Genetic Identification of Cell Types Underlying Brain Complex Traits Yields Novel Insights Into the Etiology of Parkinson's Disease

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

1 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, SE-17177 Stockholm, Sweden

2 Laboratory of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, SE-17177 Stockholm, Sweden

3 UCL Institute of Neurology, Queen Square, London, UK

4 Division of Brain Sciences, Department of Medicine, Imperial College, London, UK

5 Danish Headache Center, Dept. of Neurology, Copenhagen University Hospital, Glostrup, Denmark

6 Institute of Biological Psychiatry, Copenhagen University Hospital MHC Sct. Hans, Roskilde, Denmark

7 Novo Nordic Foundations Center for Protein Research, Copenhagen University, Denmark.

8 Department of Psychiatry, University of North Carolina at Chapel Hill, North Carolina, US

9 23andMe, Inc., Mountain View, CA, 94041, USA

10 Department of Psychiatry, University of Iowa Carver College of Medicine, University of Iowa, Iowa City, Iowa.

11 Institute of Psychiatry, Psychology and Neuroscience, MRC Social, Genetic and Developmental Psychiatry (SGDP) Centre, King's College London, London, UK

12 National Institute for Health Research Biomedical Research Centre, South London and Maudsley National Health Service Trust, London, UK

13 Department of Nutrition, University of North Carolina, Chapel Hill, NC, 27599-7264, USA

14 Departments of Genetics, University of North Carolina, Chapel Hill, NC, 27599-7264, USA

* Correspond with Dr Hjerling-Leffler (jens.hjerling-leffler@ki.se) and Dr Sullivan (patrick.sullivan@ki.se).

† Equal contributions.

Abstract

Genome-wide association studies (GWAS) have discovered hundreds of loci associated with complex brain disorders, and provide the best current insights into the etiology of these idiopathic traits. However, it remains unclear in which cell types these variants may be active, which is essential for understanding disease etiology and for disease modelling. Here we integrate GWAS results with single-cell transcriptomic data from the entire nervous system to systematically identify cell types underlying psychiatric disorders, neurological conditions, and other brain complex traits. We show that psychiatric disorders are predominantly associated with excitatory neurons from the cortex/hippocampus, medium spiny neurons from the striatum, diverse sets of midbrain neurons, and inhibitory neurons from the cortex/hippocampus.

Cognitive traits were generally associated with similar cell types but their associations were driven by different genes. Neurological disorders were associated with different cell types, consistent with other lines of evidence. Notably, we found that Parkinson's disease is not only genetically associated with dopaminergic neurons but also with serotonergic neurons and cells from the oligodendrocyte lineage. Using post-mortem brain transcriptomic data, we confirmed alterations in these cells, even at the earliest stages of disease progression. Altogether, our study provides a solid framework for understanding the cellular basis of complex brain disorders and reveals a new unexpected role of oligodendrocytes in Parkinson's disease.

Introduction

Understanding factors fundamental to complex brain disorders is critical for identifying individuals at risk, designing prevention strategies, and developing rational therapeutics. Over the last 50 years, twin studies have shown that psychiatric disorders, neurological disorders and cognitive traits are strongly influenced by genetic factors, explaining on average ~50% of the variance in liability to a given trait [1]. In order to identify which genetic variants contribute to the genetic variance of complex traits, GWAS have been performed and successfully identified thousands of loci [2–5]. However, interpretation of GWAS results remains challenging for several reasons. First, more than 90% of the discovered variants are located in non-coding regions [6], which challenges the identification of risk genes. Second, identification of causal variants at genome-wide significant loci is complicated by the extensive linkage disequilibrium present in the human genome. Finally, it remains unclear in which tissues and cell types these variants are active, and how they disrupt specific biologic networks to impact disease risk.

Functional genomic studies are critical for the interpretation of GWAS findings because they aid in identification of functional regions (e.g., open chromatin, enhancers, transcription factors binding sites) and target genes (e.g., chromatin interactions and eQTLs) [7]. Gene regulation varies substantially across tissues and cell types, and hence it is critical to perform functional genomic studies in empirically identified cell types or tissues.

Previous studies have developed strategies to identify tissues associated with complex traits [8–12], but few have focused on the identification of relevant cell types. Furthermore, studies aiming to identify relevant cell types have typically used only a small number of cell types derived from one or few different brain regions [10–12]. For example, we recently showed that, among 24 brain cell types tested, four types of neurons were consistently associated with schizophrenia [10]. However, other cell types from unsampled regions could also contribute to the disorder.

Here, we integrate genetics with gene expression from tissues across the human body and single-cell gene expression data from the entire nervous system to identify tissues and cell types underlying a large number of complex traits (**Figure 1A,B**). We expand on our prior work by showing that additional cell types are associated with schizophrenia. We also 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 consistently associated with dopaminergic neurons (as expected), but also with serotonergic neurons and oligodendrocytes, the latter providing new clues into its etiology.

Results

Genetic correlations among complex traits

We selected 28 GWAS for which we attempted to identify relevant tissues and/or cell types: 19 traits with clear prior evidence that they are brain-related and 9 non-brain traits with large number of genome-wide significant loci for comparative purposes (**Methods**). For Parkinson's disease, we meta-analyzed summary statistics from a published GWAS [13] (9,581 cases, 33,245 controls, excluding the 23andMe cohort) with self-reported Parkinson's disease from 23andMe (12,657 cases, 941,588 controls) after estimating a high genetic correlation between the samples (LDSC [14] genetic correlation (r_g) = 0.87, s.e = 0.0679).

We estimated the genetic correlations (r_g) between these 28 traits. As reported previously [15,16], we confirmed that psychiatric disorders were strongly inter-correlated (e.g., high positive correlations for schizophrenia, bipolar disorder, and major depressive disorder (MDD)) and shared little overlap with neurological disorders (**Figure S1** and **Table S1**). Interestingly, Parkinson's disease was genetically correlated with intracranial volume [17] (r_g= 0.29, s.e = 0.05) and amyotrophic lateral sclerosis (ALS, r_g= 0.19, s.e = 0.08), while ALS was negatively correlated with intelligence (r_g = -0.24, s.e = 0.06) and hippocampal volume (r_g = -0.24, s.e = 0.12). Altogether, we report novel genetic correlations among brain-related complex traits and show substantial genetic heterogeneity across traits, suggesting that different tissues and cell types contribute to the different traits.

Association of traits with tissues using bulk-tissue RNA-seq

We first aimed to identify which human tissues were enriched for genetic associations with our selected traits using bulk-tissue RNA-seq from 53 tissues from GTEx [18] (**Figure 1**). To robustly identify the tissues implied by these 28 GWAS, we used two approaches (MAGMA [19] and LDSC [11,20]) which employ different assumptions (**Methods**). MAGMA tested for a positive relationship between gene expression specificity and gene-level genetic associations (**Figure 1B**), whereas LDSC tested whether the 10% most specific genes in each tissue were enriched for trait heritability (**Figure 1B**).

As a control for our methodology, we examined non-brain traits and found that they were generally associated with tissues known to be involved in their pathophysiology. For example, inflammatory bowel disease was strongly associated with immune tissues (whole blood, spleen) and the digestive tissues impacted by the disease (terminal ileum and transverse colon, **Figure 1D**). Coronary artery disease was most associated with aorta and coronary artery (**Figure 1D**). Total cholesterol was most associated with liver, the organ with a dominant role in cholesterol biosynthesis. Age

at menopause was most associated with reproductive tissues (**Figure 1D**) and type 2 diabetes with the pancreas (**Figure S2**). Thus, we show that our integrative approach appropriately highlights tissues relevant to the pathophysiology of the different traits.

We then investigated brain-related traits and found that 12 out of 19 traits were significantly associated with at least one of the GTEx brain tissues (**Table S2**). For example, schizophrenia, intelligence, educational attainment, neuroticism, and MDD were most significantly associated with brain cortex or frontal cortex while Parkinson's disease was most significantly associated with substantia nigra and spinal cord (**Figure 1C** and **Table S2**). In addition, autism and ADHD were most strongly associated with structures from the basal ganglia (putamen and nucleus accumbens) (**Table S2**).

ALS (motor neuron disease [21]), Alzheimer's disease, intracranial volume, migraine, and stroke only had significant associations with non-brain related tissues (**Table S2**), suggesting that the biological mechanisms at the core of these traits are not necessarily brain-specific. For instance, stroke was significantly associated with coronary artery and aorta, consistent with the frequent arterial causation of stroke[23].

Most brain-related traits were strongly associated with the cerebellum. Although there are theories postulating a cerebellar role in the etiology of psychiatric disorders [24,25], an alternative explanation is that cerebellum has a notably high proportion of neurons (78-81% [26,27]). We hypothesized that the relative proportion of neurons in the different brain regions could confound our tissue-trait association analysis leading to more significant associations for traits with a neuronal basis in brain regions with high neuronal proportions. To test this hypothesis, we obtained estimates of neuronal proportions in different brain regions (from mouse single cell RNA-seq [26]) and tested whether the strength of the tissue-trait association ($-\log_{10}P$) was correlated with the estimated neuronal proportions of the different brain regions. For most traits, the tissue-trait association strength was positively correlated with the proportion of neurons in the brain region (**Figure S3**), suggesting that most of the tested traits are of neuronal origin and that the cerebellar associations are partly or largely due to the high neuronal proportion of this brain region. Alzheimer's disease and Parkinson's disease had lower association strength for brain regions with a higher proportion of neurons, suggesting that non-neuronal cell types could play a role in these disorders.

In conclusion, we show that tissue-level gene expression allows us to pinpoint relevant tissues for complex traits. However, cellular heterogeneity can confound these associations, highlighting the need to test for trait-gene expression associations at the cell type level.

Association of brain complex traits with cell types from the mouse central and peripheral nervous system

We leveraged gene expression data from 231 cell types from the mouse central and peripheral nervous system [26] to systematically map brain-related traits to cell types (**Figure 2** and **S4**). In agreement with our previous findings for schizophrenia [10], we found the strongest signals for pyramidal neurons from the cortex, pyramidal neurons

from the CA1 region of the hippocampus, one type of excitatory neuron from the dorsal midbrain, D2 medium spiny neurons from the striatum, and interneurons from the hippocampus (**Figure 2** and **Table S3**). In addition, we observed that many other types of neurons were associated with schizophrenia, although less significantly (e.g., excitatory neurons from the thalamus and inhibitory neurons from the hindbrain).

The cell type with the strongest schizophrenia association (TEGLU4) is located in layer 5 of the cortex. This finding is consistent with our previous findings [10], as well as with strong schizophrenia heritability enrichment in mouse open chromatin regions of layer 5 excitatory neurons [28] and in open chromatin regions from human cortical neurons [29]. This pattern of replication from multiple single-cell transcriptomic studies as well as a different functional genomic assay (open chromatin) is notably consistent and implies an important role of excitatory neurons from the cortex in schizophrenia.

Excitatory neurons from cortical layer 5 project to the superior colliculus [30] (a region implicated in attentional target selection) and to the striatum [30]. The eighth most significant cell type (MEGLU6, a midbrain excitatory neuron) is located in the superior colliculus, while the twelfth top cell type (MSN2, a D2 medium spiny neurons) is located in the striatum. This pattern of findings may imply a functional circuit between these different cell types that is etiologically important for schizophrenia.

Educational attainment, intelligence, bipolar disorder, neuroticism, and MDD had similar cellular association patterns to schizophrenia (**Figure S4**, **S5** and **Table S3**). There were strong enrichments for projecting excitatory neurons from the cortex and hippocampus, projecting inhibitory neurons (i.e., medium spiny neurons), olfactory inhibitory neurons, inhibitory neurons from the telencephalon, and neurons from the diencephalon and mesencephalon. We did not observe any significant associations with immune or vascular cells for any psychiatric disorders or cognitive traits. While the general pattern of associations was similar across those traits, we observed potentially meaningful differences in rankings of the most associated cell types (**Table S3**). For example, granule neurons from the dentate gyrus in the hippocampus (DGGRC1) were the fifth most associated cell type with bipolar disorder but only the 61st for schizophrenia.

For body mass index (BMI), the pattern of associations contrasted with psychiatric and cognitive traits (**Figure 2** and **S4**) as it had the strongest associations with neurons located in the midbrain, hypothalamus and solitary nucleus (HBGLU3), followed by cortical excitatory neurons (**Figure 2B**). The hypothalamus is known to play a major role in weight regulation [31], and the solitary nucleus is involved in gustatory processing and projects to the hypothalamus [32]. These results suggest functional connections between these cell types.

Neurological disorders generally implicated fewer cell types, possibly because neurological GWAS were relatively small compared to cognitive, anthropometric, and psychiatric traits. Consistent with the genome-wide genetic correlations reported above, the general pattern of associations for neurological disorders was distinct from psychiatric disorders (**Figures 2**, **S4** and **S5**), again reflecting that neurological disorders have minimal functional overlap with psychiatric disorders [15] (**Figure S1**).

Stroke and migraine were significantly associated with arterial smooth muscle cells (VSMCA), indicating that vascular process play a central role in these traits. The overall pattern of association of stroke and migraine were negatively correlated with psychiatric, cognitive traits, and BMI (**Figure S5**), as stroke and migraine appear to have a predominant non-neuronal origin.

The hallmark of Parkinson's disease is the degeneration of dopaminergic neurons of the substantia nigra. Consistently, we observed that these cells (MBDOP2) were highly associated with the disorder (second top hit) (**Figure 2B**). In addition to the degeneration of dopaminergic neurons, extranigral cell loss is also a standard feature of the disease [33,34]. We found that many of the cell types reported to degenerate in Parkinson's disease also show strong associations with its genetics (**Figure 2B** and **Table S3**), including serotonergic raphe nucleus neurons [35] (HBSER2, top hit) and cholinergic neurons of the Pons [36] (HBCHO2). We also identified significant enrichments in cells of the medulla (HBINH4 and HBGLU2), the region which is associated with the earliest lesions in Parkinson's disease [33]. Therefore, our results capture expected features of Parkinson's disease and suggest that biological mechanisms intrinsic to these neuronal cell types lead to their selective loss. However, we also found association of eight different cell types from the oligodendrocyte lineage with Parkinson's disease (NFOL2, MFOL1, COP1, NFOL1, MOL1, MFOL2, COP2, MOL2) (**Figure 2** and **Table S3**), indicating a strong glial component. This finding was unexpected but consistent with the strong association of the spinal cord at the tissue level (second top hit in **Figure 1C**), as the spinal cord contains the highest proportion of oligodendrocytes (71% [26]) in the nervous system.

Cell type-specific and trait-associated genes are enriched in specific biological functions

Understanding which biological functions are dysregulated in different cell types is key to understanding the etiology of complex traits. To obtain insights into the biological functions driving cell type-trait associations, we evaluated GO term enrichment of genes that were specifically expressed (top 20% in a given cell type) and highly associated with a trait (top 10% MAGMA gene-level genetic association). Genes that were highly associated with schizophrenia and specific to excitatory neurons from the cortex (TEGLU4), excitatory neurons from the hippocampus (TEGLU24), excitatory neurons from the midbrain (MEGLU6), or medium spiny neurons (MSN2) were enriched for GO terms related to neurogenesis, synapses, and voltage-gated channels (**Table S4**), suggesting that these functions may be fundamental to schizophrenia. Similarly, genes highly associated with educational attainment, intelligence, bipolar disorder, neuroticism, and MDD and highly specific to their most associated cell types (TEGLU4 for educational attainment, TEGLU10 for intelligence, TEGLU24 for bipolar disorder, TEGLU4 for neuroticism, and TEGLU10 for MDD) were strongly enriched in neurogenesis, synaptic processes and voltage-gated channels (**Table S4**).

Genes highly associated with Parkinson's disease and highly specific to the top cell type (HBSER2) were significantly enriched in terms related to regulation of protein localization, lysosomal transport, and intracellular vesicles (**Table S4**). Genes highly

specific to the second top cell type (MBDOP2) were enriched in terms related to endosomes and synaptic vesicles (**Table S4**), while highly specific genes for the most associated type of oligodendrocytes (NFOL2) were enriched in terms related to endosomes (**Table S4**). These results support the hypothesis that intracellular trafficking, lysosomal transport, and synaptic vesicles play a role in the etiology of Parkinson's disease [37].

Taken together, we show that cell type-trait associations are driven by genes belonging to specific biological pathways, providing insight into the etiology of brain complex traits.

Distinct traits are associated to similar cell types, but through different genes

As noted above, the pattern of associations of psychiatric and cognitive traits were highly correlated across the 231 different cell types tested (**Figure S5**). For example, the Spearman rank correlation of cell type associations ($-\log_{10}P$) between schizophrenia and intelligence was 0.94 as both traits had the strongest signal in cortical 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 GO terms involving neurogenesis and synaptic processes. We evaluated two possible explanations for these findings: (a) schizophrenia and intelligence are both associated with the same genes that are specifically expressed in the same cell types or (b) schizophrenia and intelligence are associated with different sets of genes that are both highly 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) (**Figure S1** and **Table S1**), 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 cell type gene expression specificity was positively correlated with gene-level genetic association for schizophrenia controlling for the gene-level genetic association of intelligence. We found that the pattern of associations were largely unaffected by controlling the schizophrenia cell type association analysis for the gene-level genetic association of intelligence and vice versa (**Figure S6**). Similarly, we found that controlling for educational attainment had little effect on the schizophrenia association and vice versa (**Figure S7**). In other words, genes driving the cell type associations of schizophrenia appear to be distinct from genes driving the cell types associations of cognitive traits.

Multiple cell types are independently associated with brain complex traits

Many neuronal cell types passed our stringent significance threshold for multiple brain traits (**Figure 2** and **S4**). This could be because gene expression specificity profiles are highly correlated across cell types and/or because many cell types are independently associated with the different traits. In order to address this, we performed univariate conditional analysis using MAGMA, testing whether cell type associations remained significant after controlling for gene expression specificity from other cell types (**Table S5**). We observed that multiple cell types were independently

associated with educational attainment (**Figure S8**), intelligence (**Figure S9**) and BMI (**Figure S10**). No cell type remained significantly associated with bipolar disorder (**Figure S11**), neuroticism (**Figure S12**), MDD (**Figure S13**), and age at menarche (**Figure S14**) after conditioning on a single cell type (TEGLU18, TEGLU15, TEGLU10 and MEINH11 respectively). As these GWAS tend to have a lower number of genome-wide significant hits, it is possible that other cell types could independently contribute to these traits but that power is currently insufficient to detect independent effects. Multiple cell types could explain all significant associations for intracranial volume (**Figure S15**), anorexia nervosa (**Figure S16**) and autism (**Figure S17**).

