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 Kogelman[5], Hunna J. Watson[8], Eating Disorders Working Group of the Psychiatric Genomics Consortium, International Headache Genetics Consortium, 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 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, MRC Social, Genetic and Developmental Psychiatry Centre, King's College 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

† Equal contributions. * Correspond with Drs Sullivan (patrick.sullivan@ki.se) and Hjerling-Leffler (jens.hjerling-leffler@ki.se).

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 are active, which is essential for understanding etiology and subsequent experimental modeling. Here we integrate GWAS results with single-cell transcriptomic data from the entire mouse nervous system to systematically identify cell types underlying psychiatric disorders, neurological diseases, and brain complex traits. We show that psychiatric disorders are predominantly associated with cortical and hippocampal excitatory neurons, and medium spiny neurons from the striatum. Cognitive traits were generally associated with similar cell types but their associations were driven by different genes. Neurological diseases were associated with different cell types, which is 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 of the oligodendrocyte lineage. Using post-mortem brain transcriptomic data, we confirmed alterations in these cells, even at the earliest stages of disease progression. Our study provides an important framework for understanding the cellular basis of complex brain maladies, and reveals an unexpected role of oligodendrocytes in Parkinson's disease.

Introduction

Understanding the genetic basis of complex brain disorders is critical for identifying individuals at risk, designing prevention strategies, and developing rational therapeutics. In the last 50 years, twin studies have shown that psychiatric disorders, neurological diseases, and cognitive traits are strongly

influenced by genetic factors, explaining a mean of ~50% of the variance in liability [1], and GWAS have identified thousands of highly significant loci [2–5]. However, interpretation of GWAS results remains challenging. First, >90% of the identified variants are located in non-coding regions [6], complicating precise identification of risk genes and mechanisms. Second, extensive linkage disequilibrium present in the human genome confounds efforts to pinpoint causal variants and the genes they influence . Finally, it remains unclear in which tissues and cell types these variants are active, and how they disrupt specific biological networks to impact disease risk.

Functional genomic studies from brain are now seen as critical for interpretation of GWAS findings as they can identify functional regions (e.g., open chromatin, enhancers, transcription factor binding sites) and target genes (via chromatin interactions and eQTLs) [7]. Gene regulation varies substantially across tissues and cell types [8,9], and hence it is critical to perform functional genomic studies in empirically identified cell types or tissues.

Multiple groups have developed strategies to identify tissues associated with complex traits [10–14], but few have focused on the identification of salient cell types within a tissue. Furthermore, studies aiming to identify relevant cell types often used only a small number of cell types derived from one or few different brain regions [4,12–18]. For example, we recently showed that, among 24 brain cell types, four types of neurons were consistently associated with schizophrenia [12]. We were explicit that this conclusion was limited by the relatively few brain regions we studied; other cell types from unsampled regions could conceivably contribute to the disorder.

Here, we integrate a wider range of gene expression data – tissues across the human body and single-cell gene expression data from an entire nervous system – to identify tissues and cell types underlying a large number of complex traits (**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 providing new clues into its etiology.

Results

Genetic correlations among complex traits

Our goal was to use GWAS results to identify relevant tissues and cell types. Our primary focus was human phenotypes whose etiopathology is based in the central nervous system. We thus obtained 18 sets of GWAS summary statistics from European samples for brain-related complex traits. These were selected because they had at least one genome-wide significant association (as of 2018; e.g., Parkinson's disease, schizophrenia, and IQ). For comparison, we included GWAS summary statistics for 8 diseases and traits with large sample sizes whose etiopathology is not rooted in the central nervous system (e.g., type 2 diabetes). The selection of these conditions allowed contrasts of tissues and cells highlighted by our primary interest in brain phenotypes with non-brain traits. For Parkinson's disease, we meta-analyzed summary statistics from a published GWAS [19] (9,581 cases, 33,245 controls) with self-reported Parkinson's disease from 23andMe (12,657 cases, 941,588 controls) after finding a high genetic correlation (r_g) [20] between the samples (r_g=0.87, s.e=0.068). In this new meta-analysis, we identified 61 independent loci associated with Parkinson's disease (49 reported previously[18] and 12 novel) (**Figure S1**).

We estimated the genetic correlations (r_g) between these 26 traits. We confirmed prior reports [21,22] that psychiatric disorders were strongly inter-correlated (e.g., high positive correlations for

schizophrenia, bipolar disorder, and MDD) and shared little overlap with neurological disorders (**Figure S2** and **Table S1**). Parkinson's disease was genetically correlated with intracranial volume [18] (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). There is substantial genetic heterogeneity across traits, which is a necessary (but not sufficient) condition for trait associations with different tissues or cell types.

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

We first aimed to identify the human tissues showing enrichment for genetic associations using bulk-tissue RNA-seq (53 tissues) from GTEx [8] (**Figure 1A**). To robustly identify the tissues implied by these 26 GWAS, we used two approaches (MAGMA [23] and LDSC [13,24]) which employ different assumptions (**Methods**). MAGMA tested for a positive relationship between gene expression specificity and gene-level genetic associations whereas LDSC tested whether the 10% most specific genes in each tissue were enriched for trait heritability (**Figure 1B**).

Examination of most non-brain traits found, as expected, associations with salient tissues. For example, as shown in **Figure 1D** and **Table S2**, inflammatory bowel disease was strongly associated with immune tissues (whole blood, spleen) and alimentary tissues impacted by the disease (terminal ileum and transverse colon,). Coronary artery disease was most associated with aorta and coronary artery. Age at menopause was most associated with reproductive tissues, and type 2 diabetes with pancreas. Thus, our approach can identify the expected tissue associations given the pathophysiology of the different traits.

For brain-related traits (**Figure S3 and Table S2**), 12 of 18 traits were significantly associated with one or more GTEx brain regions. For example, schizophrenia, intelligence, educational attainment, neuroticism, and MDD were most significantly associated with brain cortex or frontal cortex. Schizophrenia also had several possible associations with non-brain tissues (e.g., adrenal gland, heart) (**Figure 1C**); these were less significant by 10 or more logs, and were significant by MAGMA or LDSC but not both. Interestingly, these non-brain tissues are electrically excitable and/or prominently muscular. Myocytes share some functions with neurons (e.g., maintaining a membrane potential, depolarization, repolarization). These other tissues can be expected to have minor overlap with brain neurons: adrenal gland (GTEx sampled both adrenal cortex and medulla, the latter is a neuroendocrine tissue containing sympathetic neurons); heart is muscular; esophageal muscular layer (but not mucosa); three arterial samples (which have an important muscular layer); cervix (a muscular structure, more ectocervix than the endocervical lining); and uterus (muscular).

Parkinson's disease was most significantly associated with substantia nigra and spinal cord (**Figure 1C**). Autism and ADHD were most strongly associated with basal ganglia structures (putamen and nucleus accumbens). Alzheimer's disease was associated with tissues with prominent roles in immunity (whole blood and spleen) consistent with other studies [25,26]. Stroke was associated with coronary artery and aorta (consistent with a role of arterial pathology in stroke) [27]. Traits with no or unexpected associations could occur because the primary GWAS had insufficient sample size for its genetic architecture [28] or because the tissue RNA-seq data omitted the correct tissue or cell type.

Most brain-related traits were strongly associated with the cerebellum. Although theories postulate a cerebellar role in the etiology of psychiatric disorders [29,30], an alternative explanation is that cerebellum has a high proportion of neurons (78-81% [31,32]). We hypothesized that the relative proportion of neurons in the different brain regions could confound our tissue-trait association analysis for brain disorders 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 [31]) 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 association strength was positively correlated with the proportion of neurons in the brain region (**Figure S4**), suggesting that cerebellar associations are at least partly due to its high neuronal proportion.

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

Association of brain phenotypes 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 [31] to systematically map brain-related traits to cell types (**Figures 2, S5-6**). Our use of mouse data to inform human genetic findings was carefully considered (see **Discussion**).

As in our previous study of schizophrenia based on a small number of brain regions [12], 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**). We also observed that many other types of neurons were associated with schizophrenia albeit less significantly (e.g., excitatory neurons from the thalamus and inhibitory neurons from the hindbrain) (**Figure S6**). The cell type with the strongest schizophrenia association (TEGLU4) is located in layer 5 of the cortex. The association of TEGLU4 cells is consistent with our prior report [12] and with results from other groups including significant schizophrenia heritability enrichment in open chromatin regions in layer 5 excitatory neurons in mouse [33] and in open chromatin regions from human cortical neurons [34]. This pattern of replication [12] and consistency in orthogonal data types [33 34] is notable, and implies an important role of cortical excitatory neurons in schizophrenia. Moreover, excitatory neurons from cortical layer 5 project to the superior colliculus [35] (implicated in attentional target selection) and to the striatum [35]. The eighth most significant cell type (MEGLU6, a midbrain excitatory neuron) is located in the superior colliculus, while the twelfth top cell type (MSN2, D2 medium spiny neurons) is located in the striatum. These findings may imply a brain circuit etiologically important for schizophrenia.

Educational attainment, intelligence, bipolar disorder, neuroticism, and MDD had similar cellular association patterns to schizophrenia (**Figures S5-7** and **Table S3**). Projecting neurons from the telencephalon (excitatory neurons from the cortex, hippocampus, amygdala, granule neurons and neuroblasts from the dentate gyrus and medium spiny neurons from the striatum) were significantly more associated with psychiatric and cognitive traits than any other category of cell types (**Figure S6**). We did not observe any significant associations with immune or vascular cells for any psychiatric disorder or cognitive traits.

For body mass index (BMI), the pattern of associations contrasted with psychiatric and cognitive traits (**Figure 2, S5-S7)** 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 [36], and the solitary nucleus is involved in gustatory processing and projects to the hypothalamus [37]. These results suggest functional connections between these cell types.

Neurological disorders generally implicated fewer cell types, possibly because neurological GWAS had lower signal than GWAS of cognitive, anthropometric, and psychiatric traits (**Figure S8**). Consistent with the genetic correlations reported above, the pattern of associations for neurological

disorders was distinct from psychiatric disorders (**Figures S5** and **S7**), again reflecting that neurological disorders have minimal functional overlap with psychiatric disorders [21] (**Figure S2**).

Stroke and migraine were significantly associated with arterial smooth muscle cells (VSMCA) consistent with an important role of vascular process in these traits. The overall pattern of association of stroke and migraine were negatively correlated with psychiatric, cognitive traits, and BMI (**Figure S7**), as stroke and migraine appear to have a predominant non-neuronal origin.

The hallmark of Parkinson's disease is degeneration of dopaminergic neurons in the substantia nigra, and we found that these cells (MBDOP2) were strongly associated with the disorder (**Figure 2B**). In addition to the degeneration of dopaminergic neurons, cell loss in other brain regions can occur [38,39]. We found that many of the cell types reported to degenerate in Parkinson's disease also show strong associations (**Figure 2B** and **Table S3**), including serotonergic raphe nucleus neurons [40] (HBSER2, top hit) and cholinergic neurons of the pons [41] (HBCHO2). We also identified significant associations in cells of the medulla (HBINH4 and HBGLU2), the region associated with the earliest lesions in Parkinson's disease [38]. 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 associations for 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 (**Figure 1C**), as the spinal cord contains the highest proportion of oligodendrocytes (71%) in the nervous system [31].

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

Understanding which biological functions are dysregulated in different cell types is a key component of 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 [42].

Taken together, we show that cell type-trait associations are driven by genes belonging to specific biological pathways, providing insight into the etiology of complex brain related 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 S7**). For example, the Spearman rank correlation of cell type associations (-$\log_{10}$P) between schizophrenia and intelligence was 0.94 (0.93 for educational attainment) 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 S2** 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 S9**). Similarly, we found that controlling for educational attainment had little effect on the schizophrenia association and vice versa (**Figure S10**). 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 **S5**). 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 S11**), intelligence (**Figure S12**) and BMI (**Figure S13**). No cell type remained significantly associated with bipolar disorder (**Figure S14**), neuroticism (**Figure S15**), MDD (**Figure S16**), and age at menarche (**Figure S17**) 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 S18**), anorexia nervosa (**Figure S19**) and autism (**Figure S20**).

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 S21**). 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 [12]. 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 S22**). 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 repeated these analyses in independent RNA-seq datasets. A key caveat is that these other datasets did not sample the entire nervous system as in the analyses above. First, we used a single-cell RNA-seq dataset that identified 565 cell types in 690K single cells from 9 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) [43]. We found similar patterns of association in this external dataset (**Figure S23-S24, Table S3**, and **Table S6**). Notably, for schizophrenia, we strongly replicated associations with cortical excitatory neurons, excitatory neurons from the CA1 region of the hippocampus, medium spiny neurons, and interneurons (**Figure S24**). We also observed similar cell type associations for other psychiatric traits (**Figure S24-S25**). For neurological disorders, we replicated the associations of: Parkinson's disease with dopaminergic neurons from the ventral midbrain and oligodendrocytes (serotonergic neurons were not sampled); for stroke and migraine with arterial smooth muscle cells; and Alzheimer's disease with activated microglia (**Figure S24**), a cell type previously associated with the disease [11,44].

Second, we reanalyzed these 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; **Figure 3** 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 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; Alzheimer with microglia and intracranial volume with neuronal progenitors and neuroblasts (suggesting that drivers of intracranial volume are cell types implicated in increasing cell mass). While medium spiny neurons were strongly associated with multiple brain related traits in this dataset, we did not identify any technical reasons that could explain this result (see **Table S8** for summary statistics on this dataset).

Third, we evaluated a human single-nuclei RNA-seq dataset consisting of 15 different cell types from cortex and hippocampus [45] (**Figure 4A** and **Table S9**). We replicated our findings with psychiatric and cognitive traits being associated with pyramidal neurons (excitatory) and interneurons (inhibitory) from the somatosensory cortex and from the CA1 region of the hippocampus. We also replicated the association of Parkinson's disease with oligodendrocytes (dopaminergic and serotonergic neurons were not sampled) and the association of Alzheimer's disease with microglia.

Fourth, we evaluated a human single-nuclei RNA-seq dataset consisting of 31 different cell types from 3 different brain regions (visual cortex, frontal cortex and cerebellum) (**Figure 4B** and **Table S10**). We replicated the association of psychiatric disorders and cognitive traits with excitatory and

inhibitory neurons and the association of Alzheimer's disease with microglia. However, oligodendrocytes were not significantly associated with Parkinson's disease in this dataset.

Most cell type-trait associations were attenuated using human single-nuclei data compared with mouse single-cell RNA-seq data, suggesting that the transcripts that are lost by single-nuclei RNA-seq are important for a large number of disorders and/or that the controlled condition of mouse experiments provide more accurate gene expression quantifications (see **Discussion** and **Figure S26**).

Comparison with case/control differentially expressed genes at the cell type level

We compared our findings for Alzheimer's disease with a recent study that performed differential expression analysis at the cell type level between 24 Alzheimer's cases and 24 controls [46] (prefrontal cortex, Brodmann area 10). We tested whether the top 500, top 1000 and top 2000 most differentially expressed genes (no pathology vs pathology) in six different cell types (excitatory neurons, inhibitory neurons, oligodendrocytes, oligodendrocytes precursor cells, astrocyte and microglia) were enriched in genetic associations with Alzheimer's disease using MAGMA. Consistently with our results (**Table S3**, **Figure 3** and **Figure 4**), we found that genes differentially expressed in microglia were the most associated with Alzheimer's disease genetics (**Table S11**), indicating that our approach appropriately highlight the relevant cell type at a fraction of the cost of a case-control single cell RNA-seq study. As performing case-control single cell RNA-seq studies in the entire nervous system is currently cost prohibitive, the consistency of our results with the case-control study of Alzheimer's disease suggests that our results could be leveraged to target specific brain regions and cell types in future case-control genomic studies of brain disorders.

Validation of oligodendrocyte pathology in Parkinson's disease

We investigated the role of oligodendrocyte lineage cells in Parkinson's disease. First, we tested whether oligodendrocytes were significantly associated with Parkinson's disease conditioning on the top neuronal cell type in the different replication datasets and found: (a) oligodendrocytes were associated with Parkinson's disease in a human replication dataset at a Bonferroni significant level (P=9e-5) [45]; (b) in the other replication datasets, oligodendrocytes were associated with Parkinson's disease at a nominal level (P=7e-4, P=0.014, P=0.019) [47–49]; and (c) combining the conditional evidence from all datasets, oligodendrocytes were significantly associated with Parkinson's disease independently of the top neuronal association (P=7.3e-10, Fisher's combined probability test).

Second, we used EWCE [11] to test whether genes with rare variants associated with Parkinson's disease (**Table S12**) were specifically expressed in cell types from the mouse nervous system. We found that dopaminergic neurons (MBDOP2) were the most significantly enriched (**Table S13**) 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.

Third, we applied EWCE [11] 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 5**). Three of the studies had a case-control design and measured gene expression in: (a) the substantia nigra of 9 controls and 16 cases [50], (b) the medial substantia nigra of 8 controls and 15 cases [51], and (c) the lateral substantia nigra of 7 controls and 9 cases [51]. 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 [52,53].

We also analyzed gene expression data from post-mortem human brains which had been scored by neuropathologists for their Braak stage [54]. 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 5**), 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 used gene expression data from cells sampled from the entire nervous system to systematically map cell types to GWAS results from multiple psychiatric, cognitive, and neurological complex phenotypes.

We note several limitations. First, we again emphasize that we can implicate a particular cell type but it is premature to exclude cell types for which we do not have data [12]. Second, we used gene expression data from mouse to understand human phenotypes. We believe our approach is appropriate for several reasons. (A) Crucially, the key findings replicated in human data. (B) Single-cell RNA-seq is achievable in mouse but difficult in human neurons (where single-nuclei RNA-seq is typical [45,48,70,71]). In brain, differences between single-cell and single-nuclei RNA-seq are important as transcripts that are missed by sequencing nuclei are important for psychiatric disorders, and we previously showed that dendritically-transported transcripts (important for schizophrenia) are specifically depleted from nuclei datasets [12] (we confirmed this finding in four additional datasets, **Figure S26**). (C) Correlations in gene expression for cell type across species is high (median correlation 0.68**,** **Figure S27**), and as high or higher than correlations across methods within cell type and species (single-cell vs single-nuclei RNA-seq, median correlation 0.6) [72]. (D) We evaluated protein-coding genes with 1:1 orthologs between mouse and human. These constitute most human protein-coding genes, and these genes are generally highly conserved particularly in the nervous system. We did not study genes present in one species but not in the other. (E) More specifically, we previously showed that gene expression data cluster by cell type and not by species [12], indicating broad conservation of core brain cellular functions across species. (F) We used a large number of genes to map cell types to traits (~1500 genes with LDSC for each cell type, ~15,000 genes with MAGMA), minimizing potential bias due to individual genes differentially expressed across species. (G) 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. However, it remains possible that some cell types have different gene expression patterns between mouse and human, are only present in one species, have a different functions or are involved in different brain circuits.

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

For schizophrenia, we replicated and extended our previous findings [12]. In two independent datasets, we found significant associations with excitatory neurons from the cortex/hippocampus, D2 medium spiny neurons from the striatum, and interneurons (e.g., for MSN2-schizophrenia, the approximate probability of this degree of replication is P<1e-29, Fisher's combined probability test). These results are consistent with the strong schizophrenia heritability enrichment observed in open chromatin regions from: human dorsolateral prefrontal cortex [55]; human cortical, striatal and hippocampal neurons [34]; and mouse open chromatin regions from cortical excitatory and inhibitory neurons [33]. This degree of replication in independent transcriptomic datasets from multiple groups along with consistent findings using orthogonal open chromatin data is notable, and strongly implicates these cell types in the etiology of schizophrenia.

As in mouse open chromatin data [33], 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 5-6), the subiculum, the CA1 region of the hippocampus, and the piriform cortex. This is intriguing as layer 5 excitatory neurons project to the striatum and the superior colliculus [35], 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 [56]. This suggests the potential salience of a circuit between independently associated cell types in schizophrenia.

Moreover, we found that brain phenotypes with 30 or more associations (e.g., MDD, bipolar disorder, educational attainment, intelligence, and neuroticism) implicated largely similar cell types as schizophrenia with the strongest signal for excitatory neurons from the cortex and hippocampus, medium spiny neurons, interneurons and specific types of midbrain neurons. These biological findings are consistent with genetic and epidemiological evidence of a general psychopathy factor underlying diverse clinical psychiatric disorders [21,57,58]. 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 some psychiatric disorders [59,60]. Our results did not implicate any brain immune cell types in psychiatric disorders. We interpret these negative findings cautiously as we did not fully sample the immune system. It is also possible that a small number of genes are active in immune cell types and that these cell types play an important role in the etiology of psychiatric disorders. Finally, if immune functions are salient for a small subset of patients, GWAS may not identify these loci without larger and more detailed studies.

Our findings for neurological disorders were strikingly different from psychiatric disorders. In contrast to previous studies that either did not identify any cell type associations with Parkinson's disease [61] or identified significant associations with cell types from the adaptive immune system [53], we found that dopaminergic neurons and oligodendrocytes were consistently and significantly associated with the disease. It is well established that loss of dopaminergic neurons in the substantia nigra is a hallmark of Parkinson's disease. Our findings suggest that dopaminergic neuron loss in Parkinson's disease is at least partly due to biological mechanisms intrinsic to dopaminergic neurons. In addition, we found significant associations for other cell types that degenerate in Parkinson's disease (e.g., raphe nucleus serotonergic neurons and cholinergic neurons of the pons), suggesting that specific pathological mechanisms may 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 [62]. Two theories for the selective

vulnerability of neuronal populations in Parkinson's disease currently exist: the "spread Lewy pathology model" assumes cell-to-cell contacts enabling spreading of prion-like α-synuclein aggregates [63]; and the "threshold theory" [64,65] 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 can co-exist, our findings support the existence of cell autonomous mechanisms contributing to selective vulnerability. We caution that we have not tested all cell types which are known to degenerate (e.g., noradrenergic cells of the locus coeruleus), 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 plausible common pathogenic mechanism.

