nature genetics

SUPPLEMENTARY INFORMATION

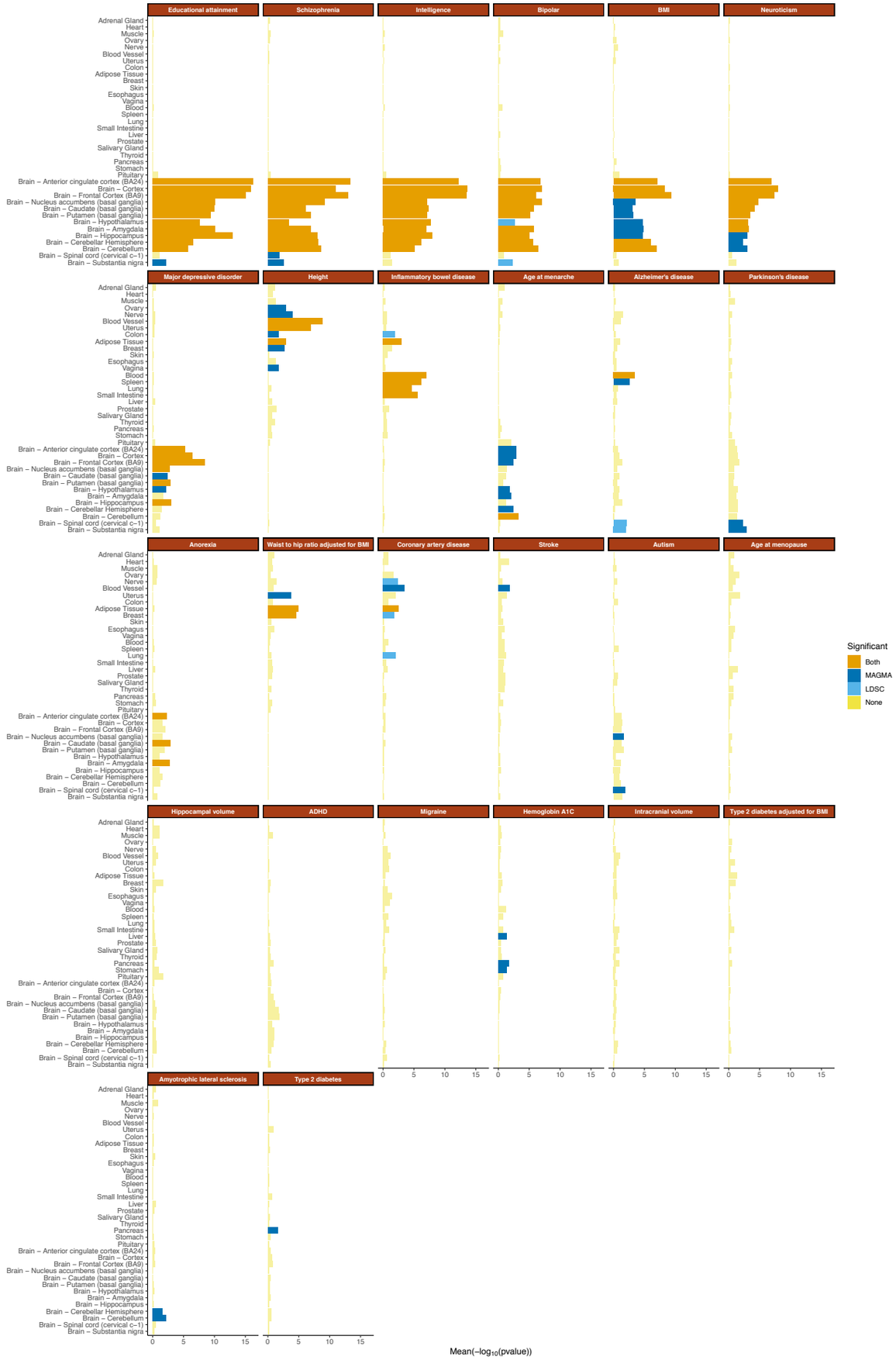
https://doi.org/10.1038/s41588-020-0610-9

In the format provided by the authors and unedited.

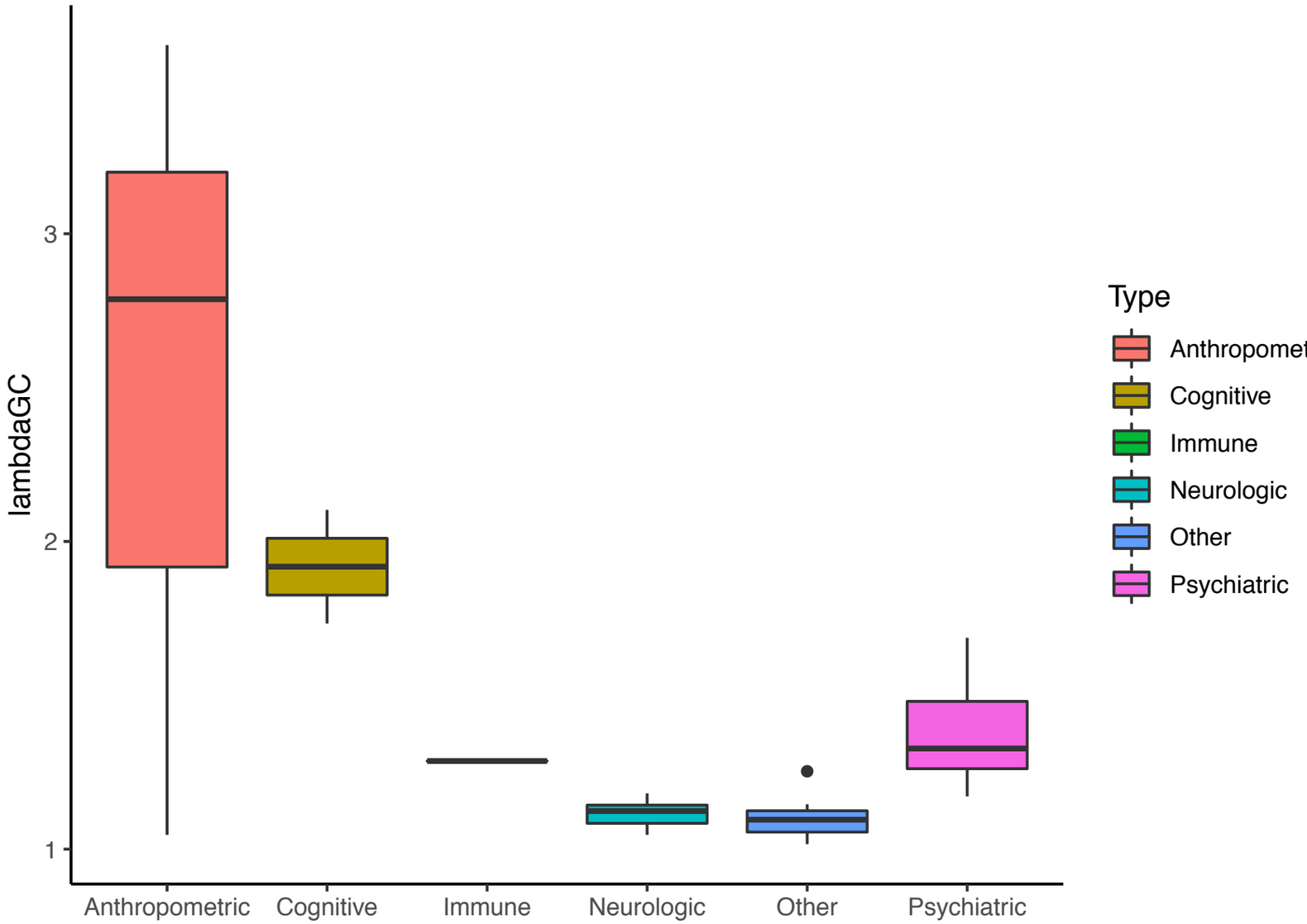
Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson's disease

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

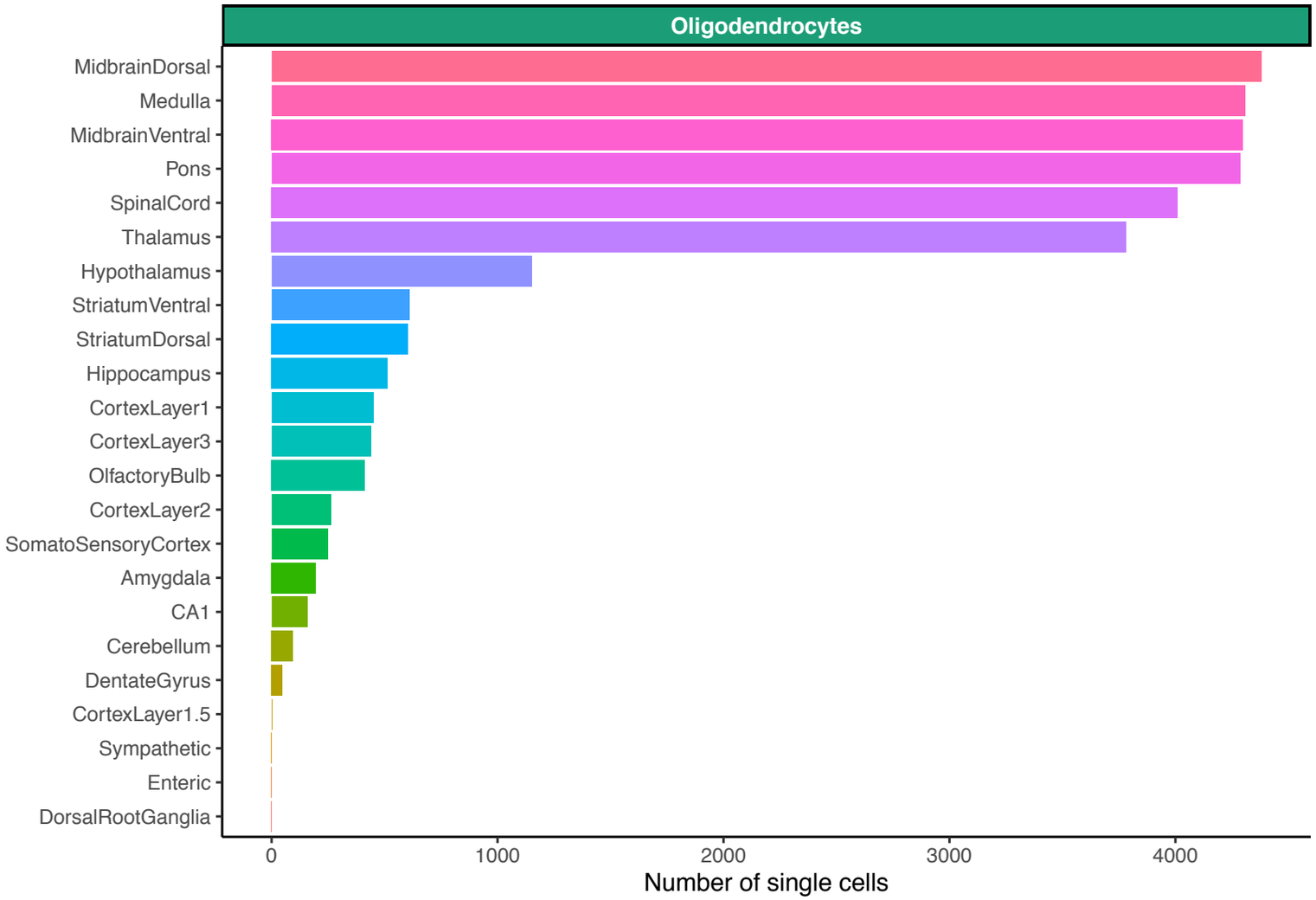
[1]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [2]Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden. [3]UCL Institute of Neurology, Queen Square, London, UK. [4]Division of Brain Sciences, Department of Medicine,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