We observed that seven types of projecting excitatory neurons from the telencephalon were independently associated with schizophrenia (TEGLU20, TEGLU9, TEGLU13, TEGLU15, TEGLU18, TEGLU4 and TEGLU24), while one type of medium spiny neurons (MSN2) was sufficient to explain the signal of telencephalon projecting inhibitory neurons (**Figure 3**). In addition, two types of telencephalon inhibitory interneurons were independently associated with schizophrenia (DEINH1 in thalamus and TEINH13 in hippocampus). TEINH13 is a *Reln* and *Ndnf* expressing interneuron subtype and corresponds to 'Cortical Interneuron 16' which was the most associated interneuron subtype in our previous study [10]. We also observed independent signals in olfactory inhibitory neurons, cholinergic and monoaminergic neurons, peptidergic neurons, diencephalon and mesencephalon neurons, hindbrain neurons, cerebellum neurons, glutamatergic neuroblasts, and oligodendrocytes.

For Parkinson's disease, the association of HBSER2 (serotonergic neurons located in the raphe nuclei) was not independent of MBDOP2 (dopaminergic neurons from the ventral midbrain) and vice-versa. HBSER2 and MBDOP2 remained significant after conditioning on different types of oligodendrocytes. Conditioning on HBSER2 attenuated the oligodendrocyte signal to below our significance threshold, while conditioning on MBDOP2 resulted in one type of oligodendrocytes (COP1) remaining significant (**Figure S18**). Therefore, it seems that while the associations with monoaminergic neurons (MBDOP2 and HBSER2) depend on each another, those with MBDOP2 and COP1 are independent, suggesting a critical involvement of these two different cell types in Parkinson's disease.

Replication in other single-cell RNA-seq datasets

To assess the robustness of our results, we attempted to replicate our findings in independent datasets. An important caveat is that other datasets did not sample cell types from the entire nervous system as in the results described above. First, we analyzed a large single-cell RNA-seq dataset that identified 565 cell types (~690k single cells sequenced) from 9 different mouse 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) [38]. We found relatively similar patterns of association in this external dataset (**Figure S19** and **Table S6**). Notably, we replicated strongly the association of excitatory neuron from the cortex, the CA1 region of the hippocampus, medium spiny neurons, and interneurons for schizophrenia (**Figure S20**). We also observed high similarity in cell type associations in other psychiatric traits (**Figure**

S21). For neurological disorders, we replicated the association of: Parkinson's disease with dopaminergic neurons from the ventral midbrain as well as oligodendrocytes (serotonergic neurons were not sampled); for stroke and migraine with arterial smooth muscle cells; and Alzheimer's disease was significantly associated with activated microglia (Figure S20), a cell type previously associated with the disease [9,22].

Second, we reanalyzed these new GWAS datasets using our previous single-cell RNA-seq dataset (24 cell types from the neocortex, hippocampus, striatum, hypothalamus midbrain, and specific enrichment for oligodendrocytes, serotonergic neurons, dopaminergic neurons and cortical parvalbuminergic interneurons, 9970 single cells sequenced in total) (Figure 4 and Table S7). We again found strong associations of pyramidal neurons from the somatosensory cortex, pyramidal neurons from the CA1 region of the hippocampus, and medium spiny neurons from the striatum with psychiatric and cognitive traits. We further replicated the associations of: Parkinson's disease with adult dopaminergic neurons and oligodendrocytes (serotonergic neurons replicated with LDSC only); migraine with vascular cells; stroke with endothelial cells; and intracranial volume with neuronal progenitors and neuroblasts (suggesting that drivers of intracranial volume are cell types implicated in increasing cell mass).

We then evaluated a human single-nuclei RNA-seq dataset consisting of 15 different cell types from 2 brain regions [39] (cortex and hippocampus) (Figure 5 and Table S8). We largely replicated our findings with psychiatric and cognitive traits being associated with pyramidal neurons from the somatosensory cortex and from the CA1 region of the hippocampus. We also replicated the association of Parkinson's disease with oligodendrocytes in human (dopaminergic and serotonergic neurons were not sampled). We did not find any significant association for other neurological disorders in this dataset.

Validation of oligodendrocyte pathology in Parkinson's disease

We further investigated the role of oligodendrocyte lineage cells in Parkinson's disorder. First we used EWCE [9] to test whether genes with rare variants associated with Parkinsonism (listed in Table S9, obtained from OMIM) were specifically expressed in cell types from the mouse nervous system. We found that dopaminergic neurons (MBDOP2) were the most significantly enriched ($p<0.0001/231$, Table S10) and enrichment were again found for cholinergic (HBCHO4 and HBCHO3) and serotonergic (HBSER1 and HBSER3) neurons. However, we did not observe any significant enrichments in the oligodendrocyte lineage for genes associated with rare variants in Parkinsonism.

Second, we applied EWCE [9] to test whether genes that are up/down-regulated in human post-mortem Parkinson's disease brains (from six separate cohorts) were enriched in specific cell types (Figure 6). Three of the studies had a case-control design and measured gene expression in: (a) the substantia nigra of 9 controls and 16 cases [40], (b) the medial substantia nigra of 8 controls and 15 cases [41], and (c) the lateral substantia nigra of 7 controls and 9 cases [41]. 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 [42].

We also analyzed gene expression data from post-mortem human brains which had been scored by neuropathologists for their Braak stage [43]. Differential expression was calculated between brains with Braak scores of zero (controls) and brains with Braak scores of 1—2, 3—4 and 5—6. At the latter stages (Braak scores 3—4 and 5—6), downregulated genes were specifically expressed in dopaminergic neurons, while upregulated genes were specifically expressed in oligodendrocytes (**Figure 6**), as observed in the case-control studies. Moreover, Braak stage 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 indicate that their involvement precedes the emergence of pathological changes in the substantia nigra.

Discussion

In this study, we leveraged cell type gene expression data from the entire nervous system to systematically map cell types to GWAS results for multiple psychiatric, cognitive, and complex neurological disorders.

We note several limitations. First, we used gene expression data from mouse to understand human disorders. We believe our approach is appropriate for several reasons: 1) single-cell RNAseq is achievable in mouse but difficult in human neurons and the transcripts that are lost are important for psychiatric disorders [10]; 2) we only evaluated protein coding genes with 1:1 orthologs between mouse and human, which are highly conserved; 3) we previously showed that gene expression data clusters by cell type and not by species [10], indicating broad conservation of core brain cellular functions across species; 4) we used a large number of genes to map cell types to traits, negating the potential effect of individual genes that would be differentially expressed across species; 5) we replicated key findings in human data; and 6) if there were strong differences in cell type gene expression between mouse and human, we would not expect that specific genes in mouse cell types would be enriched in genetic associations with human disorders. Second, gene expression data were obtained from adolescent mice (6-8 weeks old). While this age range is in line with the first manifestations of many human psychiatric and neurological disorders, many people have earlier or later onsets. It is possible that some cell types are particularly vulnerable at specific times during development or only after a certain age. Further studies mapping cell types over nervous system development and during aging will be necessary to address this. Finally, we note that we can implicate a particular cell type but it is premature to exclude cell types for which we do not have data.

By using two independent data sets, we replicated and extended our previous findings [10] that implicated excitatory neurons from the cortex/hippocampus, D2 medium spiny

neurons from the striatum, and interneurons in schizophrenia. These results are consistent with the strong schizophrenia heritability enrichment observed in open chromatin regions from: (a) human dorsolateral prefrontal cortex [44], (b) human cortical, striatal and hippocampal neurons [29], and (c) mouse open chromatin regions from cortical excitatory and inhibitory neurons [28]. This degree of replication across different single cell RNA-seq datasets from multiple groups along with consistent findings using open chromatin is rather notable, and strongly implicates these cell types in the etiology of schizophrenia.

As in mouse open chromatin data [28], we observed the strongest schizophrenia association for excitatory neurons located in the layer 5 of the cortex (TEGLU4). In addition, other types of projecting excitatory neurons were independently associated with schizophrenia (TEGLU20, TEGLU9, TEGLU13, TEGLU15, TEGLU18, TEGLU24) and were located in the deep layers of the cortex (layer 5 or 6), in the subiculum, in the CA1 region of the hippocampus and in the piriform cortex. This is intriguing as layer 5 excitatory neurons project to the striatum and the superior colliculus [30], where some of the most schizophrenia associated cell types are located (MSN2, a type of D2 medium spiny neurons and MEGLU6, a type of excitatory neuron). In addition, excitatory neurons from the CA1 region of the hippocampus are known to primarily project to the subiculum, which then projects to many cortical and subcortical regions [45]. This suggests the potential salience of a circuit between independently associated cell types in schizophrenia.

We observed that psychiatric disorders implicated largely similar cell types with the strongest signal for excitatory neurons from the cortex and hippocampus, medium spiny neurons, interneurons and specific types of midbrain neurons. These finding are consistent with epidemiological and genetic evidence of a general factor underlying diverse clinical psychiatric disorders [15,46,47]. 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.

A number of studies have argued that the immune system plays a causal role in psychiatric disorders in at least some cases [48,49]. Our results did not implicate any brain immune cell types in psychiatric disorders. We interpret this negative finding cautiously as we did not fully sample the immune system. It also remains possible that a minority of loci are active in immune cell types and that these cell types play an important role in the etiology of psychiatric disorders. In addition, if immune functions are salient for a small subset of patients, GWAS may not identify loci implicated in immune functions until much larger sample sizes are available.

Our findings for neurological disorders were strikingly different from psychiatric disorders. While a previous study did not identify any cell type associations with Parkinson's disease [50], we found that dopaminergic neurons and oligodendrocytes were consistently and significantly associated with the disease. It is well known that dopaminergic neurons in the substantia nigra are lost in individuals affected by Parkinson's disease. Our study suggests that the dopaminergic neuron loss

characteristic of Parkinson's disease is at least partly due to biological mechanisms intrinsic to dopaminergic neurons. In addition, we observed that other cell types that degenerate in Parkinson's disease (including raphe nucleus serotonergic neurons and cholinergic neurons of the Pons) were significantly associated with the disorder, suggesting that specific pathological mechanisms could be shared across these neuronal cell types and lead to their degeneration. The best characterized trait shared between the vulnerable cell types is that they all have long, highly branched, unmyelinated and relatively thin axons [54]. Two theories for the selective vulnerability of neuronal populations in Parkinson's disease currently exist: 1) the 'spread Lewy pathology' model which assumes cell-to-cell contacts enabling spreading of prion-like α-synuclein aggregates [51]; and 2) the 'threshold theory' [52,53], which proposes that the vulnerable cell types degenerate due to molecular/functional biological similarities in a cell-autonomous fashion. While both theories are compatible and both mechanisms can co-exist, our findings support the existence of cell autonomous mechanisms contributing to selective vulnerability. We caution though that we have not tested all cell types which are known to degenerate (for instance, noradrenergic cells of the locus coeruleus were not tested), nor do we know if all the cell types we found to be associated show degeneration or functional impairment. However, analysis of the cellular mechanisms in the three main cell types associated to Parkinson's disease (dopaminergic neurons, serotonergic neurons and oligodendrocytes) revealed endosomal and lysosomal trafficking as a plausible common pathogenic mechanism.

The strong association of oligodendrocytes with Parkinson's disease was an unexpected finding. A possible explanation is that the association of oligodendrocytes with Parkinson's disease could be due to a related disorder (e.g., multiple system atrophy, characterized by parkinsonism and accumulation of α-synuclein in glial cytoplasmic inclusions [55]). However, this explanation is unlikely as multiple system atrophy is a very rare disorder. Hence, only few patients could have been included in the GWAS of Parkinson's disease, which could not have affected the GWAS results. 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 Parkinson's disease [13]. In addition, a previous study found that the diagnosis of 100% of a subset of 23andMe Parkinson's disease self-reported cases (50/50 participants) could be confirmed by a neurologist [56].

We did not find an association of oligodendrocytes with Parkinsonism for genes affected by rare variants. However, we believe that this result likely reflects the low power of the analysis (insufficient number of genes). On the other hand, prior evidence has suggested an involvement of oligodendrocytes in Parkinson's disease. For example, α-synuclein-containing inclusions have been reported in oligodendrocytes in Parkinson's disease brains [57]. These inclusions, defined as "coiled bodies", are typically found throughout the brainstem nuclei and fiber tracts [58]. Even though 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 identify intrinsic genetic alterations in oligodendrocytes at an early stage of

disease, which precedes the emergence of neurodegeneration in the substantia nigra, arguing for a key role of this cell type in Parkinson's disease etiology.

Altogether, we integrated genetics and single cell gene expression data from the entire nervous system to systematically identify cell types underlying brain complex traits. We believe that this a critical step in the understanding of the etiology of brain disorders and that these results will guide modelling of brain disorders and functional genomic studies.

Methods:

GWAS data

We collected GWAS summary statistics for 19 brain complex traits along with summary statistics for 9 non-brain traits with large sample sizes: schizophrenia [2], educational attainment [3], intelligence [59], body mass index [5], bipolar disorder [60], neuroticism [4], major depressive disorder [61], age at menarche [62], autism [63], migraine [64], amyotrophic lateral sclerosis [65], ADHD [66], Alzheimer's disease [67], epilepsy [68], age at menopause [69], coronary artery disease [70], height [5], hemoglobin A1c [71], hippocampal volume [72], inflammatory bowel disease [73], intracranial volume [74], stroke [75], total cholesterol [76], type 2 diabetes mellitus [77], type 2 diabetes adjusted for BMI [77], and waist-hip ratio adjusted for BMI [78].

We performed an inverse-variance weighted meta-analysis [79] of Parkinson's disease using summary statistics from Nalls et al. [13] (9,581 cases, 33,245 controls) and summary statistics from 23andMe (12,657 cases, 941,588 controls). Summary statistics from both studies were highly genetically correlated (LDSC [14] r_g= 0.87, s.e = 0.0679) and had little evidence of sample overlap (LDSC [14] bivariate intercept = 0.0288, s.e = 0.0066).

Anorexia nervosa (AN) GWAS summary statistics are from the second freeze of the Eating Disorders Working Group of the Psychiatric Genomics Consortium (PGC-ED) (Watson et al., under review). Samples were combined from PGC-ED Freeze 1 [80], the Anorexia Nervosa Genetics Initiative (ANGI) [81,82], the Genetic Consortium for Anorexia Nervosa (GCAN)/Wellcome Trust Case Control Consortium 3 (WTCCC-3) [83], and the UK Biobank [84]. Case definitions established a lifetime diagnosis of AN via hospital or register records, structured clinical interviews, or online questionnaires based on standardized criteria (DSM-III-R, DSM-IV, ICD-8, ICD-9, or ICD-10), whereas in the UK Biobank cases self-reported a diagnosis of AN. Controls were carefully matched for ancestry, and some, but not all control cohorts were screened for lifetime eating and/or some or all psychiatric disorders. Given the relative rarity of AN, large unscreened control cohorts were deemed appropriate for inclusion [85]. The meta-analysis comprised 16,992 AN cases and 55,525 controls and identified eight independent genome-wide significant loci. Typical of complex trait GWAS, we observed test statistic inflation (λ = 1.22) consistent with polygenicity, with no evidence of significant population stratification according to the LD intercept and attenuation ratio.

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 500K single cells from the entire mouse nervous system (19 regions) and identified 265 cell types [26] (note that after filtering on unique molecular identifier (UMI) counts, described below, only 231 of these cell types are used). The replication datasets were: 1) a mouse study that sampled 690K single cells from 9 brain regions and identified 565 cell types[86] (note that after filtering on UMI counts, described below, only 414 of these cell types are used); 2) our prior mouse study of ~10K 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 [10]; 3) a study that sampled 19,550 nuclei from frozen adult human post-mortem hippocampus and prefrontal cortex and identified 16 cell types [39].

We obtained bulk tissues RNA-seq gene expression data from 53 tissues from the GTEx consortium [18] (v7, 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 statistics was not provided by the authors). We used the pre-computed median expression across individuals for the GTEx dataset. 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 mouse and human (Mouse Genome Informatics, The Jackson laboratory, version 11/22/2016). In addition, we filtered out any cell type with less than 200'000 total UMIs. Gene expression was then scaled to a total of 1M UMIs (or transcript 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). Specificity measures were than transformed to a normal distribution so that each cell type had a comparable specificity distribution using the *rntransform* function from the GenABEL R package [87].

MAGMA primary and conditional analyses

MAGMA (v1.06b) [19] is a program 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-value. MAGMA can then be used to test whether a gene set (binary variable) or a continuous variable (e.g. specificity measures of all genes) are predictors of the gene-level association statistic of the trait (Z-value) 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 statistics, we excluded any SNPs with INFO score <0.6 or with MAF < 1%. We set a window of 35kb upstream to 10kb downstream of the gene coordinates to compute gene-level association statistics and used the European reference panel from the phase 3 of the 1000 genomes project [88] as the reference population. For each trait, we then used MAGMA to test whether the standard normalized gene expression specificity per cell type was associated with gene-level genetic association with the trait. We performed a one-sided test as we were only interested in finding whether an increase in gene expression specificity was associated with an increase in gene-level genetic association with the trait. The significance threshold was set to 0.005 divided by the number of tissues/cell types for each dataset (53 for GTEx, 231 for the mouse nervous system, 414 for the mouse study from 9 different brain regions, 24 for the mouse study from 5 different brain regions + specific enrichments and 15 for the human study).

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-value from MAGMA .out file) or to look for associations of cell types conditioning on gene expression specificity from other cell types by adding these variables as covariate in the model.

To test whether MAGMA was well-calibrated, we randomly permuted the gene labels of the schizophrenia gene-level association statistic file (.raw) a thousand times. We then looked for association between cell type specific gene expression and the randomized gene-level schizophrenia association statistics. We observed that MAGMA was slightly conservative with less than 50 random sampling having a P-value < 0.05 **(Figure S22)** for all cell types.

LD score regression analysis

We used partitioned LD score regression [89] to test whether the top 10% most specific genes of each cell type (based on our specificity metric described above) were enriched in heritability for the diverse traits. In order to capture most regulatory elements that could contribute to the effect of the region on the trait, we extended the gene coordinates by 100kb upstream and by 100kb downstream of each gene as previously [11]. SNPs located in 100kb regions surrounding the top 10% most specific genes in each cell type were added to the baseline model (consisting of 53 different annotations) independently for each cell type (one file for each cell type). We then selected the heritability enrichment p-value as a measure of the association of the cell type with the traits. The significance threshold was set to 0.005 divided by the number of tissues/cell types for each dataset (53 for GTEx, 231 for the mouse nervous system, 414 for the mouse study from 9 different brain regions, 24 for the mouse study from 5 different brain regions and 15 for the human study). 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.