The strong association of oligodendrocytes with Parkinson's disease was unexpected. A possible explanation is that this association could be due to a related disorder (e.g., multiple system atrophy, characterized by Parkinsonism and accumulation of α-synuclein in glial cytoplasmic inclusions [66]). However, this explanation is unlikely as multiple system atrophy is a very rare disorder; hence, only a few patients are likely to have been included in the Parkinson's disease GWAS 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 clinically-ascertained Parkinson's disease [19]. In addition, self-report of Parkinson's disease in 23andMe subjects was confirmed by a neurologist in all 50 cases evaluated [67].

We did not find an association of oligodendrocytes with Parkinsonism for genes affected by rare variants although this result may reflect the low power and insufficient number of genes. However, 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 [68]. These inclusions ("coiled bodies") are typically found throughout the brainstem nuclei and fiber tracts [69]. Although the presence of coiled bodies in oligodendrocytes is a common, specific, and well-documented neuropathological feature of Parkinson's disease, the importance of this cell type and its early involvement in disease has not been fully recognized. Our findings suggest that intrinsic genetic alterations in oligodendrocytes occur 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.

Taken together, 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 results

Our goal was to use GWAS results to identify relevant tissues and cell types. Our primary focus was human phenotypes whose etiopathology is based in the central nervous system. We thus obtained 18 sets of GWAS summary statistics from European samples for brain-related complex traits. These were selected because they had at least one genome-wide significant association (as of 2018; e.g., Parkinson's disease, schizophrenia, and IQ). For comparison, we included GWAS summary statistics for 8 diseases and traits with large sample sizes whose etiopathology is not rooted in the central nervous system (e.g., type 2 diabetes). The selection of these conditions allowed contrasts of tissues and cells highlighted by our primary interest in brain phenotypes with non-brain traits.

The phenotypes were: schizophrenia [2], educational attainment [3], intelligence [15], body mass index [5], bipolar disorder [73], neuroticism [4], major depressive disorder [74], age at menarche [75], autism [76], migraine [77], amyotrophic lateral sclerosis [78], ADHD [79], Alzheimer's disease [80], age at menopause [81], coronary artery disease [82], height [5], hemoglobin A1c [83], hippocampal volume [84], inflammatory bowel disease [85], intracranial volume [86], stroke [87], type 2 diabetes mellitus [88], type 2 diabetes adjusted for BMI [88], waist-hip ratio adjusted for BMI [89], and anorexia nervosa [90].

For Parkinson's disease, we performed an inverse variance-weighted meta-analysis [91] using summary statistics from Nalls et al. [19] (9,581 cases, 33,245 controls) and summary statistics from 23andMe (12,657 cases, 941,588 controls). We found a very high genetic correlation (r_g) [20] between results from these cohorts (r_g=0.87, s.e=0.068) with little evidence of sample overlap (LDSC bivariate intercept=0.0288, s.e=0.0066). The P-values from the meta-analysis strongly deviated from the expected (**Figure S28**) but was consistent with polygenicity (LDSC intercept=1.0048, s.e=0.008) rather than uncontrolled inflation [20].

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 [31] (note that after filtering on unique molecular identifier (UMI) counts, described below, only 231 of these cell types are used). The 231 cell types that passed quality control expressed a median of 13736 genes, had a median UMI total count of ~1M and summed the expression of a median of 691 single cells (**Table S14**). The replication datasets were: 1) a mouse study that sampled 690K single cells from 9 brain regions and identified 565 cell types[92] (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 [12]; 3) a study that sampled 19,550 nuclei from frozen adult human post-mortem hippocampus and prefrontal cortex and identified 16 cell types [45]; 4) a study that generated 36,166 single-nuclei expression measurements (after quality control) from the human visual cortex, frontal cortex and cerebellum [48]. We obtained bulk tissues RNA-seq gene expression data from 53 tissues from the GTEx consortium [8] (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 within each cell type so that each cell type had a comparable specificity distribution using the *rntransform* function from the GenABEL R package (**Figure S29**) [93]. Our genome-wide specificity metrics were highly correlated for related tissues/cell types and lowly correlated for unrelated tissues/cell types (**Table S15** and **Table S16**), which allowed us to cluster related tissues as expected (**Figure S30** and **Figure S31**). Similarly, the top 10% most

specific genes (**Table S17** and **Table S18**) in each tissue/cell partially overlapped for related tissues/cell types, did not overlap for unrelated tissue/cell types and allowed to cluster related tissues/cell types as expected (**Figure S32** and **Figure S33**).

MAGMA primary and conditional analyses

MAGMA (v1.06b) [23] 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, with MAF < 1% or with estimated odds ratio > 25 or smaller than 1/25, the MHC region (chr6:25-34 Mb) for all GWAS and the *APOE* region (chr19:45020859–45844508) for Alzheimer's GWAS. 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 [94] 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.

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 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 5% of the random samplings having a P-value <0.05 (**Figure S34**) for all cell types. Our significance threshold (0.005/231=2.2e-5) was lower than the minimum P-value obtained with the permuted gene labels (1000 permutations*231 cell types=231,000 tests, minimum P-value=2.5e-5).

We also evaluated the effect of varying window sizes (for the SNPs to gene assignment step of MAGMA) on the schizophrenia cell type associations strength ($-\log_{10}(P)$). We observed strong Pearson correlations in cell type associations strength ($-\log_{10}(P)$) across the different window sizes tested (**Figure S35**). Our selected window size (35kb upstream to 10 kb downstream) had Pearson correlations ranging from 0.94 to 0.98 with the other window sizes, indicating that our results are robust to this parameter.

LD score regression analysis

We used partitioned LD score regression [95] 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. Only genes with at least 1TPM or 1 UMI per million in at least one cell type were used for this analysis. 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 [13]. 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. 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.

We evaluated the effect of varying window sizes and varying the percentage of most specific genes on the schizophrenia cell type associations strength (-$\log_{10}$P). We observed strong Pearson correlations in cell type associations strength (-$\log_{10}$P) across the different percentage and window sizes tested (**Figure S36**). Our selected window size (100 kb upstream to 100 kb downstream, top 10% most specific genes) had Pearson correlations ranging from 0.95 to 1 with the other window sizes and percentage, indicating that our results are robust to these parameters.

MAGMA vs LDSC ranking

In order to test whether the cell type ranking obtained using MAGMA and LDSC in the Zeisel et al. dataset [31] 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 test statistics from the expected) (Spearman ρ=0.9) (**Figure S37**) and with the average number of cell types below our stringent significance threshold using LDSC and MAGMA (Spearman ρ=0.75), indicating that the overall ranking of the cell types is very similar between the two methods, provided that the GWAS is well powered (**Figure S38**). In addition, we found that λ_{GC} was strongly correlated with the strength of association of the top tissue (-$\log_{10}$P) in the GTEx dataset (Pearson correlation=0.84) (**Figure S39**), indicating that cell type – trait associations are stronger for well powered GWAS.

Dendritic depletion analysis

This analysis was performed as previously described [12]. In brief, all datasets were reduced to a set of six common cell types: pyramidal neurons, interneurons, astrocytes, microglia and oligodendrocyte precursors. Specificity was recalculated using only these six cell types. Comparisons were then made between pairs of datasets (denoted in the graph with the format 'X versus Y'). The difference in specificity for a set of dendrite enriched genes is calculated between the datasets. Differences in specificity are also calculated for random sets of genes selected from the background gene set. The probability and z-score for the difference in specificity for the dendritic genes is thus estimated. Dendritically enriched transcripts were obtained from Supplementary Table 10 of Cajigas et al. [96]. For the KI dataset [12], we used S1 pyramidal neurons. For the Zeisel 2018 dataset [31] we used all ACTE* cells as astrocytes, TEGLU* as pyramidal neurons, TEINH* as interneurons, OPC as oligodendrocyte precursors and MGL* as microglia. For the Saunders dataset [43], we used all Neuron.Slc17a7 cellt ypes from FC, HC or PC as pyramidal neurons; all Neuron.Gad1Gad2 cell types from FC, HC or PC as interneurons; Polydendrocye as OPCs; Astrocyte as astrocytes, and Microglia as microglia. The Lake datasets both came from a single publication [48] which had data from frontal cortex, visual cortex and cerebellum. The cerebellum data was not used here. Data from frontal and visual cortices were analyzed separately. All other datasets were used as described in our previous publication [12]. The code and data for this analysis are available as an R package (see code availability below).

GO term enrichment

We tested whether genes that were highly specific to a trait-associated cell type (top 20% in a given cell type) AND highly associated with the genetics of the traits (top 10% MAGMA gene-level genetic association) were enriched in biological functions using the *topGO* R package [97]. 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 S11**). We used the EWCE R package [11] (see code availability below) 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 [51] 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 [98], 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 [99]. Differential expression analysis was performed using limma [100] taking age and gender as covariates. The Lesnick dataset [50] 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 [54] 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 [11] analysis. We computed a new specificity matrix based on the substantia nigra or ventral midbrain cells using the EWCE [11]. The EWCE analysis was performed on the 500 most up or down regulated genes using 10,000 bootstrapping replicates.

Code availability

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

Data availability

All single-cell expression data are publicly available. Most summary statistics used in this study are publicly available. The migraine GWAS can be obtained by contacting the authors [77]. The Parkinson's disease summary statistics from 23andMe can be obtained under an agreement that protects the privacy of 23andMe research participants (https://research.23andme.com/collaborate/#publication).

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). EA was supported by the Swedish Research Council (VR 2016-01526), Swedish Foundation for Strategic Research (SLA SB16-0065), Karolinska Institutet (SFO Strat. Regen., Senior grant 2018), Cancerfonden (CAN 2016/572), Hjärnfonden (FO2017-0059) and Chen Zuckeberg Initiative: Neurodegeneration Challenge Network (2018-191929-5022). 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 correlations across traits

Table S2: Association P-value between GTEx tissues and all traits

Table S3: Association P-value between cell types from the entire mouse nervous system and all traits (Zeisel et al. 2018)

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 cell types from 9 mouse brain regions and all traits (Saunders et al. 2018)

Table S7: Association P-value between cell types from 5 mouse brain regions and all traits (Skene et al. 2018)

Table S8: Summary statistics for 24 cell types from 5 mouse brain regions (Skene et al. 2018)

Table S9: Association P-value between cell types from 2 human brain regions and all traits (Habib et al. 2017)

Table S10: Association P-value between cell types from 3 human brain regions and all traits (Lake et al. 2018)

Table S11: Association of Alzheimer's disease differentially expressed genes in 6 different cell types with Alzheimer's common variant genetics.

Table S12: Rare and familial genetic mutations associated with Parkinsonism

Table S13: Cell type enrichment results using rare and familial genetic mutations associated with Parkinsonism

Table S14: Summary statistics for cell types (Zeisel et al. 2018)

Table S15: Spearman rank correlations of the genome-wide specificity metrics for tissues of the GTEx dataset

Table S16: Spearman rank correlations of the genome-wide specificity metric for the cell types from Zeisel et al. 2018

Table S17: Top 10% most specific genes per tissue for the GTEx dataset

Table S18: Top 10% most specific genes per cell type for the Zeisel dataset

Figures

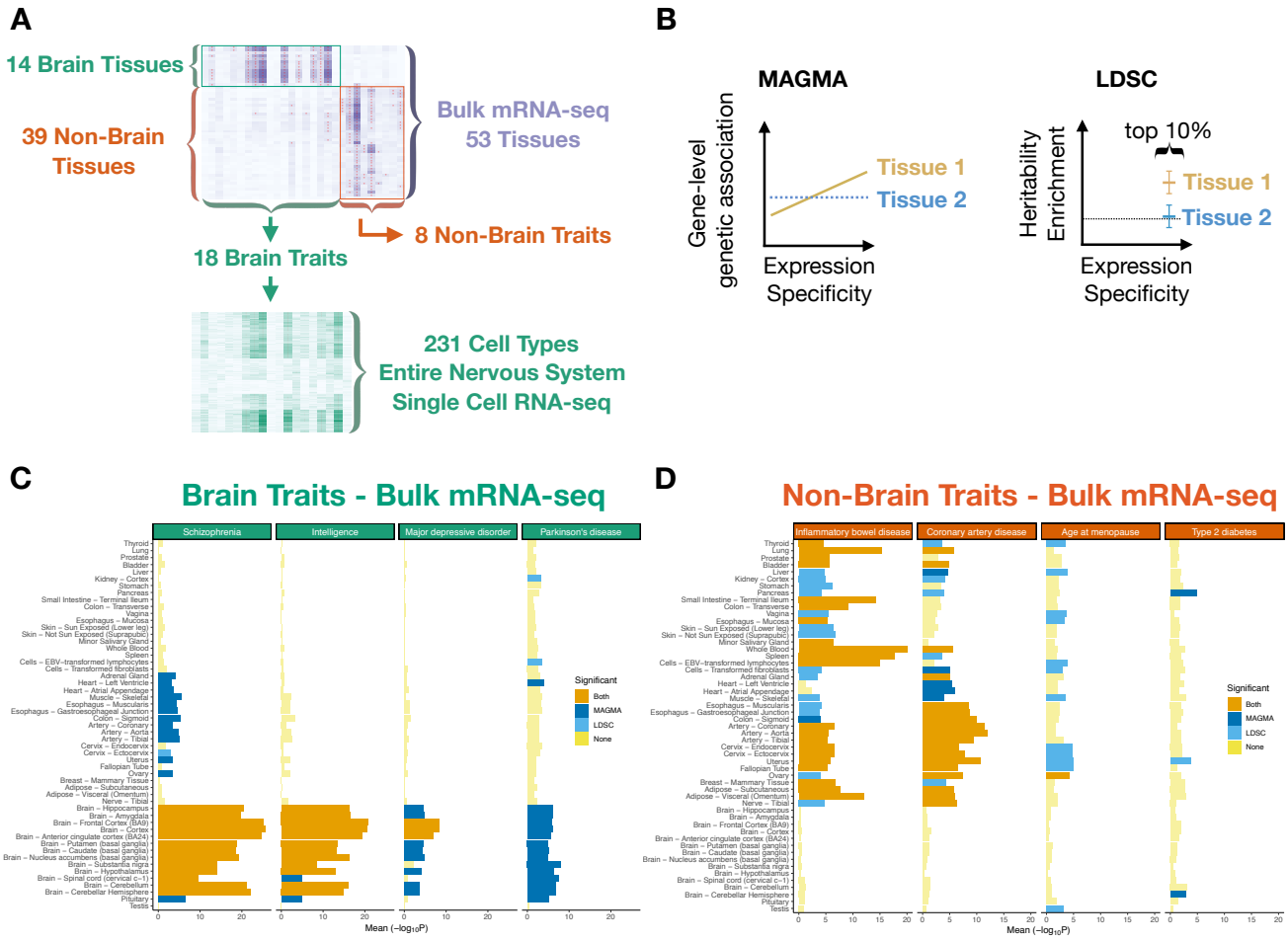

Figure 1: Study design and tissue-level associations. Heat map of trait – tissue/cell types associations ($-\log_{10}P$) for the selected traits. (**A**) Trait – tissue/cell types associations were performed using MAGMA (testing for a positive correlation between gene expression specificity and gene-level 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. Tissue label in panel C and D might require zooming on the pdf.

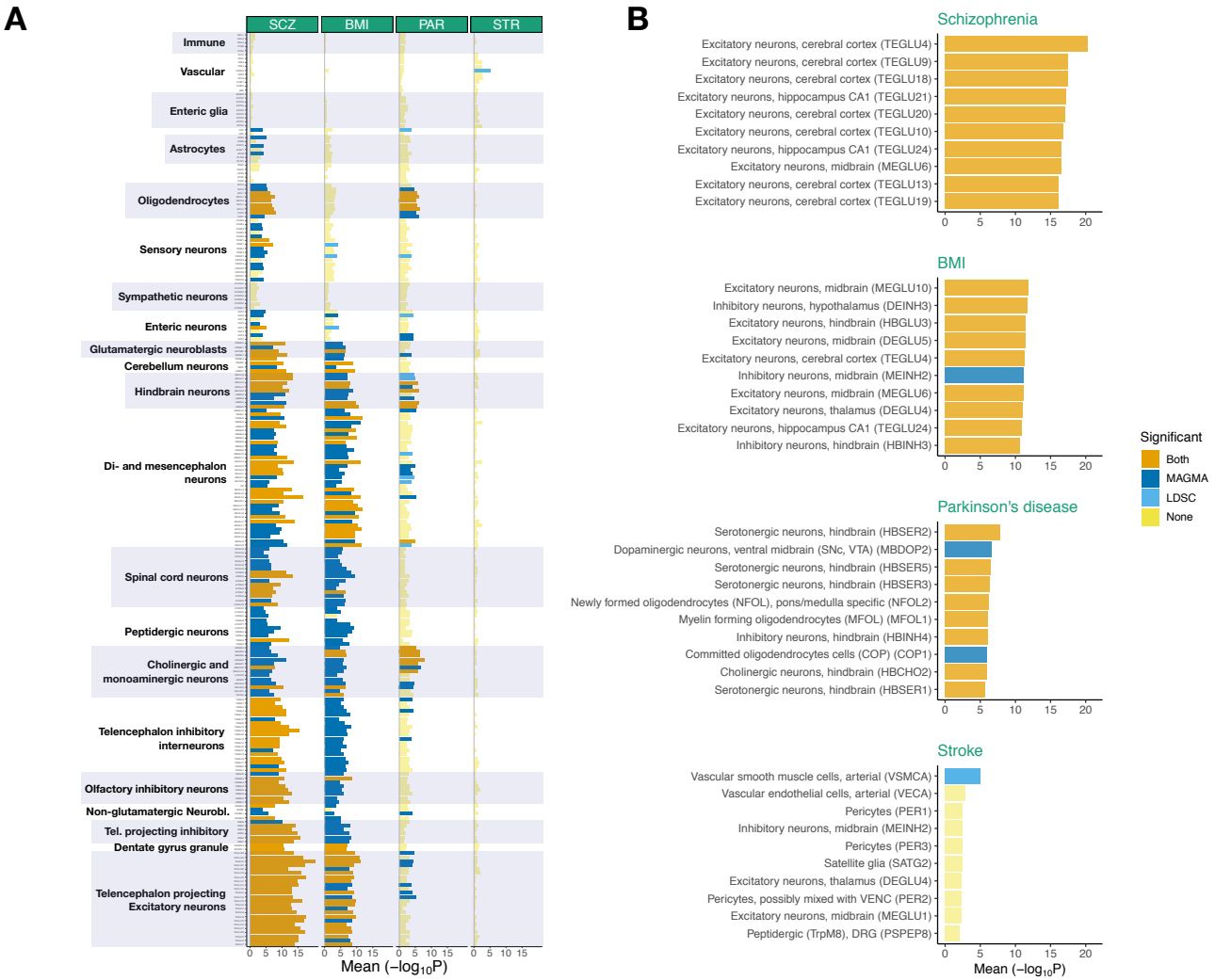

Figure 2: Association of brain related traits with cell types from the entire nervous system. Associations between all cell types and the following traits are shown: schizophrenia (SCZ), body mass index (BMI), Parkinson's disease (PAR) and stroke (STR). (**A**) The ten most associated cell types are shown for the 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). Cluster label corresponding to Zeisel et al. [31] are shown on the left of panel A (requires zooming on the pdf).

Figure 3: 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=0.05/24).

