randomly permuted the gene labels of the schizophrenia gene-level association statistic file a thousand times. We then looked for association between the 10% most specific genes in each cell type and the randomized gene-level schizophrenia association statistics. We observed that MAGMA was slightly conservative with less than 5% of the random samplings having $P < 0.05$ (Supplementary Fig. 13).

We also evaluated the effect of varying the window size (for the SNP to gene assignment step of MAGMA) on the schizophrenia cell-type association strength ($-\log_{10}[P]$). We observed strong Pearson correlations in the cell-type association strength ($-\log_{10}[P]$) across the different window sizes tested (Supplementary Fig. 14). Our selected window size (35 kb upstream to 10 kb downstream) had Pearson correlations ranging from 0.94 to 0.98 with the other window sizes, indicating that our results are robust to this parameter.

In a recent paper, Watanabe et al.[76] introduced a different methodology to test for association between cell type and complex traits based on MAGMA. Their proposed methodology tests for a positive relationship between gene expression levels and gene-level genetic associations with a complex trait (using all genes). Their method uses the average expression of each gene in all cell types in the dataset as a covariate. We examined the method of Watanabe et al. in detail, and decided against its use for multiple reasons.

First, Watanabe et al. hypothesize that genes with higher levels of expression should be more associated with a trait. In extended discussions among our team (which includes multiple neuroscientists), we have strong reservations about the appropriateness and biological meaningfulness of this hypothesis; it is a strong requirement and is at odds with decades of neuroscience research where molecules expressed at low levels can have a profound biological impact. For instance, many cell-type-specific genes that are disease relevant are expressed at moderate levels (for example, *Drd2* is in the 10% most specific genes in telencephalon projecting inhibitory neurons but in the bottom 30% of expression levels). Our method does not make this hypothesis.

Second, the method of Watanabe et al. corrects for the average expression of all cell types in a dataset. This practice is, in our view, problematic as it necessarily forces dependence on the composition of a scRNA-seq dataset. For instance, if a dataset consists mostly of neurons, this amounts to correcting for neuronal expression and necessarily erodes power to detect trait enrichment in neurons. Alternatively, if a dataset is composed mostly of non-neuronal cells, this will impact the detection of enrichment in non-neuronal cells.

Third, preliminary results indicate that the method of Watanabe et al. is sensitive to scaling. As different cell types express different numbers of genes, scaling to the same total read counts affects the average gene expression across cell types (which they use as a covariate), leading to different results with different choices of scaling factors (for example, scaling to 10,000 versus 1 million reads). Our method is not liable to this issue.

LD score regression analysis. We used partitioned LD score regression[19] to test whether the 10% most specific genes of each cell type (based on our specificity metric described above) were enriched in heritability for the diverse traits. Only genes with at least 1 TPM or 1 UMI per million in the tested cell type were used for this analysis. To capture most regulatory elements that could contribute to the effect of the region on the trait, we extended the gene coordinates by 100 kb

upstream and by 100 kb downstream of each gene as previously[12]. SNPs located in 100-kb regions surrounding the 10% most specific genes in each cell type were added to the baseline model (consisting of 53 different annotations) independently for each cell type (1 file for each cell type). We then selected the coefficient *z*-score *P* value as a measure of the association of the cell type with the traits. The significance threshold was set to a 5% false discovery rate across all tissues/cell types and traits within each dataset. All plots show the mean $-\log_{10}[P]$ of partitioned LDscore regression and MAGMA. All results for MAGMA or LDSC are available in supplementary data files (Supplementary Tables 1, 2 and 5–9).

We evaluated the effect of varying the window size and varying the percentage of most specific genes on the schizophrenia cell-type association strength ($-\log_{10}[P]$). We observed strong Pearson correlations in the cell-type association strength ($-\log_{10}[P]$) across the different percentages and window sizes tested (Supplementary Fig. 15). Our selected window size (100 kb upstream to 100 kb downstream, 10% most specific genes) had Pearson correlations ranging from 0.96 to 1 with the other window sizes and percentages, indicating that our results are robust to these parameters.

MAGMA versus LDSC ranking. To test whether the cell-type rankings obtained using MAGMA and LDSC in the Zeisel et al. dataset[23] were similar, we computed the Spearman rank correlation of the cell-type association strength ($-\log_{10}[P]$) between the two methods for each complex trait. The Spearman rank correlation was strongly correlated with λ_{GC} (a measure of the deviation of the GWAS test statistics from the expected; Spearman correlation = 0.89; Supplementary Fig. 16) and with the average number of cell types below our stringent significance threshold (Spearman correlation = 0.92), indicating that the overall ranking of the cell types is very similar between the two methods, provided that the GWAS is well powered (Supplementary Fig. 17). In addition, we found that λ_{GC} was strongly correlated with the strength of association of the top tissue ($-\log_{10}[P]$; Spearman correlation = 0.88; Supplementary Fig. 18), as well as with the effect size (beta) of the top tissue (Spearman correlation = 0.9), indicating that the associations between cell type and trait are stronger for well-powered GWASs. The significance level ($-\log_{10}[P]$) was also strongly correlated with the effect size (Spearman correlation = 0.996; Supplementary Fig. 18) for the top cell type of each trait.

Dendritic depletion analysis. This analysis was performed as previously described[11]. In brief, all datasets were reduced to a set of six common cell types: pyramidal neurons, interneurons, astrocytes, microglia and oligodendrocyte precursors. Specificity was recalculated using only these six cell types. Comparisons were then made between pairs of datasets (denoted in the graph with the format '*X* versus *Y*'). The difference in specificity for a set of dendrite-enriched genes is calculated between the datasets. Differences in specificity are also calculated for random sets of genes selected from the background gene set. The probability and *z* score for the difference in specificity for the dendritic genes is thus estimated. Dendritically enriched transcripts were obtained from Supplementary Table 10 of Cajigas et al.[77]. For the KI dataset[11], we used S1 pyramidal neurons. For the Zeisel 2018 dataset[23], we used all ACTE* cells as astrocytes, TEGLU* as pyramidal neurons, TEINH* as interneurons, OPC as oligodendrocyte precursors and MGL* as microglia. For the Saunders dataset[34], we used all Neuron.Slc17a7 cell types from FC, HC or PC as pyramidal neurons; all Neuron.Gad1Gad2 cell types from FC, HC or PC as interneurons; Polydendrocye as OPCs; Astrocyte as astrocytes, and Microglia as microglia. The Lake datasets both came from a single publication[36] that had data from the frontal cortex, visual cortex and cerebellum. The cerebellum data were not used here. Data from frontal and visual cortices were analyzed separately. All other datasets were used as described in our previous publication[11]. The code and data for this analysis are available as an R package (see 'Code availability' below).