MAGMA vs LDSC ranking

In order to test whether the cell type ranking obtained using MAGMA and LDSC in the Zeisel et al. dataset [26] were similar, we computed the Spearman rank correlation of the cell types 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 P-value distribution from the expected) (Spearman $\rho = 0.92$) (**Figure S23**) and with the average number of cell types below our stringent significance threshold using LDSC and MAGMA (Spearman $\rho = 0.83$), indicating that the overall ranking of the cell types is very similar between the two methods, provided that the GWAS is well powered (**Figure S24**).

GO term enrichments

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 [90]. 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 **Table S9**). We used the EWCE R package [9] (available from github.com/NathanSkene/EWCE) to test for cell type enrichment using the same specificity matrix used for the LDSC and MAGMA analysis. Ten thousand bootstrapping replicates were used.

Parkinson's disease post-mortem transcriptomes

The Moran dataset [41] was obtained from GEO (accession GSE8397). Processing of the U133a and U133b Cel files was done separately. The data was read in using the ReadAffy function from the R affy package [91], then Robust Multi-array Averaging (RMA) was applied. The U133a and U133b array expression data were merged after applying RMA. Probe annotations and mapping to HGNC symbols was done using the biomaRt R package [92]. Differential expression analysis was performed using limma [93] taking age and gender as covariates. The Lesnick dataset [40] was obtained from GEO (accession GSE7621). Data was processed as for the Moran dataset: however, age was not available to use as a covariate. The Disjkstra dataset [43] was obtained from GEO (accession GSE49036) and processed as above: the gender and RIN values were used as covariates.

As the transcriptome datasets measured gene expression in the substantia nigra, we only kept cell types that are present in the substantia nigra or ventral midbrain for our EWCE [9] analysis. We computed a new specificity matrix based on the substantia nigra or ventral midbrain cells using the EWCE [9] R package (available from

github.com/NathanSkene/EWCE). The EWCE [9] analysis was performed on the 500 most up or down regulated genes using 10,000 bootstrapping replicates.

Code availability

Code used in this study will be made available upon publication.

Data availability

All single-cell gene expression data are publicly available. All summary statistics used in this study are publicly available with the exception of the migraine GWAS, which can be obtained by contacting the authors [64]. The Parkinson's disease summary statistics from 23andMe will be made available to qualified researchers under an agreement with 23andMe that protects the privacy of 23andMe research participants. Please visit research.23andMe.com/collaborate/#publication for more information and to apply to access the data.

Acknowledgments

JB 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 program (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). PFS was supported by the Swedish Research Council (Vetenskapsrådet, award D0886501), the Horizon 2020 Program of the European Union (COSYN, RIA grant agreement n° 610307), and US NIMH (U01 MH109528 and R01 MH077139). KH was supported by The Michael J. Fox Foundation for Parkinson's Research (grant MJFF12737). CMB acknowledges funding from the Swedish Research Council (Vetenskapsrådet, award: 538-2013-8864) and the Klarman Family Foundation. We thank the research participants from 23andMe and other cohorts for their contribution to this study, and KH for comments on the manuscript.

Potential conflicts of interest

PFS 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).
CM Bulik reports: Shire (grant recipient, Scientific Advisory Board member); Pearson and Walker (author, royalty recipient).

Tables:

Table S1: Genetic correlation across traits
Table S2: Association P-value between GTEx and all traits
Table S3: Association P-value between mouse nervous system and all traits (Zeisel et al.)
Table S4: GO term enrichment of genes highly specific to cell type and diseases
Table S5: Univariate conditional analysis results using MAGMA
Table S6: Association P-value between 9 mouse brain regions and all traits (Saunders et al.)
Table S7: Association P-value between 5 mouse brain regions and all traits (Skene et al.)
Table S8: Association P-value between 2 human brain regions and all traits (Habib et al.)
Table S9: Rare and familial genetic mutations associated with Parkinsonism
Table S10: Cell type enrichment results using rare and familial genetic mutations associated with Parkinsonism

Figures:

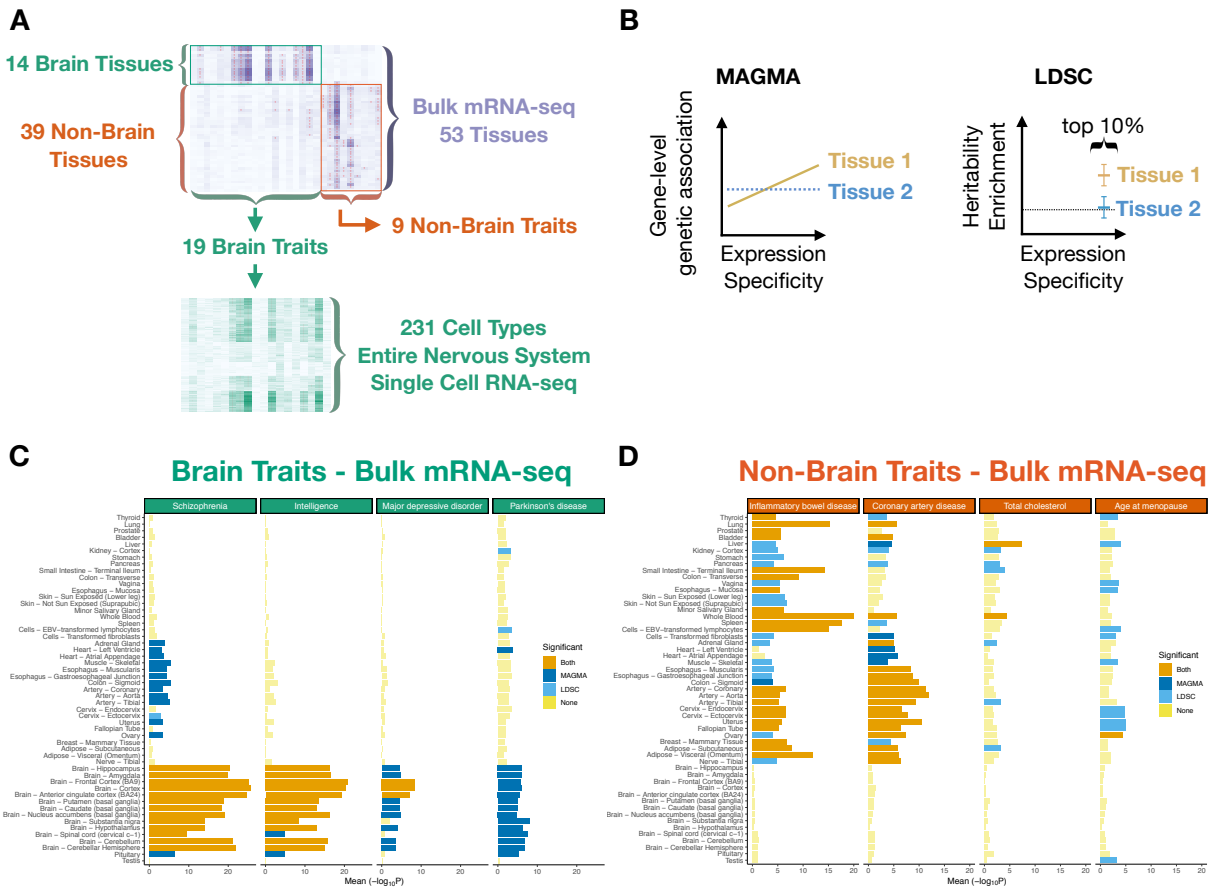

Figure 1: Study design and tissue-level associations.

Heatmap of trait – tissue/cell types associations ($-\log_{10}P$) for selected traits (**A**). Trait – tissue/cell types associations were performed using MAGMA (testing for a positive correlation between gene expression specificity and gene-level genetic associations) and LDSC (testing for heritability enrichment of the top 10% most specific genes in each tissue/cell type) (**B**). Tissue – trait associations for selected brain related traits (**C**). Tissue – trait associations for selected non-brain related traits (**D**). 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: $P = 0.005/53$). Traits are ordered based on hierarchical clustering of the gene expression specificities.

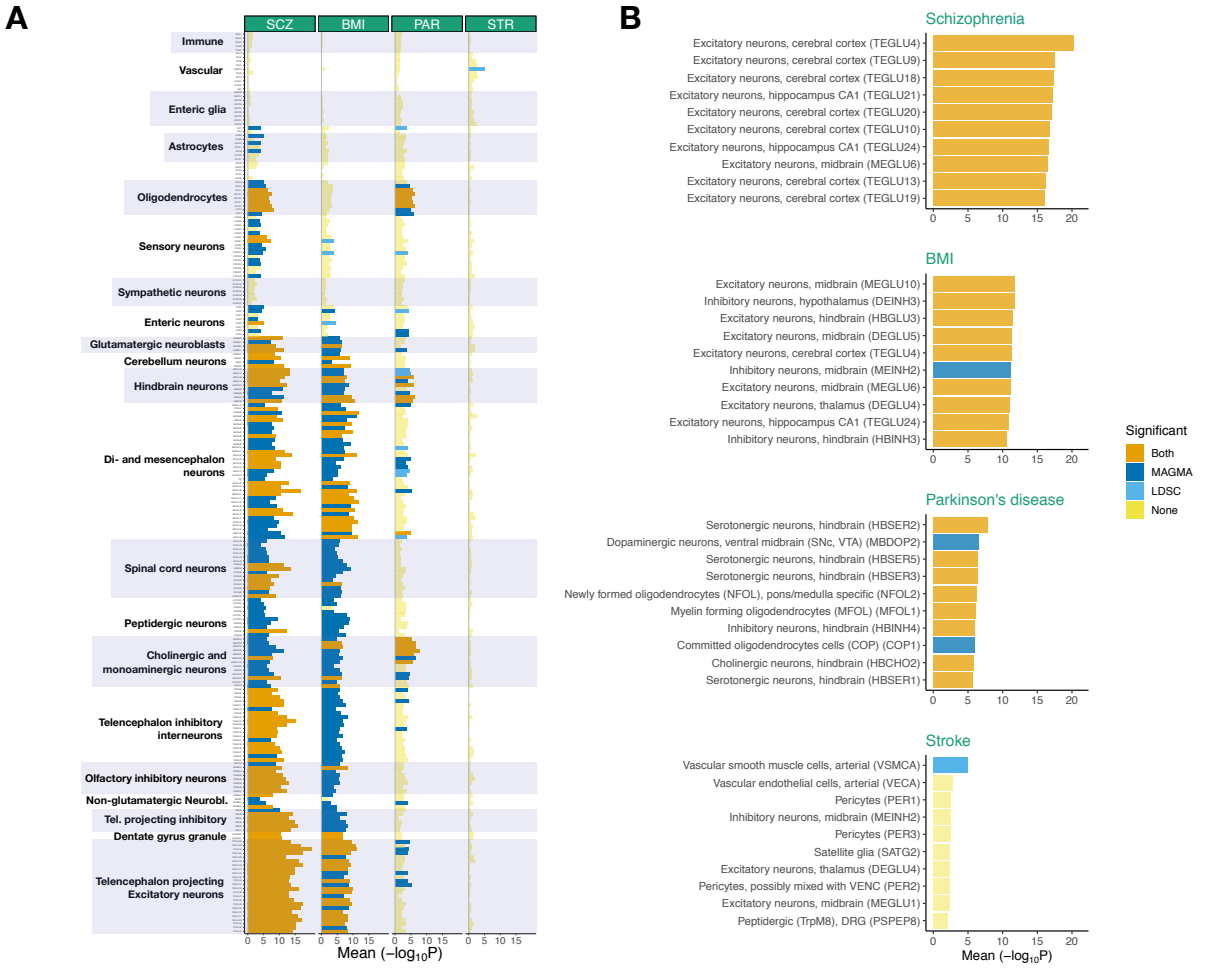

Figure 2: Association of brain related traits with cell types from the entire nervous system.

Associations between all cell types and the following traits are shown: schizophrenia (SCZ), body mass index (BMI), Parkinson's disease (PAR) and stroke (STR) (**A**).The ten most associated cell types are shown for these same traits (**B**). 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: P = 0.005/231).

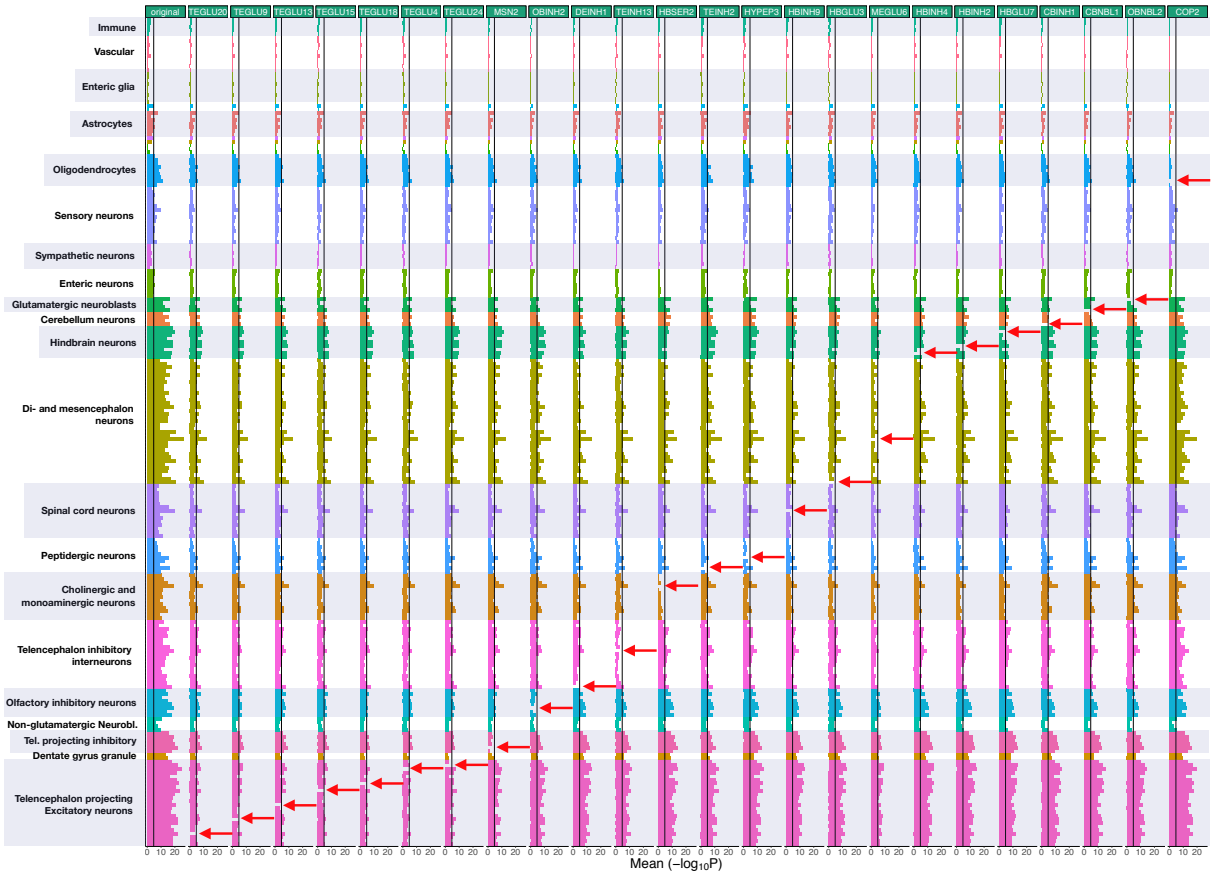

Figure 3: Conditional cell type associations for schizophrenia.

The MAGMA cell types association results are shown before (first column) and after conditioning on gene expression specificity from different cell types. The red arrow indicates the cell types that are being conditioned on. The black bar represents the significance threshold (P=0.005/231).

Figure 4: Replication of cell type – trait associations in 24 cell types from 5 different brain regions.

Associations for the top 10 cell types for all traits with at least one significant cell type with at least one method 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: $P = 0.005/24$).

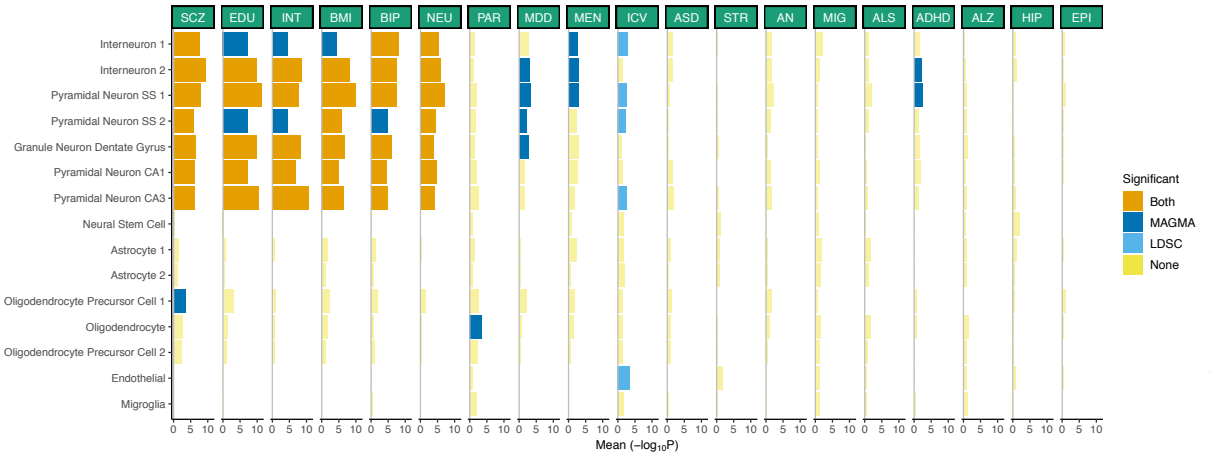

Figure 5: Human replication of cell type – trait associations in 15 cell types from 2 different brain regions.

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: P = 0.005/15).