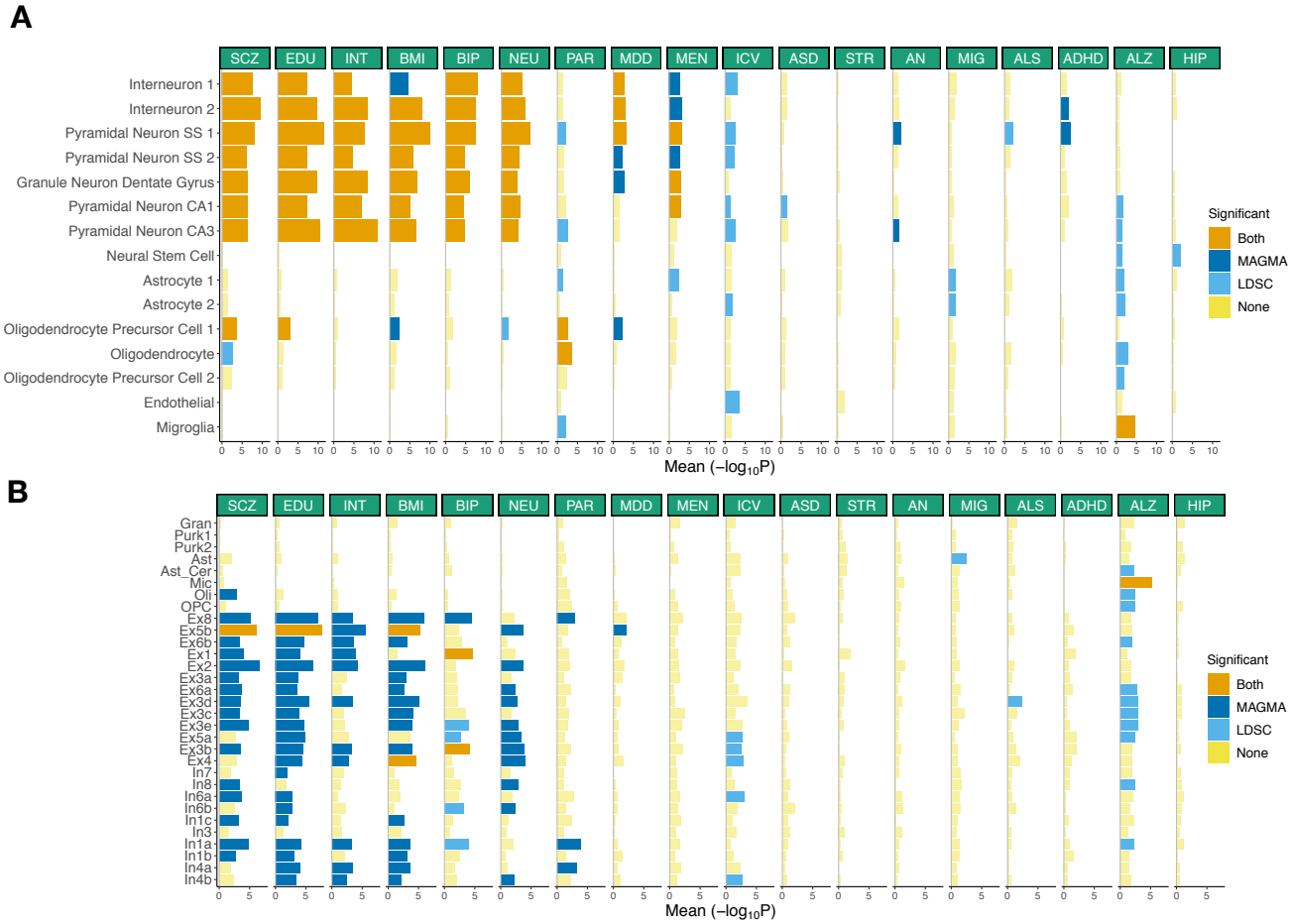

Figure 4: Human replication of cell type – trait associations. Cell type - trait associations for 15 cell types (derived from single-nuclei RNA-seq) from 2 different brain regions (cortex, hippocampus). (**A**) Cell type - trait associations for 31 cell types (derived from single-nuclei RNA-seq) from 3 different brain regions (frontal cortex, visual cortex and cerebellum). (**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=0.05/15 for panel A and P=0.05/31 for panel B).

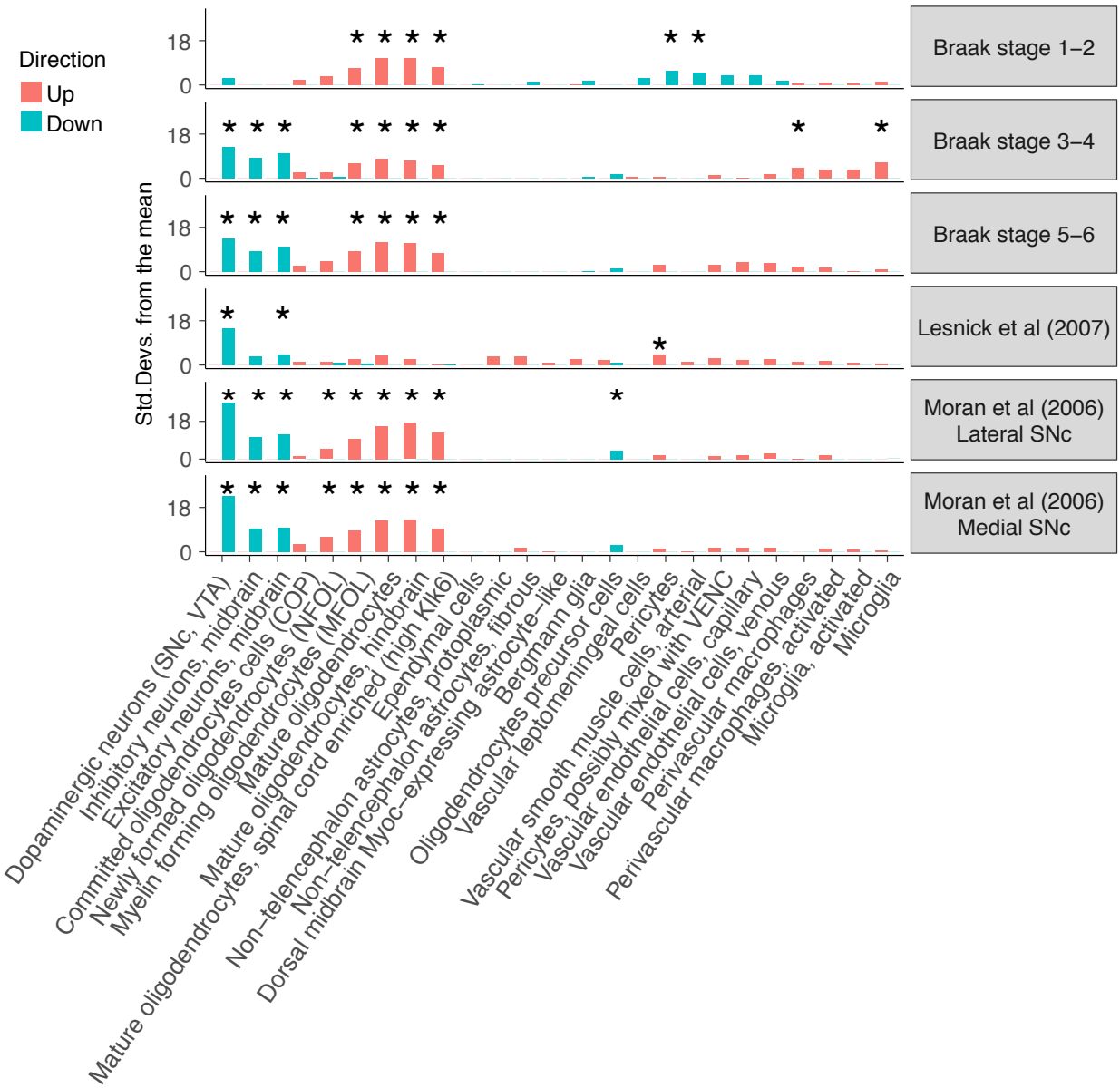

Figure 5: Enrichment of Parkinson's disease differentially expressed genes in cell types from the substantia nigra. Enrichment of the 500 most up/down regulated genes (Braak stage 0 vs Braak stage 1—2, 3—4 and 5—6, as well as cases vs controls) in postmortem human substantia nigra gene expression samples. The enrichments were obtained using EWCE[11]. A star shows significant enrichments after multiple testing correction ($P<0.05/(25*4)$).

Supplementary Figures

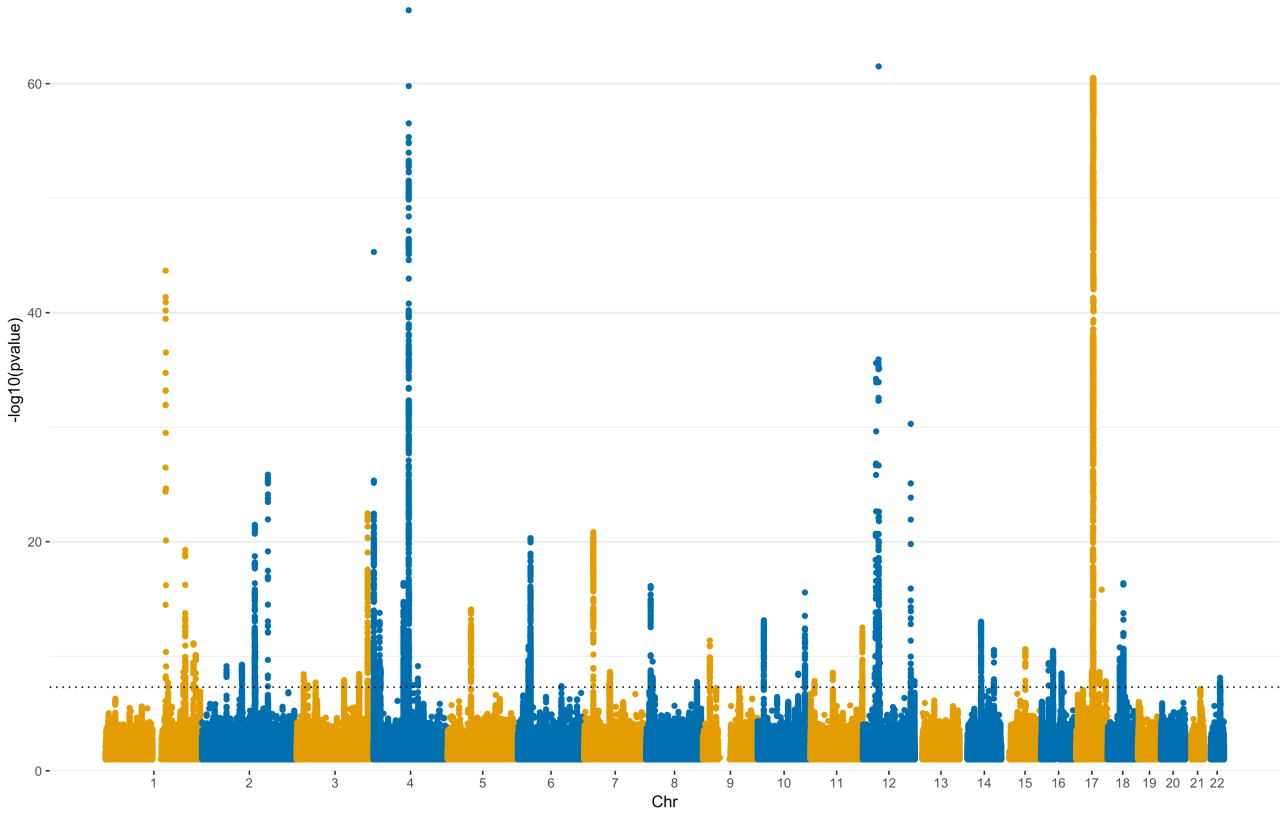

Figure S1: Manhattan plot of Parkinson's disease meta-analysis. The black dotted line represents the genome-wide significance threshold ($5x10^{-8}$).

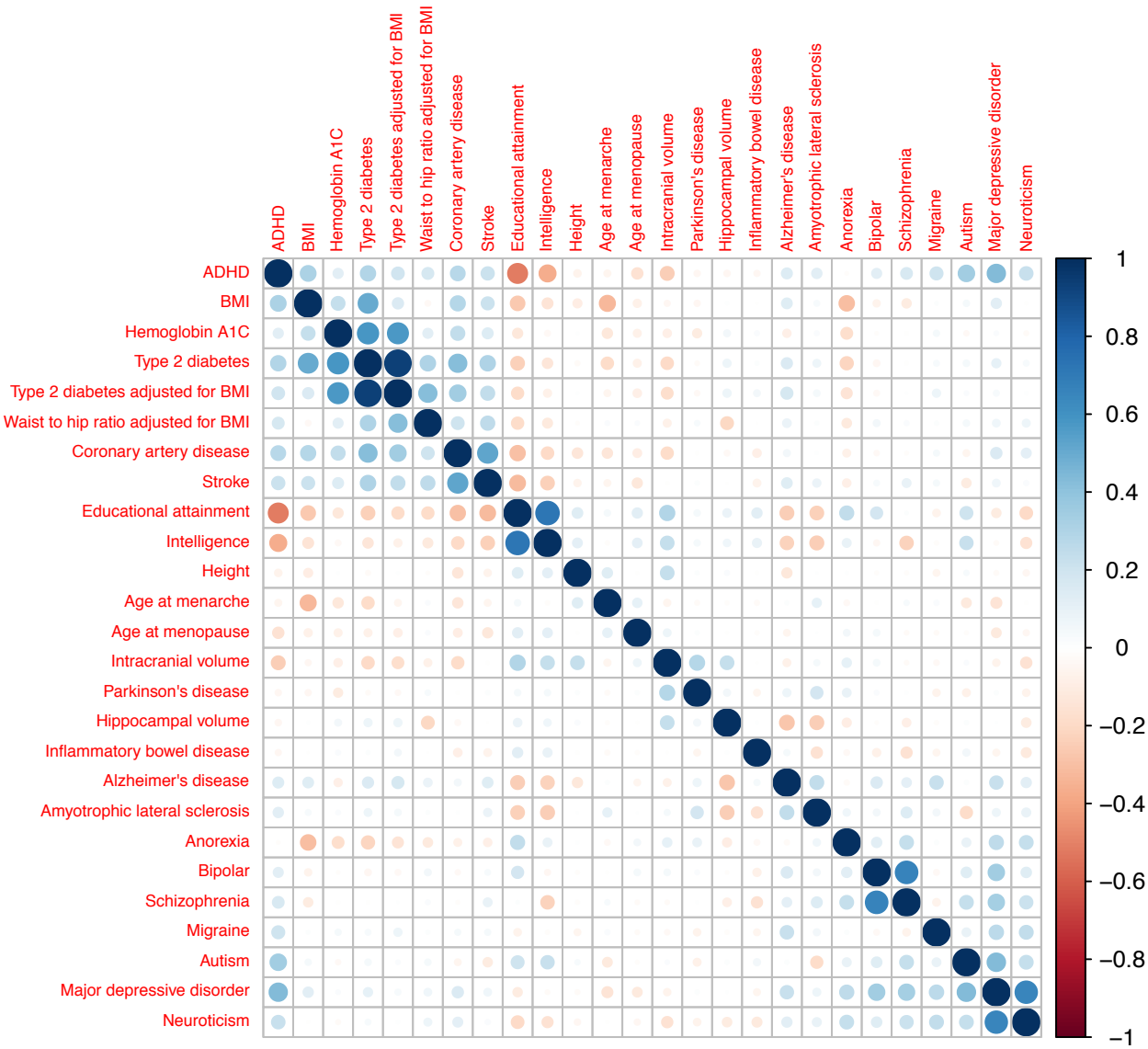

Figure S2: Genetic correlation across traits. The genetic correlation across traits were computed using LDSC[101]. Traits are ordered based on hierarchical clustering.

Figure S3: Tissue – trait associations for all 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). Tissues are ordered based on hierarchical clustering of the gene expression specificities.

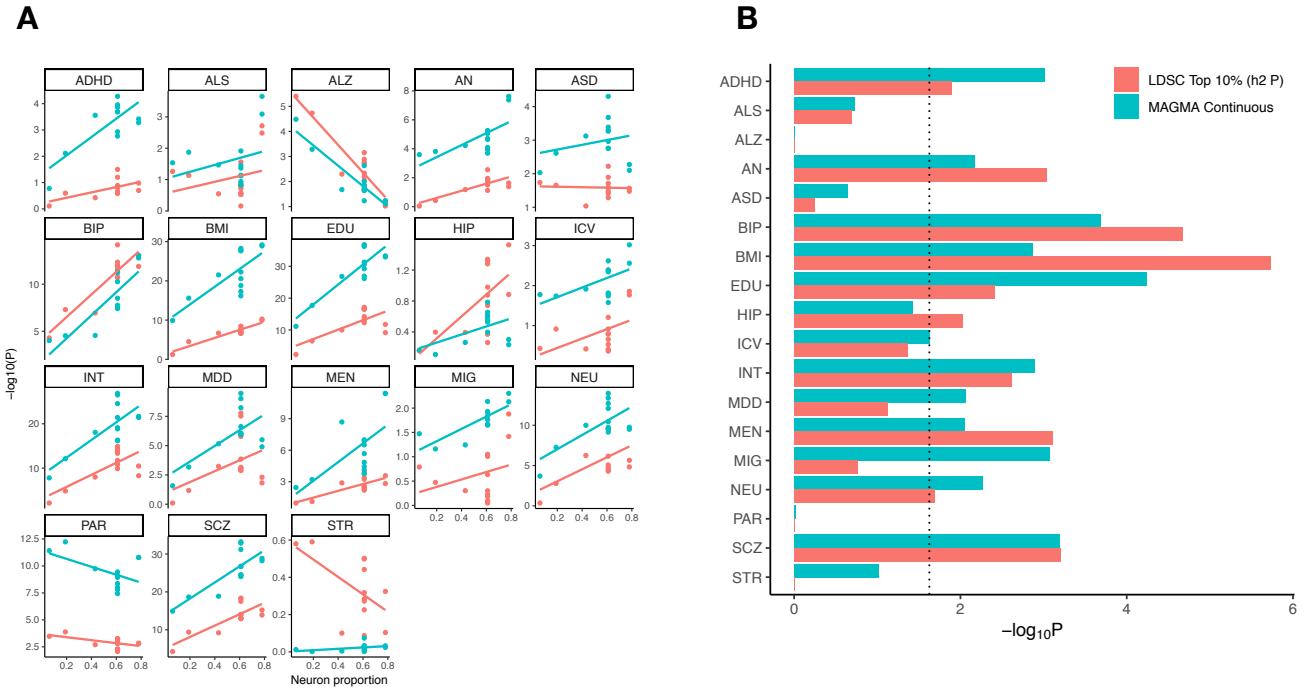

Figure S4: Correlation between the proportion of neurons in diverse brain regions and strength of tissue – trait associations. Correlation between the proportion of neurons in diverse brain regions and strength of tissue – trait associations ($-\log_{10}P$) are shown for LDSC and MAGMA (**A**). The strength 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).

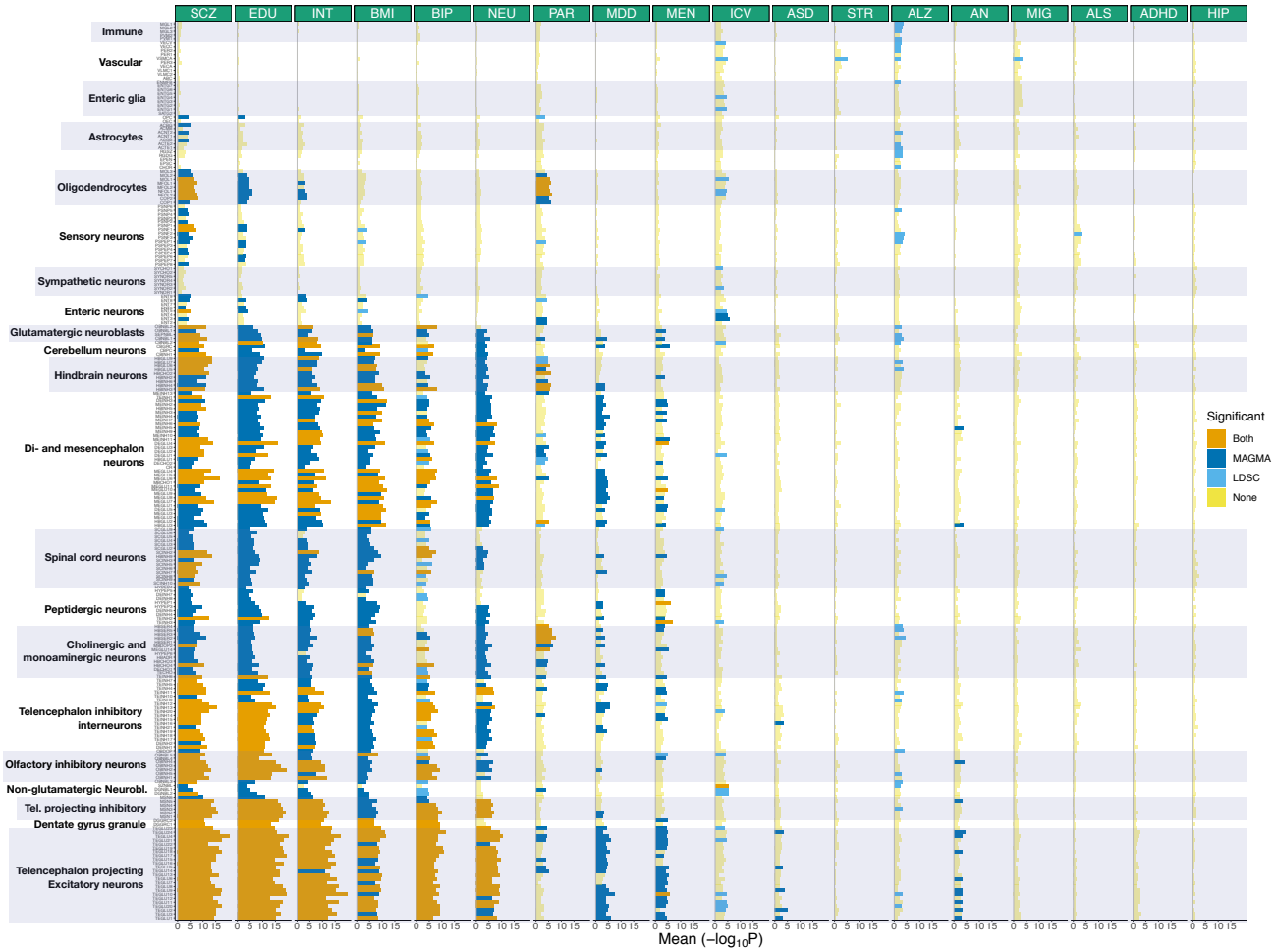

Figure S5: Associations of brain related traits with cell types from the entire nervous system. Association between cell types and the following traits are shown: SCZ (schizophrenia), EDU (educational attainment), INT (intelligence), BMI (body mass index), BIP (bipolar disorder), NEU (neuroticism), PAR (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). 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).