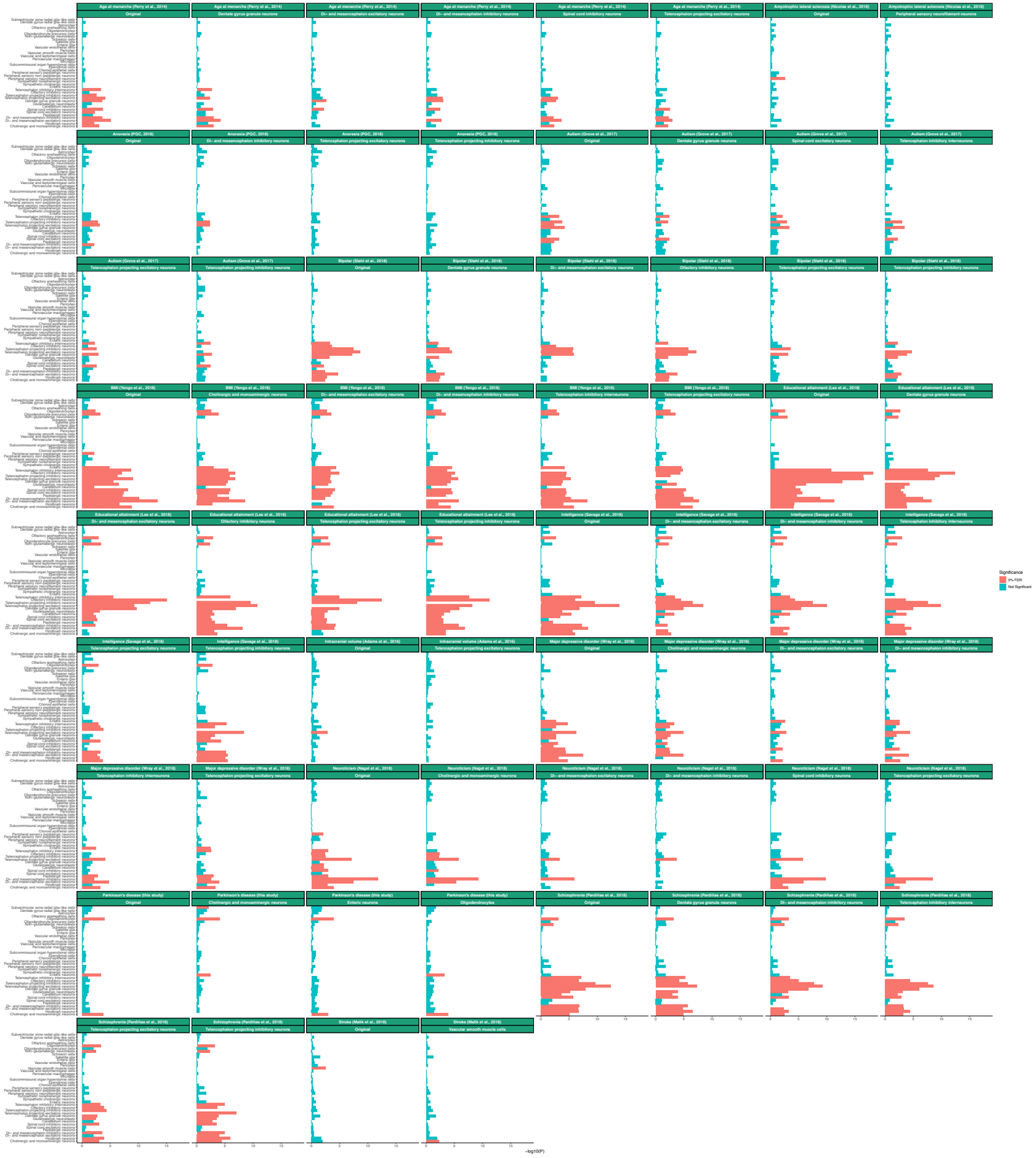
Supplementary Figure 1: Tissue – trait associations for all traits. The mean strength of association ($-\log_{10}P$) of MAGMA and LDSC is shown and the bar color indicates whether the tissue is significantly associated with both methods, one method or none (significance threshold: 5% false discovery rate).



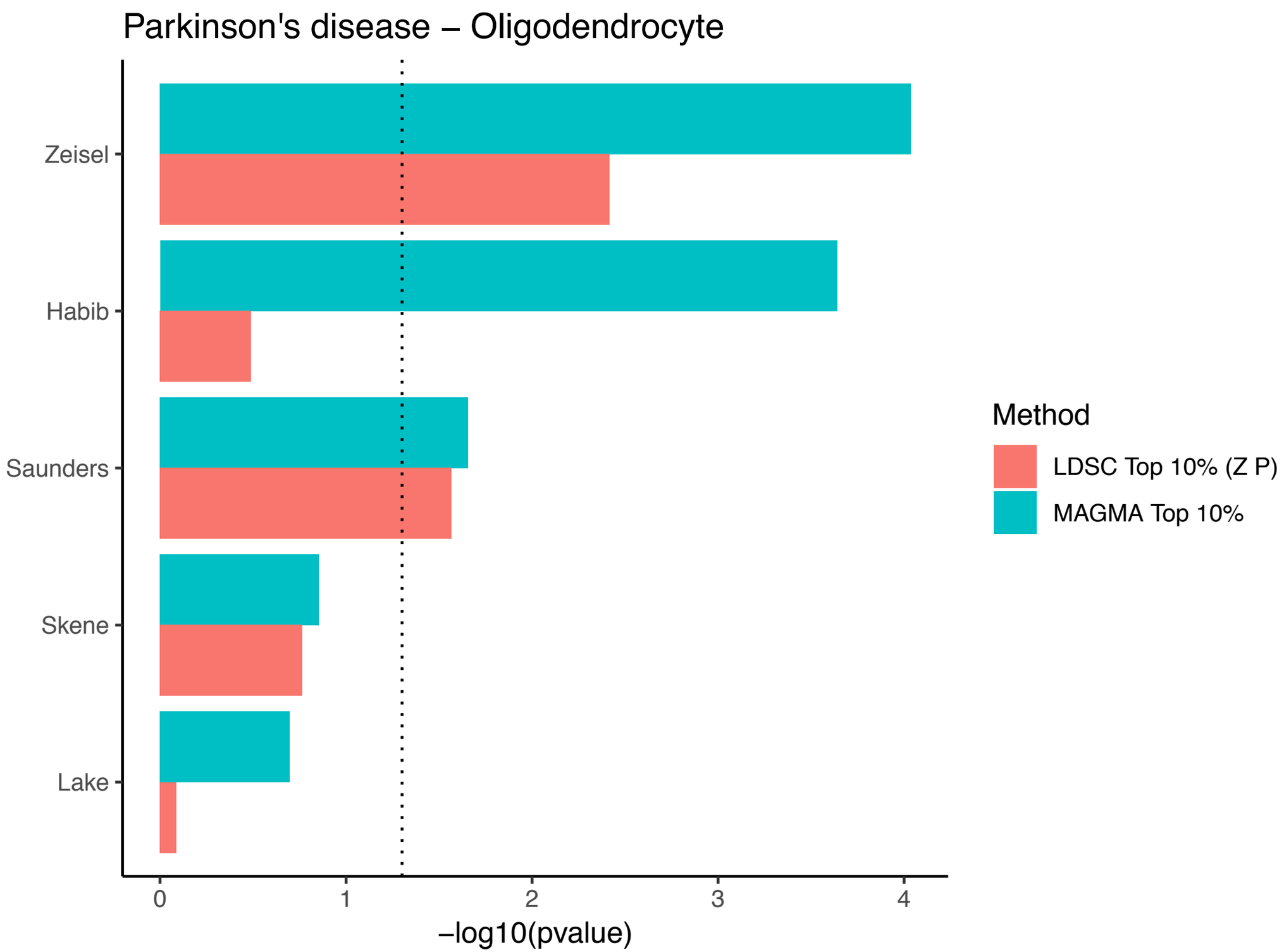
Supplementary Figure 2: GWAS signal to noise ratio (λ_{GC}) by category of GWAS trait. Boxplot of the λ_{GC} of the different GWAS by category of trait. λ_{GC} was estimated using LDSC for each GWAS. The box represents the median, upper and lower quartile, while the whiskers extend to the most extreme value up to 1.5 time the interquartile range. Data beyond the whiskers are called “outlying” points and are plotted individually