GO term enrichment. We tested whether genes that were highly specific to a trait-associated cell type (top 20% in a given cell type) and highly associated with the genetics of the traits (top 10% MAGMA gene-level genetic association) were enriched in biological functions using the topGO R package[78]. As background, we used genes that were highly specific to the cell type (top 20%) or highly associated with the trait (top 10% MAGMA gene-level genetic association).

Parkinson's disease rare variant enrichments. We searched the literature for genes associated with parkinsonism on the basis of rare and familial mutations. We found 66 genes (listed in Supplementary Table 10). We used linear regression to test whether the *z*-scaled specificity metrics (per cell type) of the 66 genes were greater than 0 in the different cell types.

Parkinson's disease post-mortem transcriptomes. The Moran dataset[38] was obtained from GEO (accession code GSE8397). Processing of the U133a and U133b Cel files was performed separately. The data were read in using the ReadAffy function from the R affy package[79]; then robust multi-array averaging was applied. The U133a and U133b array expression data were merged after applying robust multi-array averaging. Probe annotation and mapping to HUGO Gene Nomenclature Committee symbols were performed using the biomaRt R package[80]. Differential expression analysis was performed using limma[81] taking age and sex as covariates. The Lesnick dataset[37] was obtained from GEO (accession code GSE7621). Data were processed as for the Moran dataset: however, age was not available to use as a covariate. The Disjkstra dataset[41] was obtained from GEO (accession code GSE49036) and processed as above: the sex and RNA integrity number values were used as covariates. As the transcriptome datasets measured gene expression in the substantia nigra, we kept only cell types that are present in the substantia nigra or ventral midbrain for our EWCE[10] analysis. We computed a new specificity matrix based on the substantia nigra or ventral midbrain cells from the Zeisel dataset (level 5) using EWCE[10]. The EWCE analysis was performed on the 500 most upregulated or downregulated genes using 10,000 bootstrapping replicates.

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

Data availability

All single-cell expression data are publicly available. Most summary statistics used in this study are publicly available. The migraine GWAS[60] can be obtained by contacting the authors of that study. The full Parkinson's disease summary statistics from 23andMe can be obtained under an agreement that protects the privacy of 23andMe research participants (https://research.23andme.com/collaborate/#publication). The 10,000 most associated SNPs from the 23andMe cohort are available in Supplementary Table 12.

Code availability

The code used to generate these results is available at https://github.com/jbryois/scRNA_disease. An R package for performing cell-type enrichments using MAGMA is also available from https://github.com/NathanSkene/MAGMA_Celltyping.

References

56. Stahl, E. A. et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* **51**, 793–803 (2019).
57. 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).
58. Perry, J. R. B. et al. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92–97 (2014).
59. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).
60. Gormley, P. et al. Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. *Nat. Genet.* **48**, 1296 (2016).
61. Van Rheenen, W. et al. Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. *Nat. Genet.* **48**, 1043–1048 (2016).
62. Demontis, D. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
63. Day, F. R. et al. Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat. Genet.* **47**, 1294–1303 (2015).
64. Nelson, C. P. et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat. Genet.* **49**, 1385–1391 (2017).
65. Wheeler, E. et al. Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations. *PLoS Med.* **14**, 1–30 (2017).
66. Hibar, D. P. et al. Novel genetic loci associated with hippocampal volume. *Nat. Commun.* **8**, 13624 (2017).
67. de Lange, K. M. et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* **49**, 256–261 (2017).
68. Adams, H. H. H. et al. Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nat. Neurosci.* **19**, 1569–1582 (2016).
69. Malik, R. et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50**, 524–537 (2018).
70. Scott, R. A. et al. An expanded genome-wide association study of type 2 diabetes in Europeans. *Diabetes* **66**, 2888–2902 (2017).
71. Shungin, D. et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
72. Watson, H. J. et al. Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat. Genet.* **51**, 1207–1214 (2019).
73. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
74. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).

75. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
76. 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**, 3222 (2019).
77. Cajigas, I. J. et al. The local transcriptome in the synaptic neuropil revealed by deep sequencing and high-resolution imaging. *Neuron* **74**, 453–466 (2012).
78. Alexa, A. & Rahnenfuhrer, J. topGO: enrichment analysis for Gene Ontology v.2.36 (2016).
79. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. Affy - analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315 (2004).
80. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
81. Ritchie, M. E. et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

Acknowledgements

J.B. was funded by a grant from the Swiss National Science Foundation (P400PB_180792). N.G.S. was supported by the Wellcome Trust (108726/Z/15/Z). N.G.S. and L.B. performed part of the work at the Systems Genetics of Neurodegeneration summer school funded by BMBF as part of the e:Med programme (FKZ 01ZX1704). J.H.-L. was funded by the Swedish Research Council (Vetenskapsrådet, award 2014-3863), StratNeuro, the Wellcome Trust (108726/Z/15/Z) and the Swedish Brain Foundation (Hjärnfonden). P.F.S. was supported by the Swedish Research Council (Vetenskapsrådet, award D0886501), the Horizon 2020 Program of the European Union (COSYN, RIA grant agreement no. 610307) and US NIMH (U01 MH109528 and R01 MH077139). E.A. was supported by the Swedish Research Council (VR 2016-01526), Swedish Foundation for Strategic Research (SLA SB16-0065), Karolinska Institutet (SFO Strat. Regen., Senior grant 2018), Cancerfonden (CAN 2016/572), Hjärnfonden (FO2017-0059) and Chen Zuckeberg Initiative: Neurodegeneration Challenge Network (2018-191929-5022). C.M.B. acknowledges funding from the Swedish Research Council (Vetenskapsrådet, award: 538-2013-8864) and the Klarman Family Foundation. This work is supported by the UK Dementia Research Institute, which receives its funding from UK DRI Ltd, funded by the UK Medical Research Council, Alzheimer's Society and Alzheimer's Research UK. We thank the research participants from 23andMe and other cohorts for their contribution to this study.

Author contributions

J.B., N.G.S., J.H.-L. and P.F.S. designed the study, and wrote and reviewed the manuscript; J.B. performed the analyses pertaining to Figs. 1–4, Extended Data Figs. 1–10, Supplementary Figs. 1–5 and 7–20, and Supplementary Tables 1–9 and 12–17; N.G.S performed the analyses pertaining to Fig. 5, Supplementary Fig. 6 and Supplementary Tables 10 and 11; T.F.H., L.J.A.K. and the I.H.G.C. provided the migraine GWAS summary statistics; H.J.W., the E.D.W.G.P.G.C., G.B. and C.M.B. performed the anorexia GWAS; Z.L. contributed to the revision of the manuscript, the 23andMe R.T. provided GWAS summary statistics for Parkinson's disease in the 23andMe cohort. L.B. contributed to the post-mortem differential expression analysis (Fig. 5); E.A. provided expert knowledge on Parkinson's disease and reviewed the manuscript.