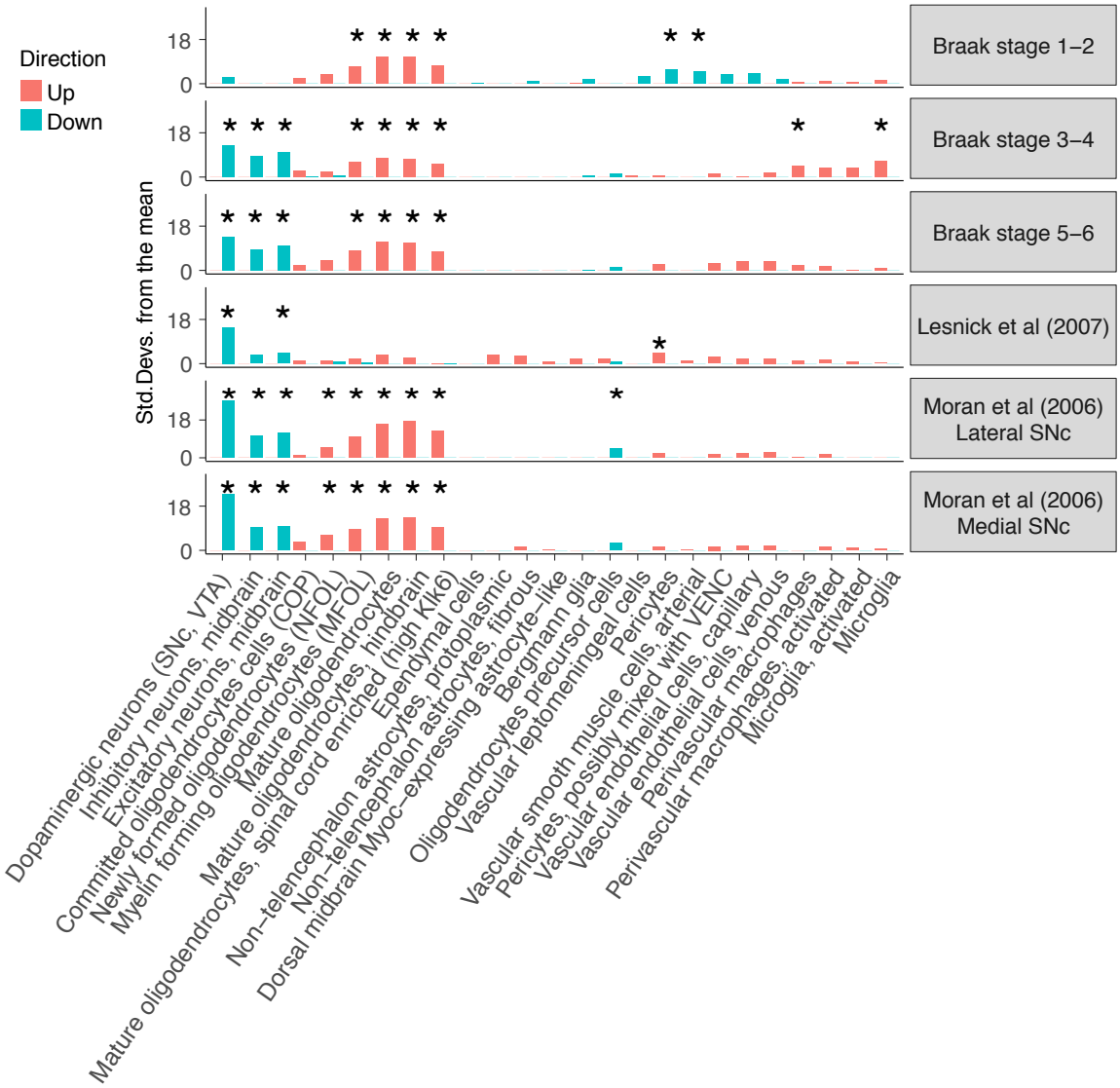

Figure 6: Enrichment of Parkinson differentially expressed genes in cell types from the substantia nigra

Enrichment of the 500 most up/down regulated genes (Braak stage 0 vs Braak stage 1—2, 3—4 and 5—6, as well as cases vs controls) in postmortem human substantia nigra gene expression samples. The enrichments were obtained using EWCE[9]. A star shows significant enrichments after multiple testing correction ($P<0.05/(25*4)$).

Supplementary figures:

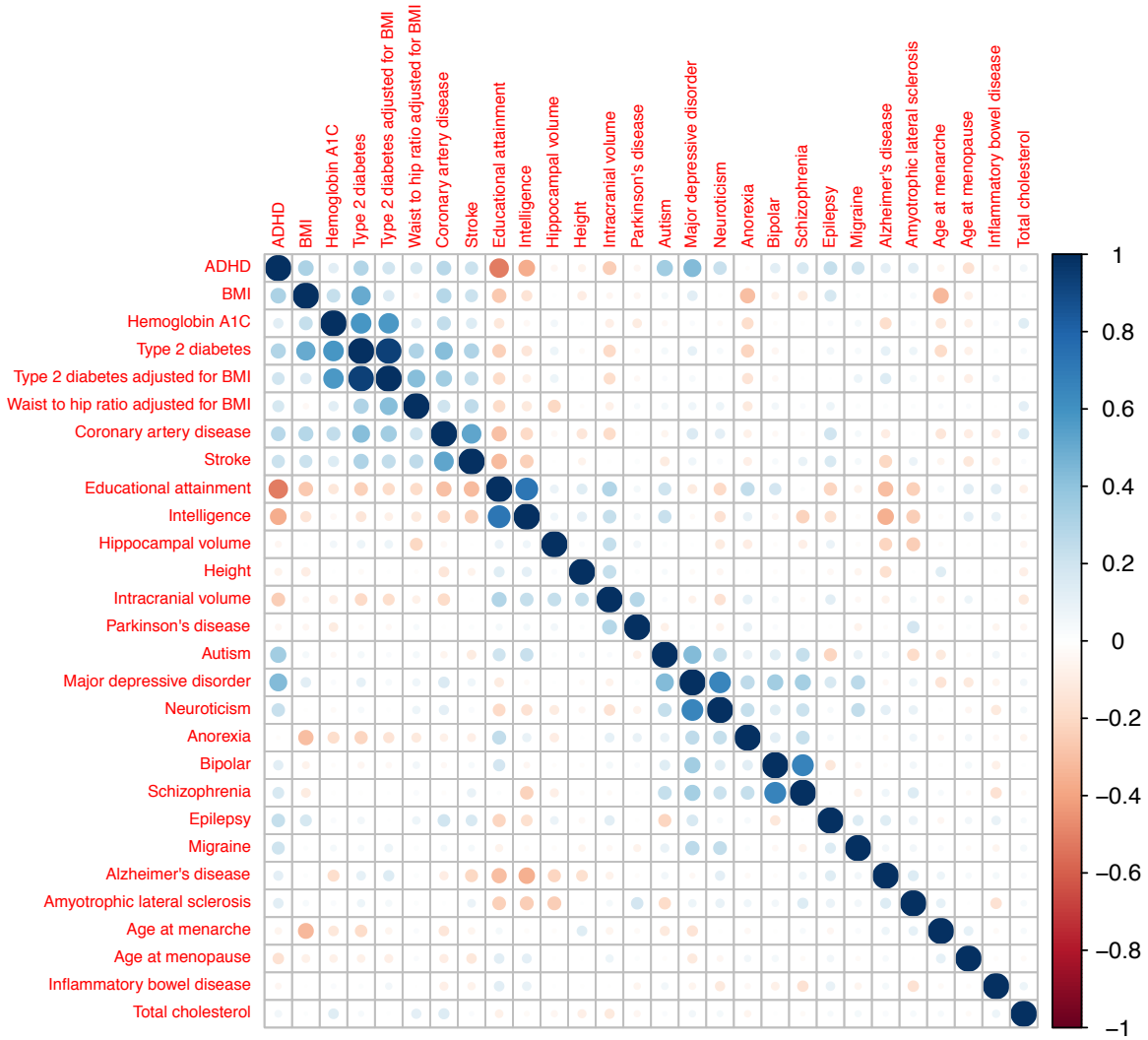

Figure S1: Genetic correlation across traits.

The genetic correlation across traits were computed using LDSC[94]. Traits are ordered based on hierarchical clustering.

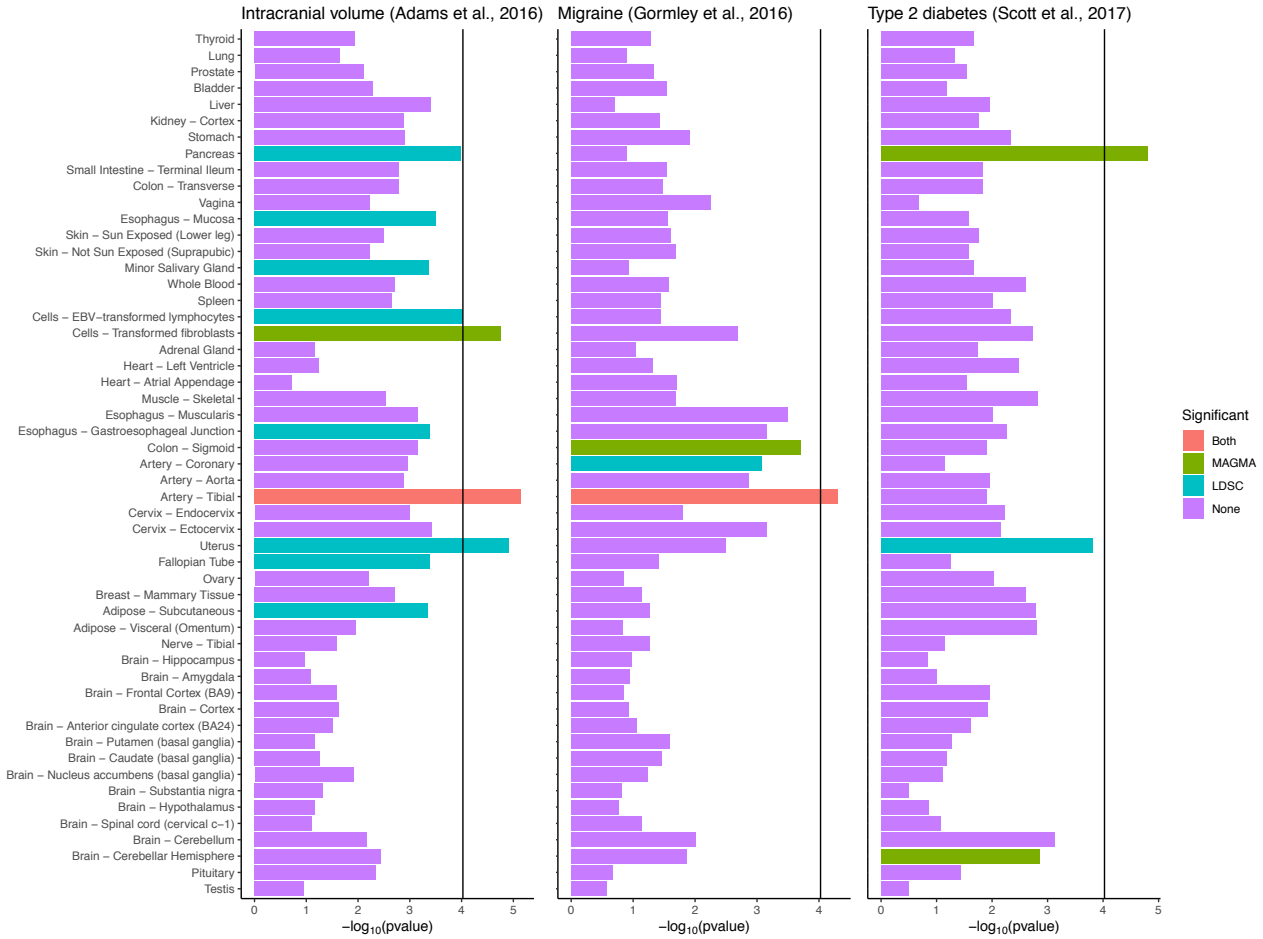

Figure S2: Tissue – trait associations for selected traits.

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: P = 0.005/53). Traits are ordered based on hierarchical clustering of the gene expression specificities.

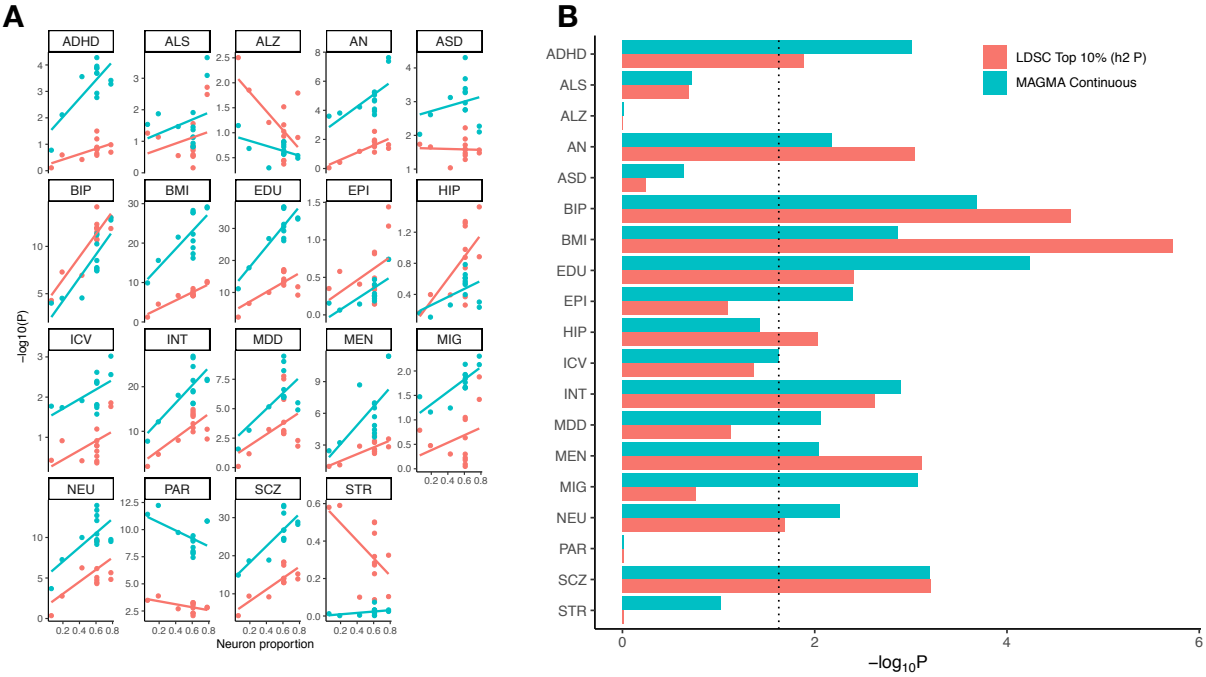

Figure S3: Correlation between the proportion of neurons in diverse brain regions and strength of tissue – trait associations.

Correlation between the proportion of neurons in diverse brain regions and strength of tissue – trait associations ($-\log_{10}P$) are shown for LDSC and MAGMA (**A**). The strength of associations ($-\log_{10}P$) are shown (**B**) with a dotted line representing the 5% false discovery rate threshold. 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), EPI (epilepsy).

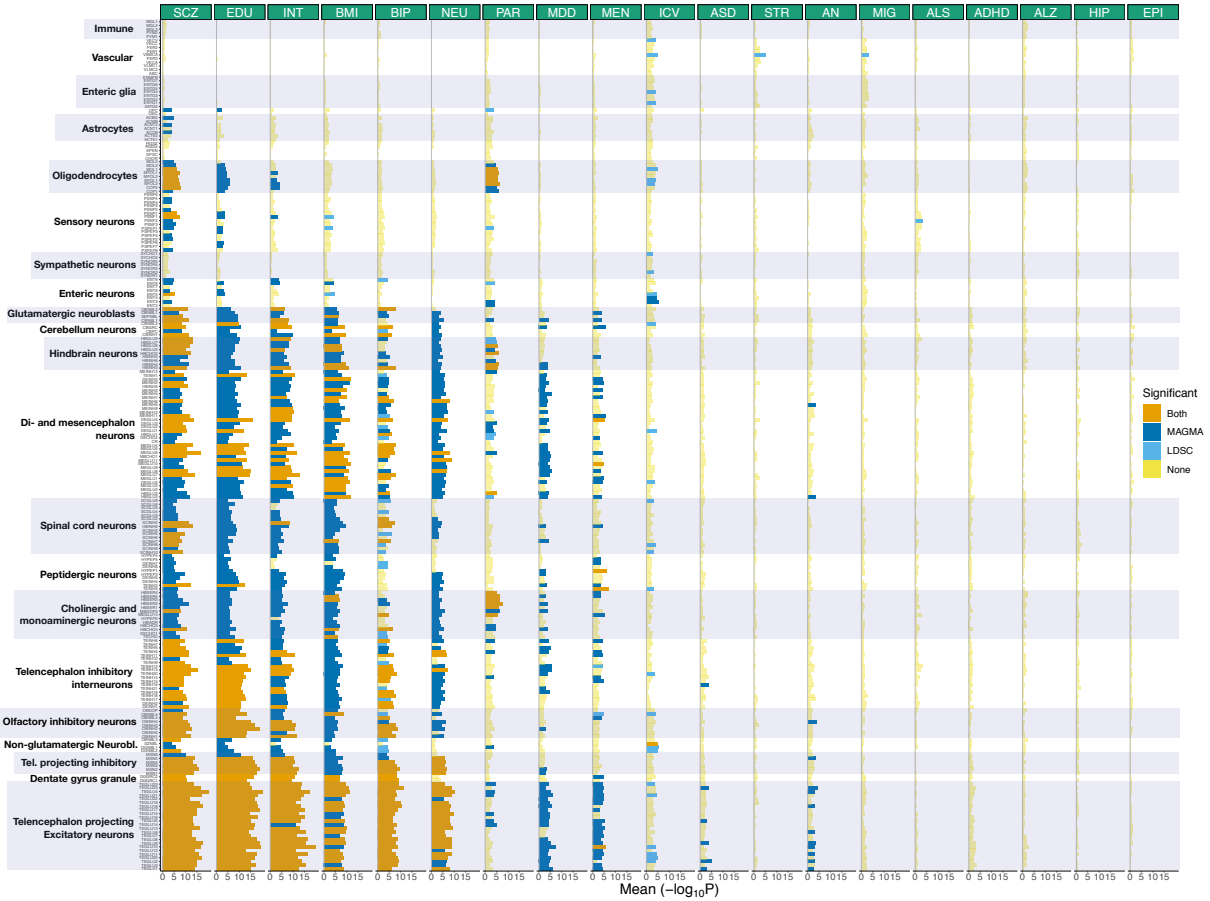

Figure S4: Association of brain related traits with cell types from the entire nervous system.

Association between cell types and the following 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), EPI (epilepsy). 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: P = 0.005/231).