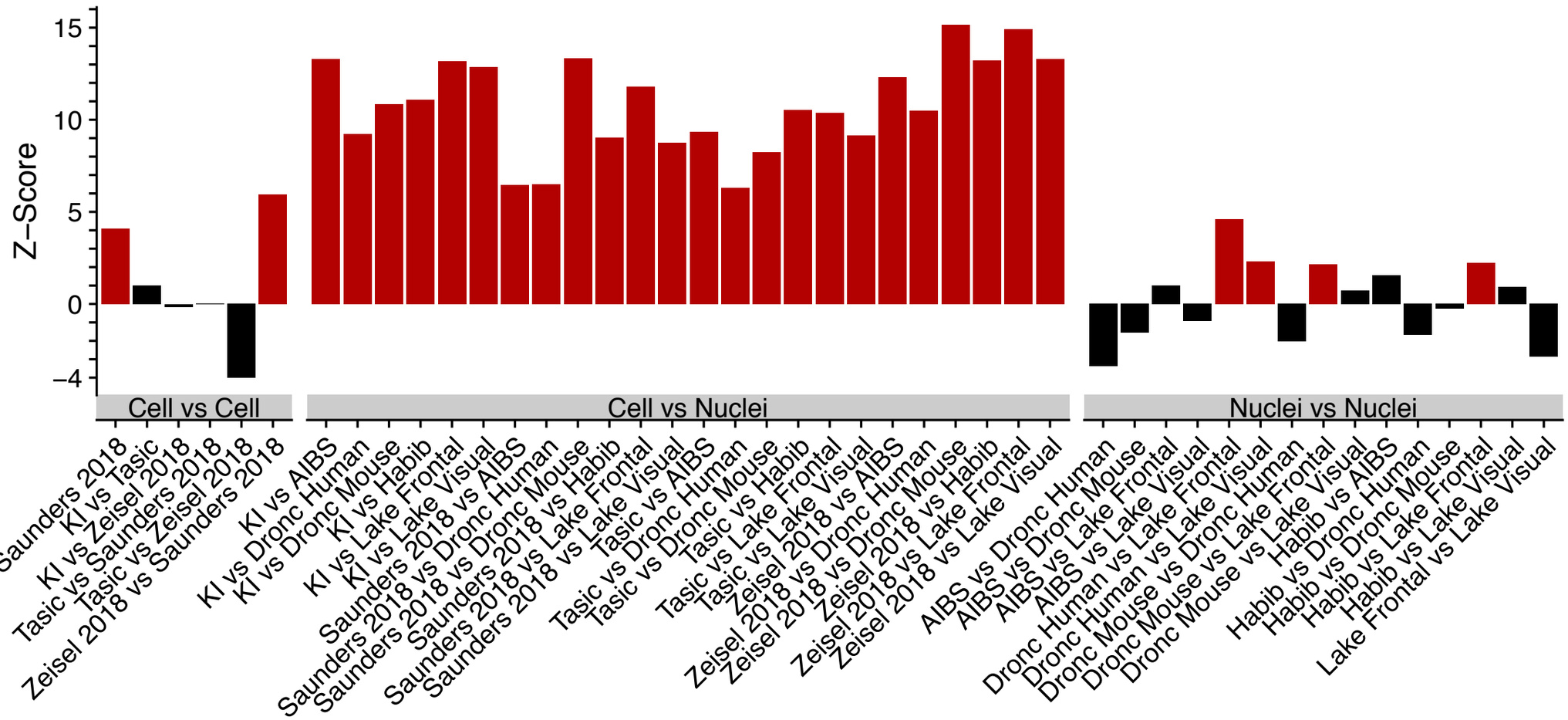
Supplementary Figure 3: Number of single cells forming the oligodendrocyte cluster. Number of single cells per region of the mouse nervous system used to estimate the average gene expression of oligodendrocytes.



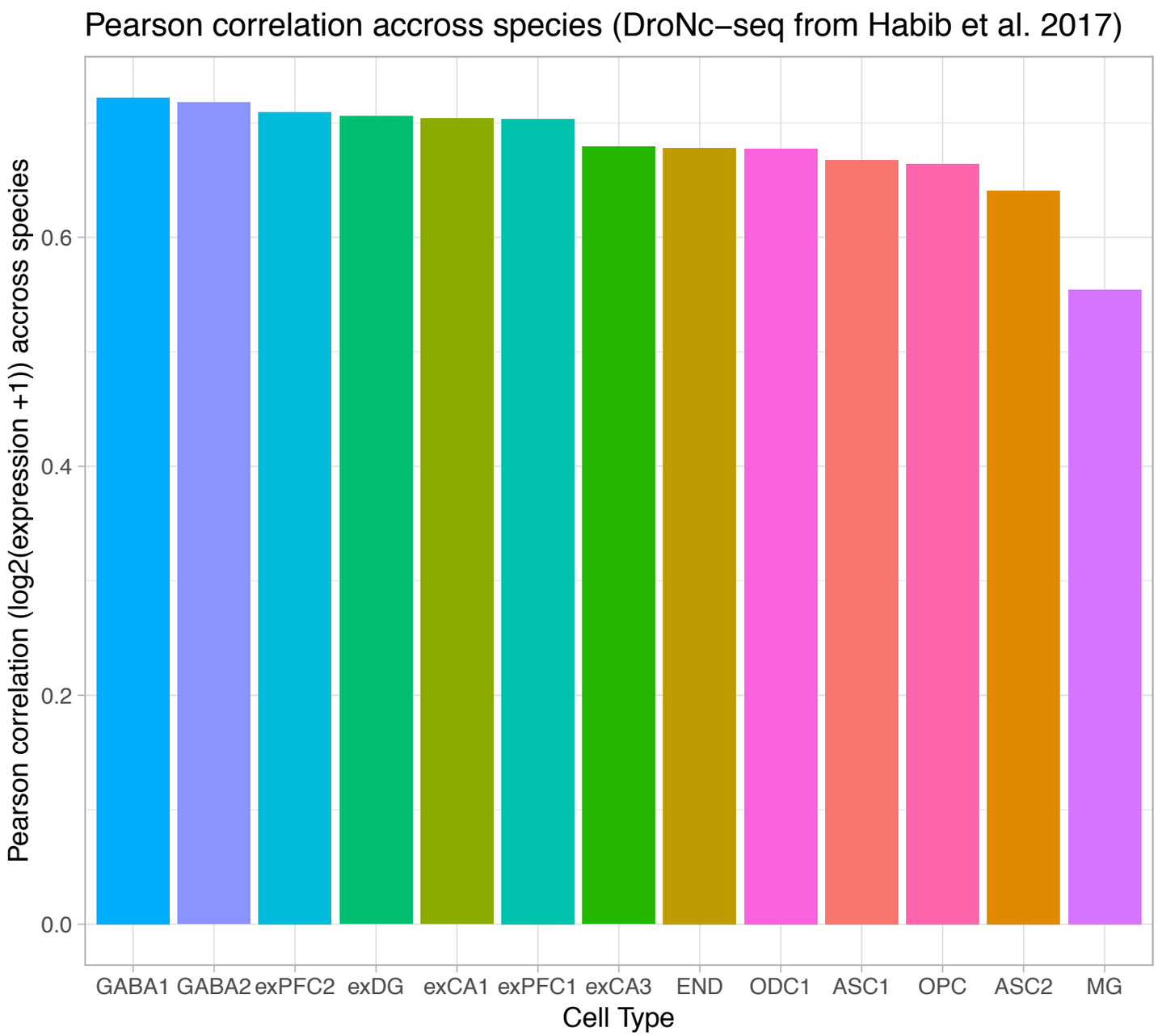
Supplementary Figure 4: Conditional analysis results for brain related traits. Conditional analysis results using MAGMA are shown for up to the 5 most associated cell types (if at least 5 cell types were significant at a 5% false discovery rate in the original analysis. The color indicates if the cell type is significant at a 5% false discovery rate and the label indicates the cell type the association analysis is being conditioned on.



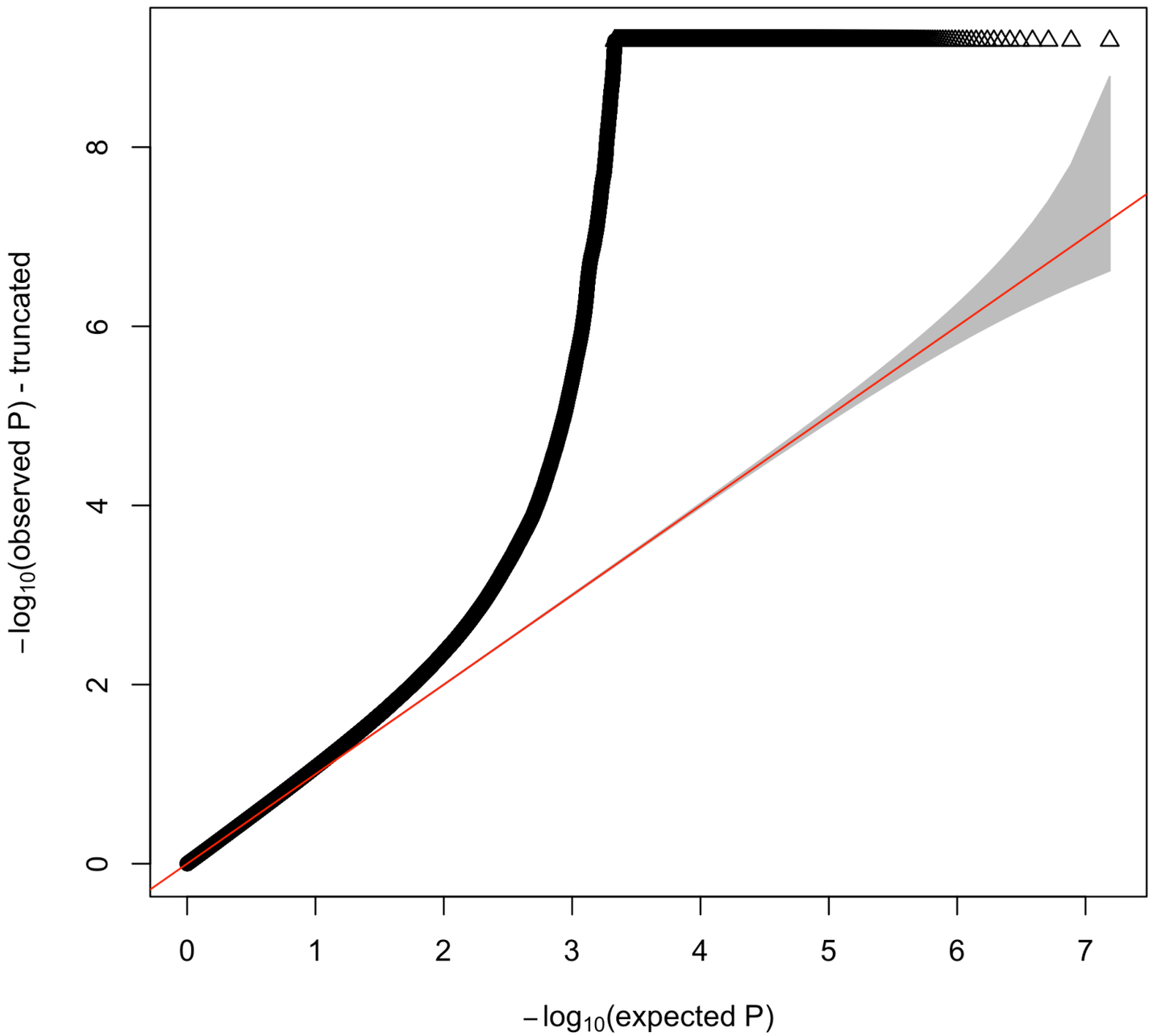
Supplementary Figure 5: Association of Parkinson's disease with oligodendrocytes in the different datasets. The dotted line indicated the nominal significance threshold (P=0.05)