Competing interests

P.F.S. reports the following potentially competing financial interests: current—Lundbeck (advisory committee, grant recipient); past three years—Pfizer (scientific advisory board), Element Genomics (consultation fee) and Roche (speaker reimbursement). C.M.B. reports: Shire (grant recipient, Scientific Advisory Board member); Pearson and Walker (author, royalty recipient).

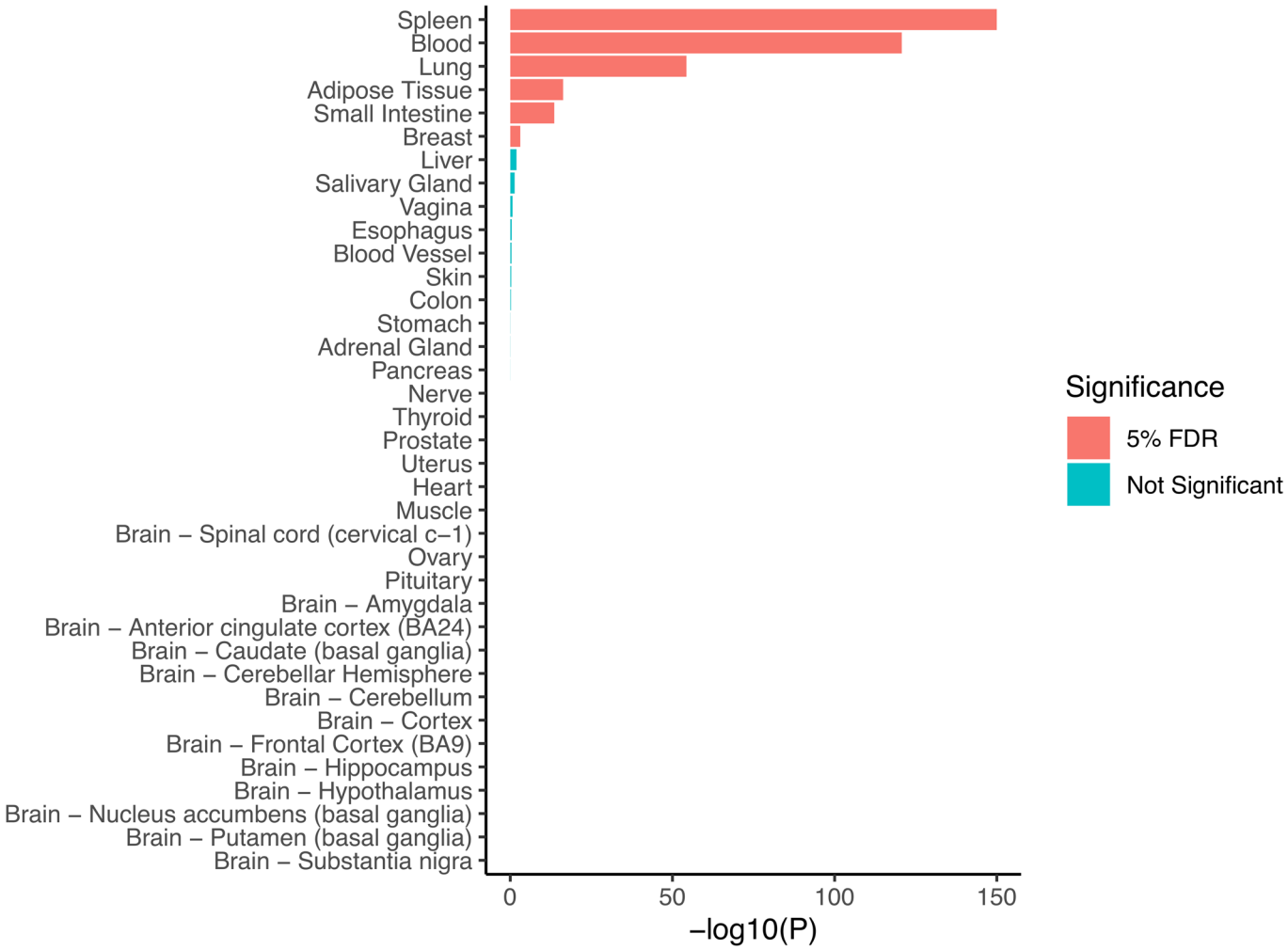
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-020-0610-9.

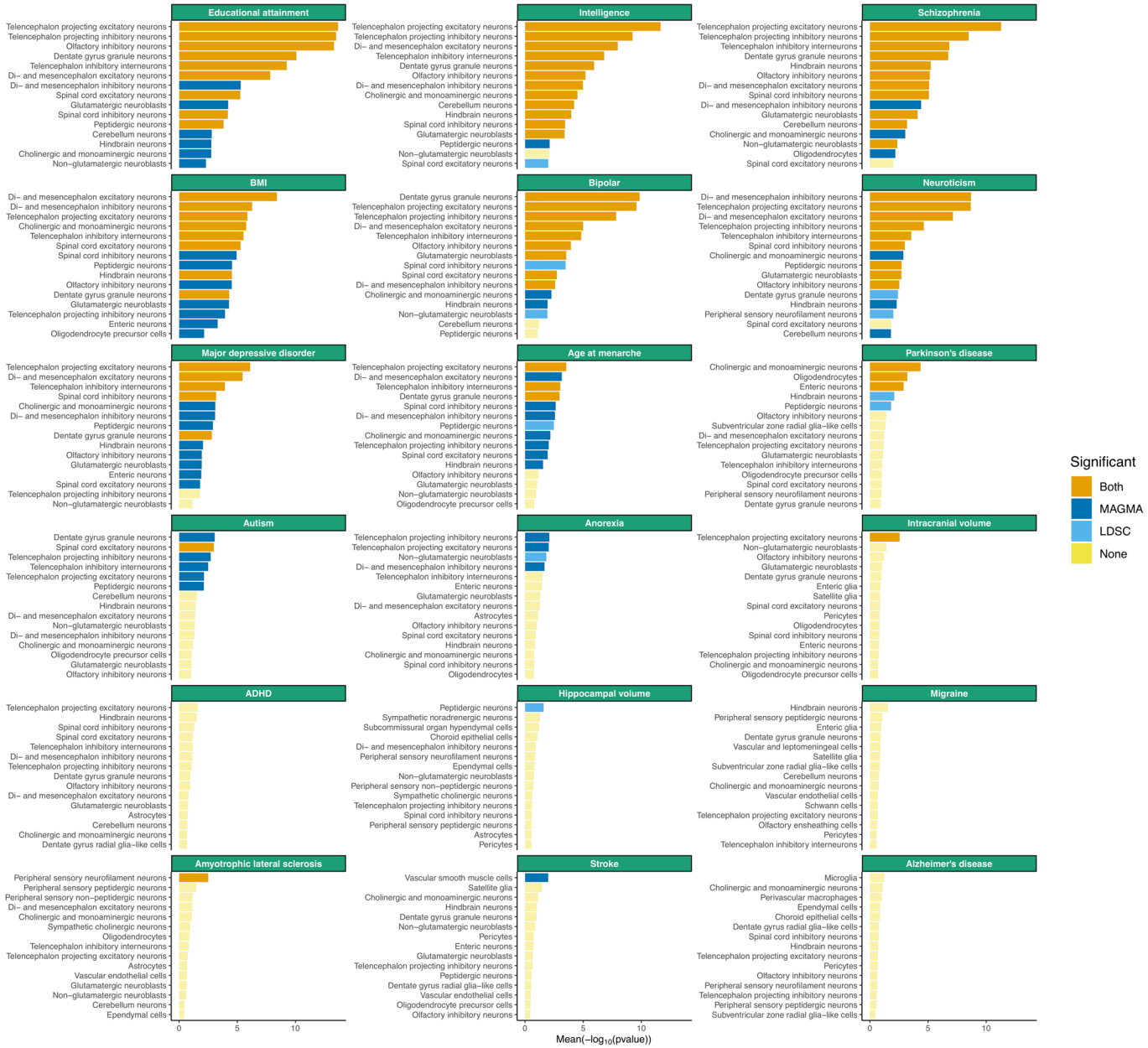
Supplementary information is available for this paper at https://doi.org/10.1038/s41588-020-0610-9.