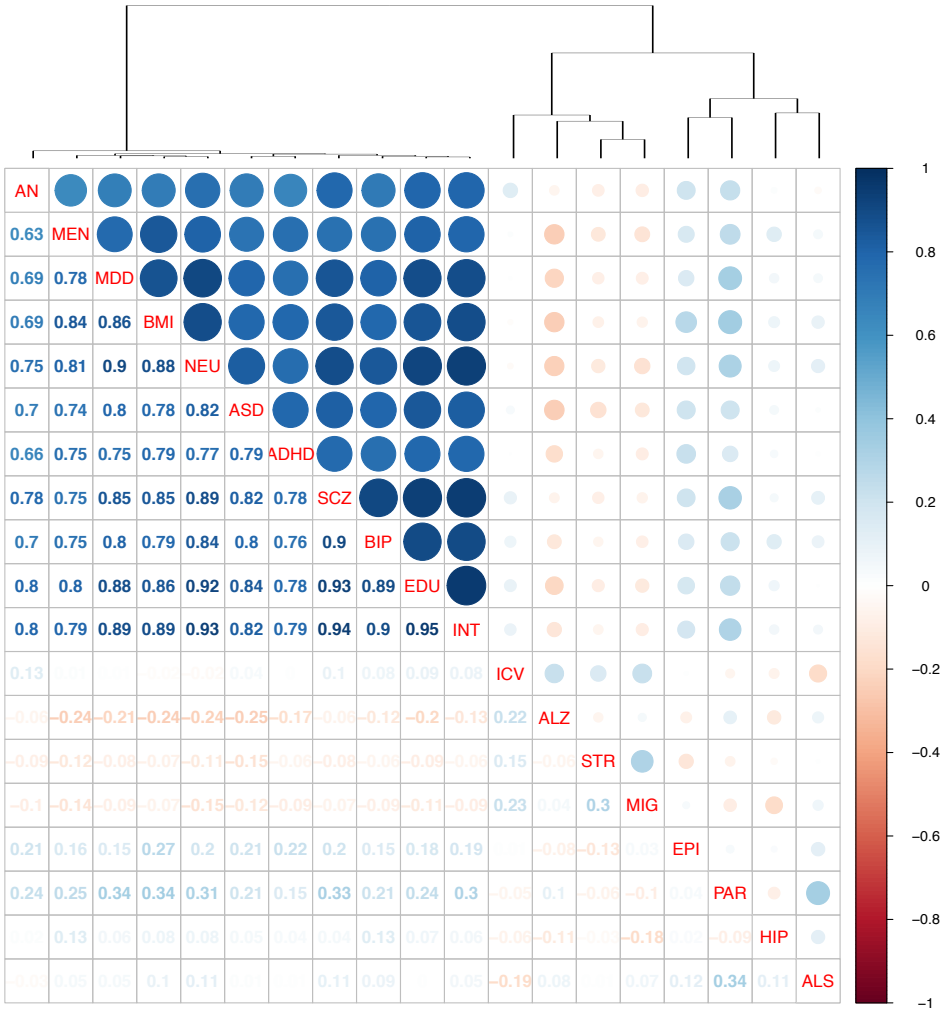

Figure S5: Correlation in cell type associations across traits.

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), EPI (epilepsy).

Figure S6: Association of cell types with schizophrenia/intelligence conditioning on gene-level genetic association of intelligence/schizophrenia.

MAGMA association strength for each cell types before and after conditioning on gene-level genetic association for another trait. The black bar represents the significance threshold (P=0.005/231). SCZ (schizophrenia), INT (intelligence).

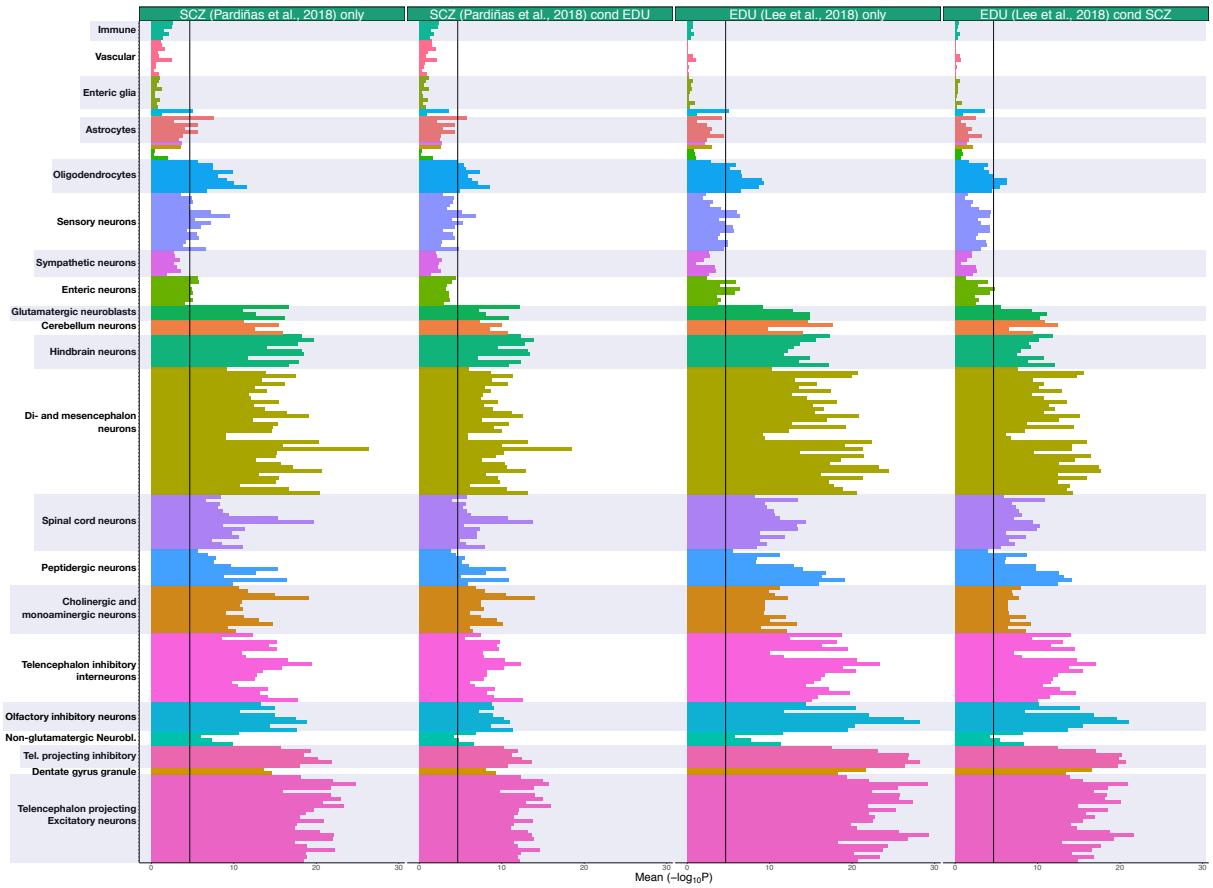

Figure S7: Association of cell types with schizophrenia/educational attainment conditioning on gene-level genetic association of educational attainment/schizophrenia.

MAGMA association strength for each cell types before and after conditioning on gene-level genetic association for another trait. The black bar represents the significance threshold (P=0.005/231). SCZ (schizophrenia), EDU (educational attainment).

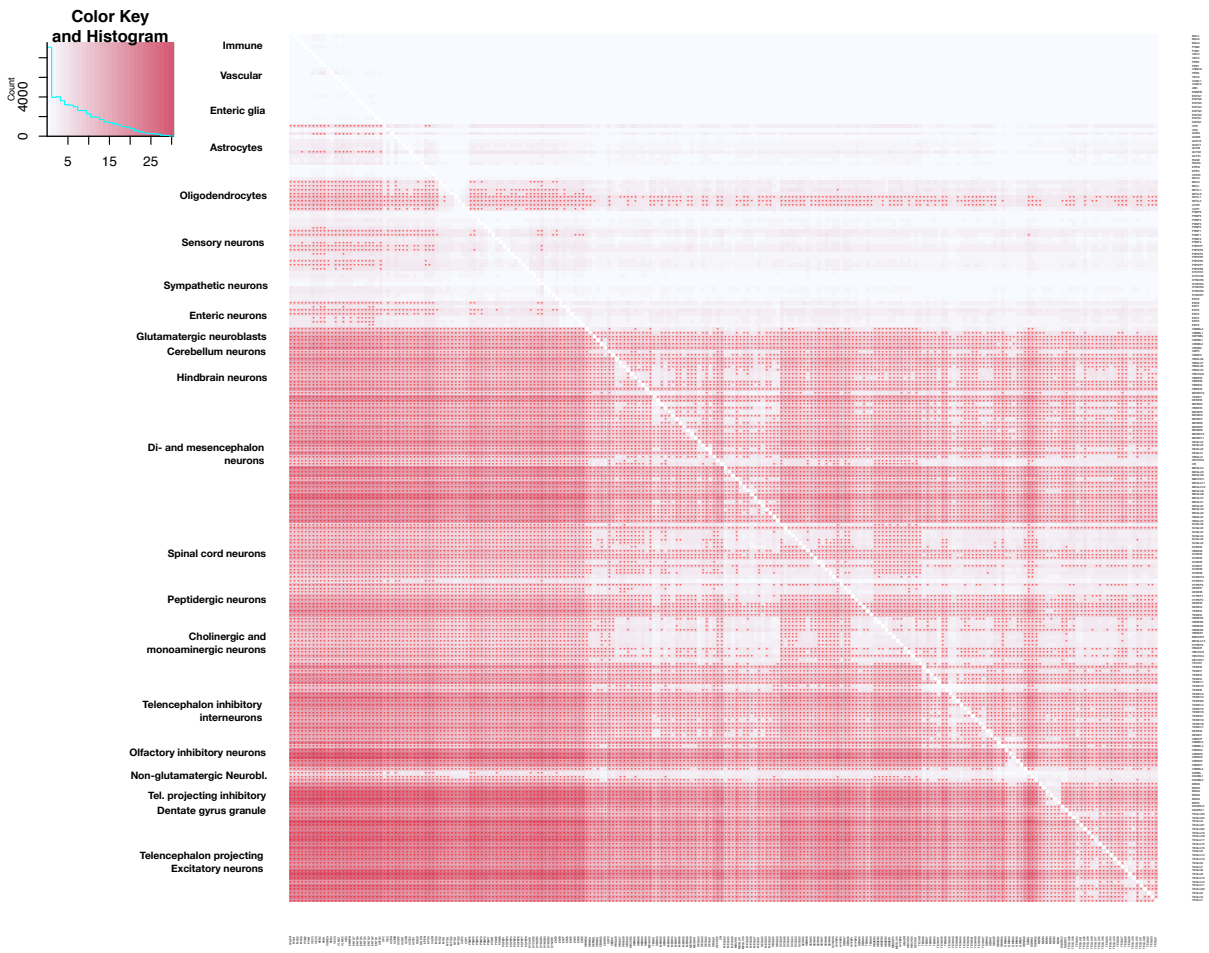

Figure S8: Univariate conditional analysis for educational attainment.

MAGMA association strength for each cell type before (1st column) and after conditioning on the 231 cell types from the nervous system. A red star is shown if the association is significant ($P<0.005/231$). The histogram indicates the scale ($-\log_{10}P$).

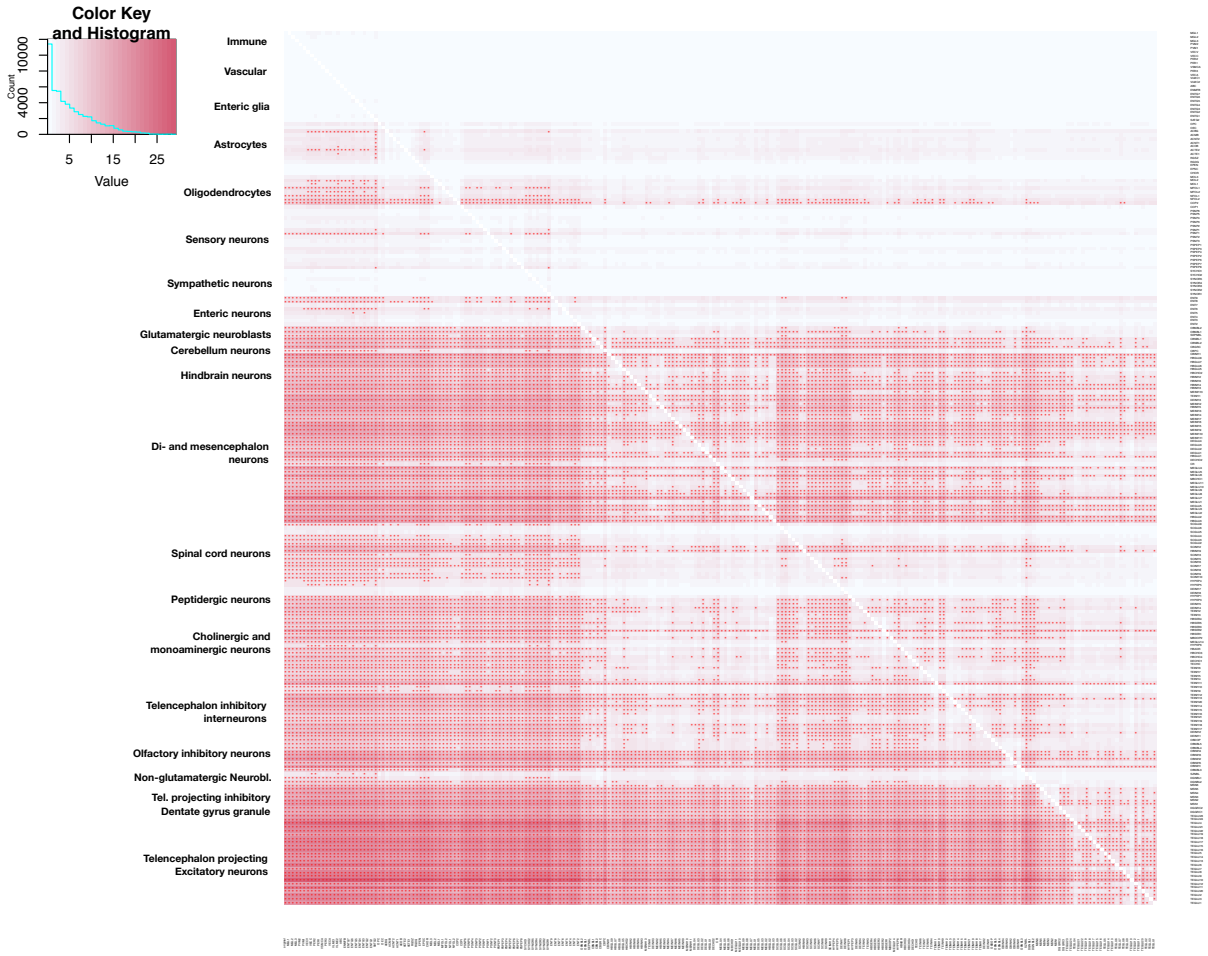

Figure S9: Univariate conditional analysis for intelligence.

MAGMA association strength for each cell type before (1st column) and after conditioning on the 231 cell types from the nervous system. A red star is shown if the association is significant ($P<0.005/231$). The histogram indicates the scale ($-\log_{10}P$).

Figure S10: Univariate conditional analysis for body mass index.

MAGMA association strength for each cell type before (1st column) and after conditioning on the 231 cell types from the nervous system. A red star is shown if the association is significant ($P<0.005/231$). The histogram indicates the scale ($-\log_{10}P$).

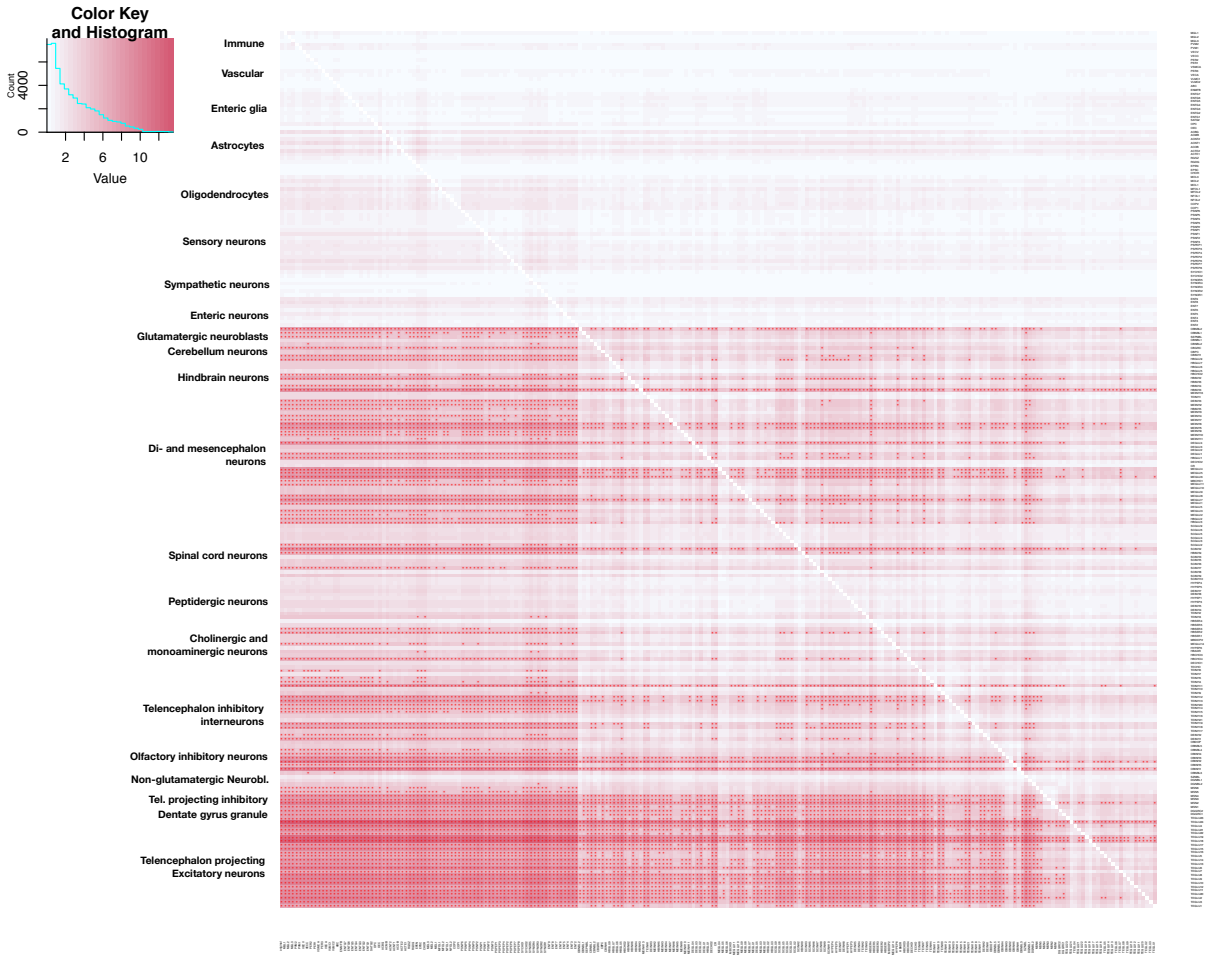

Figure S11: Univariate conditional analysis for bipolar disorder.