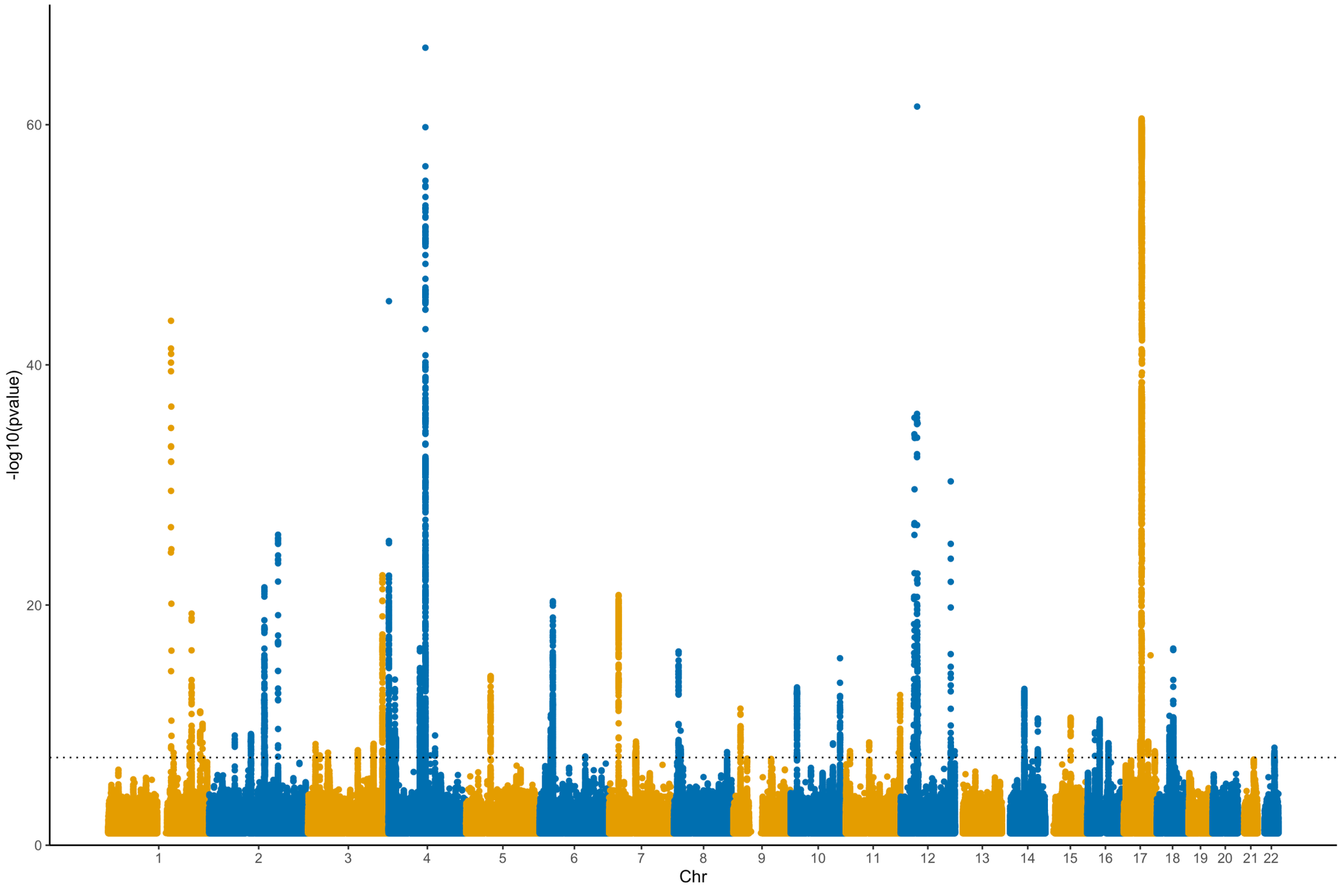
Supplementary Figure 6: Single nuclei datasets are systematically depleted of dendritically enriched transcripts relative to single-cell datasets. Each bar represents a comparison between two datasets (X versus Y), with the bootstrapped z-scores representing the extent to which dendritically enriched transcripts 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).



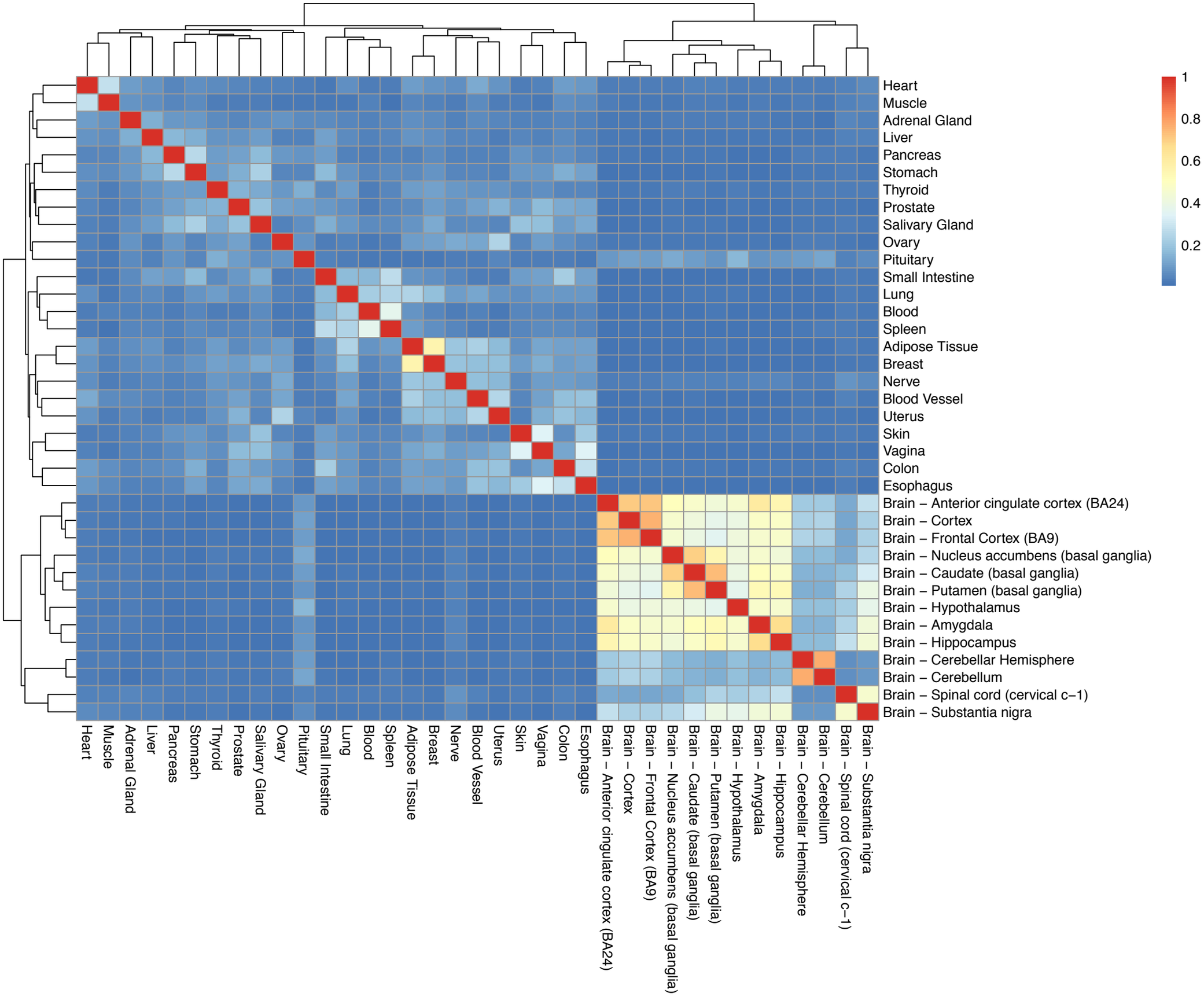
Supplementary Figure 7: Gene expression correlation within cell type across species. Pearson correlation of gene expression (log_2(expression) +1) between mouse and human cell types with matching names (from Habib et al. 2017).



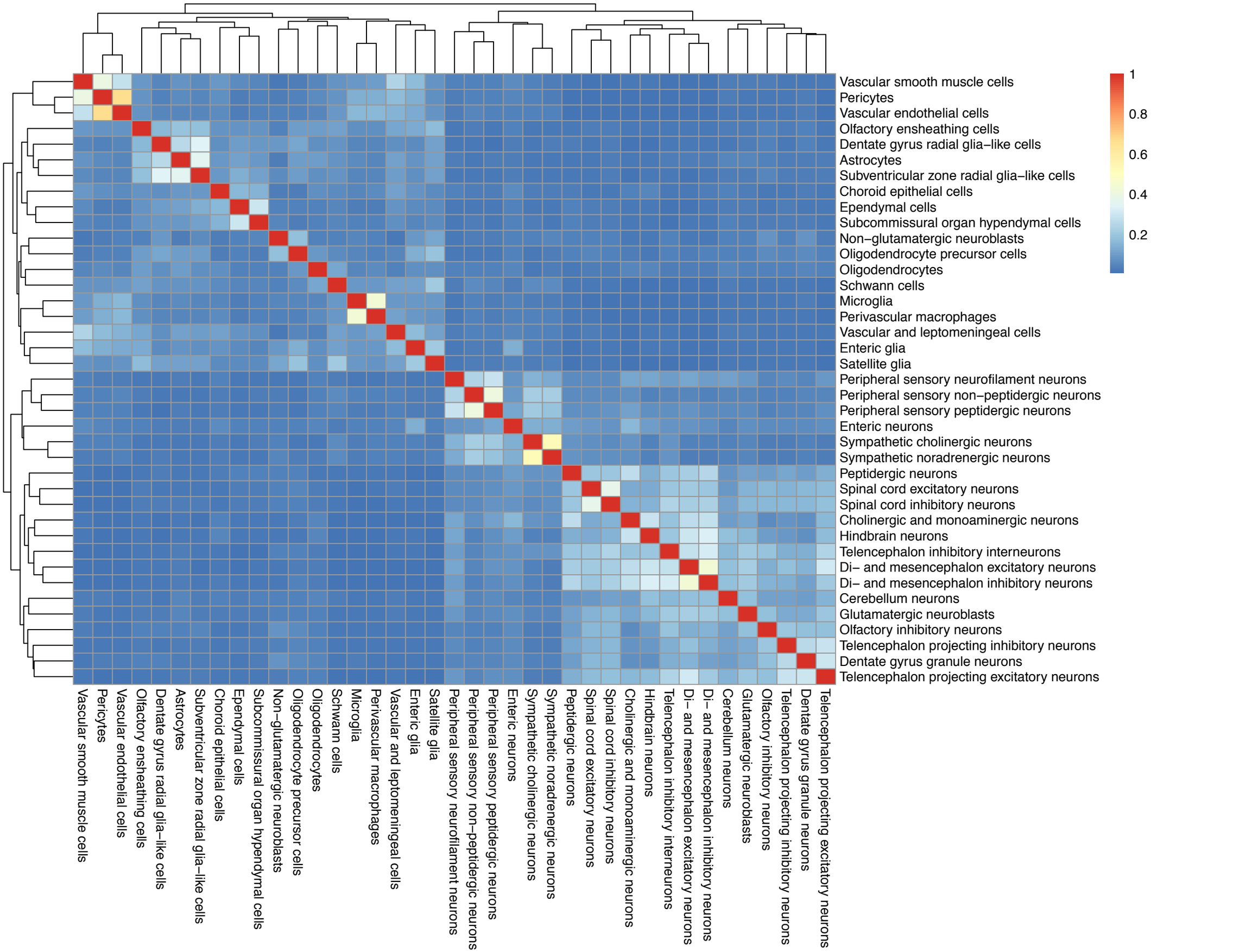
Supplementary Figure 8: Quantile-quantile plot of Parkinson's disease meta-analysis. Quantile-quantile plot of the meta-analyzed pvalues for Parkinson's disease. The y-axis is truncated for clarity. The grey zone around the red line represents the 95% confidence interval for the null distribution.