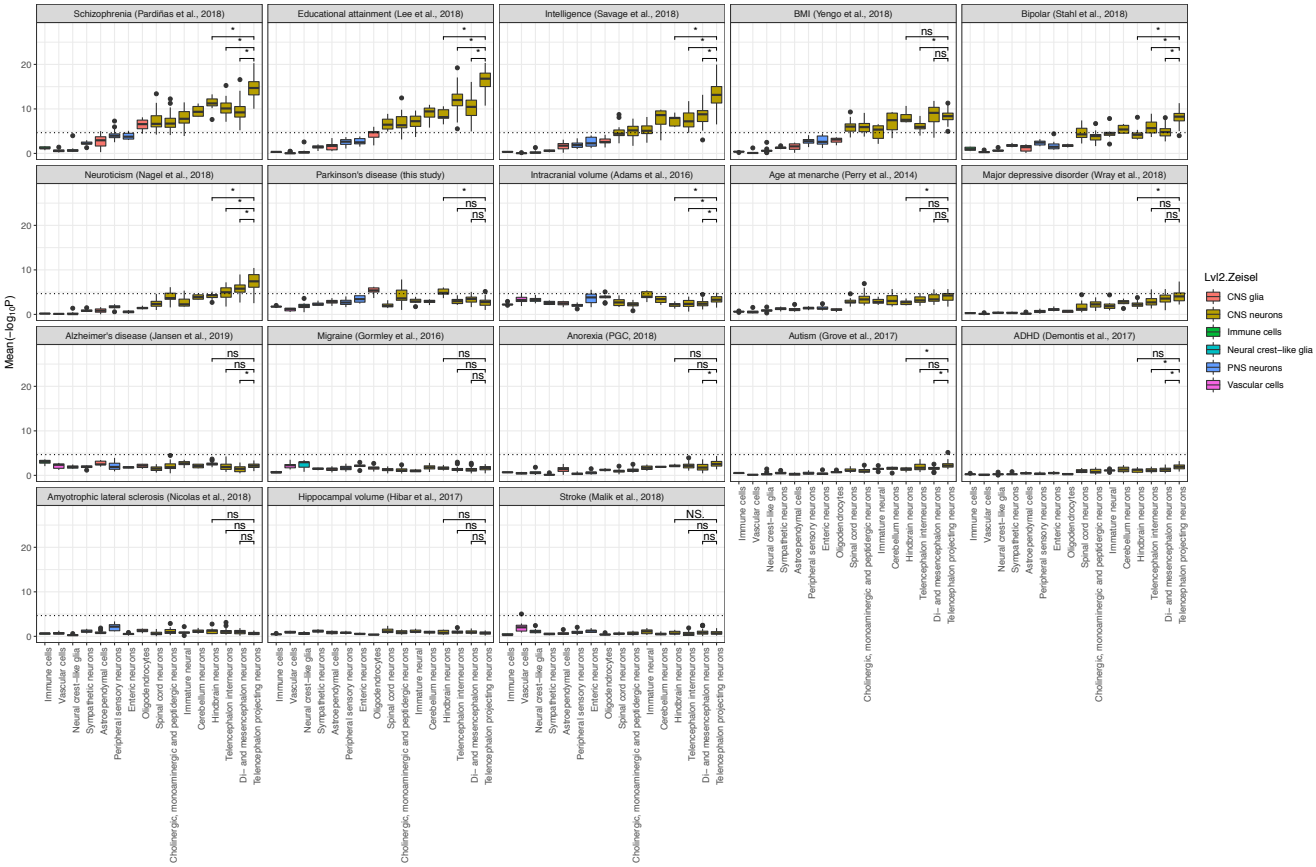

Figure S6: Cell type – trait associations grouped by broad categories of cell types. Boxplot of the mean association strength ($-\log_{10}P_{MAGMA}$, $-\log_{10}P_{LDSC}$) between gene expression specificity and brain related traits by broad categories of cell types (level 3 from Zeisel et al. 2018[31]). The black dotted horizontal bar represents our significance threshold: 0.005/231 (231 cell types tested). Statistical difference in the mean association strength are indicated by * ($P<0.05/(3*18)$, Wilcoxon test) (3 tests for 18 traits).

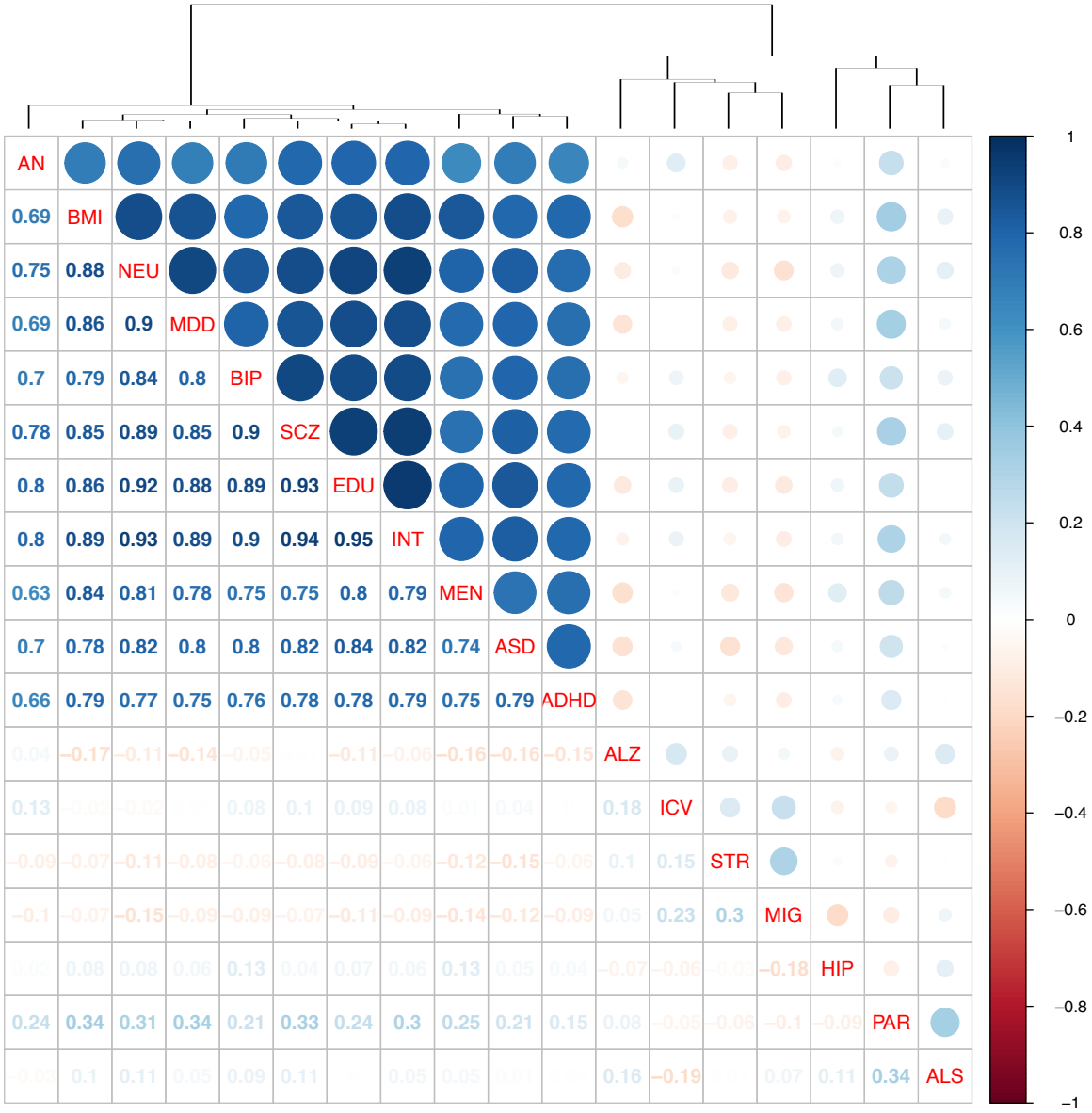

Figure S7: Correlation in cell type associations across traits. The Spearman rank correlations between the cell types associations across traits (-$\log_{10}$P) are shown. SCZ (schizophrenia), EDU (educational attainment), INT (intelligence), BMI (body mass index), BIP (bipolar disorder), NEU (neuroticism), 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).

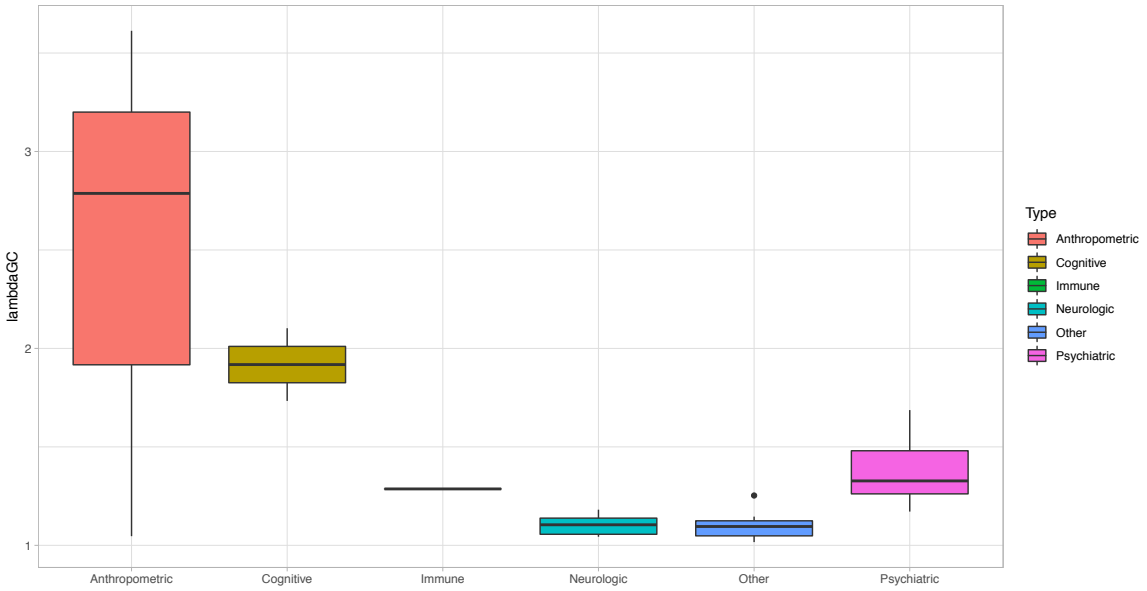

Figure S8: GWAS signal to noise ratio (λ_{GC}) by category of GWAS trait. Boxplot of the λ_{GC} of the different GWAS by category of trait. λ_{GC} was estimated using LDSC for each GWAS.

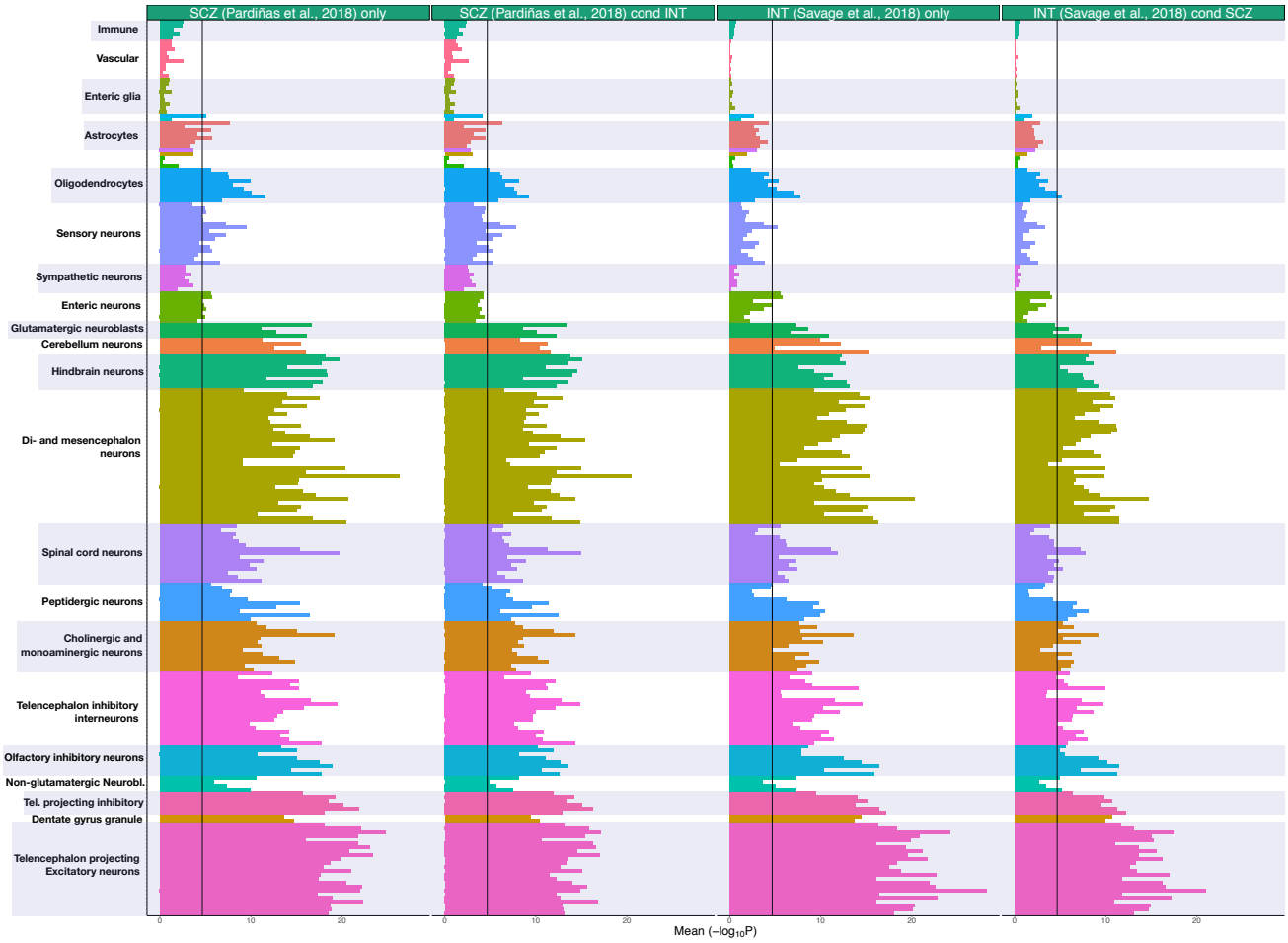

Figure S9: Associations of cell types with schizophrenia/intelligence conditioning on gene-level genetic association of intelligence/schizophrenia. MAGMA association strength for each cell type before and after conditioning on gene-level genetic association for another trait. The black bar represents the significance threshold (P=0.005/231). SCZ (schizophrenia), INT (intelligence). The different color distinguishes high-level grouping of the cell types (ex. immune cells).

Figure S10: Associations of cell types with schizophrenia/educational attainment conditioning on gene-level genetic association of educational attainment/schizophrenia. MAGMA association strength for each cell type before and after conditioning on gene-level genetic association for another trait. The black bar represents the significance threshold (P=0.005/231). SCZ (schizophrenia), EDU (educational attainment). The different color distinguishes high-level grouping of the cell types (ex. immune cells).

Figure S11: 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 ($-\log_{10}P$). The label of each cell type that is being conditioned on is shown at the bottom (requires zooming on the pdf).

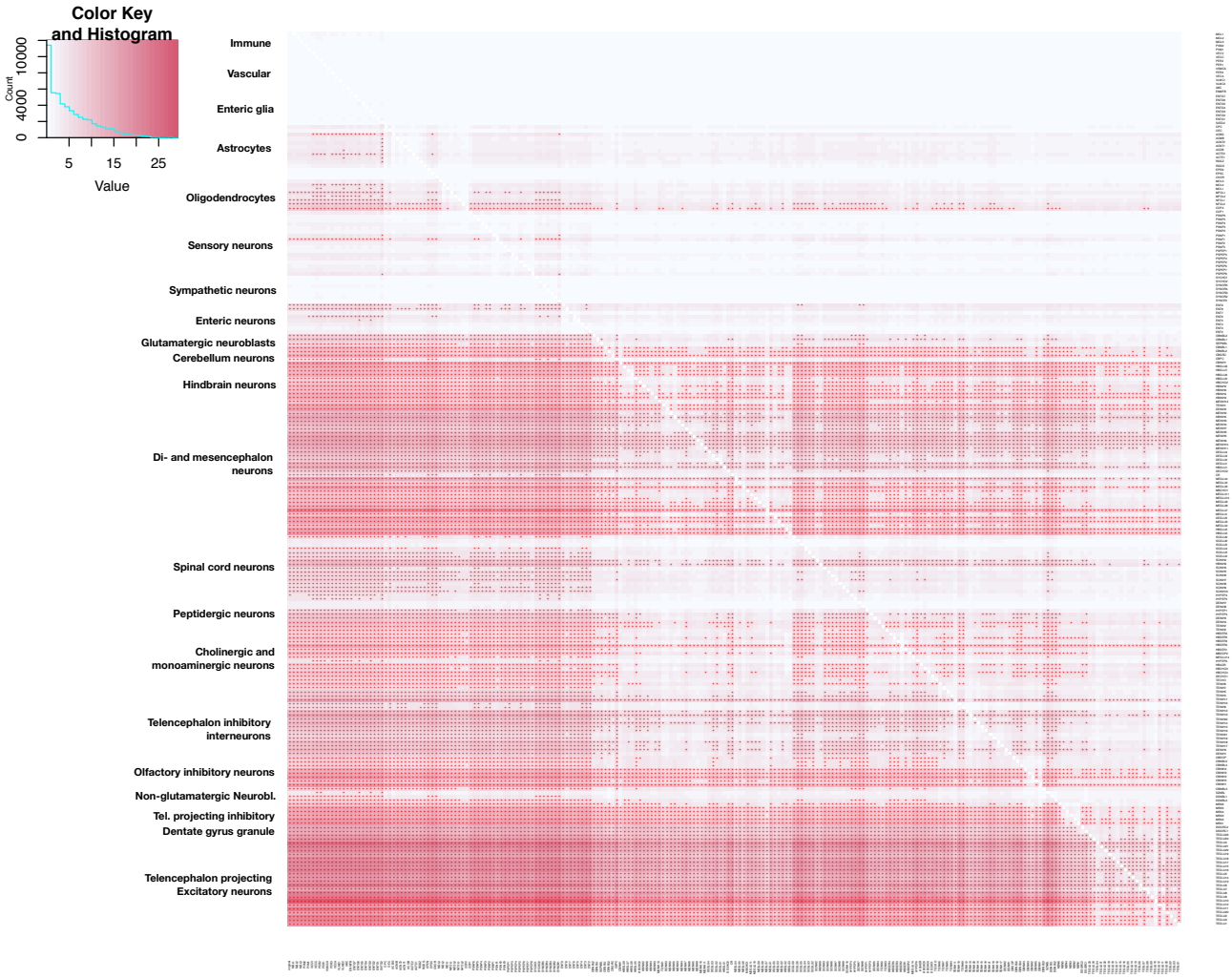

Figure S12: Univariate conditional analysis for intelligence. MAGMA association strength for each cell type before (1st column) and after conditioning on the 231 cell types from the nervous system. Red stars indicate significance (P<0.005/231). The histogram indicates the scale ($-\log_{10}P$). The label of each cell type that is being conditioned on is shown at the bottom (requires zooming on the pdf).

Figure S13: 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 ($-\log_{10}P$). The label of each cell type that is being conditioned on is shown at the bottom (requires zooming on the pdf).