MAGMA association strength for each cell type before (1st column) and after conditioning on the 231 cell types from the nervous system. A red star is shown if the association is significant ($P<0.005/231$). The histogram indicates the scale ($-\log_{10}P$).

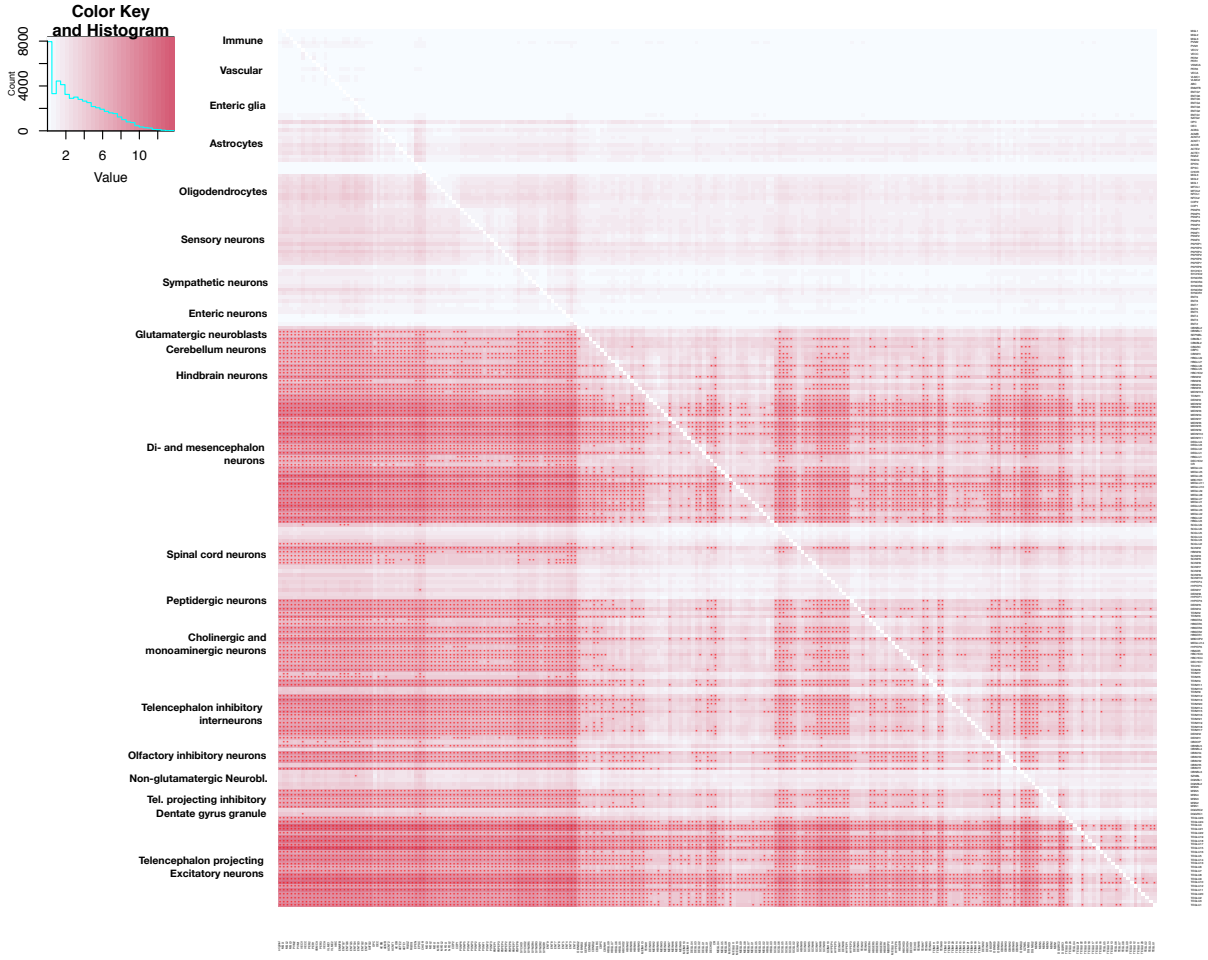

Figure S12: Univariate conditional analysis for neuroticism.

MAGMA association strength for each cell type before (1st column) and after conditioning on the 231 cell types from the nervous system. A dot is shown if the association is significant ($P<0.005/231$). The histogram indicates the scale ($-\log_{10}P$).

Figure S13: Univariate conditional analysis for major depressive disorder.

MAGMA association strength for each cell type before (1st column) and after conditioning on the 231 cell types from the nervous system. A red star is shown if the association is significant (P<0.005/231). The histogram indicates the scale ($-\log_{10}P$).

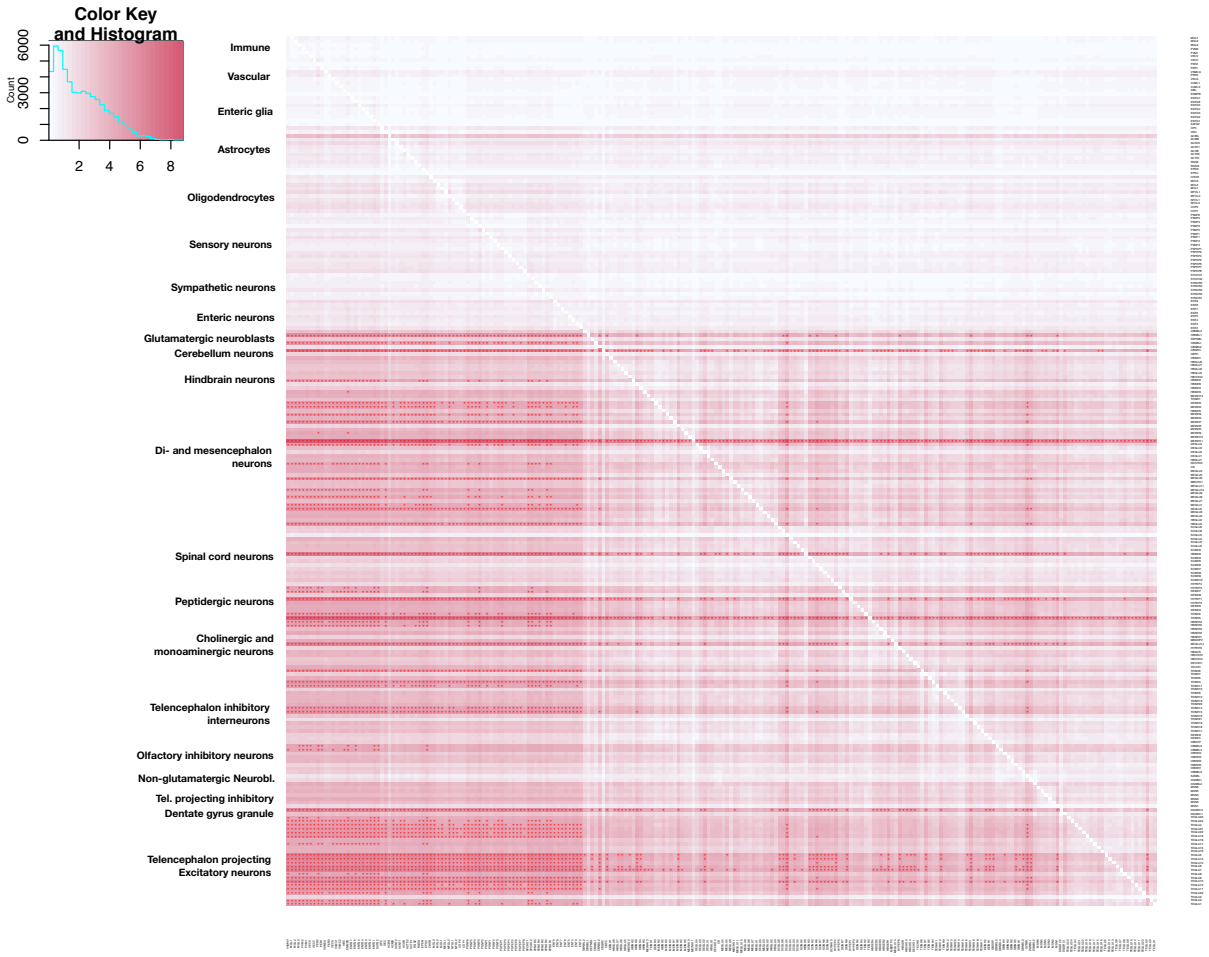

Figure S14: Univariate conditional analysis for age at menarche.

MAGMA association strength for each cell type before (1st column) and after conditioning on the 231 cell types from the nervous system. A red star is shown if the association is significant ($P<0.005/231$). The histogram indicates the scale ($-\log_{10}P$).

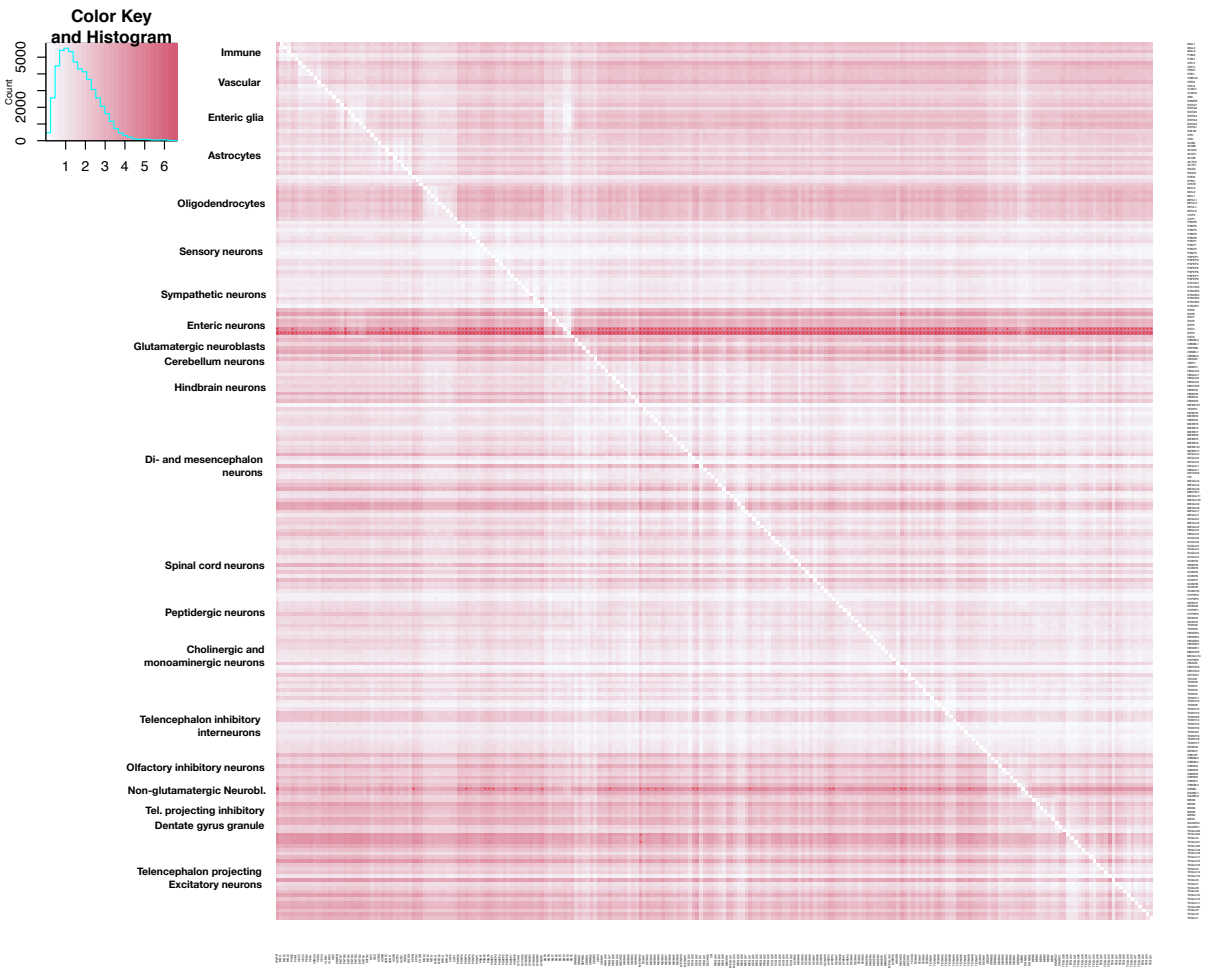

Figure S15: Univariate conditional analysis for intracranial volume.

MAGMA association strength for each cell type before (1st column) and after conditioning on the 231 cell types from the nervous system. A red star is shown if the association is significant ($P<0.005/231$). The histogram indicates the scale ($-\log_{10}P$).

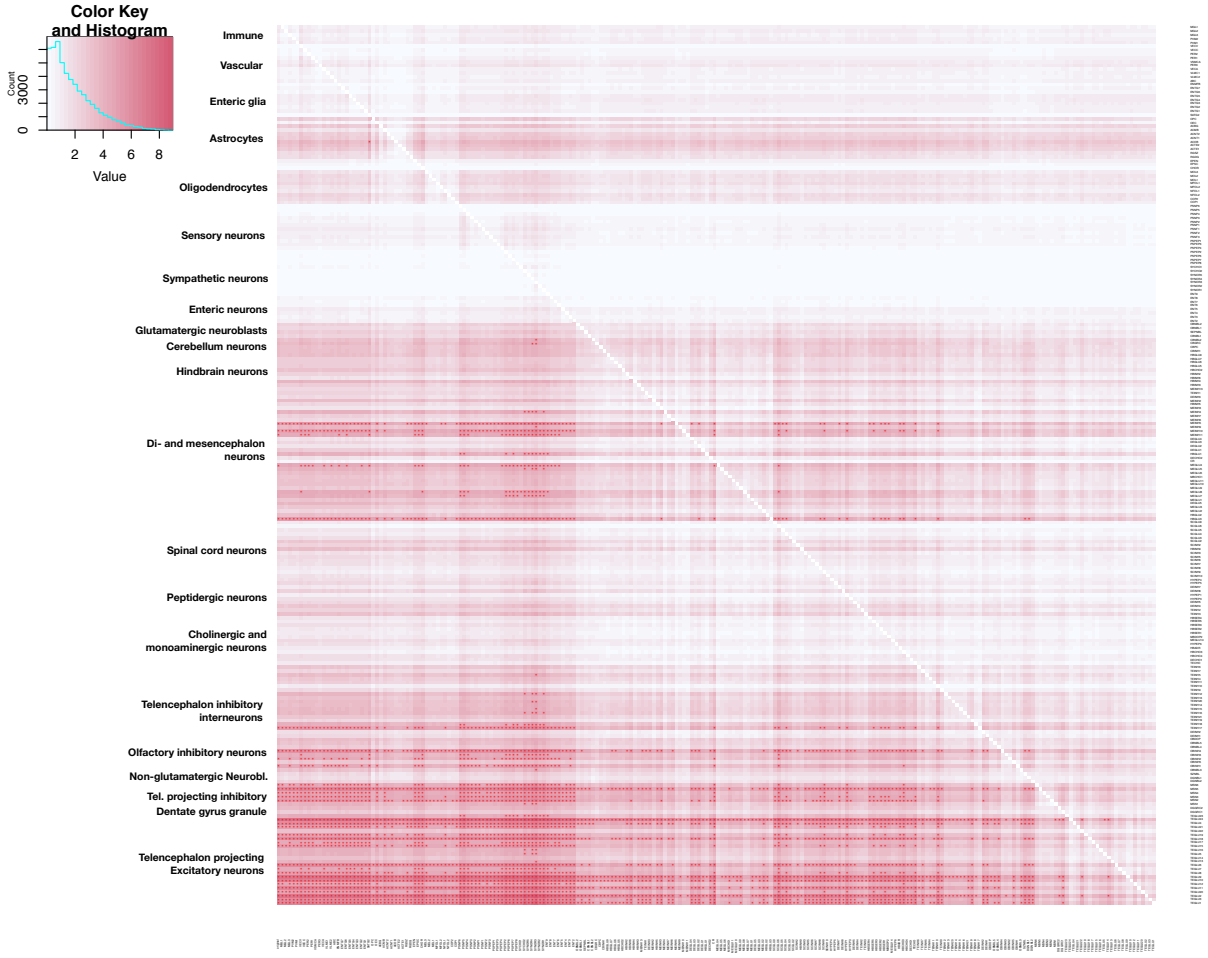

Figure S16: Univariate conditional analysis for anorexia nervosa.

MAGMA association strength for each cell type before (1st column) and after conditioning on the 231 cell types from the nervous system. A red star is shown if the association is significant ($P<0.005/231$). The histogram indicates the scale ($-\log_{10}P$).