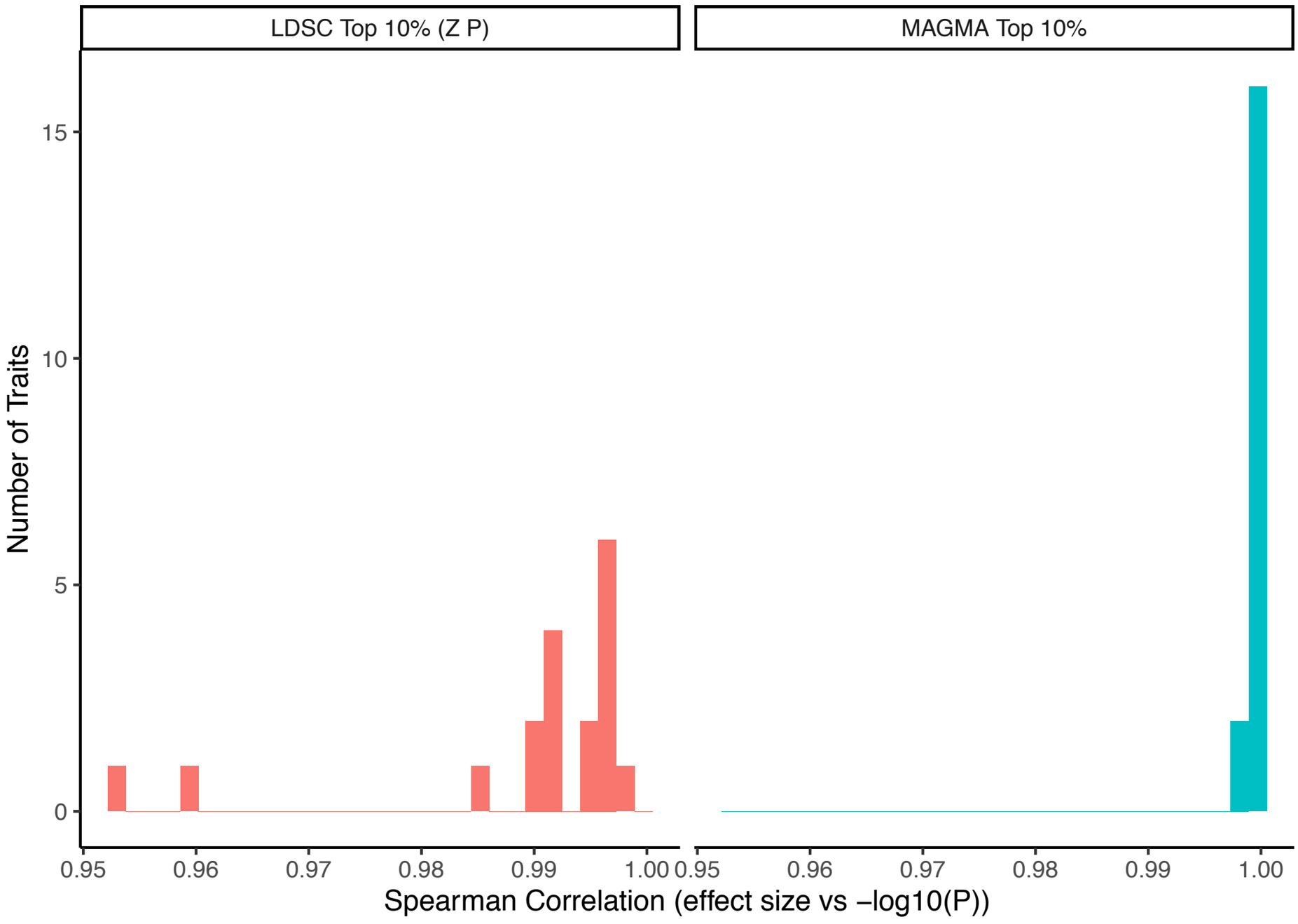
Supplementary Figure 9:: Manhattan plot of Parkinson's disease meta-analysis. The black dotted line represents the genome-wide significance threshold ($5x10^{-8}$).



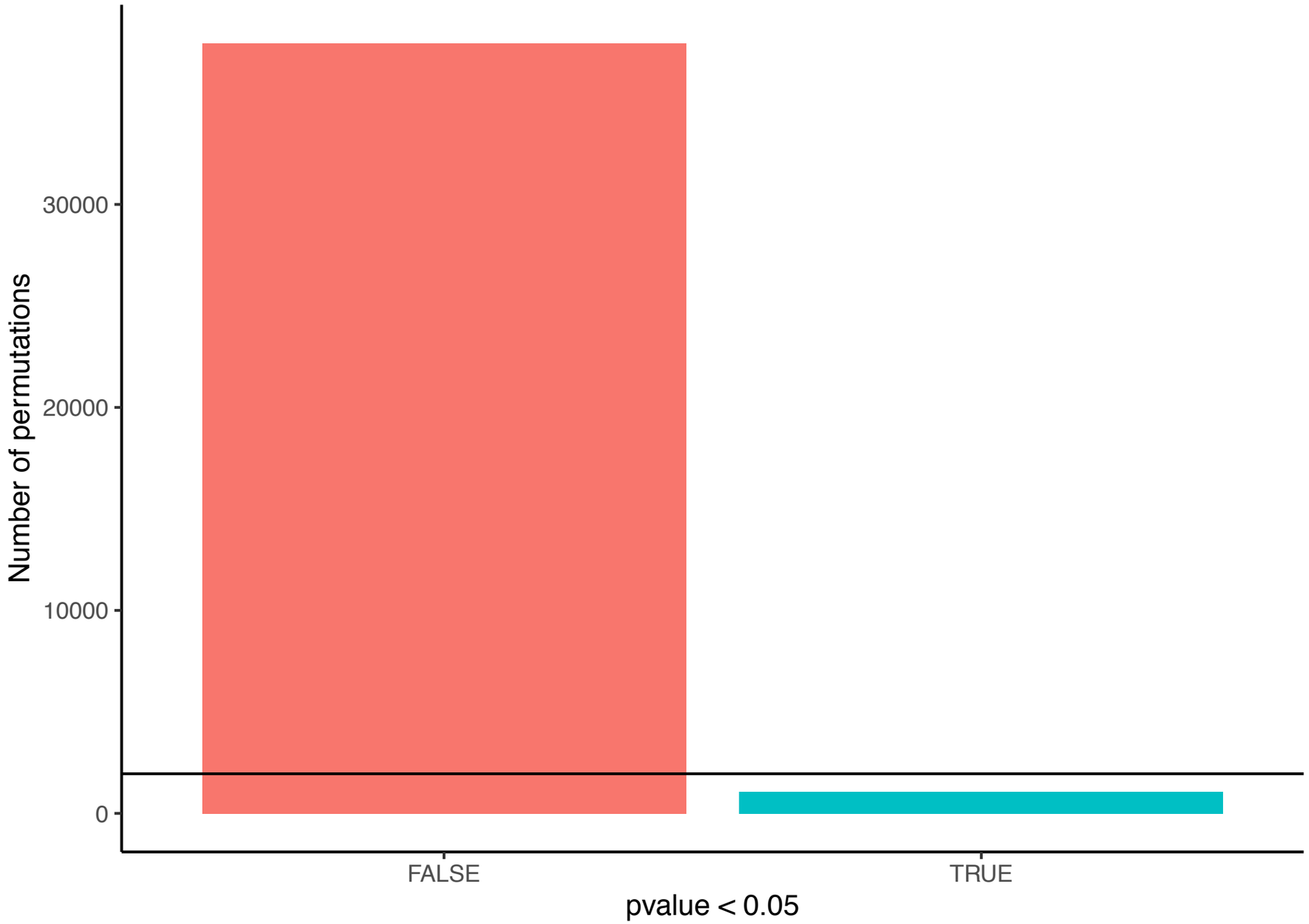
Supplementary Figure 10: Jaccard index for the top 10% most specific genes in each tissue in the GTEx dataset. Jaccard index were calculated between the top 10% most specific genes in each tissue from the GTEx dataset.



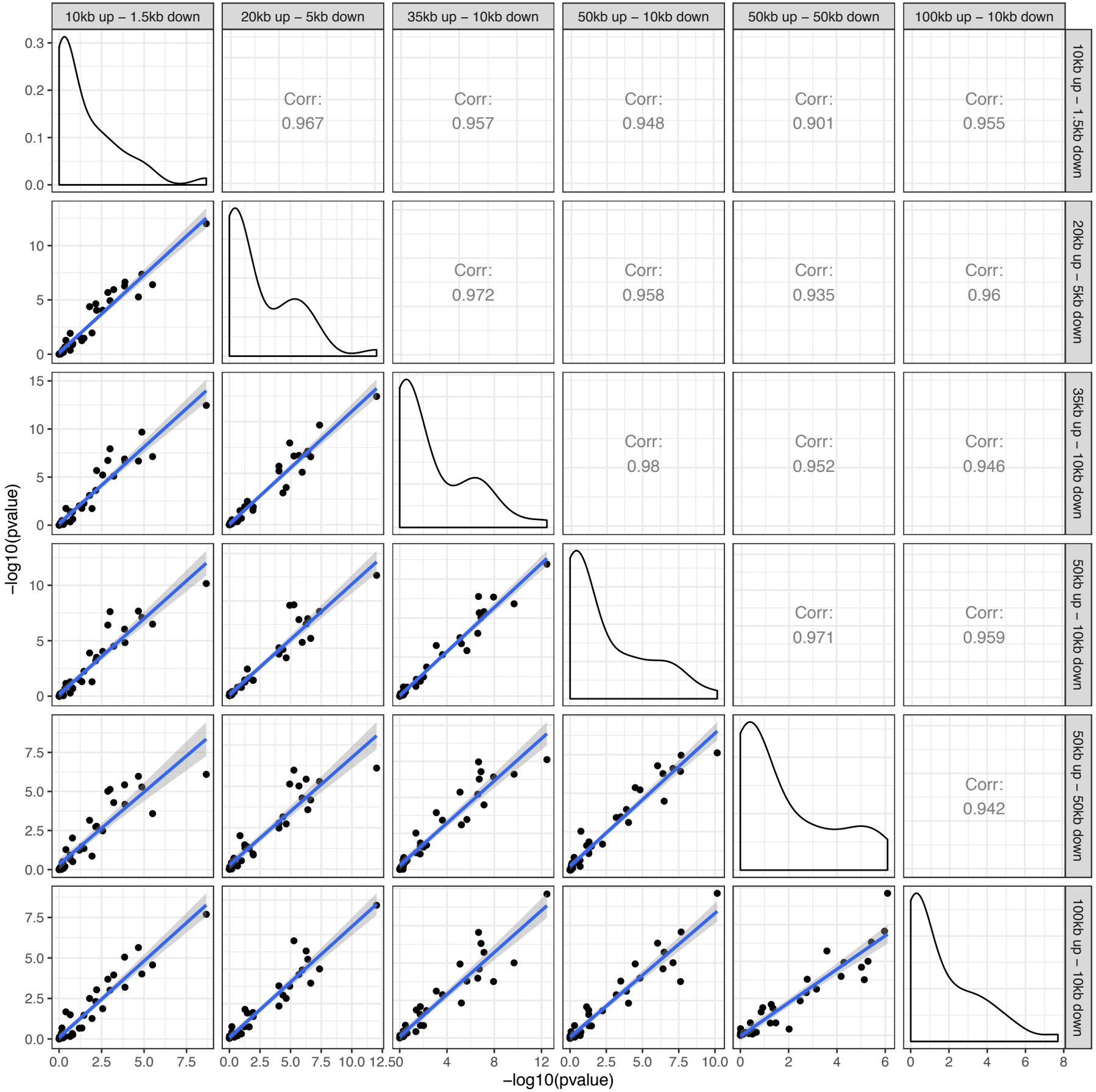
Supplementary Figure 11: Jaccard index for the top 10% most specific genes in each cell type in the mouse nervous system (Zeisel et al. 2018). Jaccard index were calculated between the top 10% most specific genes in each cell type from the mouse nervous system (Zeisel et al. 2018).