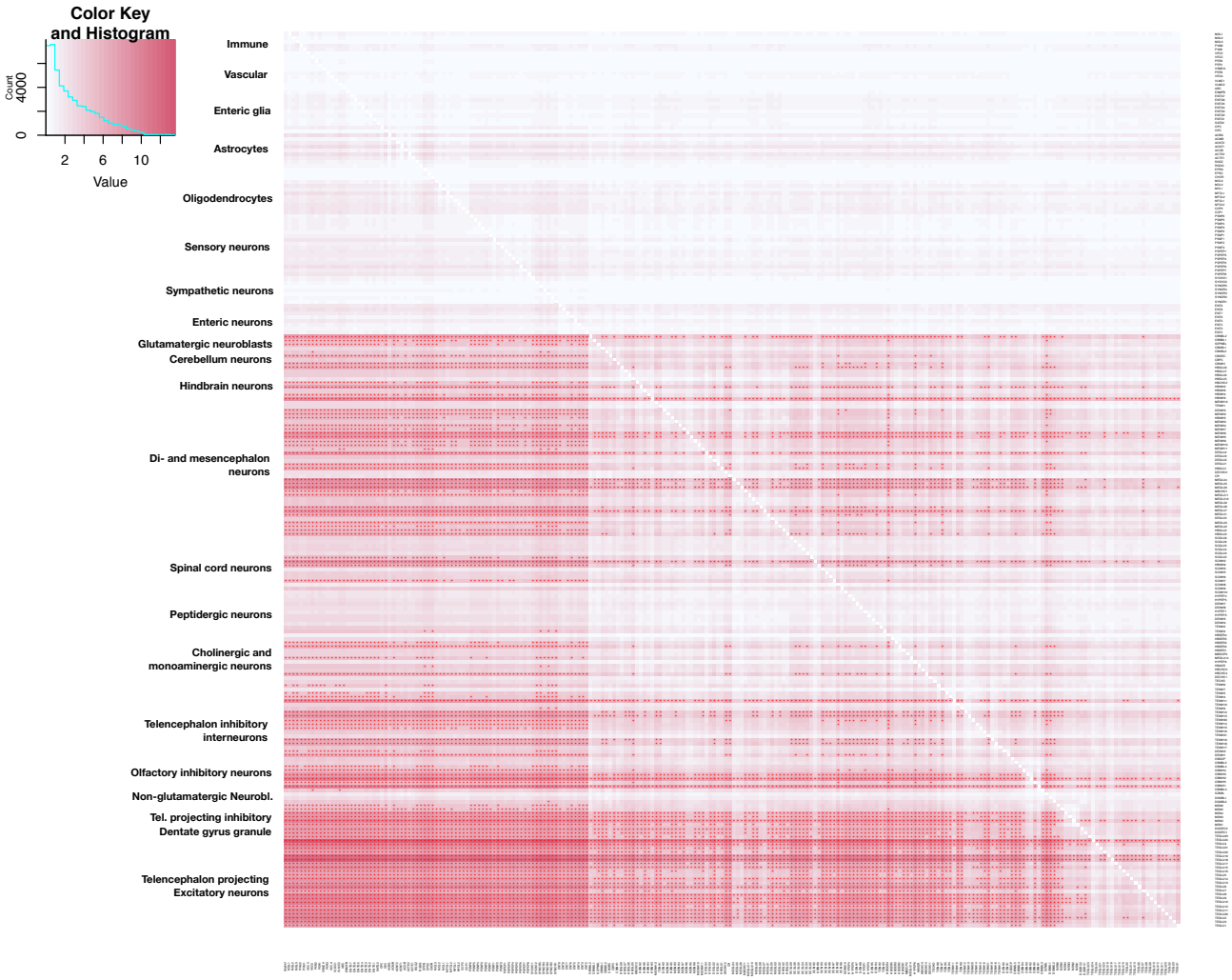

Figure S14: Univariate conditional analysis for bipolar disorder. MAGMA association strength for each cell type before (1st column) and after conditioning on the 231 cell types from the nervous system. A red star is shown if the association is significant (P<0.005/231). The histogram indicates the scale ($-\log_{10}P$). The label of each cell type that is being conditioned on is shown at the bottom (requires zooming on the pdf).

Figure S15: 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 ($-\log_{10}P$). The label of each cell type that is being conditioned on is shown at the bottom (requires zooming on the pdf).

Figure S16: 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 ($-\log_{10}P$). The label of each cell type that is being conditioned on is shown at the bottom (requires zooming on the pdf).

Figure S17: 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 ($-\log_{10}P$). The label of each cell type that is being conditioned on is shown at the bottom (requires zooming on the pdf).

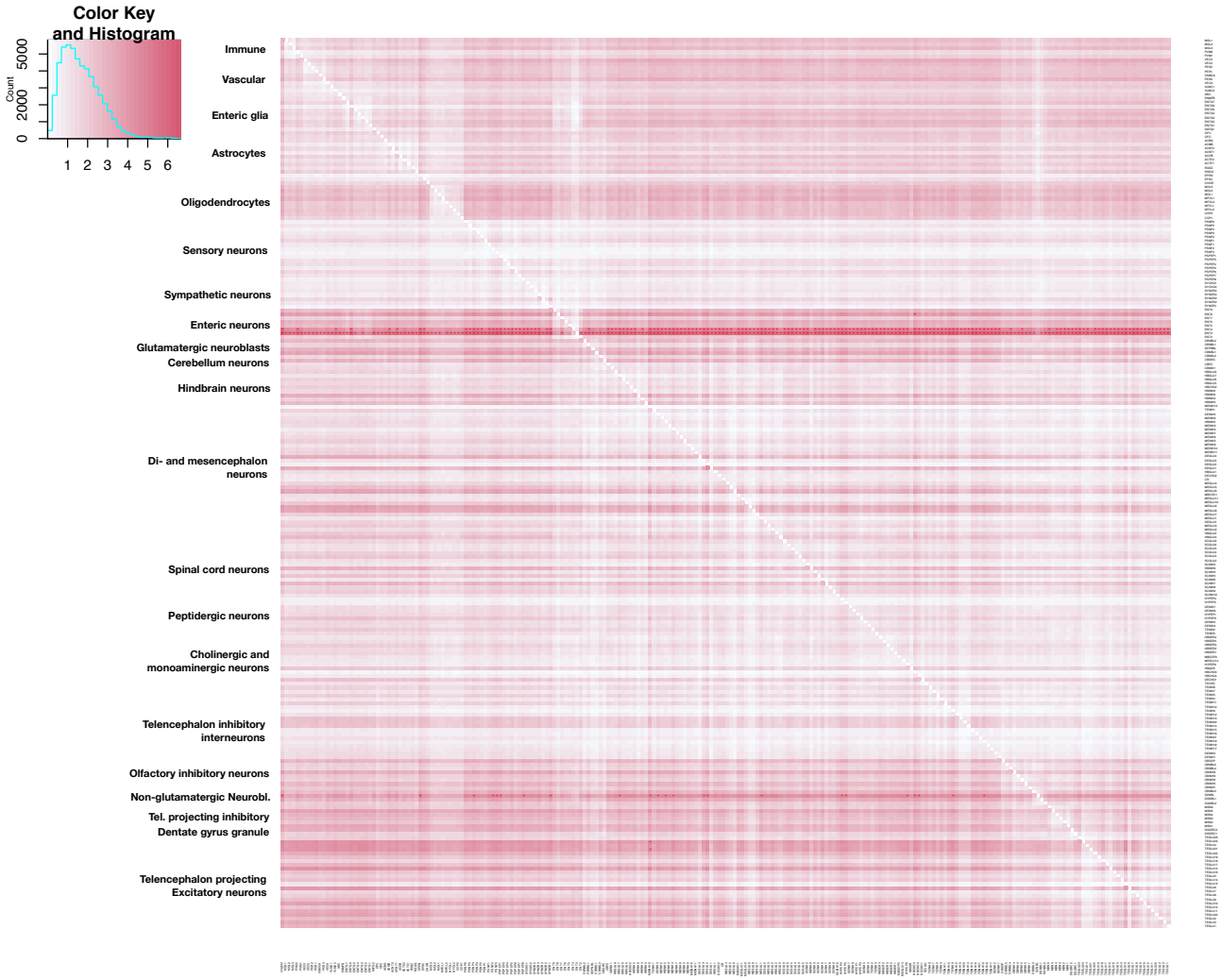

Figure S18: Univariate conditional analysis for intracranial volume. MAGMA association strength for each cell type before (1st column) and after conditioning on the 231 cell types from the nervous system. A red star is shown if the association is significant (P<0.005/231). The histogram indicates the scale ($-\log_{10}P$). The label of each cell type that is being conditioned on is shown at the bottom (requires zooming on the pdf).

Figure S19: 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 ($-\log_{10}P$). The label of each cell type that is being conditioned on is shown at the bottom (requires zooming on the pdf).

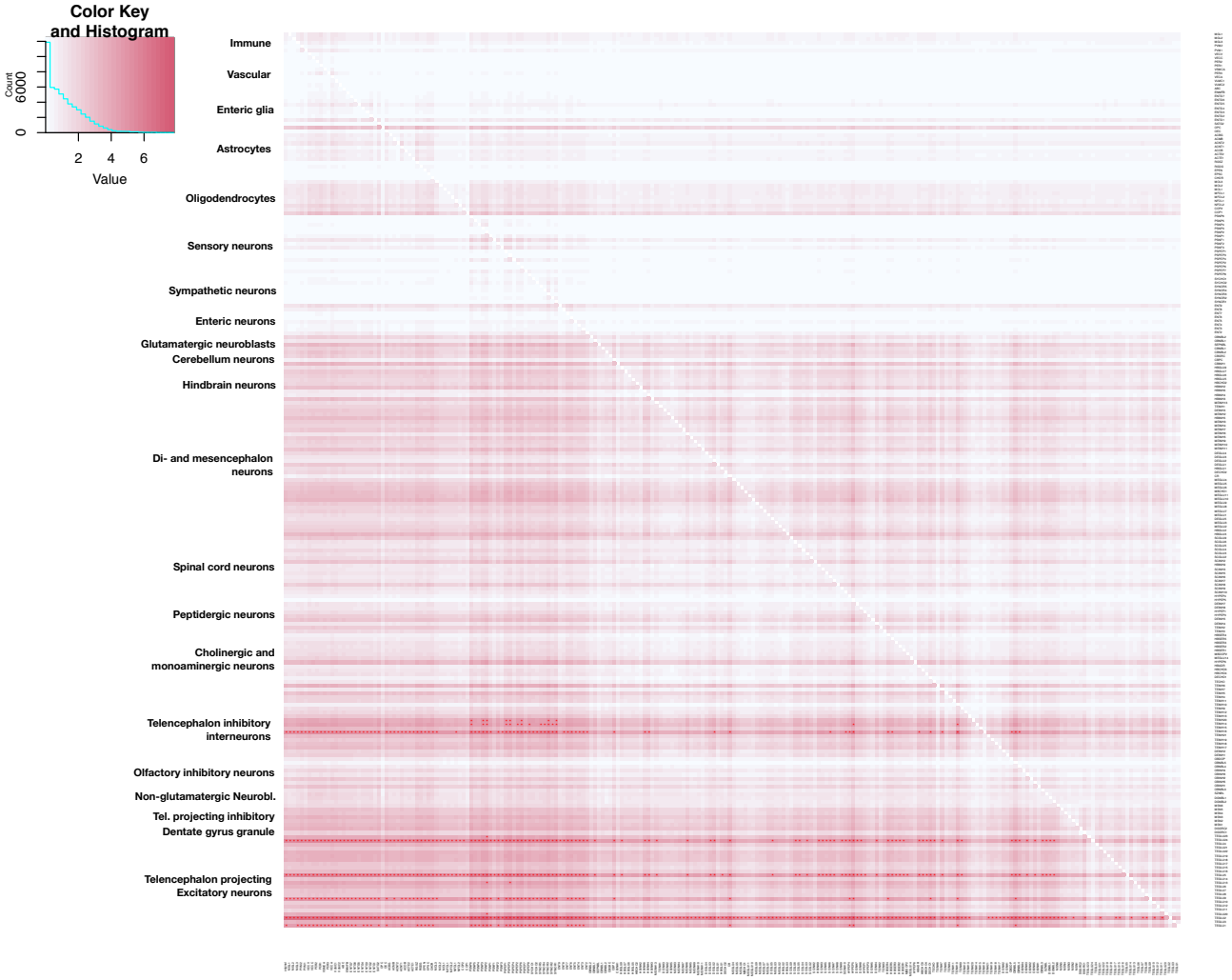

Figure S20: Univariate conditional analysis for autism. MAGMA association strength for each cell type before (1st column) and after conditioning on the 231 cell types from the nervous system. A red star is shown if the association is significant (P<0.005/231). The histogram indicates the scale ($-\log_{10}P$). The label of each cell type that is being conditioned on is shown at the bottom (requires zooming on the pdf).

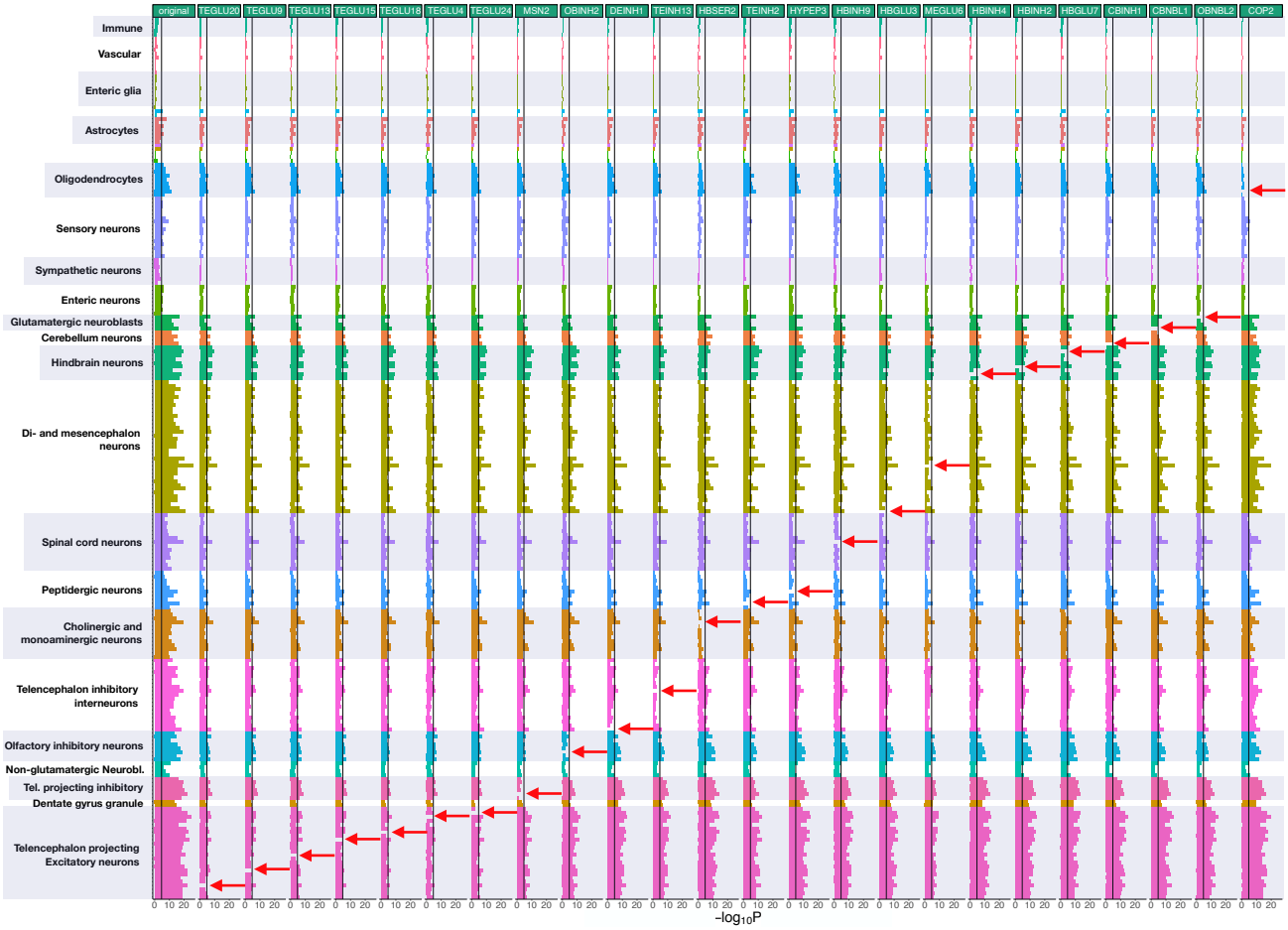

Figure S21: Conditional cell type associations for schizophrenia. The MAGMA cell types association results are shown before (first column) and after conditioning on gene expression specificity from different cell types. The red arrow indicates the cell types that are being conditioned on. The black bar represents the significance threshold (P=0.005/231). The different color represents the different classes of cell types (i.e. immune cells, vascular cells, etc.). The conditional analysis was performed by conditioning on the most significant cell type per class, followed by conditional analysis on individual cell types that remained significant within the class.

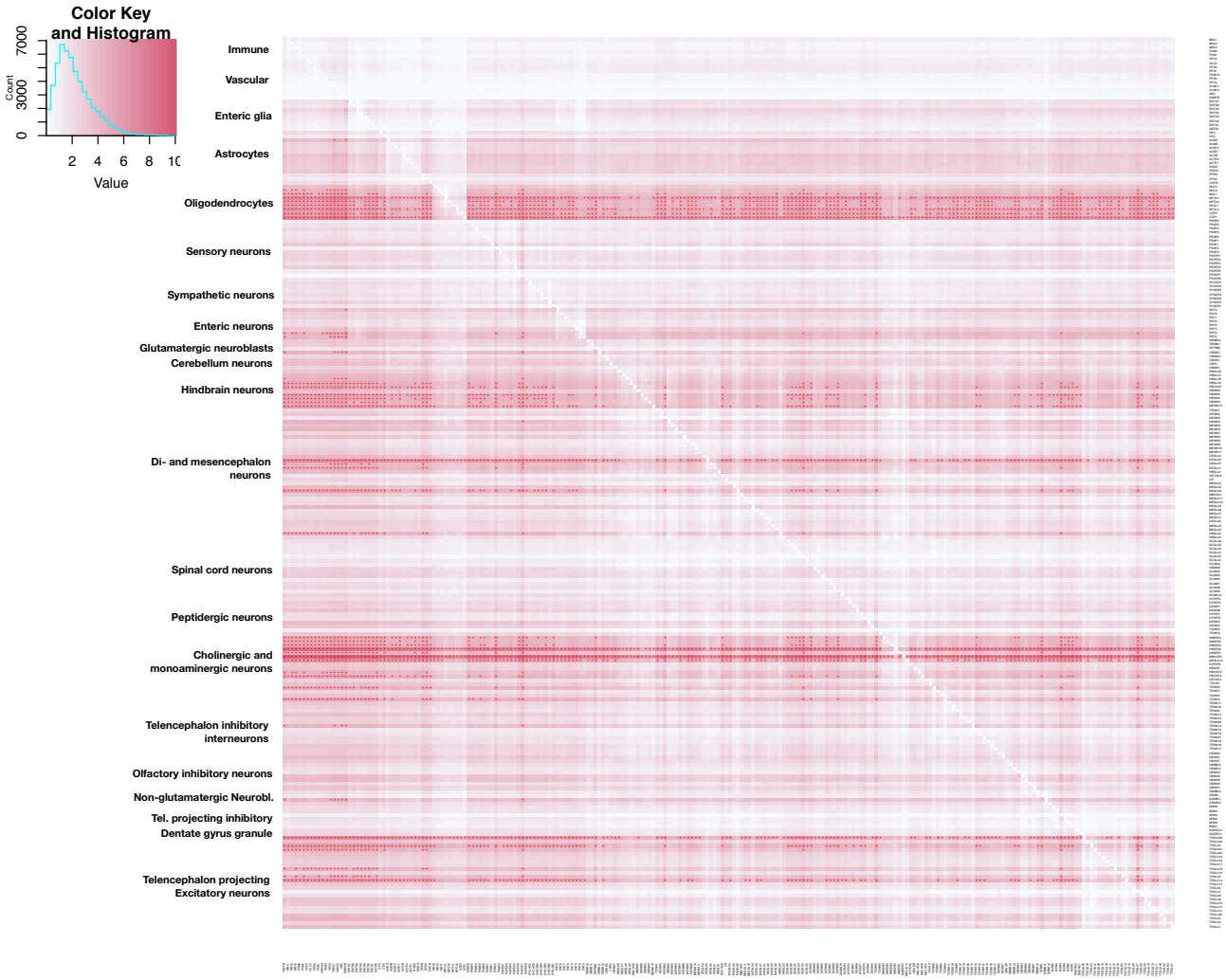

Figure S22: Univariate conditional analysis for Parkinson's disease. MAGMA association strength for each cell type before (1st column) and after conditioning on the 231 cell types from the nervous system. A red star is shown if the association is significant (P<0.005/231). The histogram indicates the scale ($-\log_{10}P$). The label of each cell type that is being conditioned on is shown at the bottom (requires zooming on the pdf).

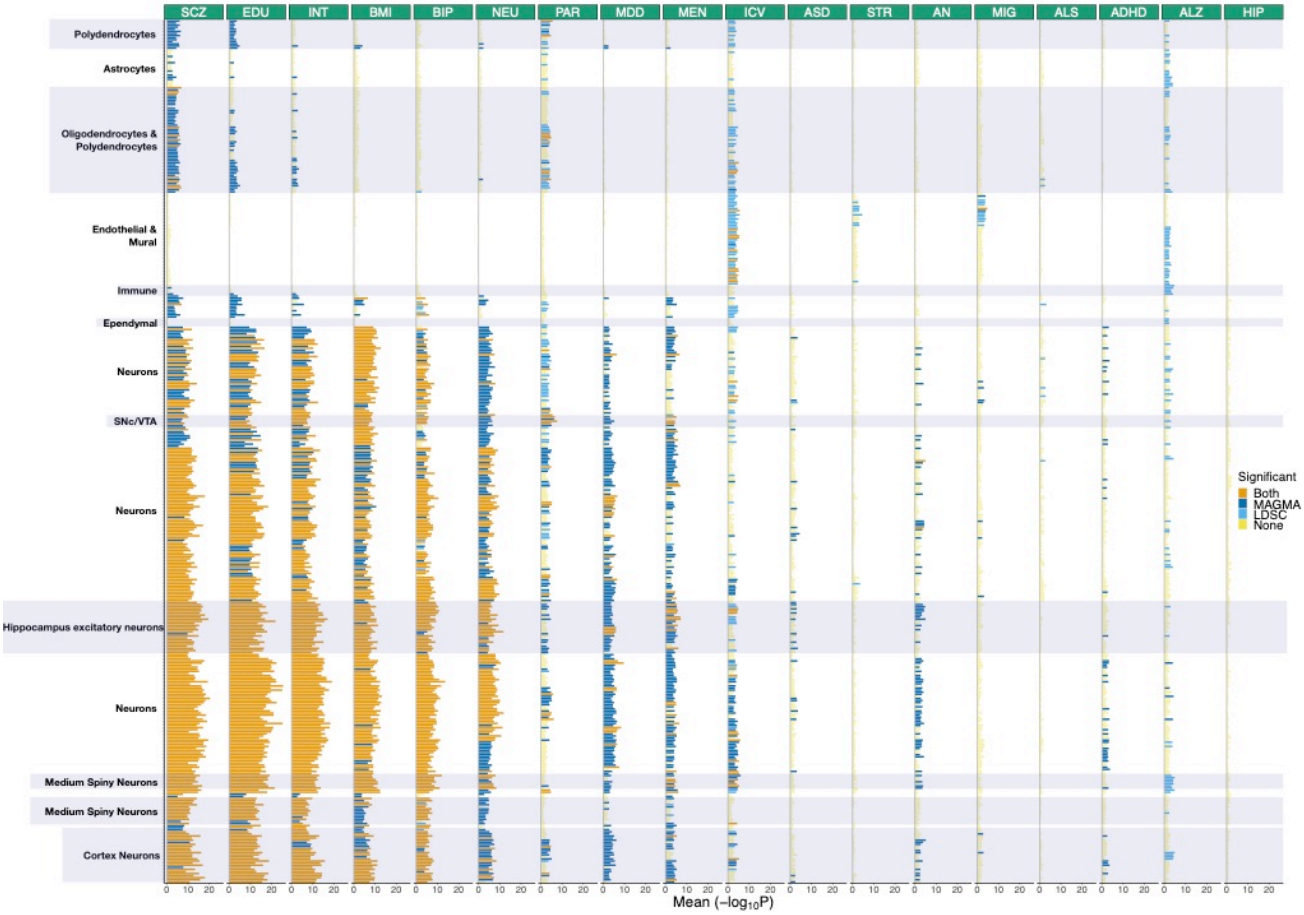

Figure S23: Replication of cell type – trait associations in 414 cell types from 9 different brain regions. The mean strength of association ($-\log_{10}P$) of MAGMA and LDSC is shown and the bar color indicates whether the cell type is significantly associated with both methods, one method or none (significance threshold: P=0.05/414). 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).