Correspondence and requests for materials should be addressed to J.H.-L. or P.F.S.

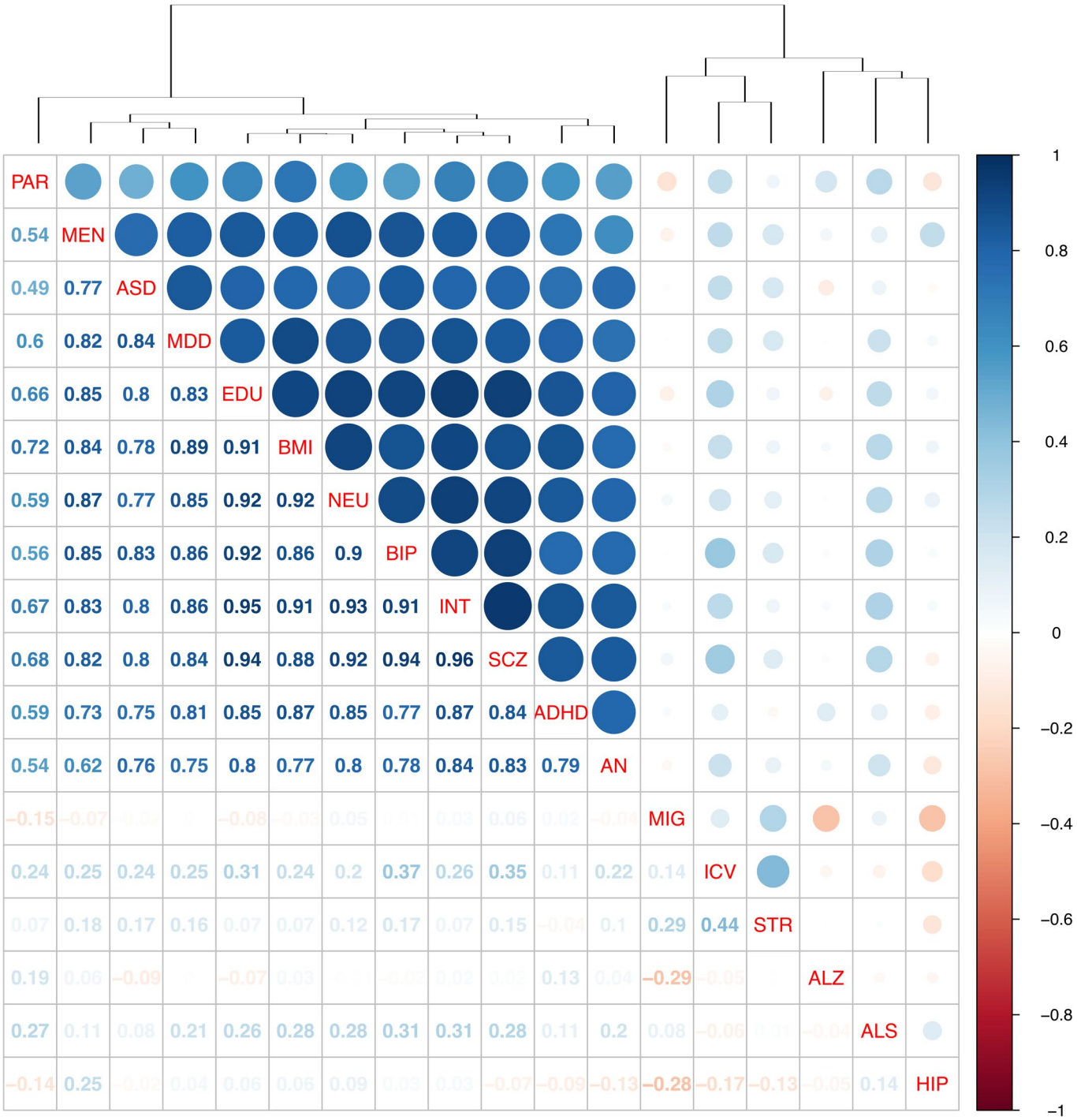
Reprints and permissions information is available at www.nature.com/reprints.