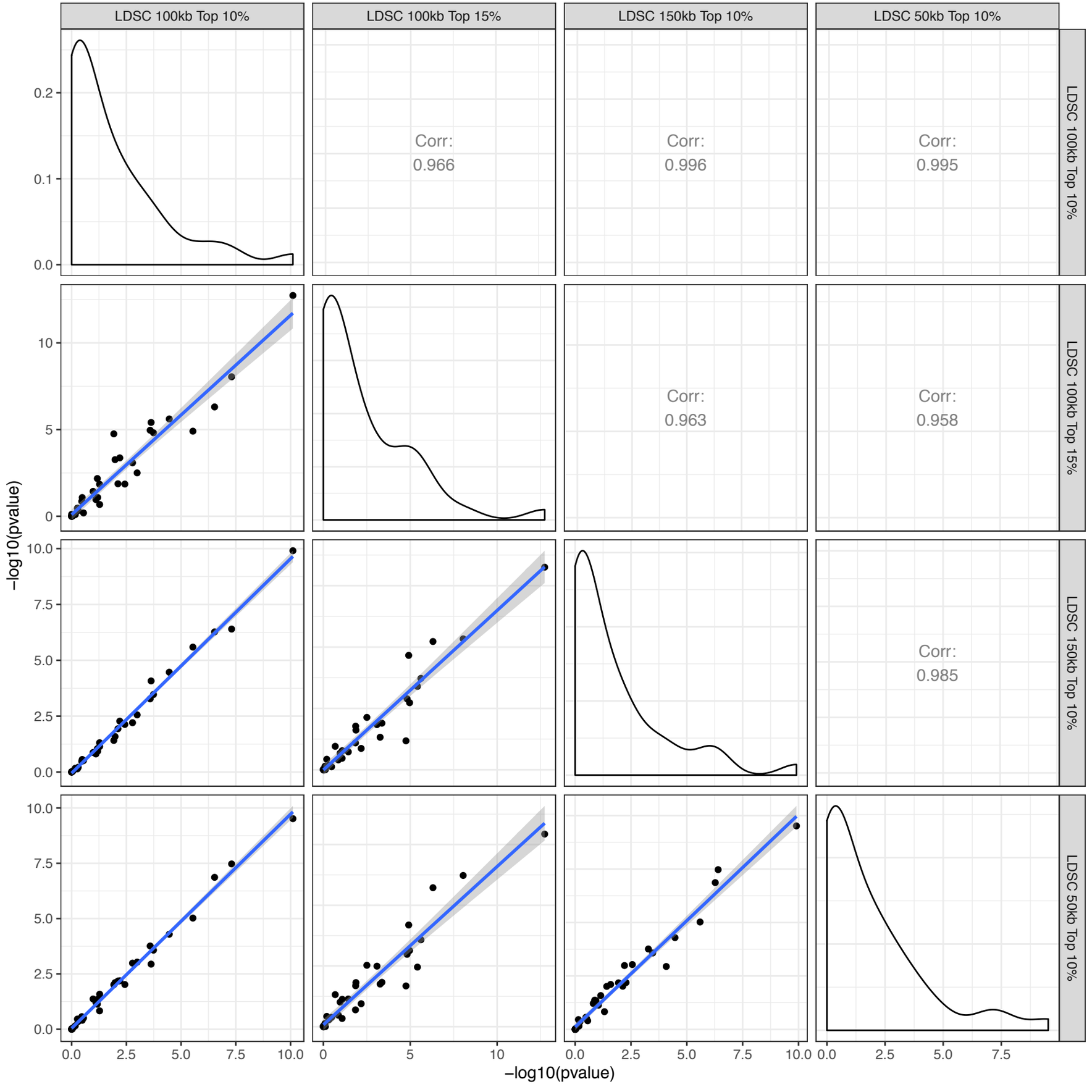
Supplementary Figure 12: Correlation between beta coefficient and significance level. Histograms of the spearman rank correlations between effect size (beta coefficient) and significance ($-\log_{10}P$) computed for each trait in the Zeisel dataset. The effect sizes are strongly correlated with the significance level of the cell type with values ranging from 0.999 to 1 using MAGMA and 0.953 to 1 with LDSC.



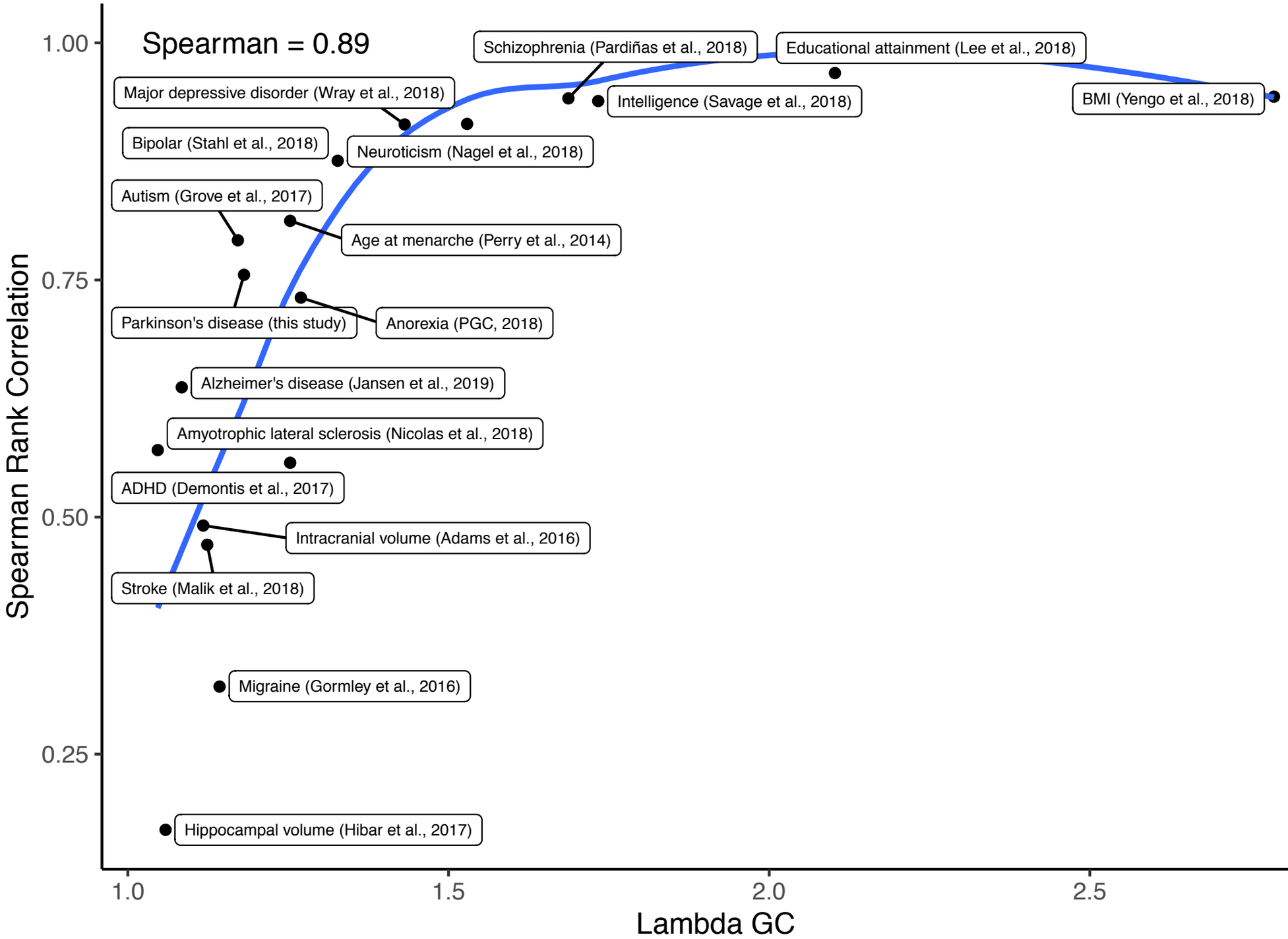
Supplementary Figure 13: Number of MAGMA associations with P<0.05 using permuted gene-level genetic associations. Gene labels were randomly permuted a thousand times for the schizophrenia MAGMA gene-level genetic associations (39 cell types * 1000 permuted labels=39,000 associations with permuted gene 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 (39,000*0.05=1950).