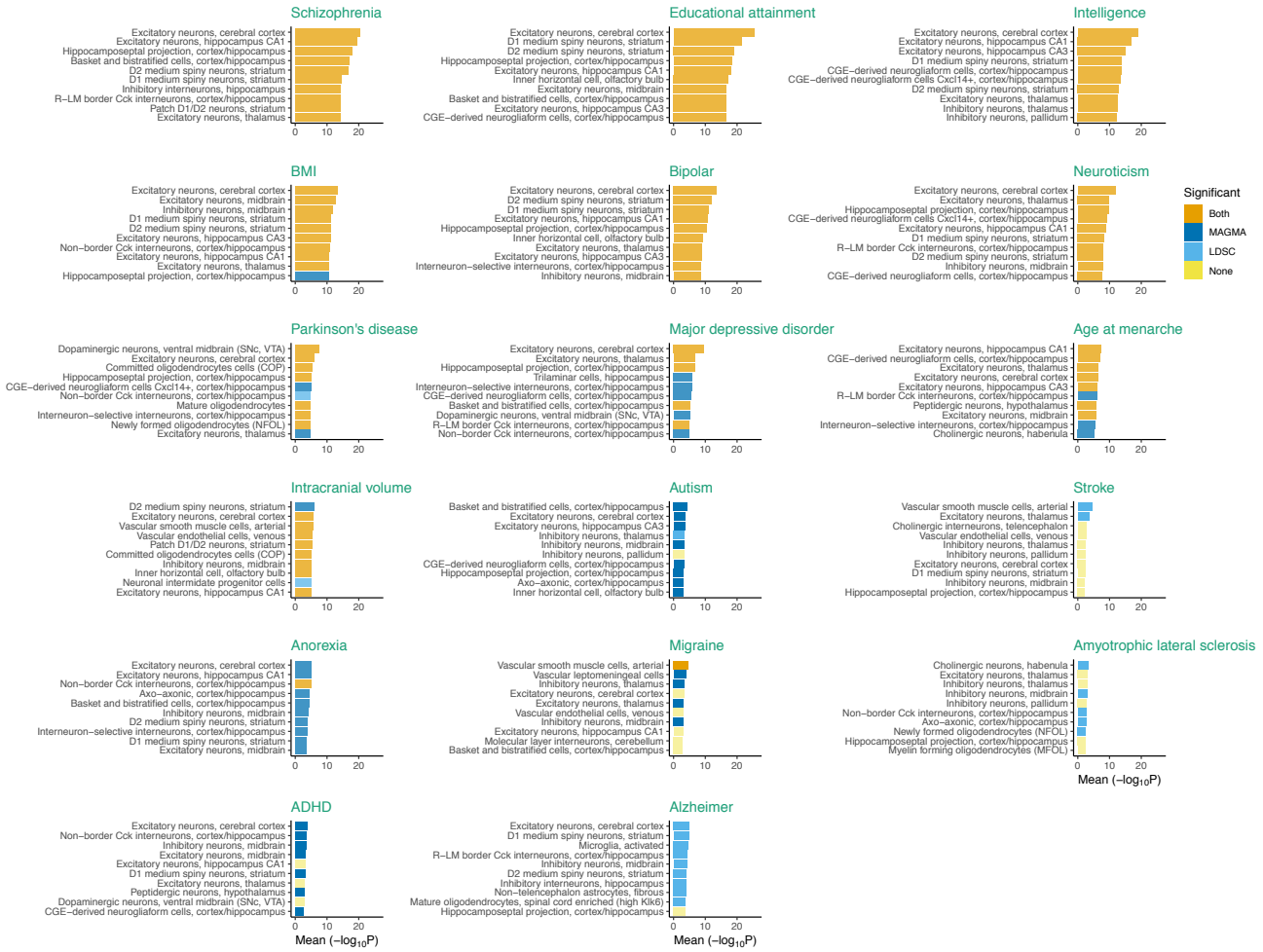

Figure S24: Top associated cell types with brain related traits among 414 cell types from 9 different brain regions. The mean strength of association (-$\log_{10}$P) of MAGMA and LDSC is shown for the ten top cell types for each trait. The bar color indicates whether the cell type is significantly associated with both methods, one method or none (significance threshold=0.05/414). 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.

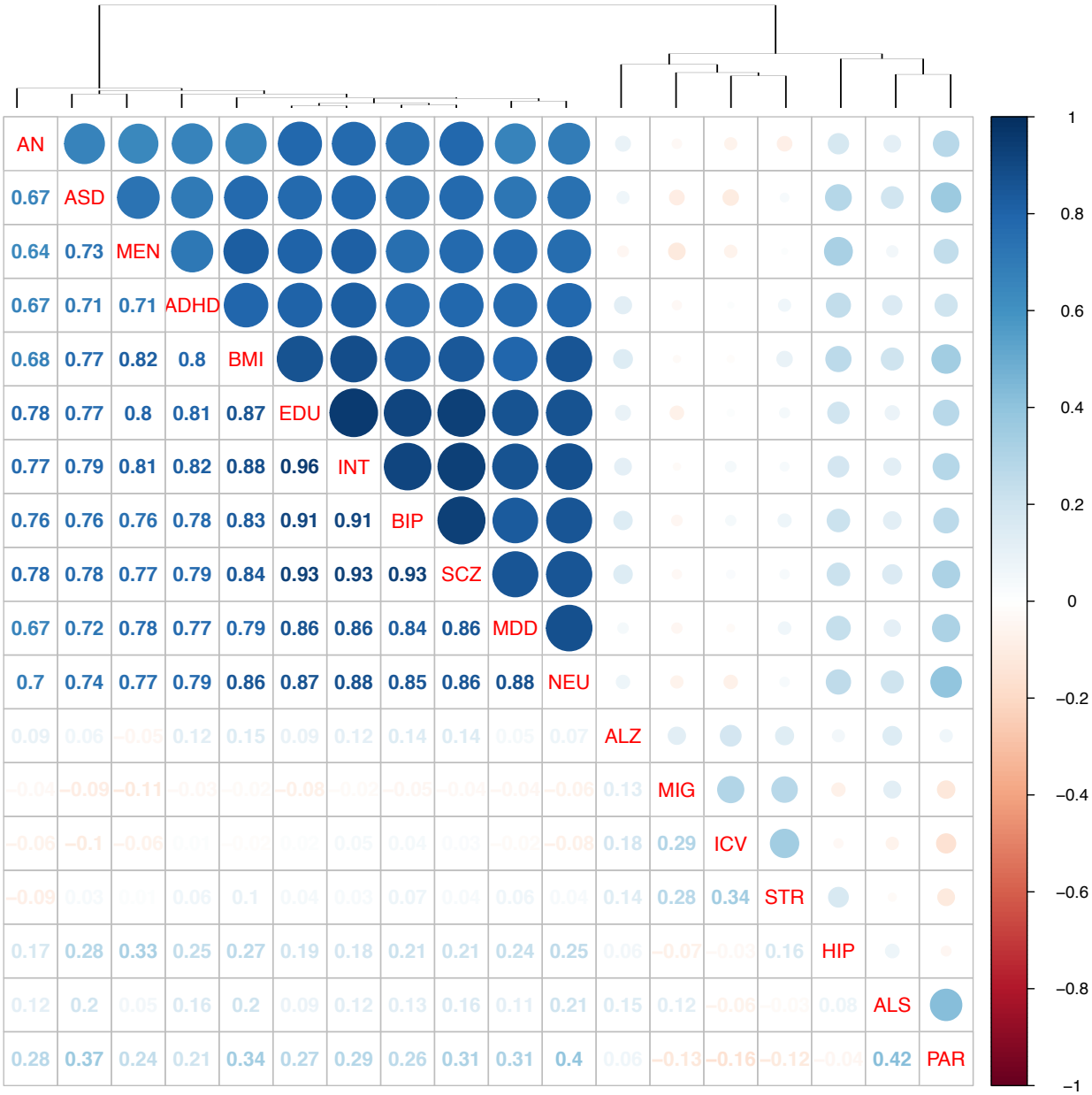

Figure S25: Correlation in cell type associations across traits in a replication data set (414 cell types, 9 brain regions). Spearman rank correlations for cell types associations ($-\log_{10}P$) across traits are shown. SCZ (schizophrenia), EDU (educational attainment), INT (intelligence), BMI (body mass index), BIP (bipolar disorder), NEU (neuroticism), PAR (Parkinson's disease), MDD (Major depressive disorder), MEN (age at menarche), ICV (intracranial volume), ASD (autism spectrum disorder), STR (stroke), AN (anorexia nervosa), MIG (migraine), ALS (amyotrophic lateral sclerosis), ADHD (attention deficit hyperactivity disorder), ALZ (Alzheimer's disease), HIP (hippocampal volume).

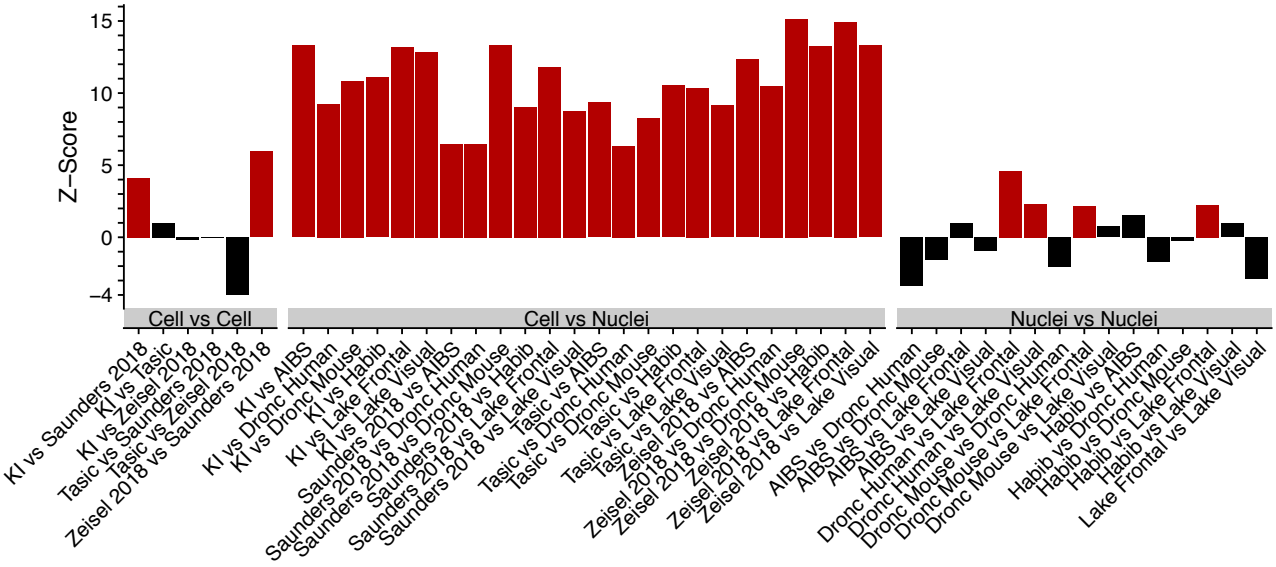

Figure S26: Single nuclei datasets are systematically depleted of dendritically enriched transcripts relative to single-cell datasets. Each bar represents a comparison between two datasets (X versus Y), with the bootstrapped z-scores representing the extent to which dendritically enriched transcripts [96] have lower specificity for pyramidal neurons in dataset Y relative to that in dataset X. Larger z-scores indicate greater depletion of dendritically enriched transcripts, and red bars indicate a statistically significant depletion ($P < 0.05$, by bootstrapping).

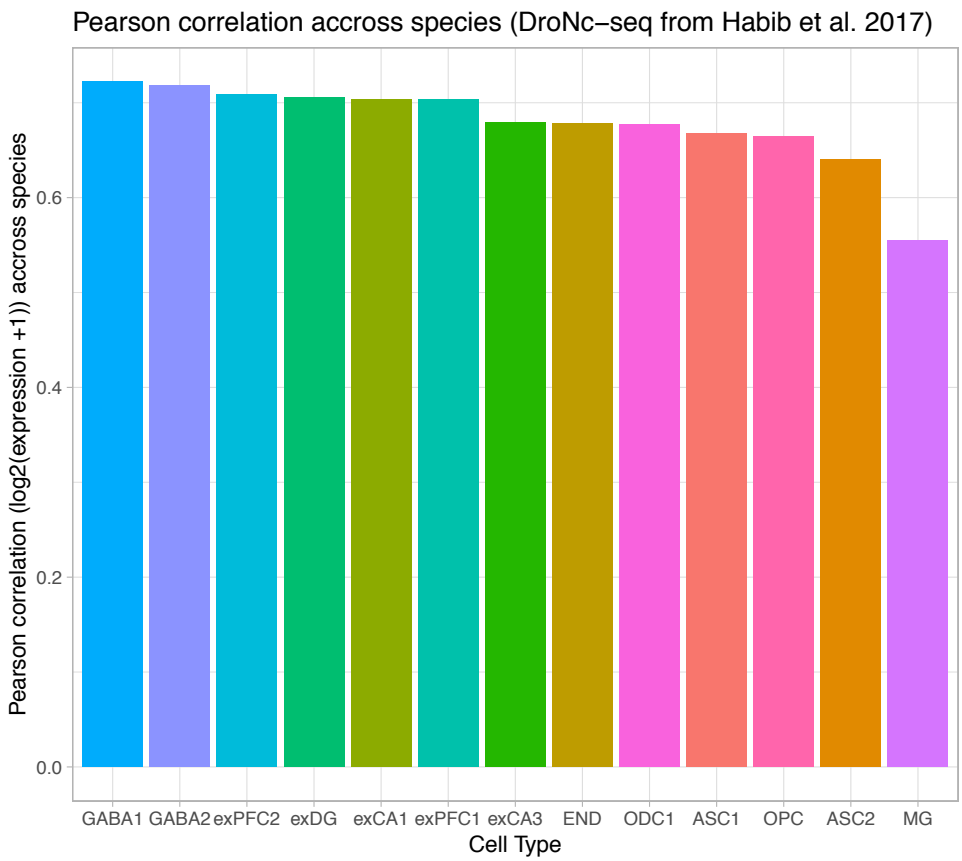

Figure S27: Gene expression correlation within cell type across species. Pearson correlation of gene expression ($\log_2$(expression) +1) between mouse and human cell types with matching names (from Habib et al. 2017 [45]).

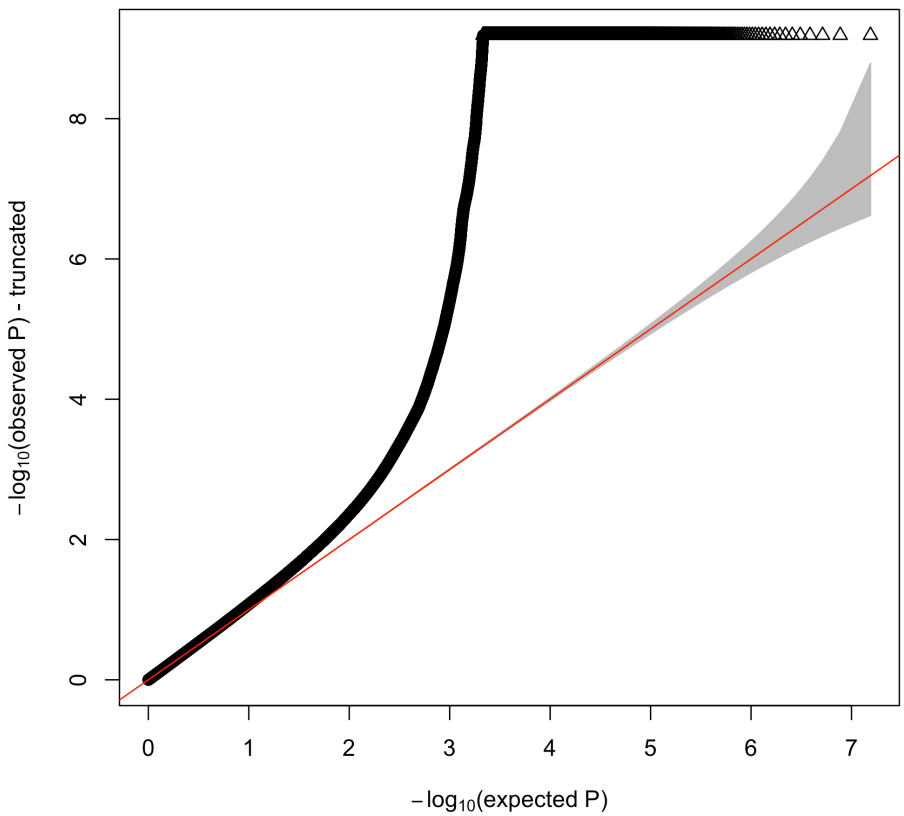

Figure S28: Quantile-quantile plot of Parkinson's disease meta-analysis. Quantile-quantile plot of the meta-analyzed pvalues for Parkinson's disease. The y-axis is truncated for clarity. The grey zone around the red line represents the 95% confidence interval for the null distribution.

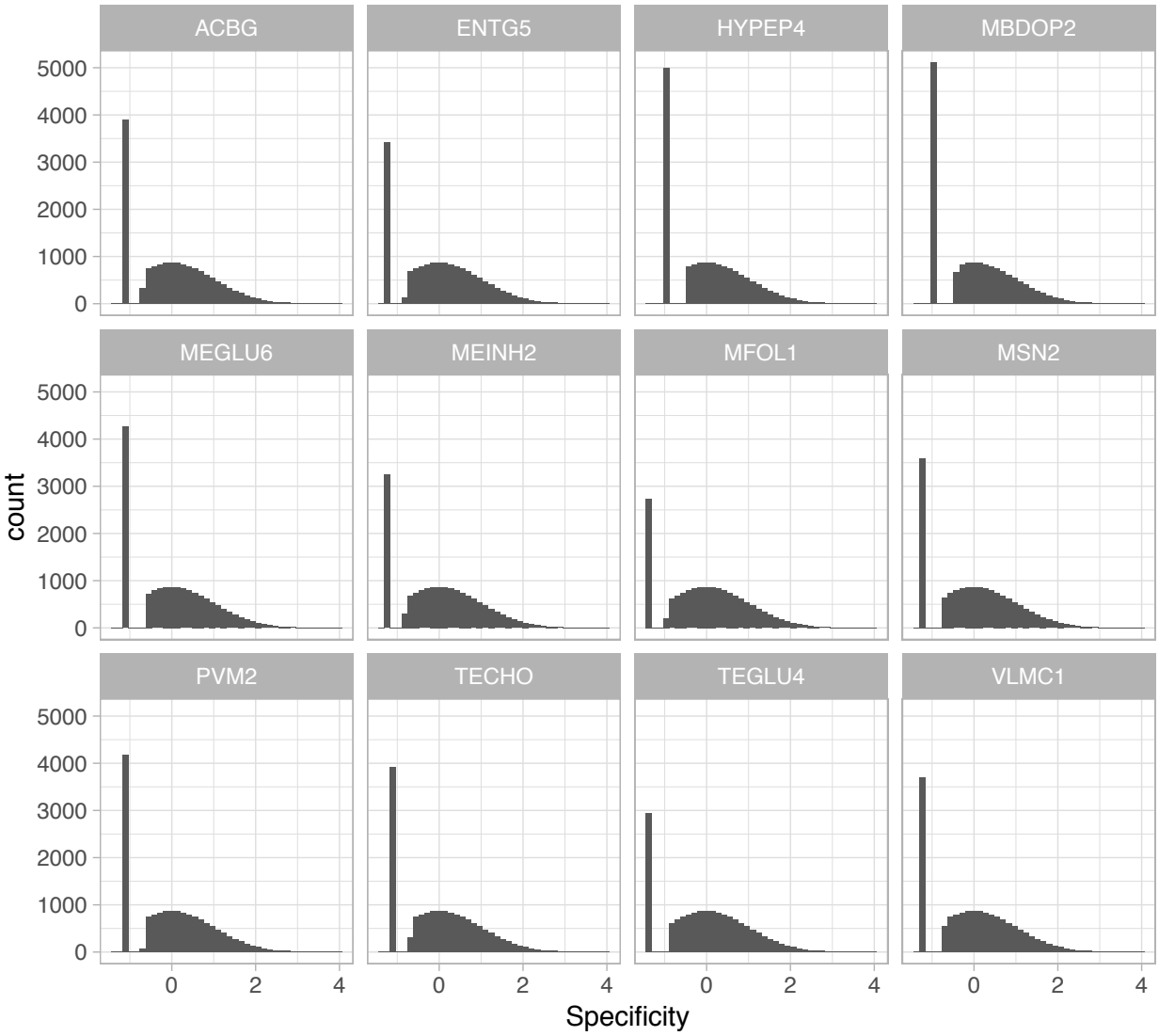

Figure S29: Distribution of specificity metrics for selected cell types from the mouse nervous system. The distribution of gene expression specificities are shown for 12 of the 231 cell types from Zeisel et al. [31]. The bar on the left of each subplots represents genes that are not expressed in the cell type.