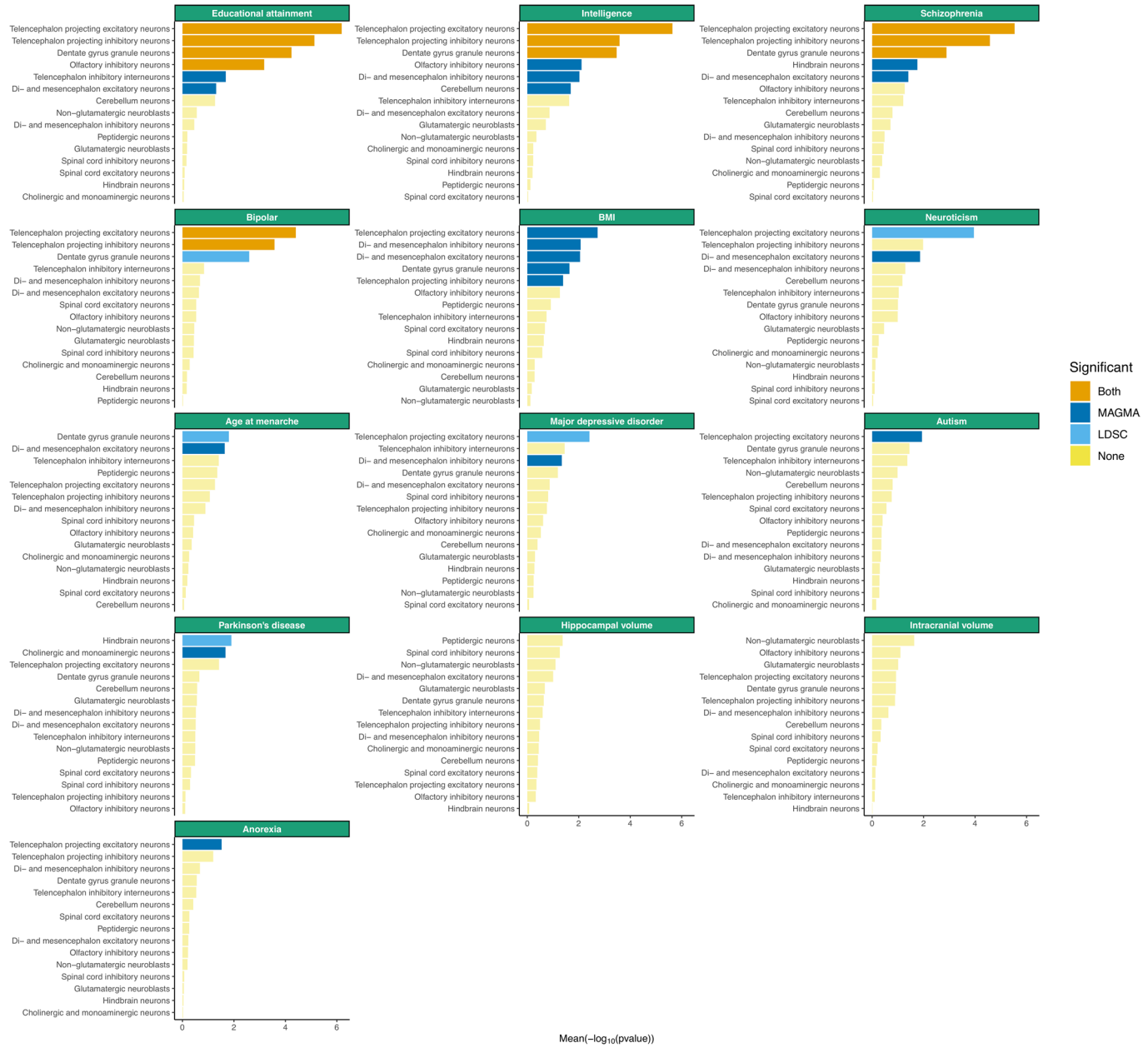
Extended Data Fig. 1 | Enrichment of immune genes in GTEx tissues. Enrichment pvalues of genes belonging to the GO term 'Immune System Process' in the 10% most specific genes in each tissue. The one-sided pvalues were computed using linear regression, testing whether the average specificity metric of the gene set was higher than 0 (z-scaled specificity metrics per tissue). The GO term was selected because it is the most associated with inflammatory bowel disease using MAGMA.



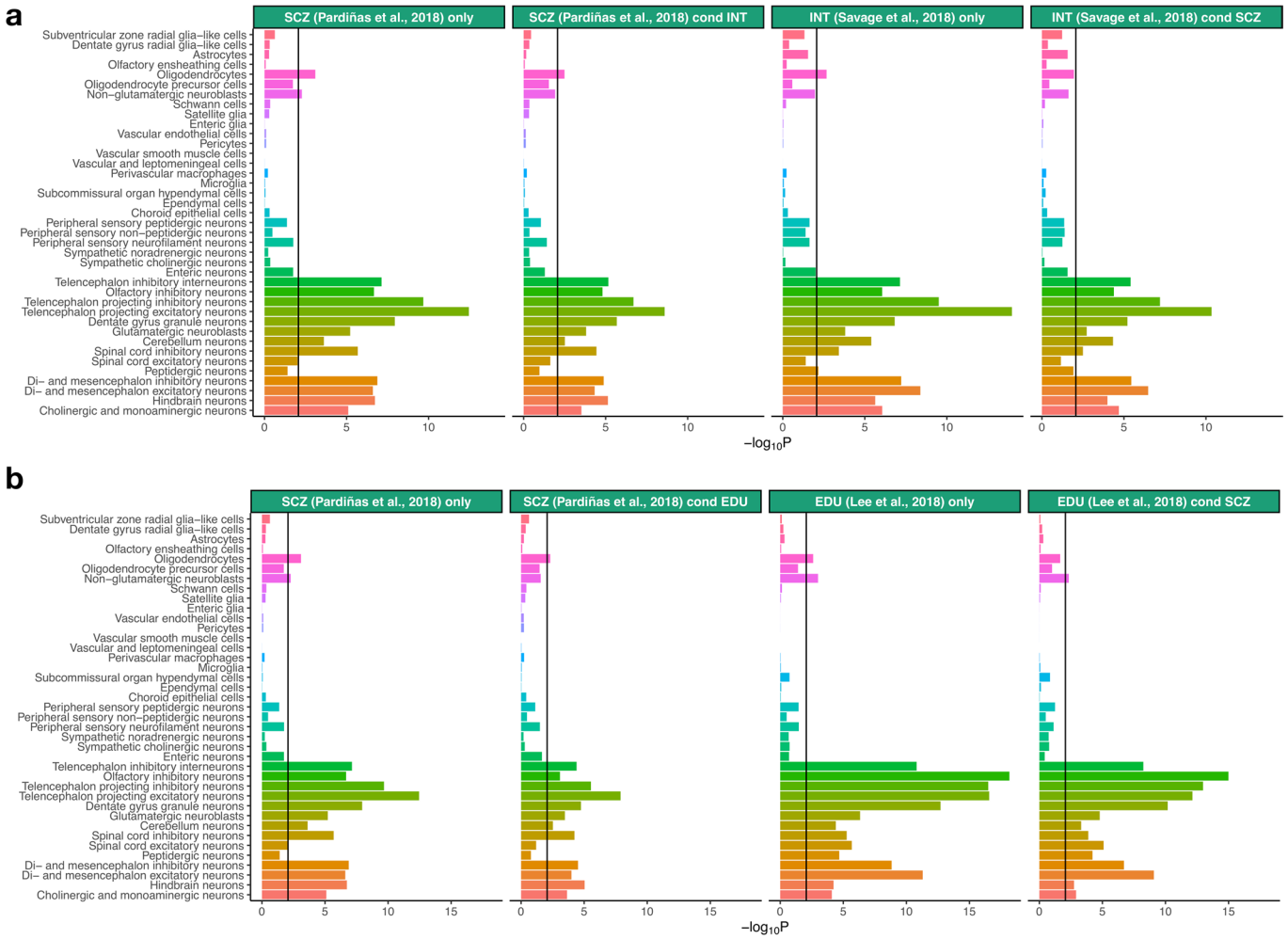
Extended Data Fig. 2 | Associations of brain related traits with cell types from the entire mouse nervous system. Associations of the top 15 most associated cell types are shown. The mean strength of association ($-\log_{10}$P) of MAGMA and LDSC is shown and the bar color indicates whether the cell type is significantly associated with both methods, one method or none (significance threshold: 5% false discovery rate).