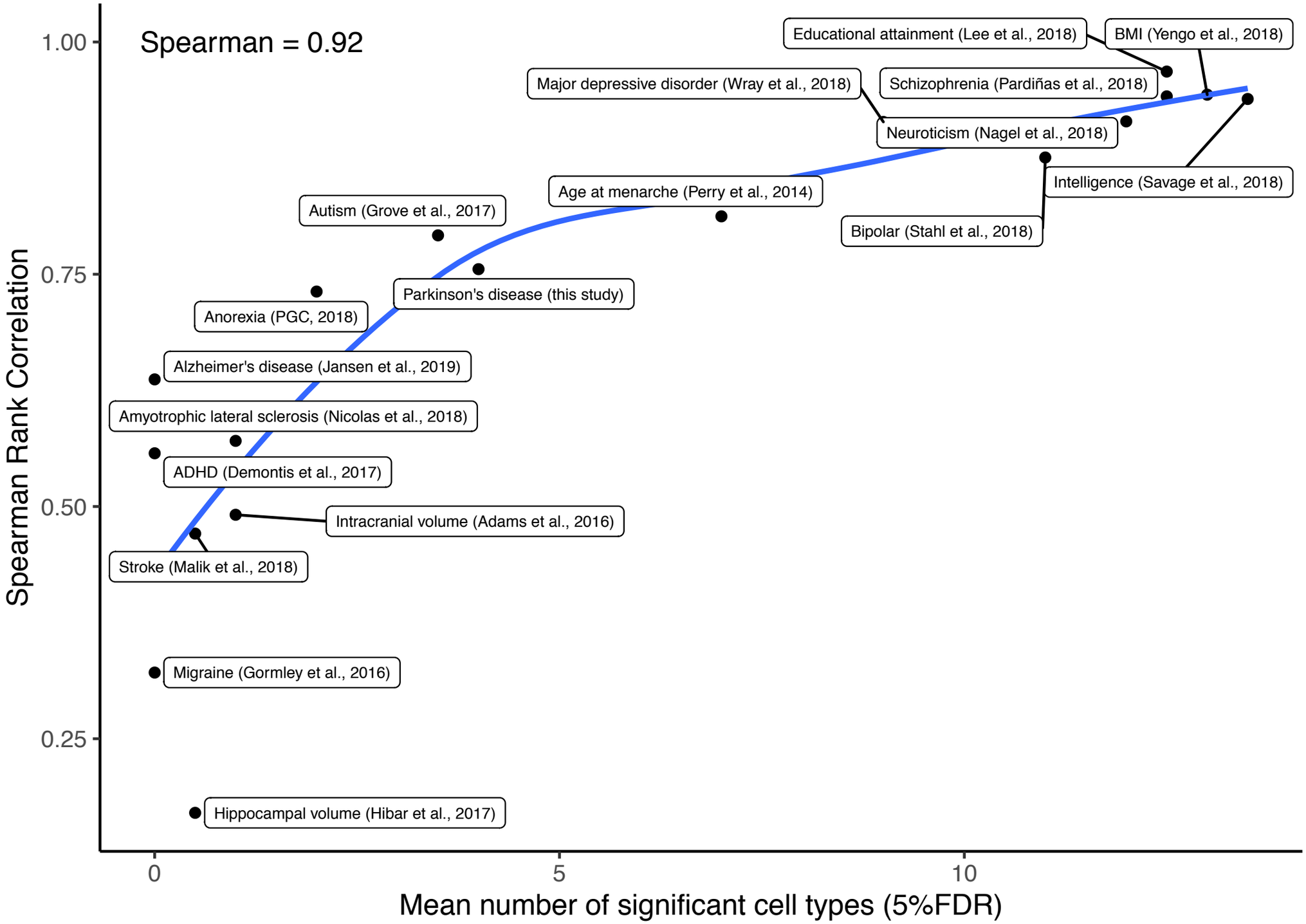
Supplementary Figure 14: Correlation in schizophrenia cell type association strengths with different window sizes using MAGMA. Pearson correlations of the cell type association strength ($-\log_{10}P$) across different window sizes using MAGMA. The diagonal shows the distribution of the ($-\log_{10}P$) for each window size.



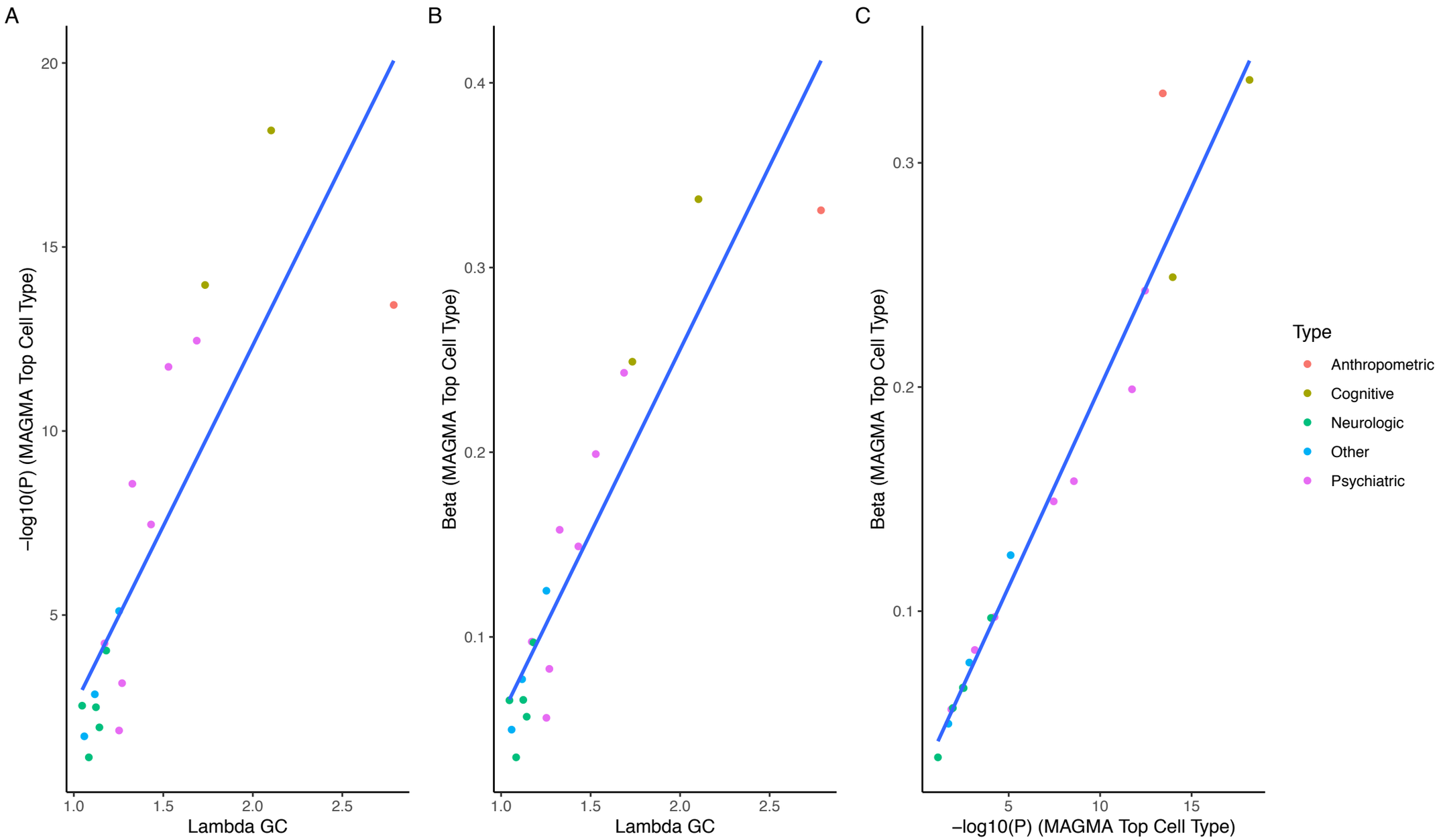
Supplementary Figure 15: Correlation in schizophrenia cell type association strengths with different window sizes and percentages of most specific genes using LDSC. Pearson correlations of the cell type association strength ($-\log_{10}P$) across different window sizes and percentages of most specific genes using LDSC. The diagonal shows the distribution of the ($-\log_{10}P$) for the cell type associations using different parameters.



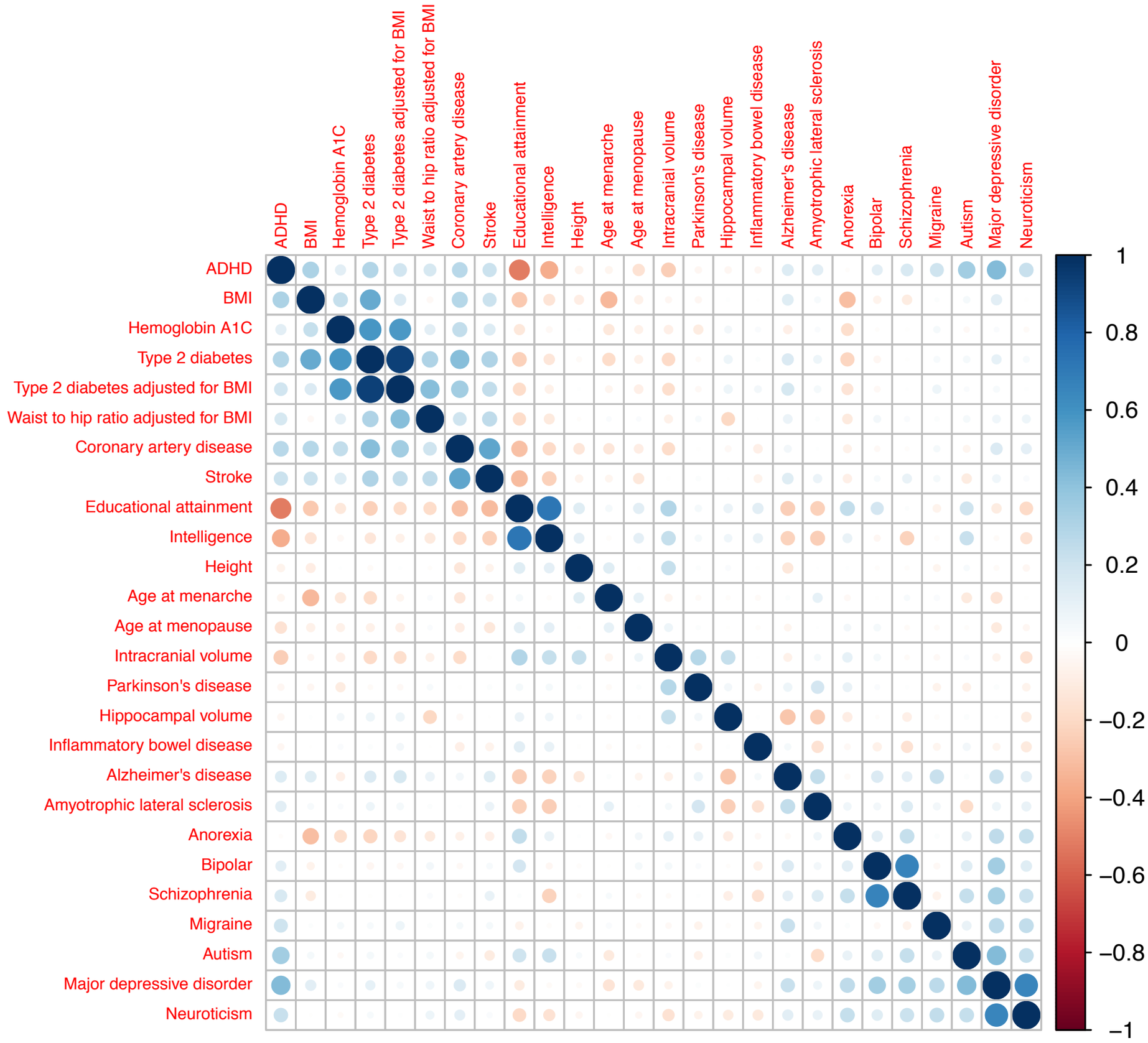
Supplementary Figure 16: Correlation between λ_{GC} and similarity in cell type ordering between MAGMA and LDSC. LDSC was used to obtain λ_{GC} (a measure of the deviation of the GWAS statistics from the expected) for each GWAS. Spearman rank correlation was used to test for similarity in association strength ($-\log_{10}P$) between MAGMA and LDSC for each GWAS among 39 cell types from the nervous system.