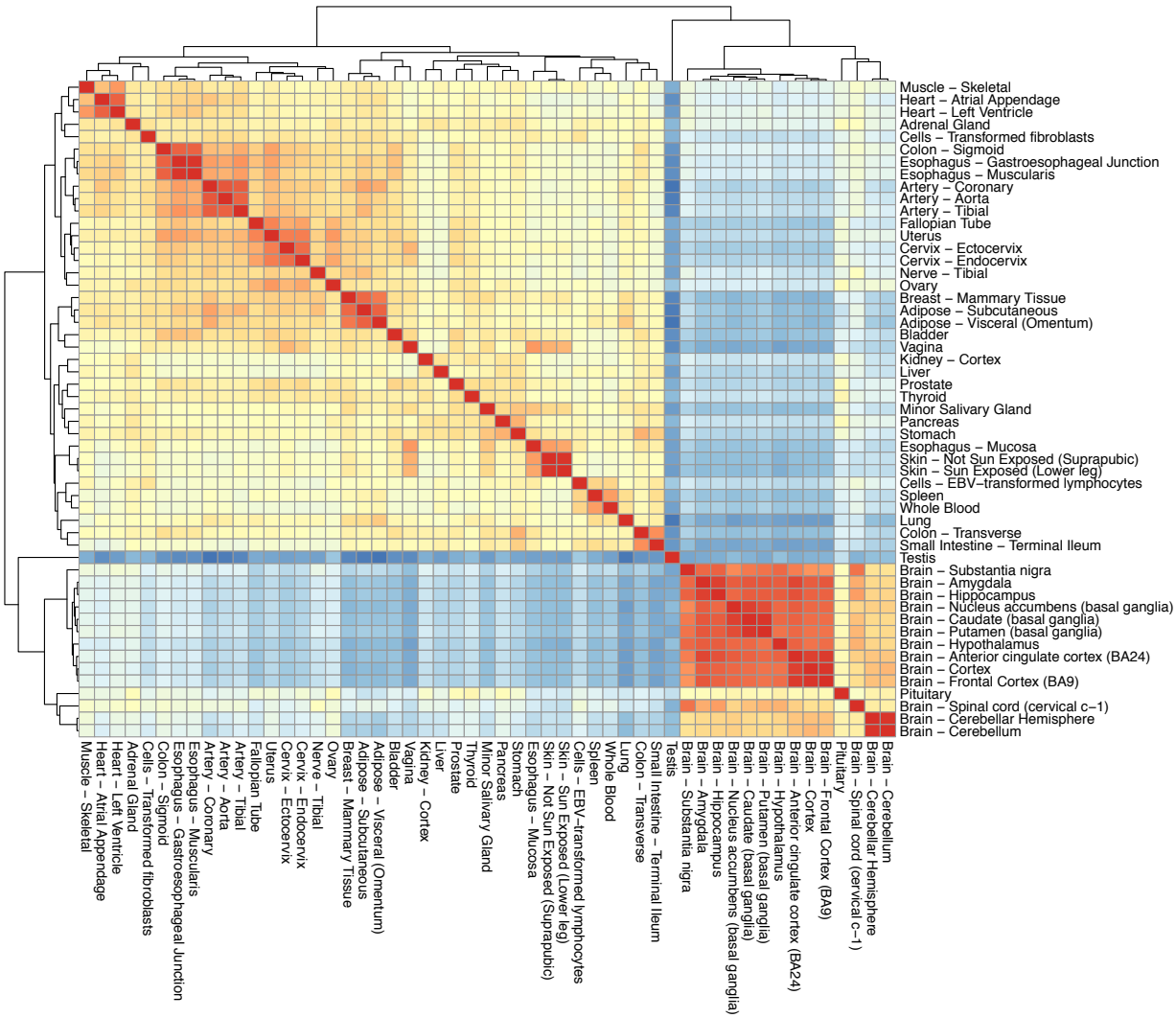

Figure S30: Correlation of the tissue specificity metrics in the GTEx dataset. Spearman rank correlations of the genome-wide gene specificity metric of each tissue are shown. These specificity metrics are used in the MAGMA approach.

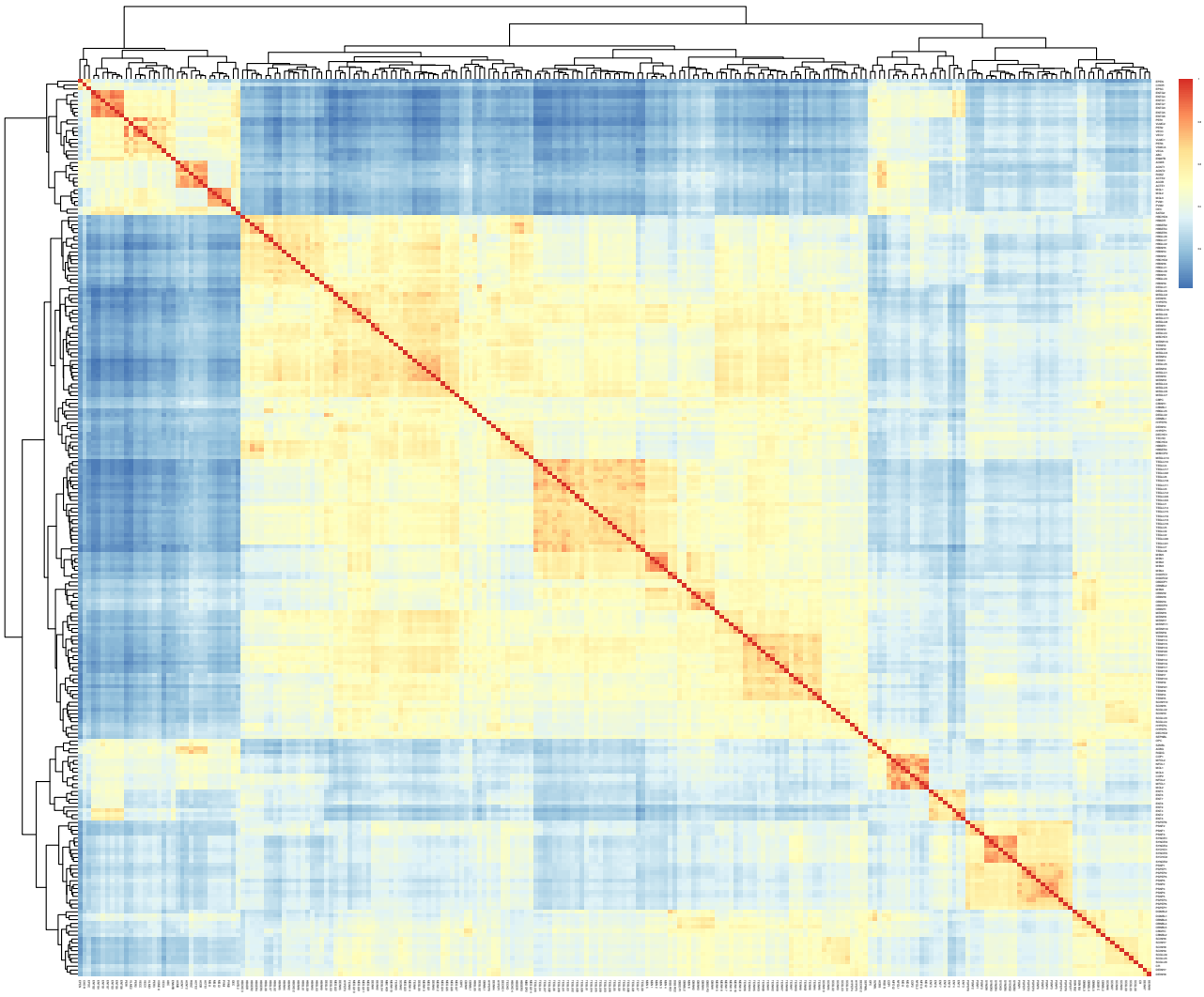

Figure S31: Correlation of the cell type specificity metrics in the mouse nervous system (Zeisel et al. 2018). Spearman rank correlations of the genome-wide gene specificity metric of each cell type are shown. These specificity metrics are used in the MAGMA approach

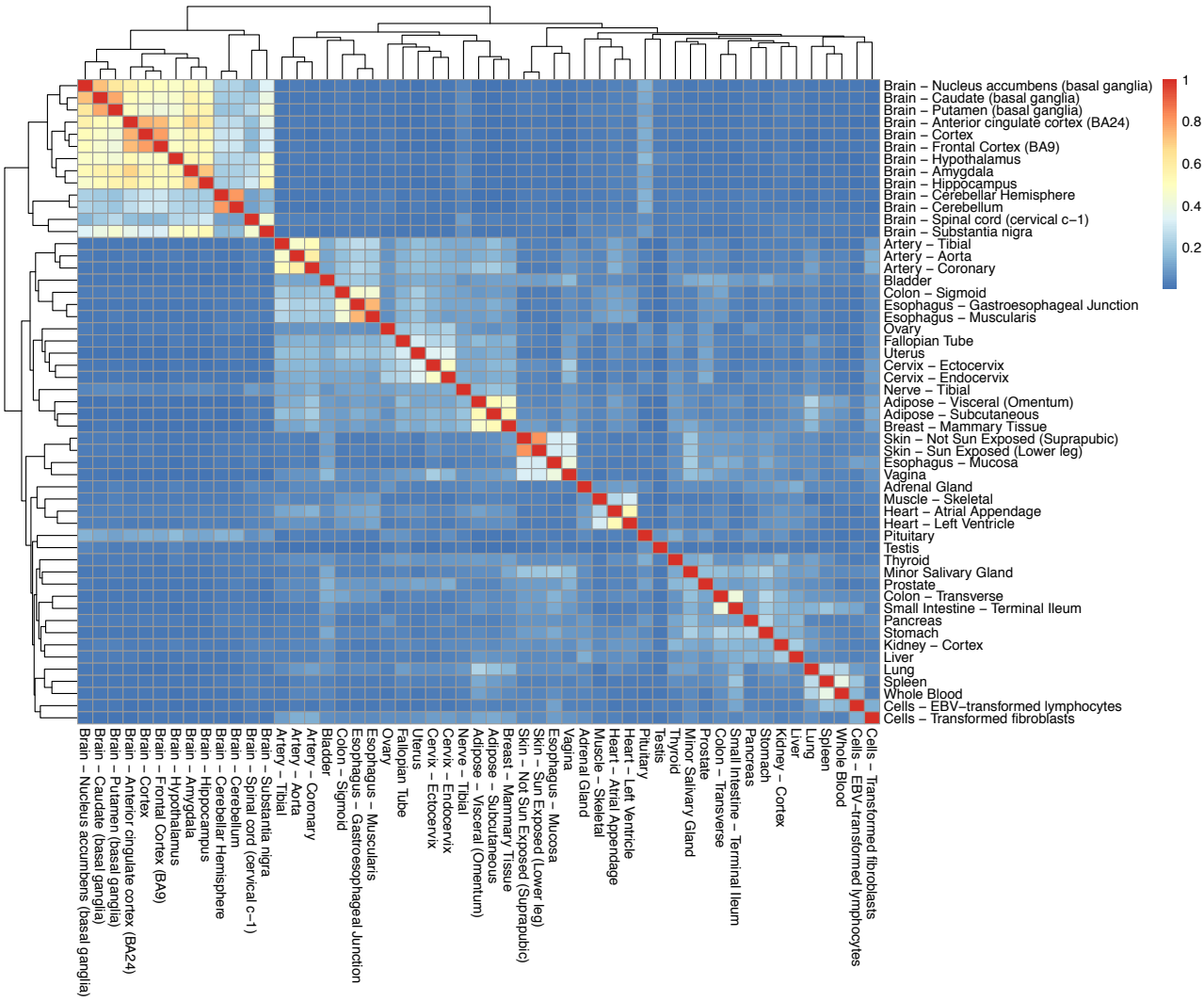

Figure S32: Jaccard index for the top 10% most specific genes in each tissue in the GTEx dataset. Jaccard index were calculated between the top 10% most specific genes in each tissue from the GTEx dataset. These gene sets are used in the LDSC approach.

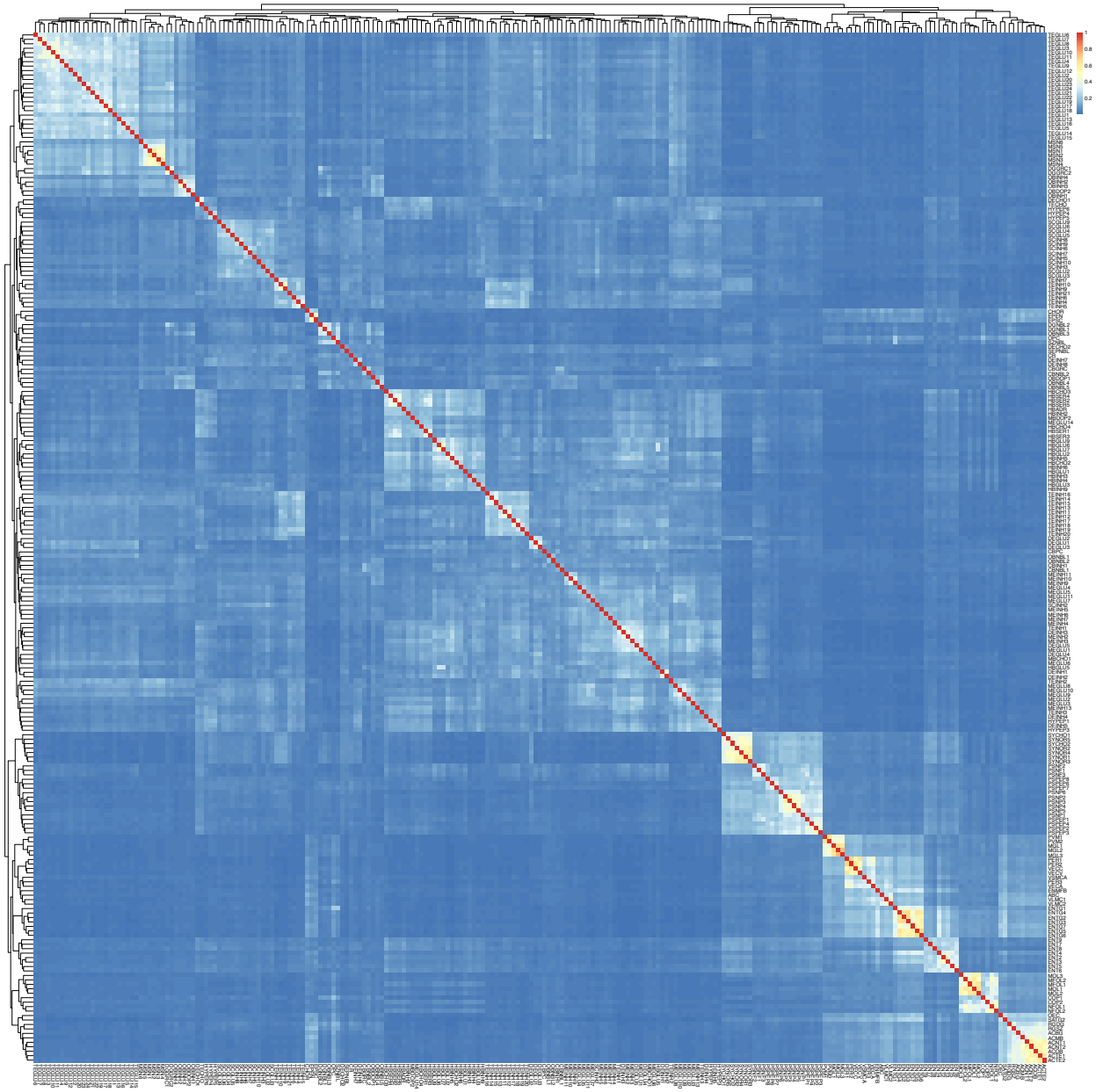

Figure S33: Jaccard index for the top 10% most specific genes in each cell type in the mouse nervous system (Zeisel et al. 2018). Jaccard index were calculated between the top 10% most specific genes in each cell type from the mouse nervous system (Zeisel et al. 2018). These gene sets are used in the LDSC approach.

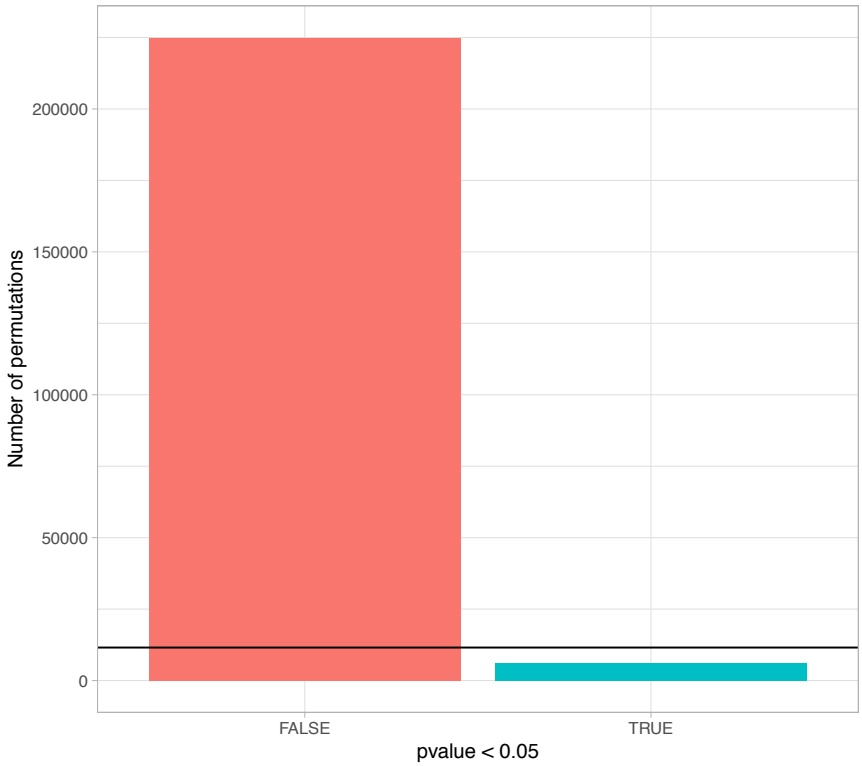

Figure S34: Number of MAGMA associations with P<0.05 using permuted gene-level genetic associations. Gene labels were randomly permuted a thousand times for the schizophrenia MAGMA gene-level genetic associations (231 cell types * 1000 permuted labels=230,000 associations with permuted gene labels). The number of permutations with P < 0.05 is shown in blue. The black horizontal bar shows expected number of random associations with P < 0.05 (231,000*0.05=11,550).