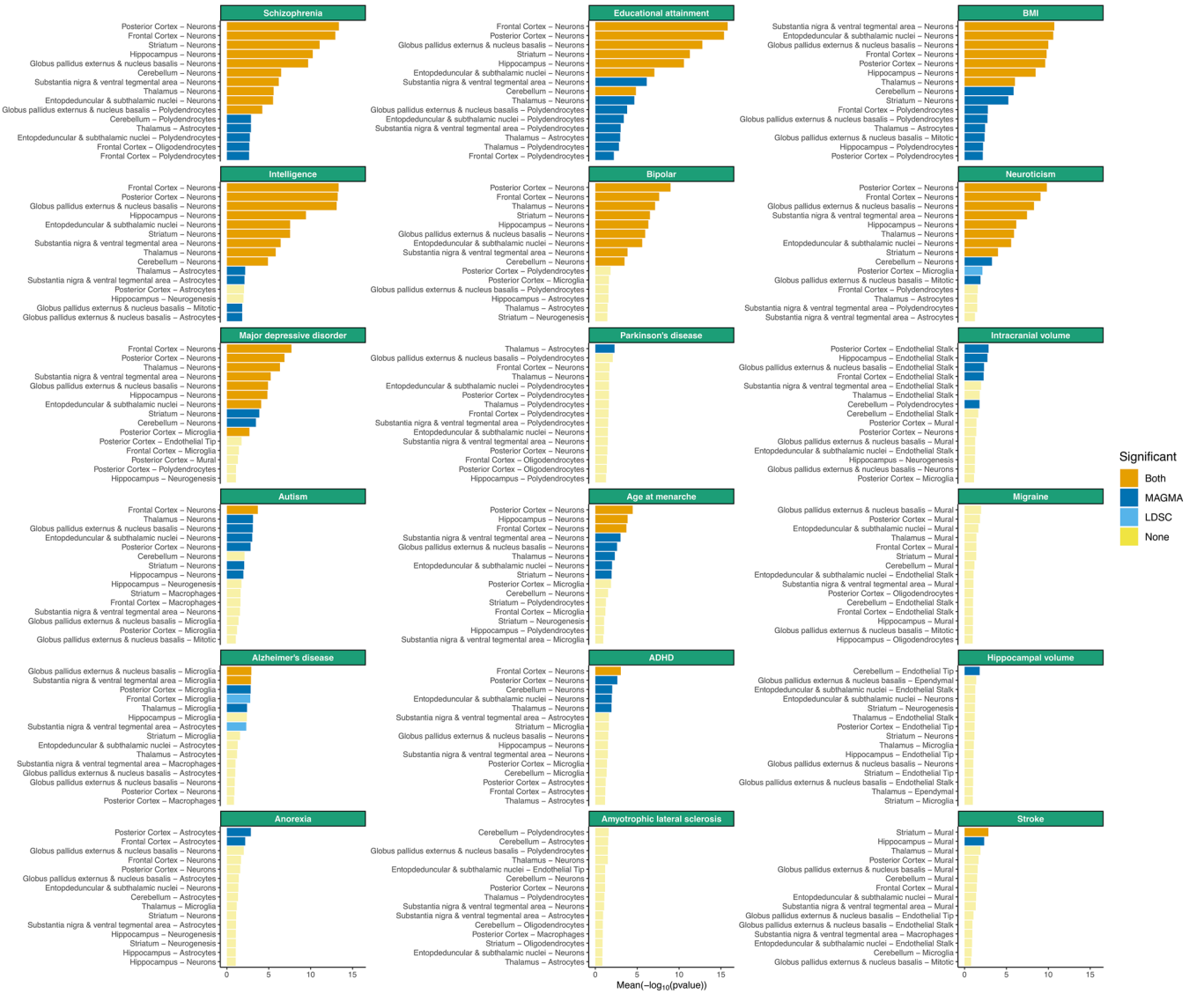
Extended Data Fig. 3 | Correlation in cell type associations across traits. The Spearman rank correlations between the cell types associations across traits (-$\log_{10}$P) are shown. SCZ (schizophrenia), EDU (educational attainment), INT (intelligence), BMI (body mass index), BIP (bipolar disorder), NEU (neuroticism), PAR (Parkinson's disease), MDD (Major depressive disorder), MEN (age at menarche), ICV (intracranial volume), ASD (autism spectrum disorder), STR (stroke), AN (anorexia nervosa), MIG (migraine), ALS (amyotrophic lateral sclerosis), ADHD (attention deficit hyperactivity disorder), ALZ (Alzheimer's disease), HIP (hippocampal volume).