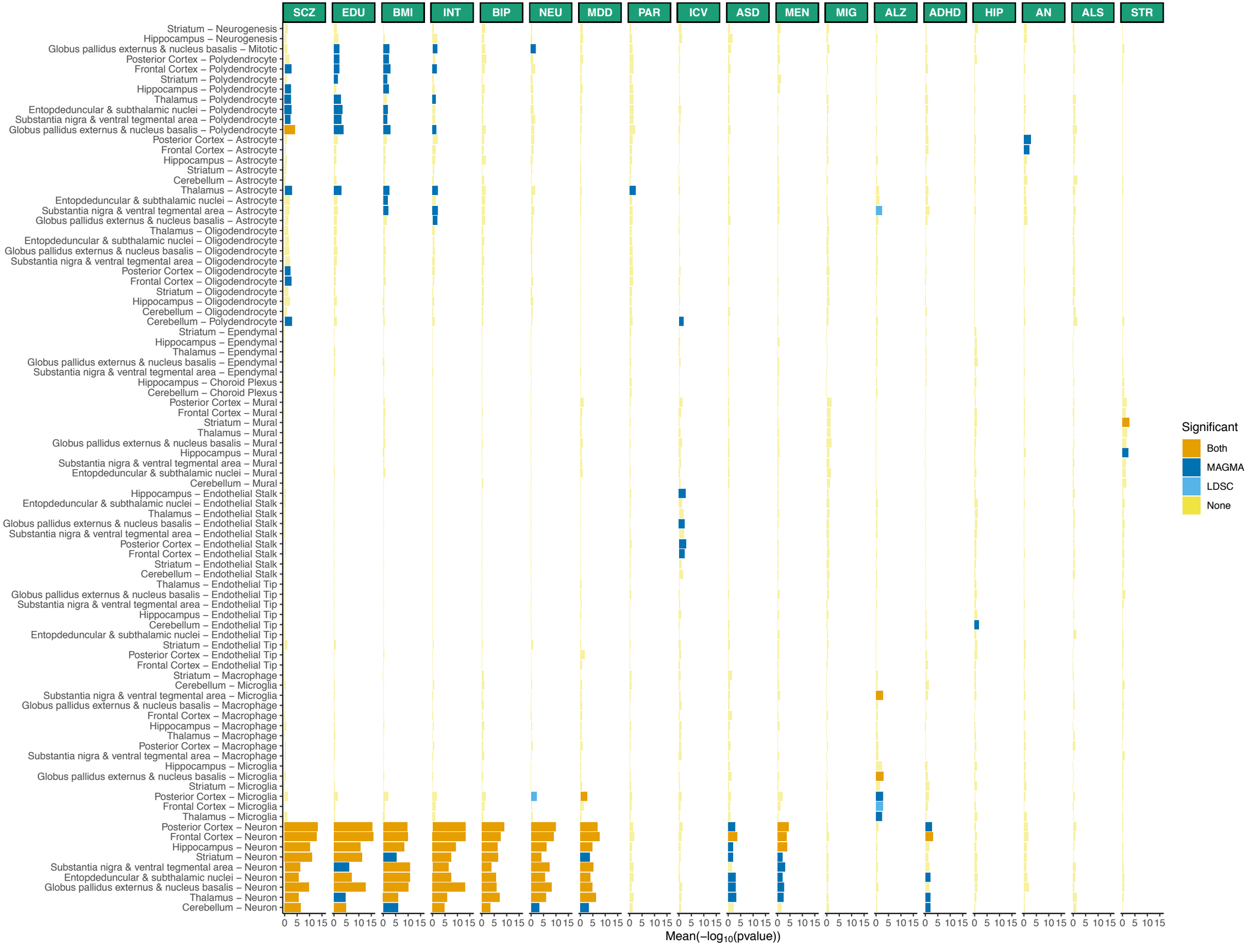
Supplementary Figure 17: Correlation between mean number of significant cell types and similarity in cell type ordering between MAGMA and LDSC. The mean number of cell types was obtained by taking the average of the number of cell types that were significantly associated with each trait (FDR<5%) using MAGMA and LDSC. Spearman rank correlation was used to test for similarity in association strength ($-\log_{10}P$) between MAGMA and LDSC among 39 cell types from the nervous system.



Supplementary Figure 18: The GWAS λ_{GC} is correlated with the strength of association of the top cell type in the Zeisel 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 Zeisel cell type associated with the trait ($-\log_{10}(P_{MAGMA})$). Spearman correlation=0.88 (**A**). Scatter plot of the λ_{GC} of each GWAS vs the effect size of the top Zeisel cell type associated with the trait ($-\log_{10}(P_{MAGMA})$). Spearman correlation=0.9 (**B**). Scatter plot of the strength of association of the top Zeisel cell type ($-\log_{10}(P_{MAGMA})$) of each GWAS vs the effect size of the top Zeisel cell type. Spearman correlation=0.996 (**C**).



Supplementary Figure 19: Genetic correlation across traits. The genetic correlation across traits were computed using LDSC. Traits are ordered based on hierarchical clustering.



Supplementary Figure 20: Replication of cell type – trait associations in 88 cell types from 9 different brain regions. The mean strength of association ($-\log_{10}P$) of MAGMA and LDSC is shown and the bar color indicates whether the cell type is significantly associated with both methods, one method or none (significance threshold: 5% false discovery rate). 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).

Supplementary Note:

Genetic correlations among complex traits

We estimated the genetic correlations (r_g) between the 26 traits. We confirmed prior reports [1,2] 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 (Supplementary Fig. 19 and Supplementary Table 4). Parkinson's disease was genetically correlated with intracranial volume [3] (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). These results indicate that there is substantial genetic heterogeneity across traits, which is a necessary (but not sufficient) condition for trait associations with different tissues or cell types.

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 telencephalon projecting excitatory neurons were enriched for GO terms related to neurogenesis, synapses, and voltage-gated channels (Supplementary Table 16), suggesting that these functions may be fundamental to schizophrenia. Similarly, genes highly associated with educational attainment, intelligence, bipolar disorder, neuroticism, BMI, anorexia and MDD and highly specific to their most associated cell types were enriched in terms related to neurogenesis, synaptic processes and voltage-gated channels (Supplementary Table 16). In contrast, genes highly associated with stroke and specific to vascular cells were enriched in terms related to vasculature development, while genes highly associated with ALS and peripheral sensory neurofilament neurons were enriched in terms related to lysosomes.

Genes highly associated with Parkinson's disease and highly specific to cholinergic and monoaminergic neurons were significantly enriched in terms related to endosomes and synapses (Supplementary Table 16). Similarly, genes highly specific to oligodendrocytes and Parkinson's disease were enriched in endosomes. These results support the hypothesis that the endosomal pathway plays an important role in the etiology of Parkinson's disease [4].

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.

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

We compared our findings for Alzheimer's disease (Supplementary Table 2, Fig. 4B, Supplementary Fig. 20) with a recent study that performed differential expression analysis at the cell type level between 24 Alzheimer's cases and 24 controls [5] (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, we found that genes differentially expressed in microglia were

the most associated with Alzheimer's disease genetics (Supplementary Table 17), 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.

Association strength between mice and human

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.

Supplementary discussion

We replicated and extended our previous findings for schizophrenia [6]. We found the most significant associations for neurons located in the cortex, hippocampus and striatum (Fig. 2A, 3) in multiple independent datasets, and showed that these neuronal cell types can be prioritized among neurons (Extended Data 4, 8 and 10). These results are consistent with the strong schizophrenia heritability enrichment observed in open chromatin regions from: human dorsolateral prefrontal cortex [7]; human cortical, striatal and hippocampal neurons [8]; and mouse open chromatin regions from cortical excitatory and inhibitory neurons [9]. This degree of replication in independent datasets from multiple groups strongly implicates these cell types in the etiology of schizophrenia.

Two theories for the selective vulnerability of neuronal populations in Parkinson's disease currently exist: the "spread Lewy pathology model" which assumes cell-to-cell contacts enabling spreading of prion-like α-synuclein aggregates [10]; and the "threshold theory" [11,12] 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 do not know if all cholinergic and monoaminergic neurons show degeneration or functional impairment. However, analysis of the cellular mechanisms driving the association of cholinergic and monoaminergic neurons with Parkinson's disease revealed endosomal trafficking as a plausible common pathogenic mechanism (Supplementary Table 16).

References

1. Anttila, V. *et al.* Analysis of shared heritability in common disorders of the brain. *Science (80-.).* **360**, (2018).
2. Lee, S. H. *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).
3. Nalls, M. A. *et al.* Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* **18**, 1091–1102 (2019).
4. Perrett, R. M., Alexopoulou, Z. & Tofaris, G. K. The endosomal pathway in Parkinson's disease. *Molecular and Cellular Neuroscience* **66**, 21–28 (2015).
5. Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).

6. Skene, N. G. *et al.* Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.* **50**, 825–833 (2018).
7. Bryois, J. *et al.* Evaluation of chromatin accessibility in prefrontal cortex of individuals with schizophrenia. *Nat. Commun.* **9**, (2018).
8. Fullard, J. F. *et al.* An atlas of chromatin accessibility in the adult human brain. *Genome Res.* **28**, 1243–1252 (2018).
9. Hook, P. W. & McCallion, A. S. Heritability enrichment in open chromatin reveals cortical layer contributions to schizophrenia. *bioRxiv* (2018).
10. Recasens, A. & Dehay, B. Alpha-synuclein spreading in Parkinson's disease. *Front. Neuroanat.* **8**, (2014).
11. Engelender, S. & Isacson, O. The threshold theory for Parkinson's disease. *Trends in Neurosciences* **40**, 4–14 (2017).
12. Surmeier, D. J., Obeso, J. A. & Halliday, G. M. Selective neuronal vulnerability in Parkinson disease. *Nature Reviews Neuroscience* **18**, 101–113 (2017).