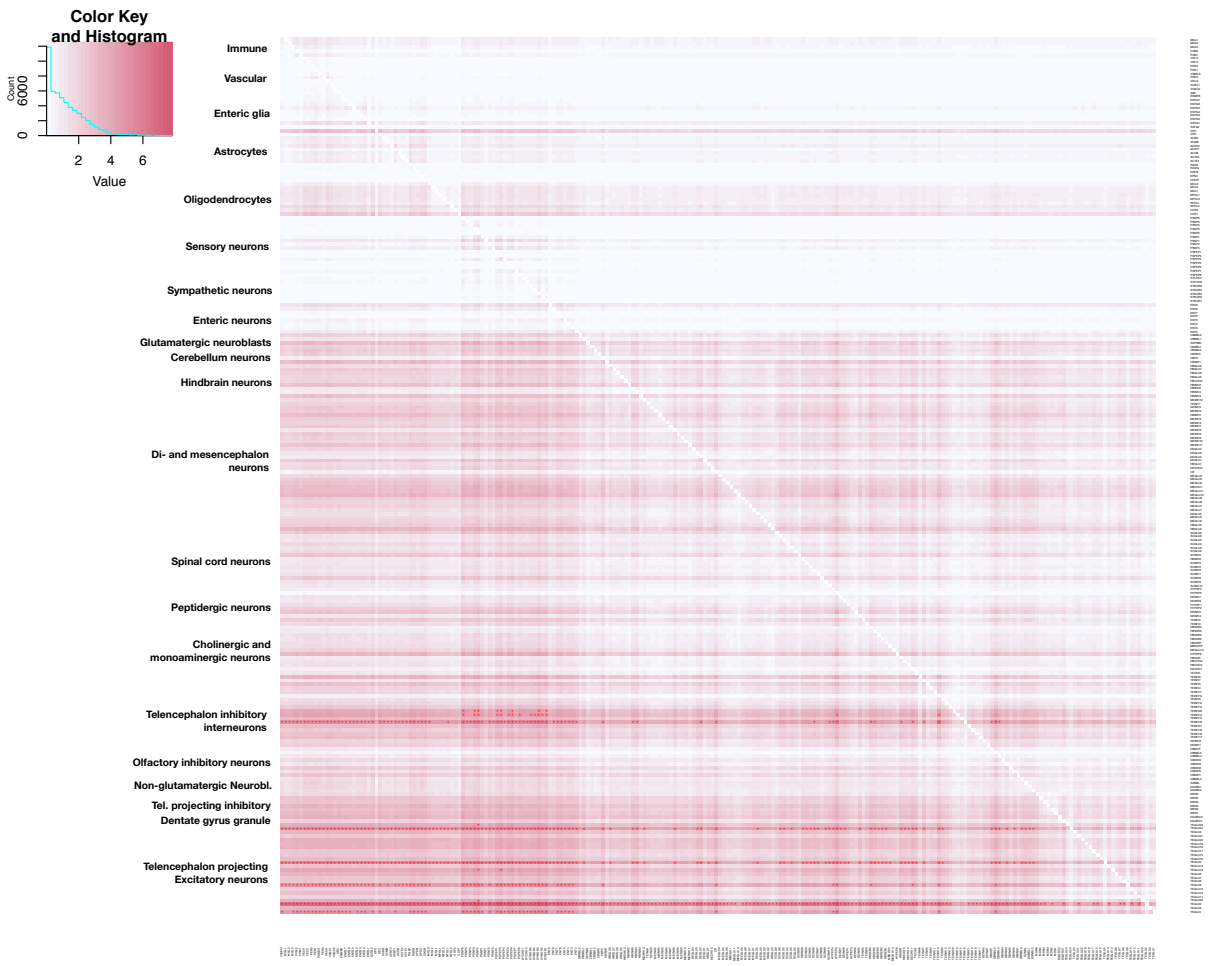

Figure S17: Univariate conditional analysis for autism.

MAGMA association strength for each cell type before (1st column) and after conditioning on the 231 cell types from the nervous system. A red star is shown if the association is significant ($P<0.005/231$). The histogram indicates the scale ($-\log_{10}P$).

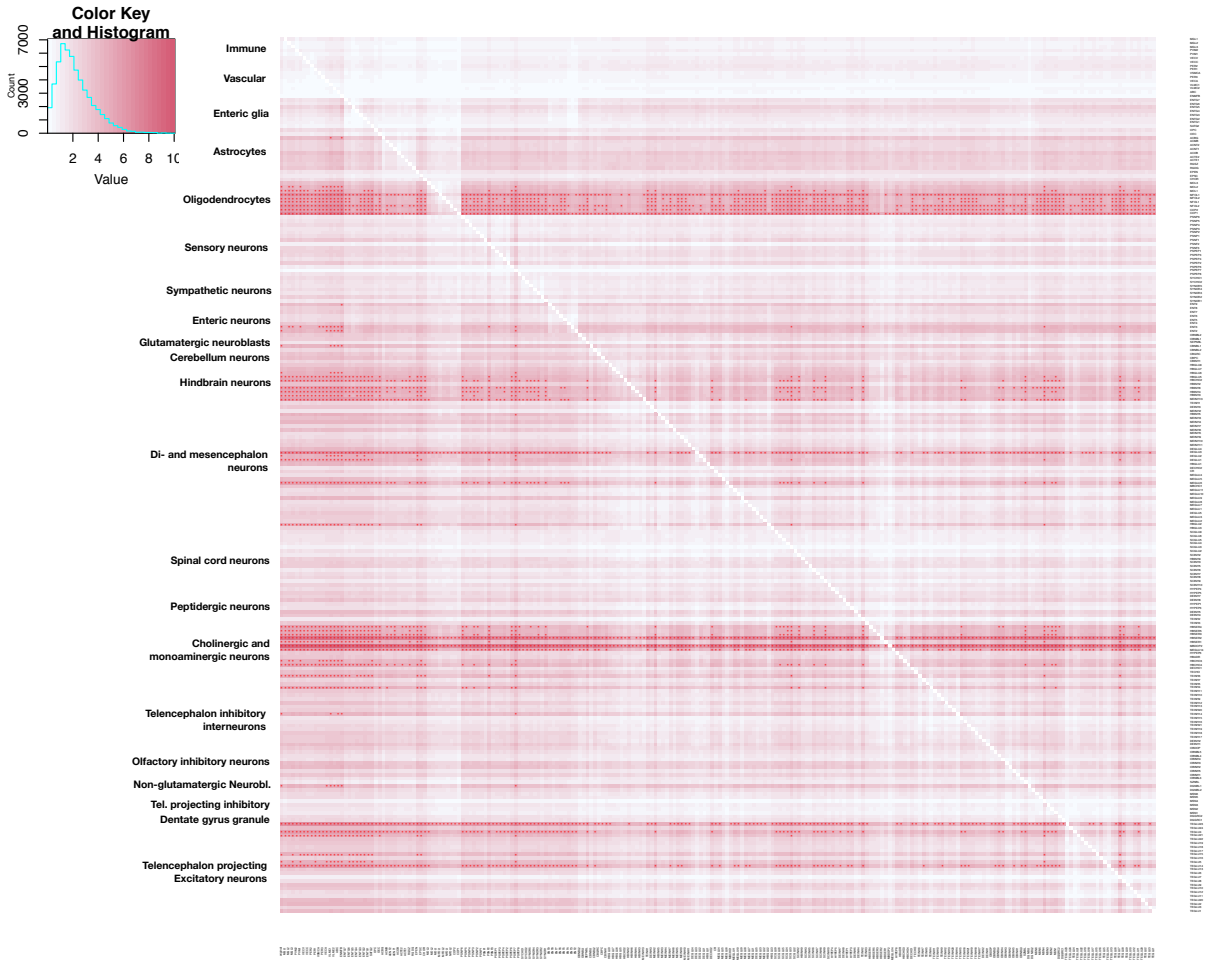

Figure S18: Univariate conditional analysis for Parkinson's disease.

MAGMA association strength for each cell type before (1st column) and after conditioning on the 231 cell types from the nervous system. A red star is shown if the association is significant (P<0.005/231). The histogram indicates the scale ($-\log_{10}P$).

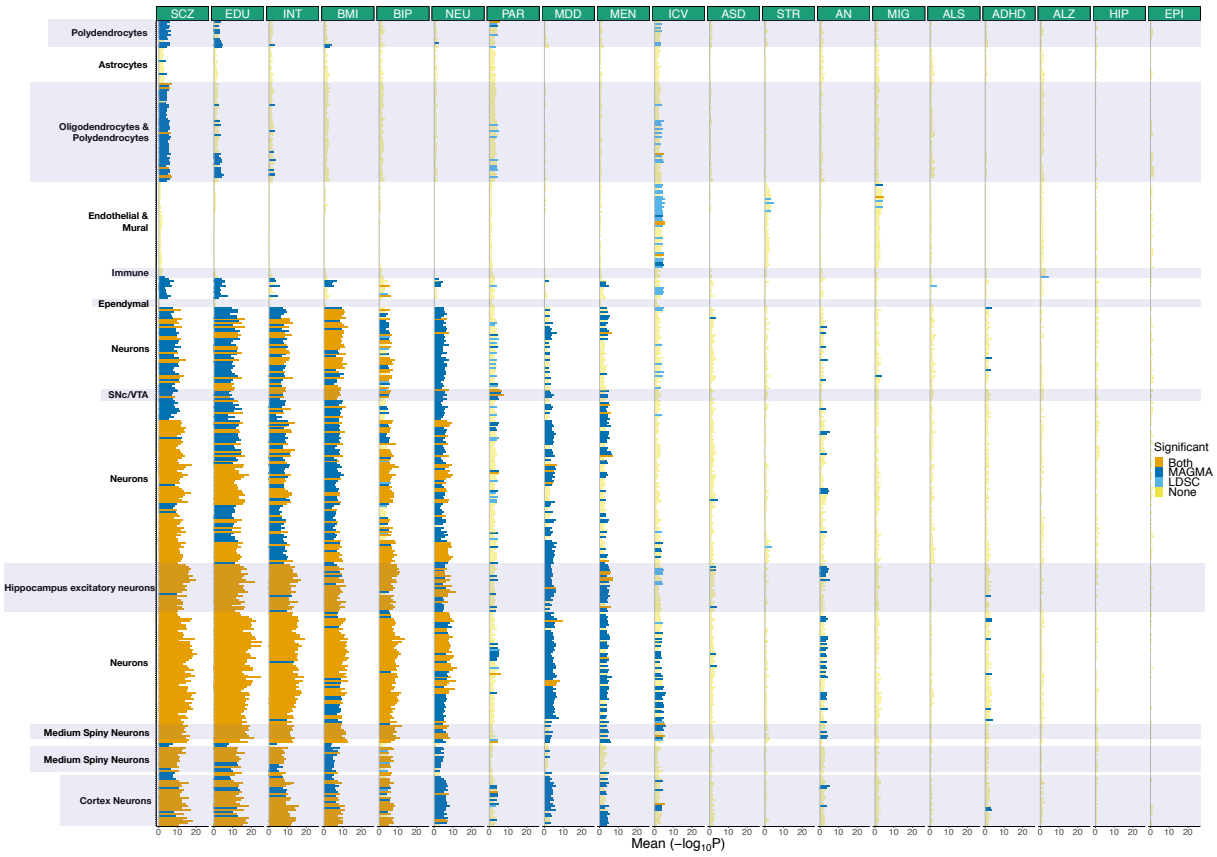

Figure S19: Replication of cell type – trait associations in 414 cell types from 9 different brain regions.

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: P = 0.005/414). 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), EPI (epilepsy).

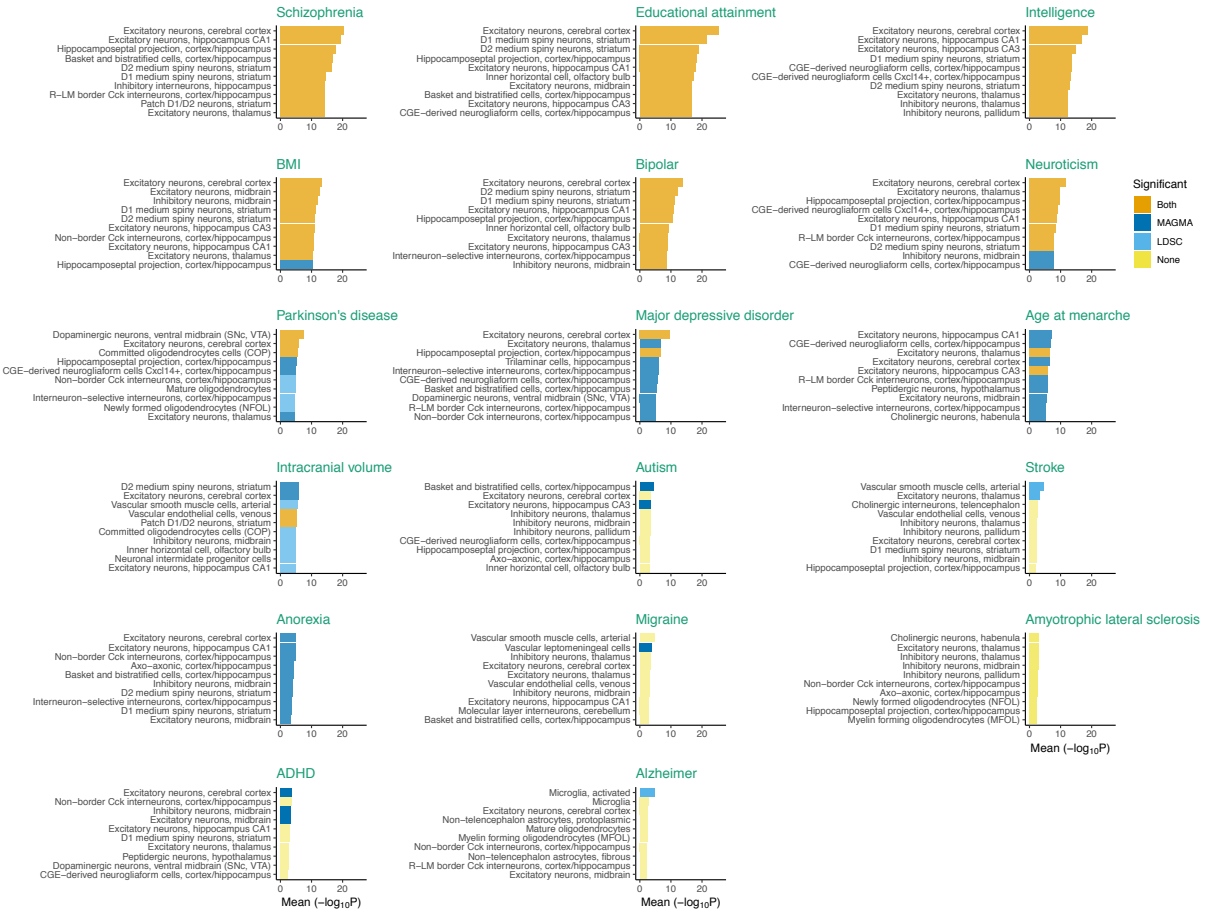

Figure S20: Top associated cell types with brain related traits among 414 cell types from 9 different brain regions.

The mean strength of association (-$\log_{10}$P) of MAGMA and LDSC is shown for the ten 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 = 0.005/414). If multiple cell types with the same label are strongly associated with a trait, only the first label is shown. The label shown correspond to the label of the most similar cell type in our main dataset (231 cell types from entire mouse nervous system). The most similar cell type is defined as the cell type with the highest Spearman rank correlation in gene expression specificity.

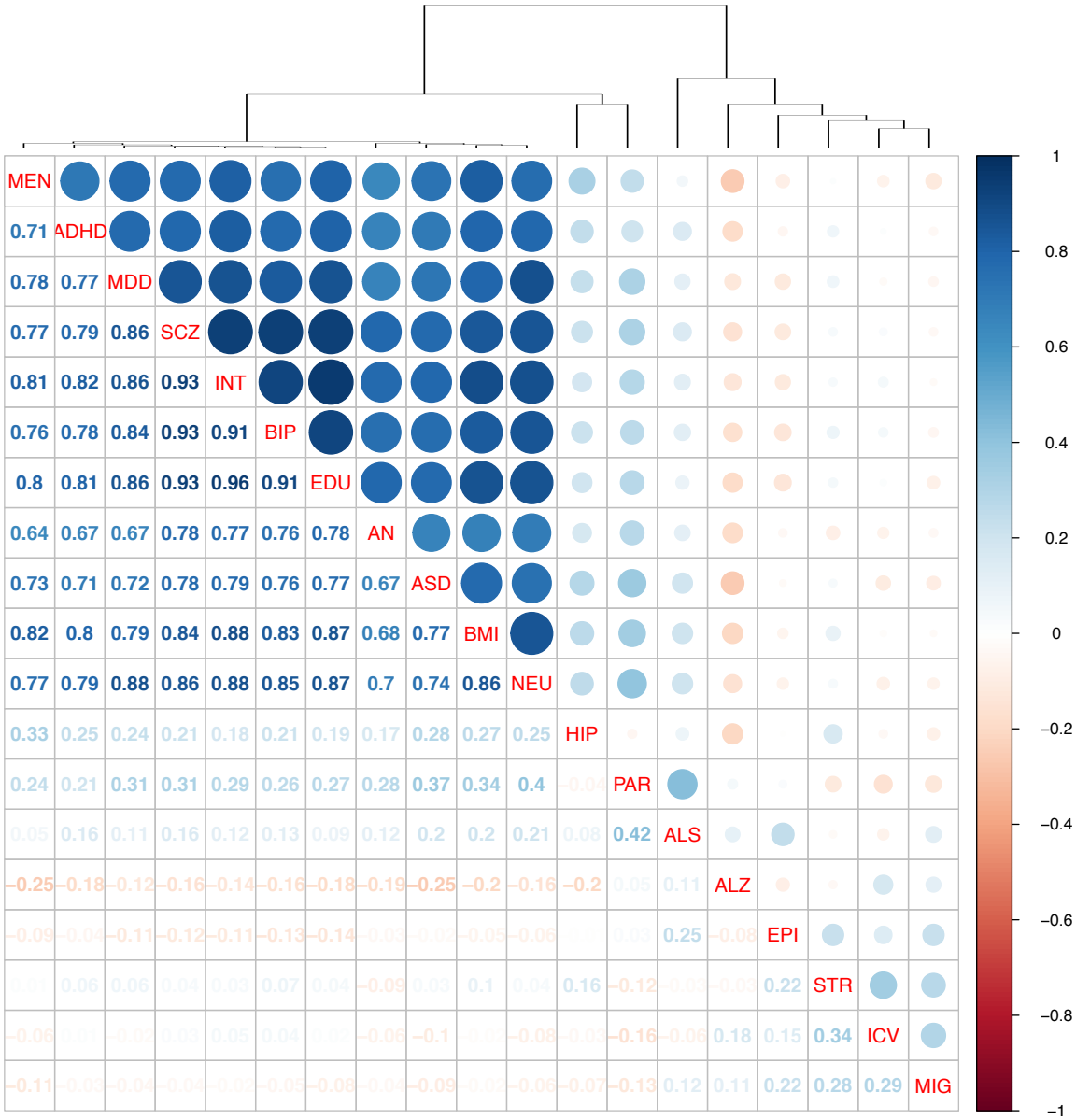

Figure S21: Correlation in cell type associations across traits in replication data set (414 cell types, 9 brain regions).

The Spearman rank correlation between the 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), EPI (epilepsy).