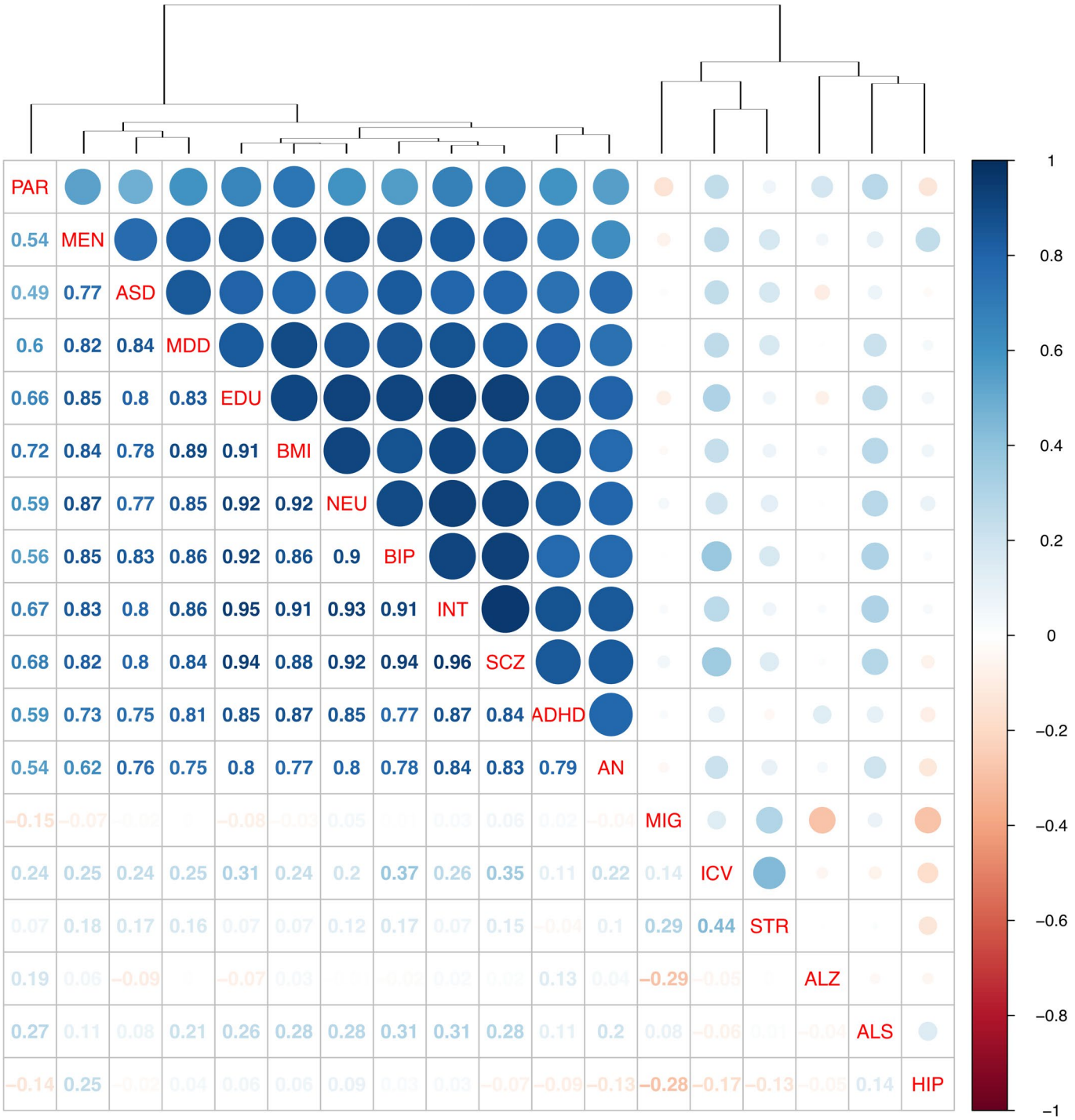
Extended Data Fig. 4 | Associations of brain related traits with neurons from the central nervous system. Associations of the 15 most associated neurons from the central nervous system (CNS) are shown. The specificity metrics were computed only using neurons from the CNS. The mean strength of association (-$\log_{10}$P) of MAGMA and LDSC is shown and the bar color indicates whether the cell type is significantly associated with both methods, one method or none (significance threshold: 5% false discovery rate).



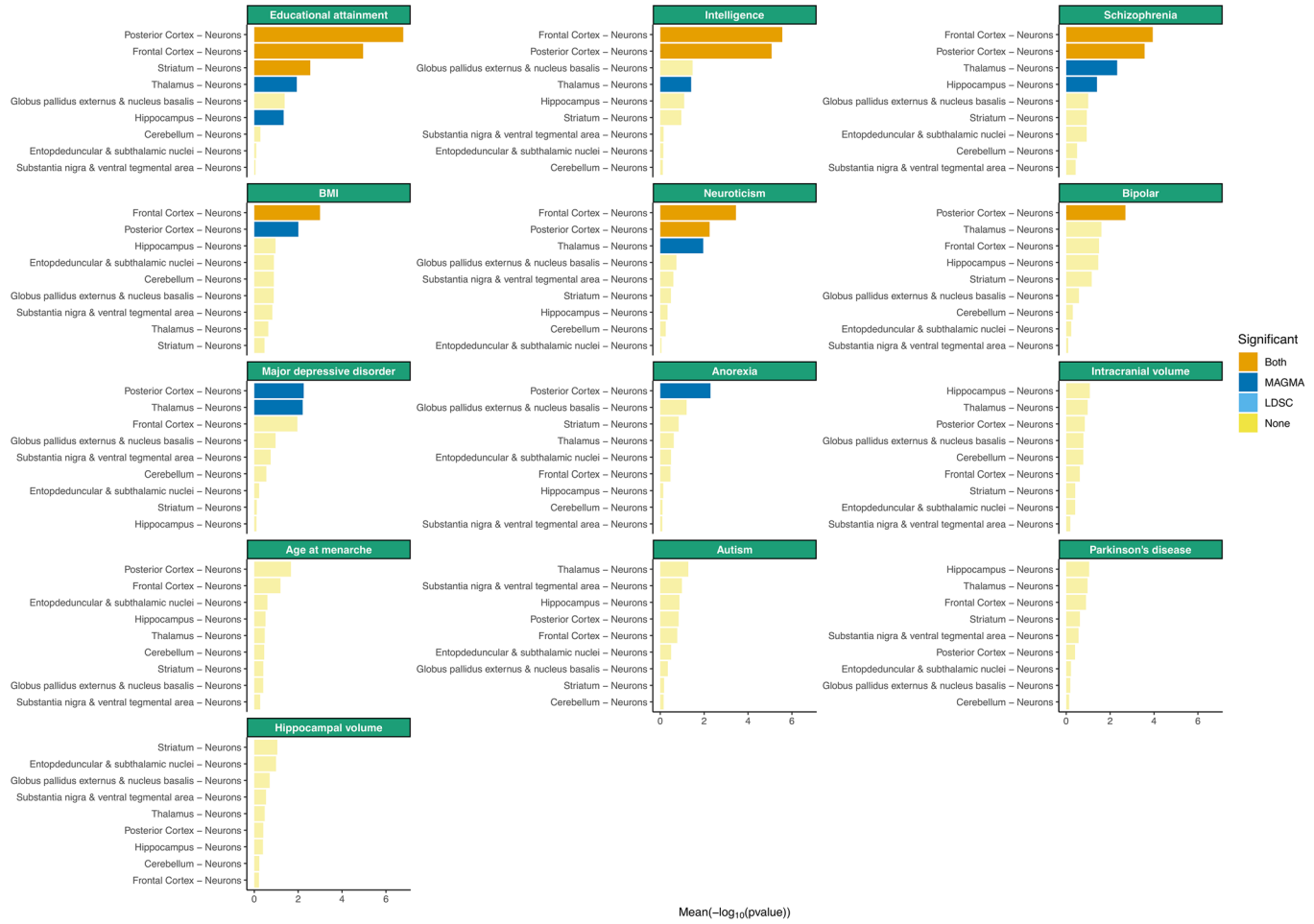
Extended Data Fig. 5 | Associations of cell types with schizophrenia/cognitive traits conditioning on gene-level genetic association of cognitive traits/schizophrenia. MAGMA association strength for each cell type before and after conditioning on gene-level genetic association for another trait. The black bar represents the significance threshold (5% false discovery rate). SCZ (schizophrenia), INT (intelligence), EDU (educational attainment).