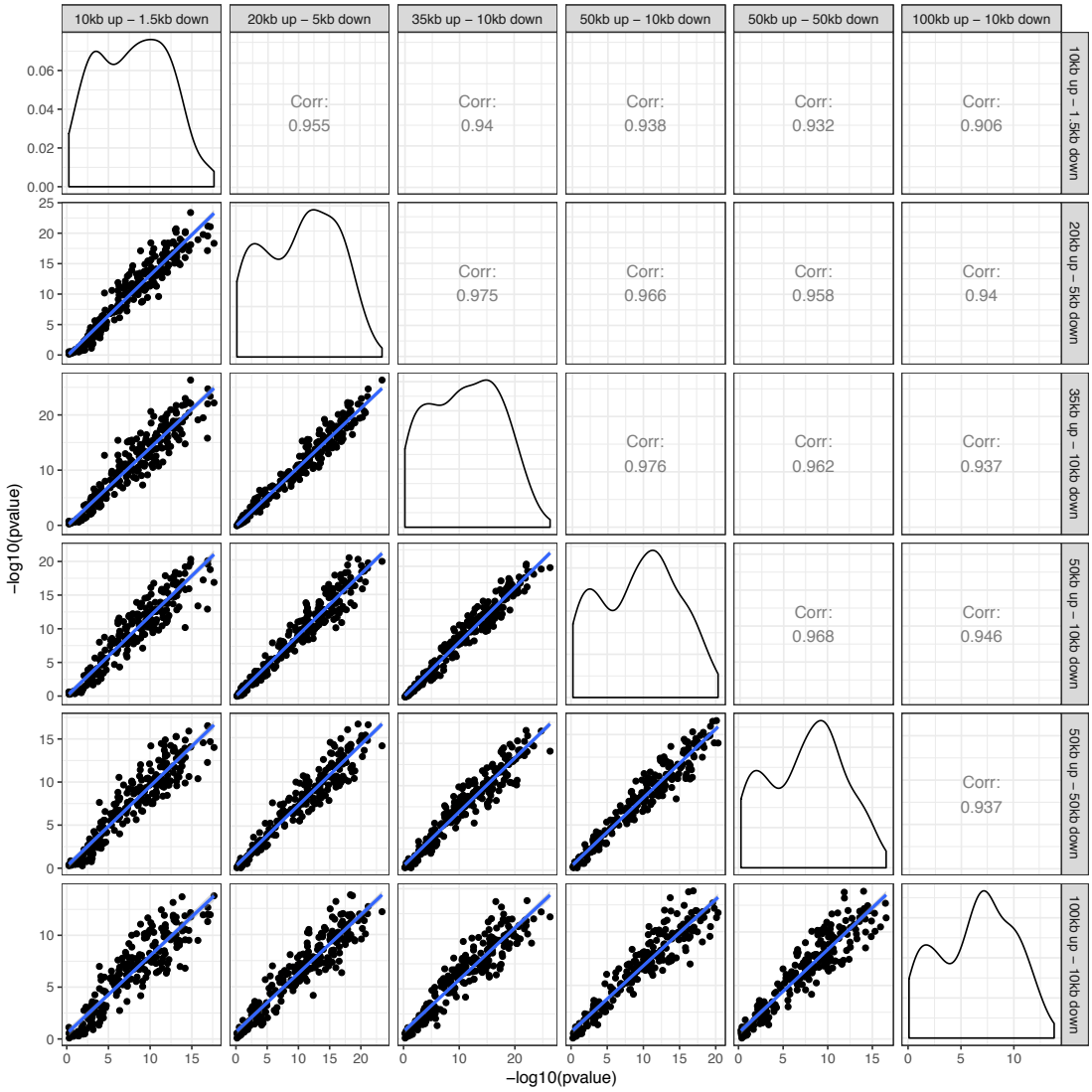

Figure S35: Correlation in schizophrenia cell type association strengths with different window sizes using MAGMA. Pearson correlations of the cell type association strength (-$\log_{10}$P) across different window sizes using MAGMA. The diagonal shows the distribution of the (-$\log_{10}$P) for each window size.

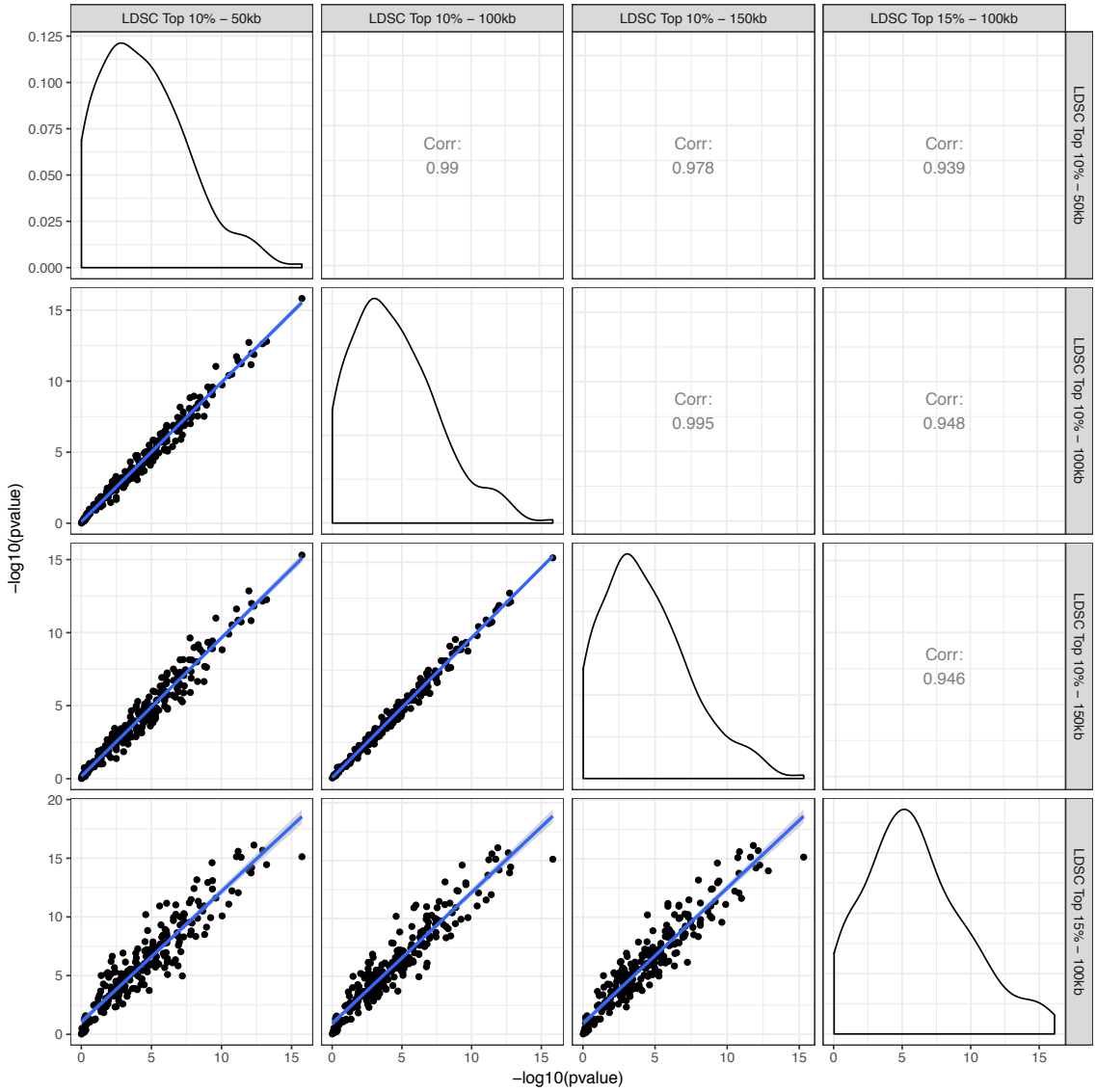

Figure S36: Correlation in schizophrenia cell type association strengths with different window sizes and percentages of most specific genes using LDSC. Pearson correlations of the cell type association strength ($-\log_{10}P$) across different window sizes and percentages of most specific genes using LDSC. The diagonal shows the distribution of the ($-\log_{10}P$) for the cell type associations using different parameters.

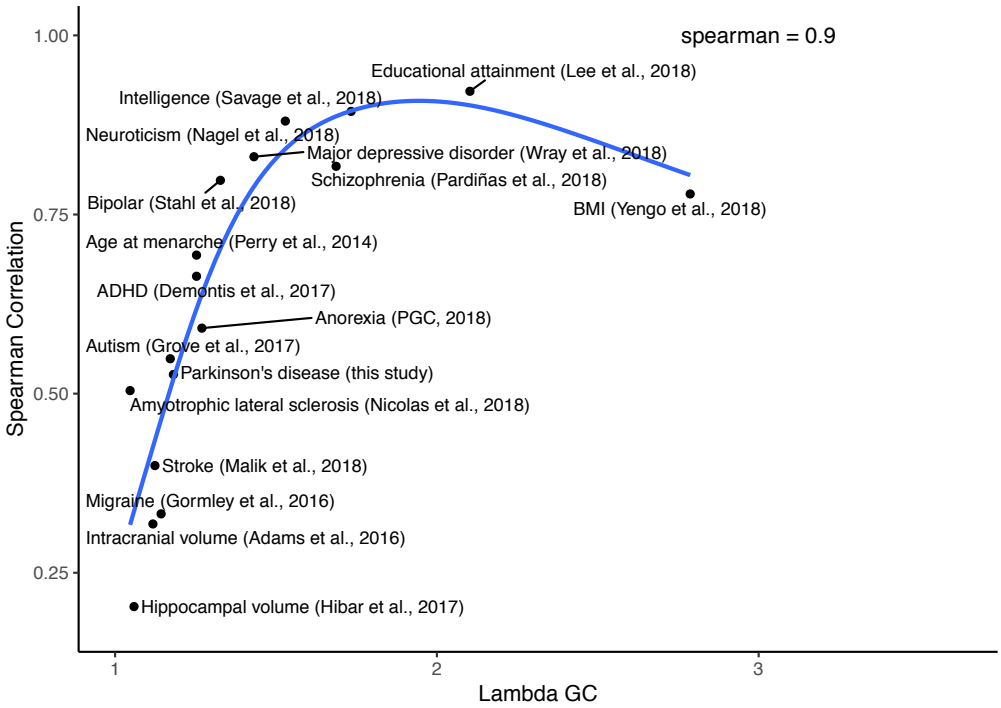

Figure S37: Correlation between λ_{GC} and similarity in cell type ordering between MAGMA and LDSC. LDSC[101] was used to obtain λ_{GC} (a measure of the deviation of the GWAS statistics from the expected) for each GWAS. Spearman rank correlation was used to test for similarity in association strength (-$\log_{10}$P) between MAGMA and LDSC for each GWAS among 231 cell types from the nervous system.

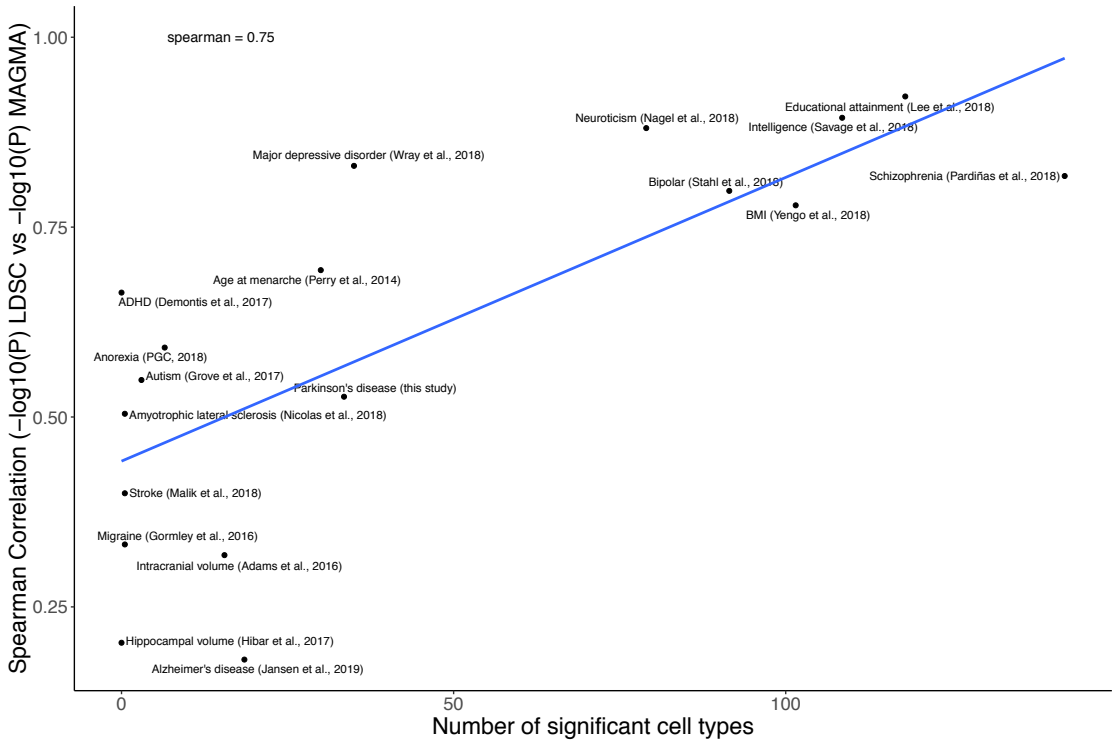

Figure S38: Correlation between mean number of significant cell types and similarity in cell type ordering between MAGMA and LDSC. The mean number of cell types was obtained by taking the average of the number of cell types that were significantly associated with each trait (P<0.005/231) using MAGMA 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.

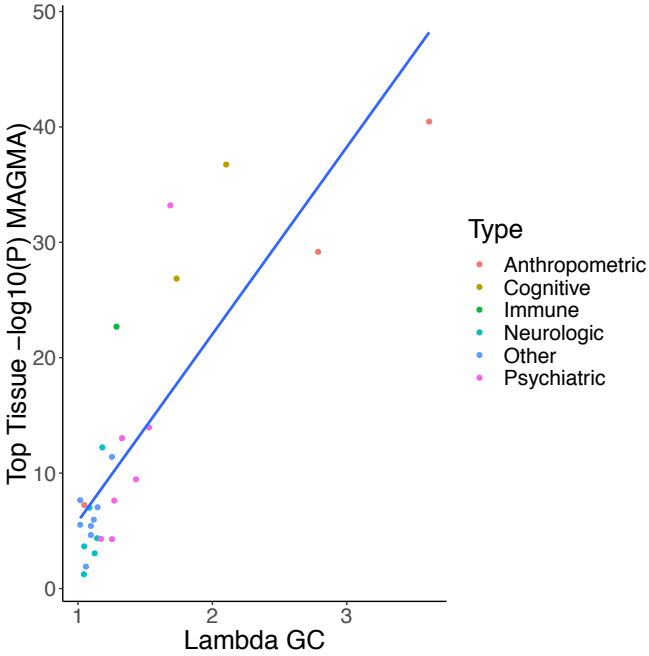

Figure S39: The GWAS λ_{GC} is correlated with the strength of association of the top tissue in the GTEx dataset. Scatter plot of the λ_{GC} (median of chi-squared test statistics divided by expected median of the chi-squared distribution) of each GWAS vs the strength of association of the top GTEx tissue associated with the trait ($-\log_{10}(P_{MAGMA})$). Pearson correlation=0.84.

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. Aguet, F. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
9. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–329 (2015).
10. Ongen, H. *et al.* Estimating the causal tissues for complex traits and diseases. *Nat. Genet.* **49**, 1676–1683 (2017).
11. 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).
12. Skene, N. G. *et al.* Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.* **50**, 825–833 (2018).
13. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
14. 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).
15. 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).
16. Coleman, J. R. I. *et al.* Biological annotation of genetic loci associated with intelligence in a meta-analysis of 87 , 740 individuals.
17. Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* doi:10.1038/s41588-018-0311-9
18. Nalls, M. A. *et al.* Expanding Parkinson's disease genetics: novel risk loci, genomic context, causal insights and heritable risk. *bioRxiv* 388165 (2019). doi:10.1101/388165
19. 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).
20. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
21. Anttila, V. *et al.* Analysis of shared heritability in common disorders of the brain. *Science (80-.).* **360**, (2018).
22. Lee, S. H. *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).
23. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, (2015).
24. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
25. Jevtic, S., Sengar, A. S., Salter, M. W. & McLaurin, J. A. The role of the immune system in Alzheimer disease: Etiology and treatment. *Ageing Research Reviews* **40**, 84–94 (2017).

26. Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).
27. 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).
28. Sullivan, P. F. & Geschwind, D. H. Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. *Cell* **177**, 162–183 (2019).
29. Andreasen, N. C. & Pierson, R. The role of the cerebellum in schizophrenia. *Biol. Psychiatry* **64**, 81–88 (2008).
30. Villanueva, R. The cerebellum and neuropsychiatric disorders. *Psychiatry Res.* **198**, 527–532 (2012).
31. Zeisel, A. *et al.* Molecular architecture of the mouse nervous system. *Cell* **174**, 999-1014.e22 (2018).
32. 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).
33. Hook, P. W. & McCallion, A. S. Heritability enrichment in open chromatin reveals cortical layer contributions to schizophrenia. *bioRxiv* (2018).
34. Fullard, J. F. *et al.* An atlas of chromatin accessibility in the adult human brain. *Genome Res.* **28**, 1243–1252 (2018).
35. 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).
36. 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
37. Katz, D. B., Nicolelis, M. A. L. & Simon, S. A. Gustatory processing is dynamic and distributed. *Current Opinion in Neurobiology* **12**, 448–454 (2002).
38. Braak, H. *et al.* Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol. Aging* **24**, 197–211 (2003).
39. Sulzer, D. & Surmeier, D. J. Neuronal vulnerability, pathogenesis, and Parkinson's disease. *Movement Disorders* **28**, 41–50 (2013).
40. Halliday, G. M. *et al.* Neuropathology of immunohistochemically identified brainstem neurons in Parkinson's disease. *Ann. Neurol.* **27**, 373–385 (1990).
41. 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
42. Perrett, R. M., Alexopoulou, Z. & Tofaris, G. K. The endosomal pathway in Parkinson's disease. *Molecular and Cellular Neuroscience* **66**, 21–28 (2015).
43. Saunders, A. *et al.* Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* **174**, 1015-1030.e16 (2018).
44. Keren-Shaul, H. *et al.* A unique microglia type associated with restricting development of alzheimer's disease. *Cell* **169**, (2017).
45. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955 (2017).
46. Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* (2019). doi:10.1038/s41586-019-1195-2
47. Saunders, A. *et al.* Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. *Cell* **174**, 1015-1030.e16 (2018).
48. Lake, B. B. *et al.* Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018).
49. Skene, N. G. *et al.* Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.* **50**, 825–833 (2018).
50. Lesnick, T. G. *et al.* A genomic pathway approach to a complex disease: axon guidance and Parkinson disease. *PLoS Genet.* **3**, 0984–0995 (2007).

51. 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
52. 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).
53. Gagliano, S. A. *et al.* Genomics implicates adaptive and innate immunity in Alzheimer's and Parkinson's diseases. *Ann. Clin. Transl. Neurol.* **3**, 924–933 (2016).
54. 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).
55. Bryois, J. *et al.* Evaluation of chromatin accessibility in prefrontal cortex of individuals with schizophrenia. *Nat. Commun.* **9**, (2018).
56. 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).
57. Caspi, A. *et al.* The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clin. Psychol. Sci.* **2**, 119–137 (2014).
58. Sullivan, P. F. & Geschwind, D. H. Defining the genetic, genomic, cellular, and diagnostic architectures of psychiatric disorders. *Submitted*
59. 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).
60. Müller, N., Weidinger, E., Leitner, B. & Schwarz, M. J. The role of inflammation in schizophrenia. *Frontiers in Neuroscience* **9**, (2015).
61. 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
62. 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).
63. Recasens, A. & Dehay, B. Alpha-synuclein spreading in Parkinson's disease. *Front. Neuroanat.* **8**, (2014).
64. Engelender, S. & Isacson, O. The threshold theory for Parkinson's disease. *Trends in Neurosciences* **40**, 4–14 (2017).
65. Surmeier, D. J., Obeso, J. A. & Halliday, G. M. Selective neuronal vulnerability in Parkinson disease. *Nature Reviews Neuroscience* **18**, 101–113 (2017).
66. Gilman, S. *et al.* Second consensus statement on the diagnosis of multiple system atrophy. *Neurology* **71**, 670–676 (2008).
67. Dorsey, E. R. *et al.* Virtual research visits and direct-to-consumer genetic testing in Parkinson's disease. *Digit. Heal.* **1**, 205520761559299 (2015).
68. 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).
69. Seidel, K. *et al.* The brainstem pathologies of Parkinson's disease and dementia with lewy bodies. *Brain Pathol.* **25**, 121–135 (2015).
70. Lake, B. B. *et al.* Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science (80-.).* **352**, 1586–1590 (2016).
71. Sathyamurthy, A. *et al.* Massively Parallel Single Nucleus Transcriptional Profiling Defines Spinal Cord Neurons and Their Activity during Behavior. *Cell Rep.* **22**, 2216–2225 (2018).
72. Lake, B. B. *et al.* A comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA. *Sci. Rep.* **7**, (2017).
73. Stahl, E. *et al.* Genomewide association study identifies 30 loci associated with bipolar disorder. *bioRxiv* 173062 (2017). doi:10.1101/173062
74. 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).
75. 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).
76. Grove, J. *et al.* Common risk variants identified in autism spectrum disorder. *bioRxiv* **33**, 42 (2017).
77. Gormley, P. *et al.* Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. *Nat. Genet.* **48**, 1296 (2016).
78. 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).
79. 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
80. Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature Genetics* **51**, 404–413 (2019).
81. 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).
82. Nelson, C. P. *et al.* Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nature Genetics* **49**, (2017).
83. 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).
84. Hibar, D. P. *et al.* Novel genetic loci associated with hippocampal volume. *Nat. Commun.* **8**, 13624 (2017).
85. 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).
86. Adams, H. H. H. *et al.* Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nat. Neurosci.* **19**, 1569–1582 (2016).
87. 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).
88. Scott, R. A. *et al.* An expanded genome-wide association study of type 2 diabetes in europeans. *Diabetes* **66**, 2888–2902 (2017).
89. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
90. Watson, H. J. *et al.* Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat. Genet.* (2019). doi:10.1038/s41588-019-0439-2
91. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
92. Saunders, A. *et al.* Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* **174**, 1015-1030.e16 (2018).
93. 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).
94. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
95. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
96. Cajigas, I. J. *et al.* The Local Transcriptome in the Synaptic Neuropil Revealed by Deep Sequencing and High-Resolution Imaging. *Neuron* **74**, 453–466 (2012).
97. Alexa, A. & Rahnenfuhrer, J. topGO: Enrichment analysis for gene ontology. (2016).
98. 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).
99. 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).
100. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

101. Bulik-Sullivan, B. *et al.* LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).