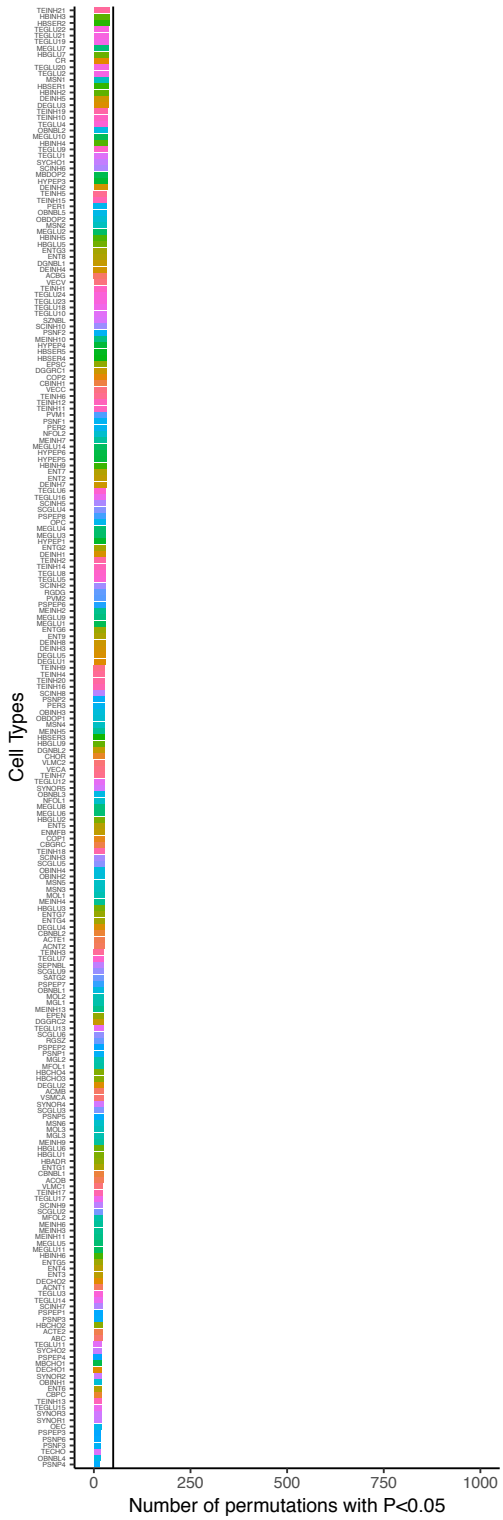

Figure S22: Number of MAGMA associations with P<0.05 using permuted gene-level genetic associations.

Gene labels were randomly permuted a thousand times for the schizophrenia MAGMA gene-level genetic association file. The number of times each cell type has a pvalue <0.05 is shown.

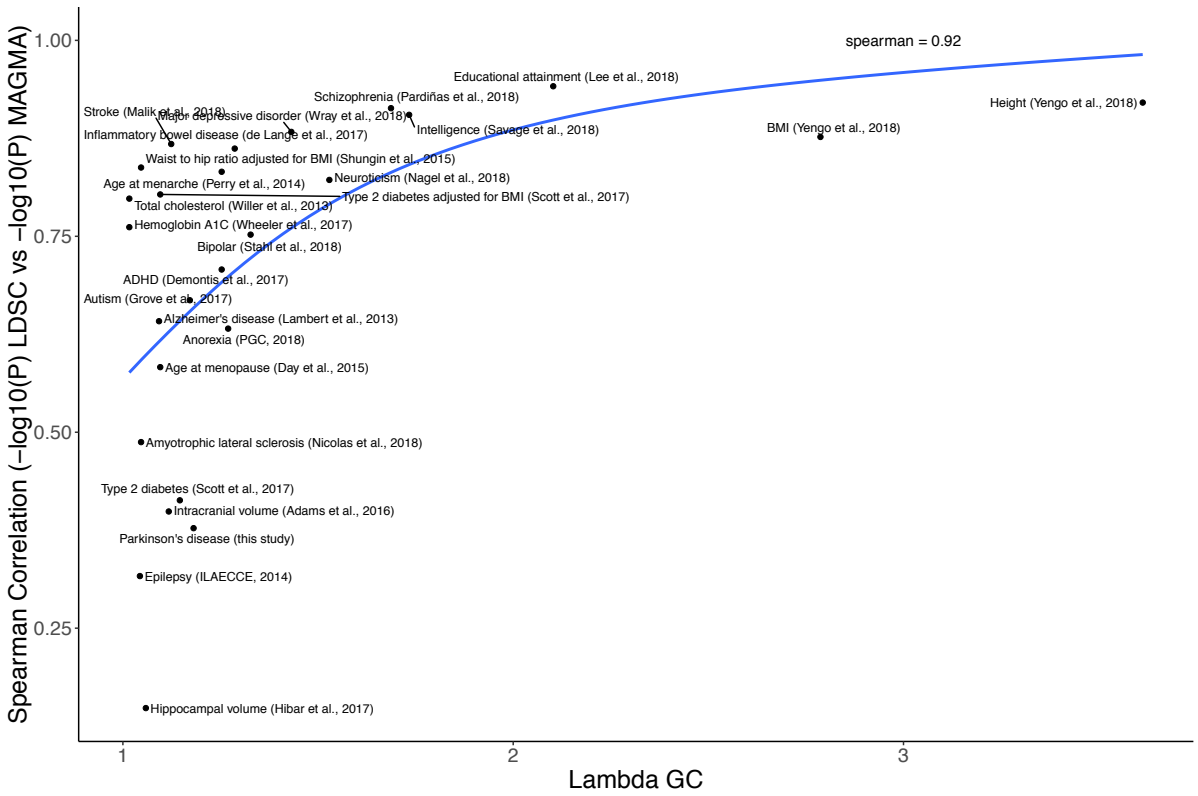

Figure S23: Correlation between λ_{GC} and similarity in cell type ordering between MAGMA and LDSC.

LDSC[94] was used to obtain λ_{GC} (a measure of the deviation of the GWAS P-value distribution from the expected) for each GWAS. Spearman rank correlation was used to test for similarity in association strength (-$\log_{10}$P) between MAGMA and LDSC for each GWAS among 231 cell types from the nervous system.

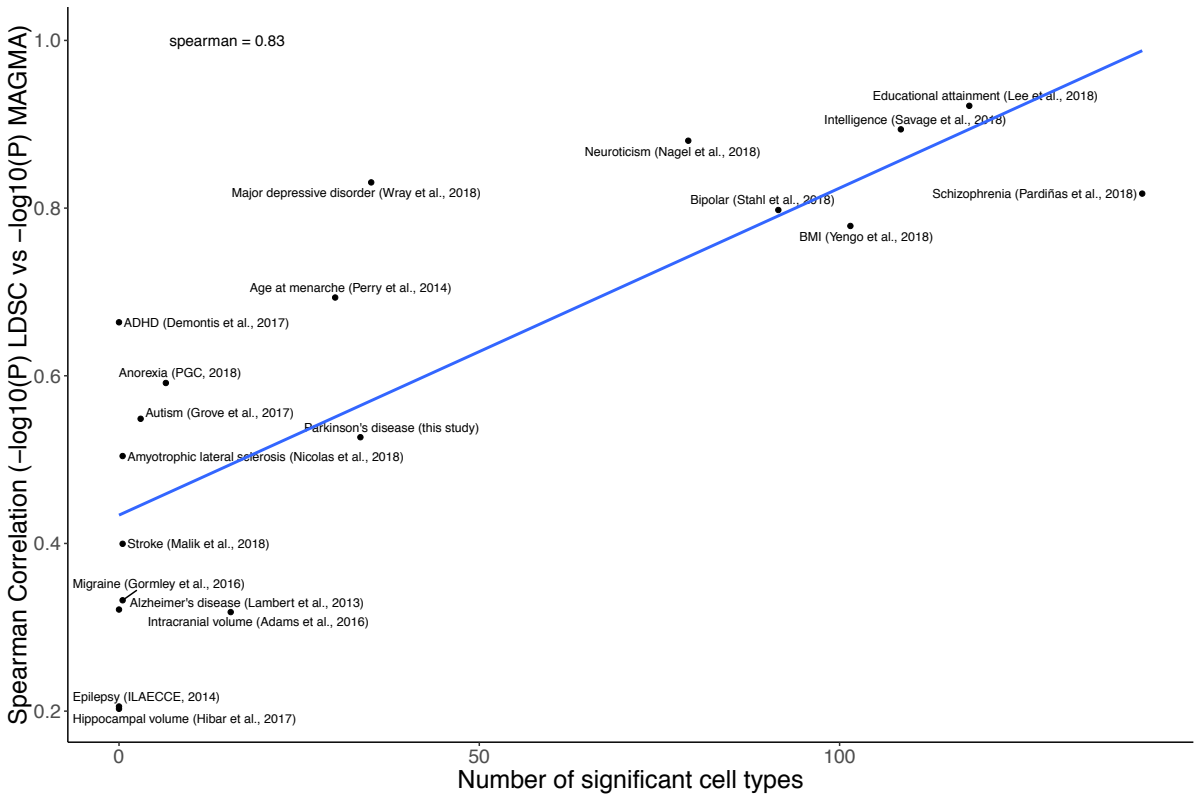

Figure S24: Correlation between mean number of significant cell types and similarity in cell type ordering between MAGMA and LDSC.

The mean number of cell types was obtained by taking the average of the number of cell types that were significantly associated with each trait (P<0.005/231) using MAGMA and LDSC. Spearman rank correlation was used to test for similarity in association strength ($-\log_{10}P$) between MAGMA and LDSC among 231 cell types from the nervous system.

References

1. Polderman, T. J. C. *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **47,** 702–709 (2015).
2. 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).
3. 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).
4. 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).
5. 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).
6. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science (80-.).* **337,** 1190–1195 (2012).
7. Akbarian, S. *et al.* The PsychENCODE project. *Nature Neuroscience* **18,** 1707–1712 (2015).
8. Ongen, H. *et al.* Estimating the causal tissues for complex traits and diseases. *Nat. Genet.* **49,** 1676–1683 (2017).
9. 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).
10. Skene, N. G. *et al.* Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.* **50,** 825–833 (2018).
11. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50,** 621–629 (2018).
12. 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).
13. 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).
14. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47,** 1236–1241 (2015).
15. Anttila, V. *et al.* Analysis of shared heritability in common disorders of the brain. *Science (80-.).* **360,** (2018).
16. Lee, S. H. *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45,** 984–994 (2013).
17. Nalls, M. A. *et al.* Parkinson's disease genetics: identifying novel risk loci, providing causal insights and improving estimates of heritable risk. *bioRxiv* 388165 (2018). doi:10.1101/388165
18. Aguet, F. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550,** 204–213 (2017).
19. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11,** (2015).

20. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47,** 1228–1235 (2015).
21. Zarei, S. *et al.* A comprehensive review of amyotrophic lateral sclerosis. *Surg. Neurol. Int.* **6,** 171 (2015).
22. Keren-Shaul, H. *et al.* A unique microglia type associated with restricting development of alzheimer's disease. *Cell* **169,** (2017).
23. 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).
24. Andreasen, N. C. & Pierson, R. The role of the cerebellum in schizophrenia. *Biol. Psychiatry* **64,** 81–88 (2008).
25. Villanueva, R. The cerebellum and neuropsychiatric disorders. *Psychiatry Res.* **198,** 527–532 (2012).
26. Zeisel, A. *et al.* Molecular architecture of the mouse nervous system. *Cell* **174,** 999–1014.e22 (2018).
27. Azevedo, F. A. C. *et al.* Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J. Comp. Neurol.* **513,** 532–541 (2009).
28. Hook, P. W. & McCallion, A. S. Heritability enrichment in open chromatin reveals cortical layer contributions to schizophrenia. *bioRxiv* (2018).
29. Fullard, J. F. *et al.* An atlas of chromatin accessibility in the adult human brain. *Genome Res.* **28,** 1243–1252 (2018).
30. Hattox, A. M. & Nelson, S. B. Layer V neurons in mouse cortex projecting to different targets have distinct physiological properties. *J. Neurophysiol.* **98,** 3330–3340 (2007).
31. Inui, A. Feeding and body-weight regulation by hypothalamic neuropeptides - Mediation of the actions of leptin. *Trends in Neurosciences* (1999). doi:10.1016/S0166-2236(98)01292-2
32. Katz, D. B., Nicolelis, M. A. L. & Simon, S. A. Gustatory processing is dynamic and distributed. *Current Opinion in Neurobiology* **12,** 448–454 (2002).
33. Braak, H. *et al.* Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol. Aging* **24,** 197–211 (2003).
34. Sulzer, D. & Surmeier, D. J. Neuronal vulnerability, pathogenesis, and Parkinson's disease. *Movement Disorders* **28,** 41–50 (2013).
35. Halliday, G. M. *et al.* Neuropathology of immunohistochemically identified brainstem neurons in Parkinson's disease. *Ann. Neurol.* **27,** 373–385 (1990).
36. 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. *Park. Relat. Disord.* (2008). doi:10.1016/j.parkreldis.2008.01.006
37. Perrett, R. M., Alexopoulou, Z. & Tofaris, G. K. The endosomal pathway in Parkinson's disease. *Molecular and Cellular Neuroscience* **66,** 21–28 (2015).
38. Saunders, A. *et al.* Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* **174,** 1015–1030.e16 (2018).
39. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14,** 955 (2017).
40. Lesnick, T. G. *et al.* A genomic pathway approach to a complex disease: axon

guidance and Parkinson disease. *PLoS Genet.* **3,** 0984–0995 (2007).
41. Moran, L. B. *et al.* Whole genome expression profiling of the medial and lateral substantia nigra in Parkinson's disease. *Neurogenetics* (2006). doi:10.1007/s10048-005-0020-2
42. Kannarkat, G. T., Boss, J. M. & Tansey, M. G. The role of innate and adaptive immunity in parkinson's disease. *Journal of Parkinson's Disease* **3,** 493–514 (2013).
43. 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,** (2015).
44. Bryois, J. *et al.* Evaluation of chromatin accessibility in prefrontal cortex of individuals with schizophrenia. *Nat. Commun.* **9,** (2018).
45. O'Mara, S. The subiculum: What it does, what it might do, and what neuroanatomy has yet to tell us. *Journal of Anatomy* **207,** 271–282 (2005).
46. Caspi, A. *et al.* The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clin. Psychol. Sci.* **2,** 119–137 (2014).
47. Sullivan, P. F. & Geschwind, D. H. Defining the genetic, genomic, cellular, and diagnostic architectures of psychiatric disorders. *Submitted*
48. Miller, A. H. & Raison, C. L. The role of inflammation in depression: From evolutionary imperative to modern treatment target. *Nature Reviews Immunology* **16,** 22–34 (2016).
49. Müller, N., Weidinger, E., Leitner, B. & Schwarz, M. J. The role of inflammation in schizophrenia. *Frontiers in Neuroscience* **9,** (2015).
50. Reynolds, R. H. *et al.* Moving beyond neurons: the role of cell type-specific gene regulation in Parkinson's disease heritability. *bioRxiv* 442152 (2018). doi:10.1101/442152
51. Recasens, A. & Dehay, B. Alpha-synuclein spreading in Parkinson's disease. *Front. Neuroanat.* **8,** (2014).
52. Engelender, S. & Isacson, O. The threshold theory for Parkinson's disease. *Trends in Neurosciences* **40,** 4–14 (2017).
53. Surmeier, D. J., Obeso, J. A. & Halliday, G. M. Selective neuronal vulnerability in Parkinson disease. *Nature Reviews Neuroscience* **18,** 101–113 (2017).
54. Braak, H., Ghebremedhin, E., Rüb, U., Bratzke, H. & Del Tredici, K. Stages in the development of Parkinson's disease-related pathology. *Cell and Tissue Research* **318,** 121–134 (2004).
55. Gilman, S. *et al.* Second consensus statement on the diagnosis of multiple system atrophy. *Neurology* **71,** 670–676 (2008).
56. Dorsey, E. R. *et al.* Virtual research visits and direct-to-consumer genetic testing in Parkinson's disease. *Digit. Heal.* **1,** 205520761559299 (2015).
57. 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).
58. Seidel, K. *et al.* The brainstem pathologies of Parkinson's disease and dementia with lewy bodies. *Brain Pathol.* **25,** 121–135 (2015).
59. 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).

60. Stahl, E. *et al.* Genomewide association study identifies 30 loci associated with bipolar disorder. *bioRxiv* 173062 (2017). doi:10.1101/173062
61. Wray, N., Sullivan, PF & PGC, M. D. D. W. G. of the. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50,** 668–681 (2018).
62. 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).
63. Grove, J. *et al.* Common risk variants identified in autism spectrum disorder. *bioRxiv* **33,** 42 (2017).
64. Gormley, P. *et al.* Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. *Nat. Genet.* **48,** 1296 (2016).
65. 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).
66. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* (2018). doi:10.1038/s41588-018-0269-7
67. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45,** 1452–1458 (2013).
68. Anney, R. J. L. *et al.* Genetic determinants of common epilepsies: A meta-analysis of genome-wide association studies. *Lancet Neurol.* **13,** 893–903 (2014).
69. 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).
70. Nelson, C. P. *et al.* Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nature Genetics* **49,** (2017).
71. 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).
72. Hibar, D. P. *et al.* Novel genetic loci associated with hippocampal volume. *Nat. Commun.* **8,** 13624 (2017).
73. 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).
74. Adams, H. H. H. *et al.* Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nat. Neurosci.* **19,** 1569–1582 (2016).
75. 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).
76. Willer, C. J. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45,** 1274–1285 (2013).
77. Scott, R. A. *et al.* An expanded genome-wide association study of type 2 diabetes in europeans. *Diabetes* **66,** 2888–2902 (2017).
78. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518,** 187–196 (2015).
79. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis

of genomewide association scans. *Bioinformatics* **26,** 2190–2191 (2010).
80. Duncan, L. *et al.* Significant locus and metabolic genetic correlations revealed in genome-wide association study of anorexia nervosa. *Am. J. Psychiatry* **174,** 850–858 (2017).
81. Kirk, K. M. *et al.* The Anorexia Nervosa Genetics Initiative: Study description and sample characteristics of the Australian and New Zealand arm. *Aust. N. Z. J. Psychiatry* **51,** 583–594 (2017).
82. Thornton, L. M. *et al.* The Anorexia Nervosa Genetics Initiative (ANGI): Overview and methods. *Contemp. Clin. Trials* **74,** 61–69 (2018).
83. Boraska, V. *et al.* A genome-wide association study of anorexia nervosa. *Mol. Psychiatry* **19,** 1085–1094 (2014).
84. Sudlow, C. *et al.* UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12,** (2015).
85. Moskvina, V., Holmans, P., Schmidt, K. M. & Craddock, N. Design of case-controls studies with unscreened controls. *Ann. Hum. Genet.* **69,** 566–576 (2005).
86. Saunders, A. *et al.* Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* **174,** 1015–1030.e16 (2018).
87. Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23,** 1294–1296 (2007).
88. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).
89. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47,** 1228–1235 (2015).
90. Alexa, A. & Rahnenfuhrer, J. topGO: Enrichment analysis for gene ontology. (2016).
91. 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).
92. 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).
93. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43,** e47 (2015).
94. Bulik-Sullivan, B. *et al.* LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47,** 291–295 (2015).