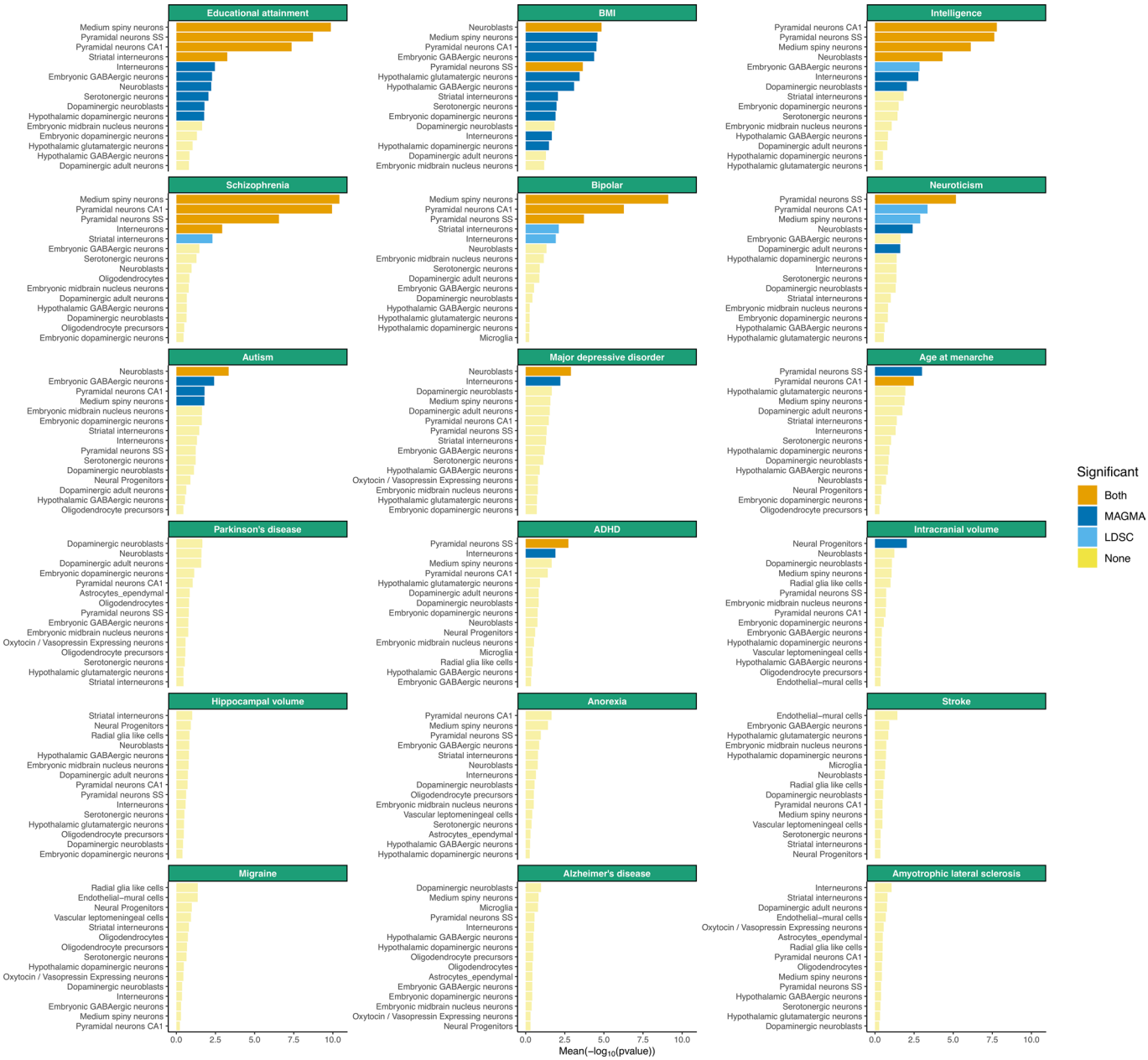
Extended Data Fig. 6 | Replication of cell type—trait associations in 88 cell types from 9 different brain regions. The mean strength of association ($-\log_{10}$P) of MAGMA and LDSC is shown for the 15 top cell types for each trait. The bar color indicates whether the cell type is significantly associated with both methods, one method or none (significance threshold: 5% false discovery rate).



Extended Data Fig. 7 | Correlation in cell type associations across traits in a replication data set (88 cell types, 9 brain regions). Spearman rank correlations for cell types associations ($-\log_{10}P$) across traits are shown. SCZ (schizophrenia), EDU (educational attainment), INT (intelligence), BMI (body mass index), BIP (bipolar disorder), NEU (neuroticism), PAR (Parkinson's disease), MDD (Major depressive disorder), MEN (age at menarche), ICV (intracranial volume), ASD (autism spectrum disorder), STR (stroke), AN (anorexia nervosa), MIG (migraine), ALS (amyotrophic lateral sclerosis), ADHD (attention deficit hyperactivity disorder), ALZ (Alzheimer's disease), HIP (hippocampal volume).



Extended Data Fig. 8 | Associations of brain related traits with neurons from 9 different brain regions. Trait—neuron association are shown for neurons of the 9 different brain regions. The specificity metrics were computed only using neurons. The mean strength of association ($-\log_{10}$P) of MAGMA and LDSC is shown and the bar color indicates whether the cell type is significantly associated with both methods, one method or none (significance threshold: 5% false discovery rate).



Extended Data Fig. 9 | Top associated cell types with brain related traits among 24 cell types from 5 different brain regions. The mean strength of association (-$\log_{10}$P) of MAGMA and LDSC is shown for the 15 top cell types for each trait. The bar color indicates whether the cell type is significantly associated with both methods, one method or none (significance threshold: 5% false discovery rate).

Extended Data Fig. 10 | Top associated neurons with brain related traits among 16 neurons from 5 different brain regions. The specificity metrics were computed only using neurons. The mean strength of association ($-\log_{10}$P) of MAGMA and LDSC is shown for the top 15 cell types for each trait. The bar color indicates whether the cell type is significantly associated with both methods, one method or none (significance threshold= 5% false discovery rate).

nature research

Corresponding author(s): Patrick F Sullivan

Last updated by author(s): Mar 4, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☒	☐ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data are publicly available. The 10,000 most associated SNPs from the 23andMe cohort are available in Supplementary Table 12.
Data analysis	Code is available at: https://github.com/jbryois/scRNA_disease

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data are publicly available. The 10,000 most associated SNPs from the 23andMe cohort are available in Supplementary Table 12.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	N/A
Data exclusions	GTEx tissues were excluded if expression was measured in less than 100 samples, the tissues were non-natural (e.g. lymphoblast cell lines) and testis (gene expression outlier).
Replication	Multiple datasets in both human and mouse were used for replication
Randomization	N/A
Blinding	N/A

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging