

In the format provided by the authors and unedited.

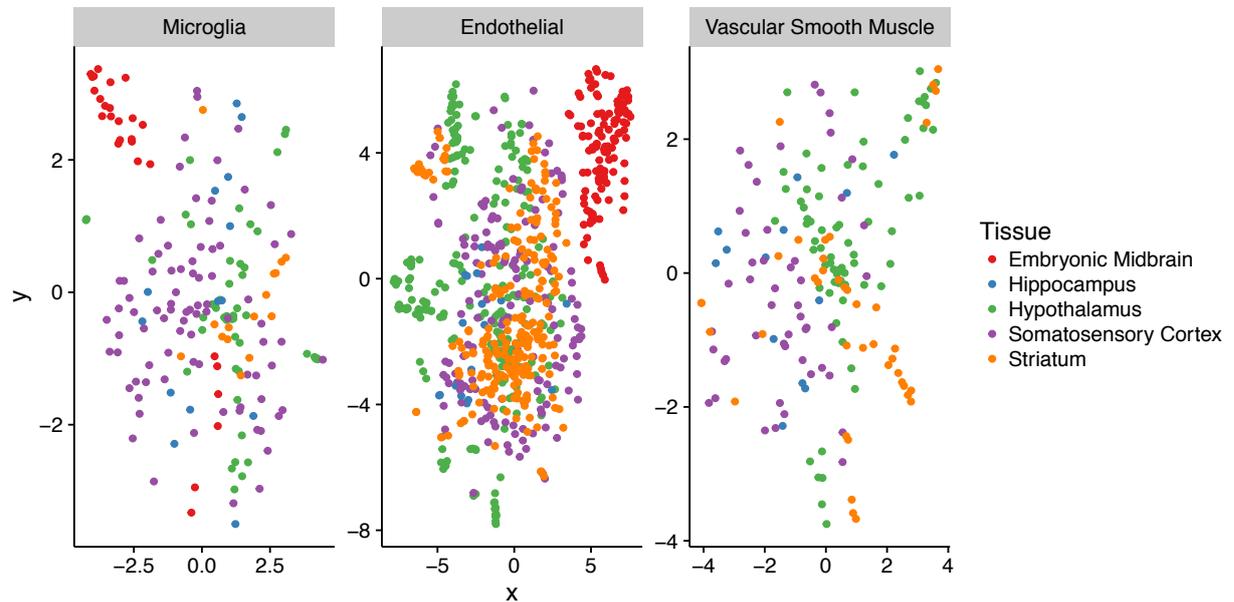
Genetic identification of brain cell types underlying schizophrenia

Nathan G. Skene^{1,2,11}, Julien Bryois^{3,11}, Trygve E. Bakken⁴, Gerome Breen^{5,6}, James J. Crowley⁷, H  l  na A. Gaspar^{5,6}, Paola Giusti-Rodr  guez⁷, Rebecca D. Hodge⁴, Jeremy A. Miller⁴, Ana B. Mu  oz-Manchado¹, Michael C. O'Donovan⁸, Michael J. Owen⁸, Antonio F. Pardi  as⁸, Jesper Ryge⁹, James T. R. Walters⁸, Sten Linnarsson¹, Ed S. Lein⁴, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium¹⁰, Patrick F. Sullivan^{3,7*} and Jens Hjerling-Leffler^{1*}

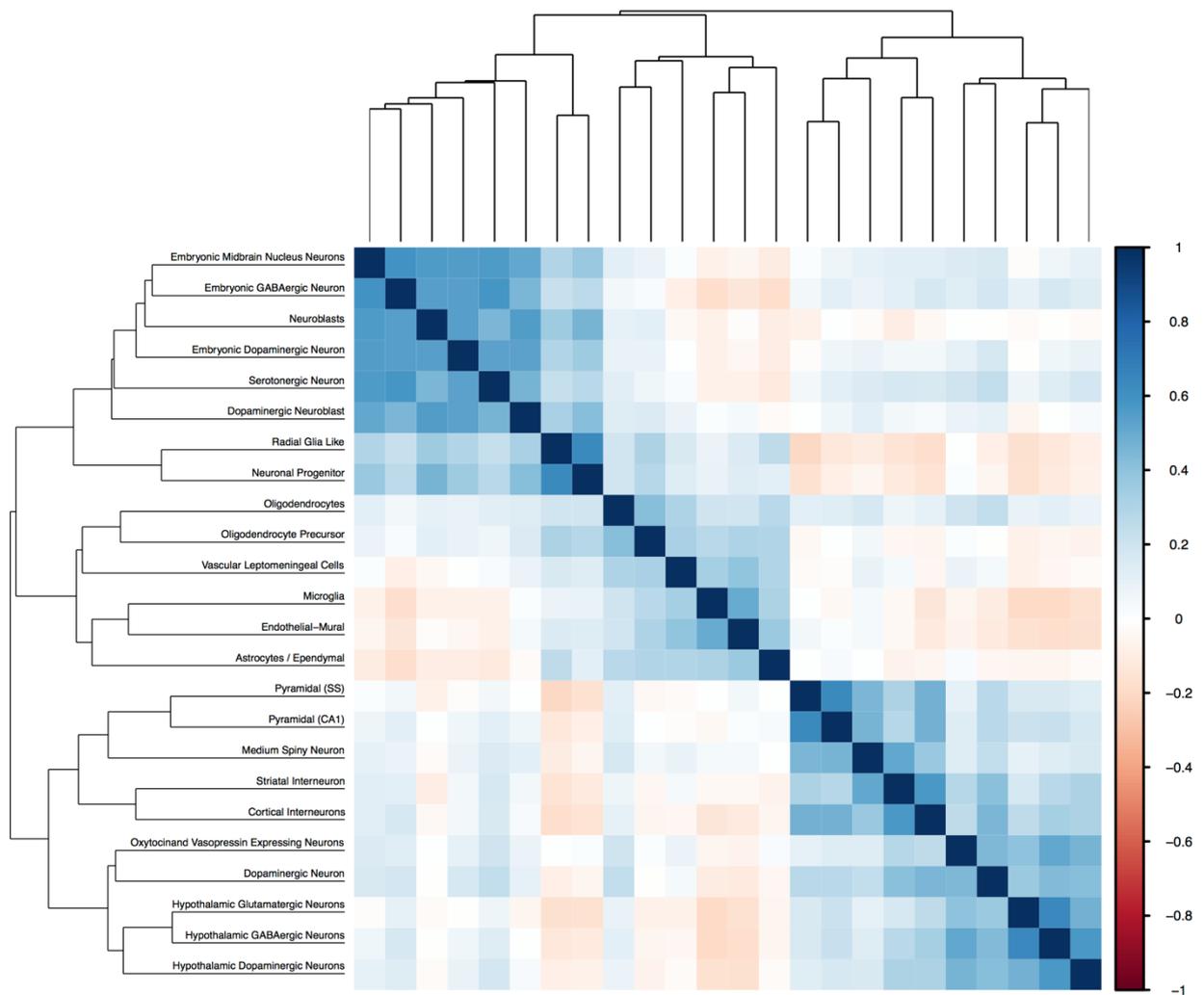
¹Laboratory of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden. ²UCL Institute of Neurology, Queen Square, London, UK. ³Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. ⁴Allen Institute for Brain Science, Seattle, WA, USA. ⁵King's College London, Institute of Psychiatry, Psychology and Neuroscience, MRC Social, Genetic and Developmental Psychiatry (SGDP) Centre, London, UK. ⁶National Institute for Health Research Biomedical Research Centre, South London and Maudsley National Health Service Trust, London, UK. ⁷Department of Genetics, University of North Carolina, Chapel Hill, NC, USA. ⁸MRC Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, UK. ⁹Brain Mind Institute,   cole Polytechnique F  d  rale de Lausanne, Lausanne, Switzerland. ¹⁰A list of members and affiliations appears in the Supplementary Note.

¹¹These authors contributed equally: Nathan G. Skene, Julien Bryois. *e-mail: patrick.sullivan@ki.se; jens.hjerling-leffler@ki.se

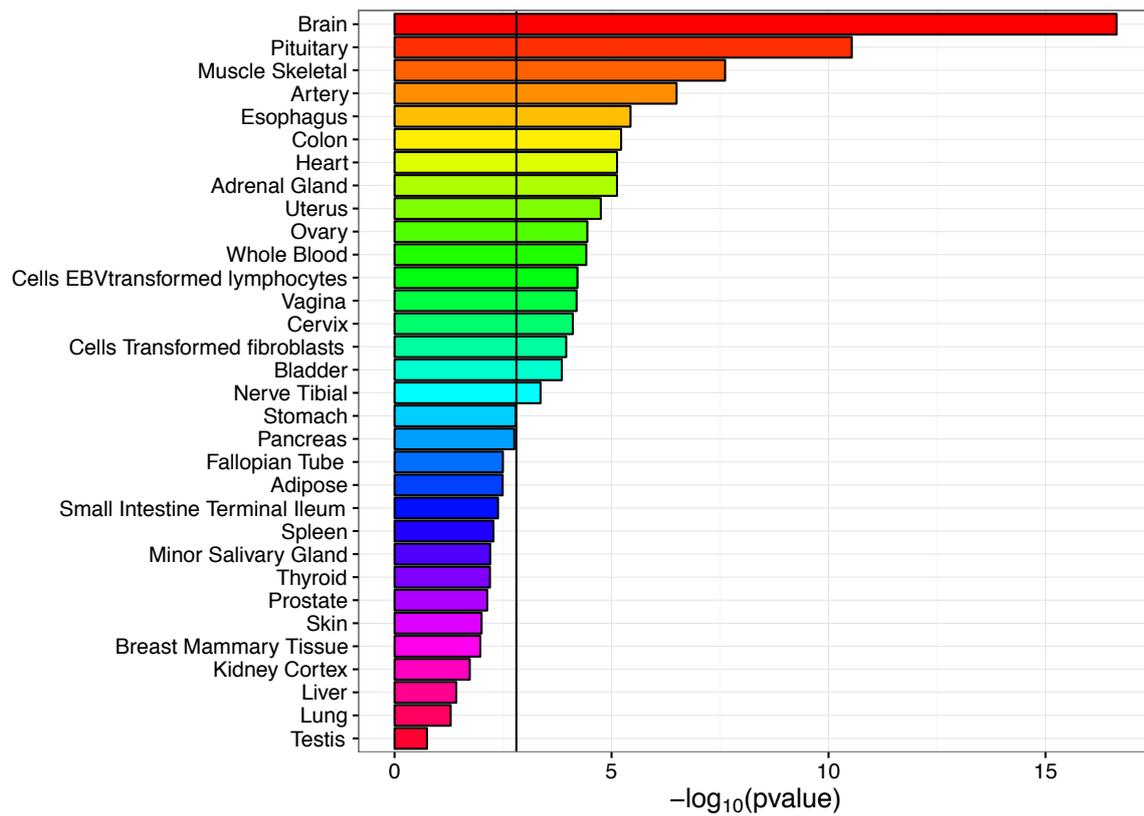
Supplementary Figures



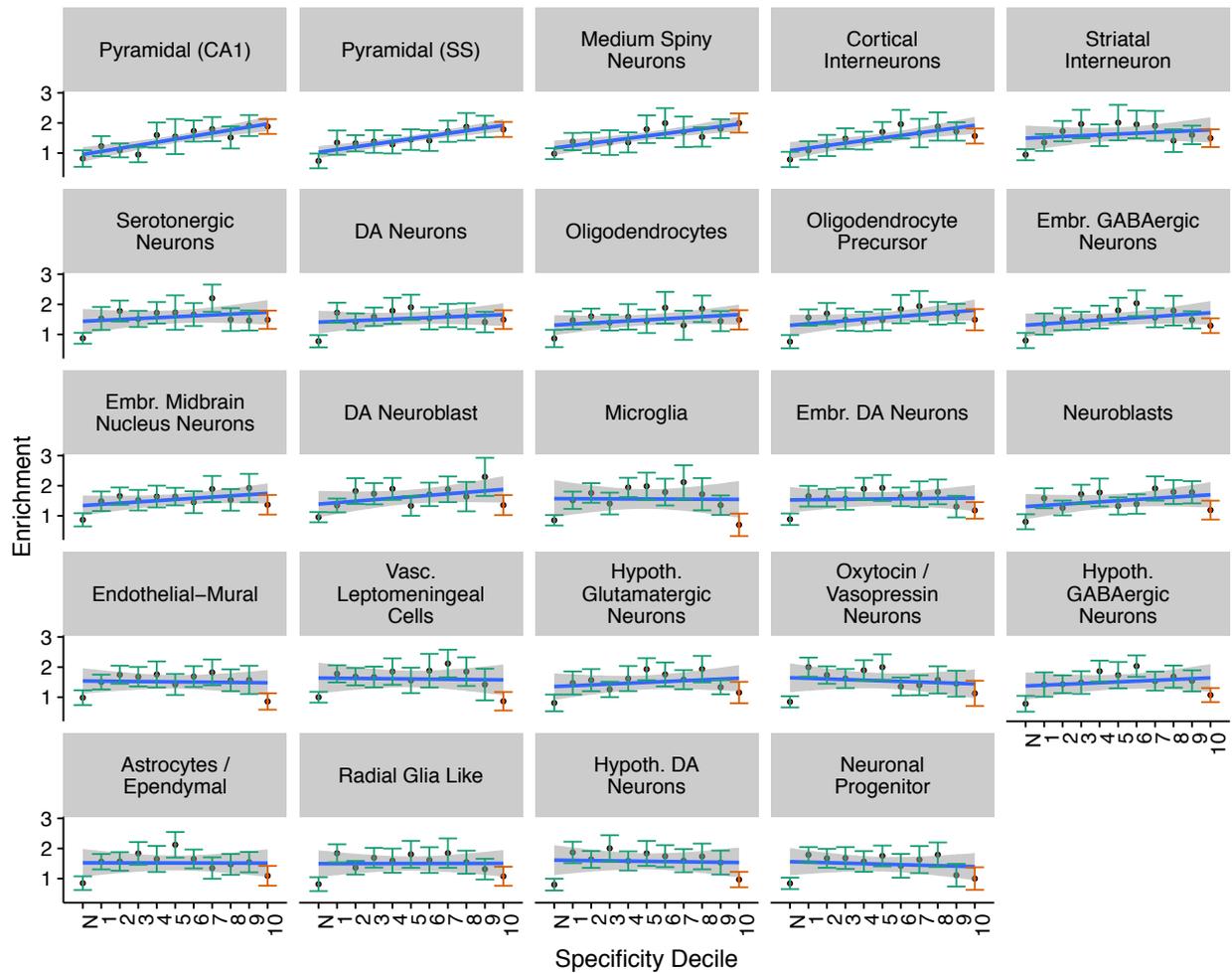
Supplementary Figure 1. tSNE plot of three cell types shared between brain regions from the KI superset. The plots show clustering across regions. Microglia, endothelial cells, and vascular smooth muscle cells were selected on the basis that there is little prior expectation for them to have regional differences in expression. We found that embryonic midbrain cells cluster separately as expected as these cells were obtained from embryonic tissue whereas all other samples were from adolescent mice. The cells from the other datasets were largely overlapping confirming that little to no batch effects exist in the data.



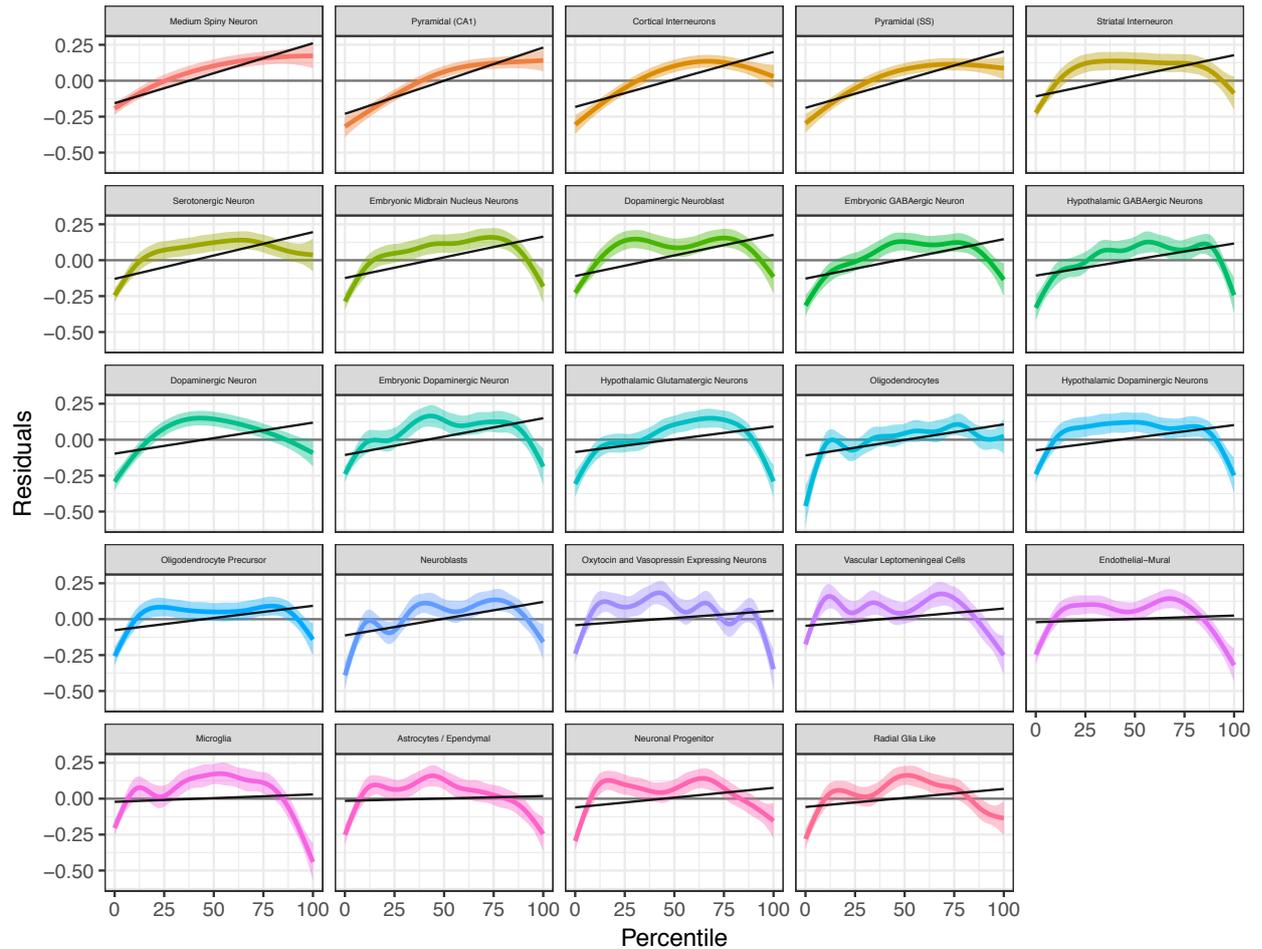
Supplementary Figure 2. Pearson correlation of the binned measure of cell type-specific expression (S , defined in Online Methods) for mouse brain cell types in the KI level 1 dataset. To illustrate the overall structure of mouse brain scRNAseq dataset, we show a heat map and clustering of the brain cell types identified in the KI Level 1 data. The major divisions are embryonic/progenitor cells (upper left), support cells (middle; e.g., oligodendrocytes and microglia), and mature cells (lower right). The major division of the mature cells include pyramidal cells/medium spiny neurons, interneurons, and “specialty” neurons (i.e., dopaminergic, GABAergic, and glutamatergic).



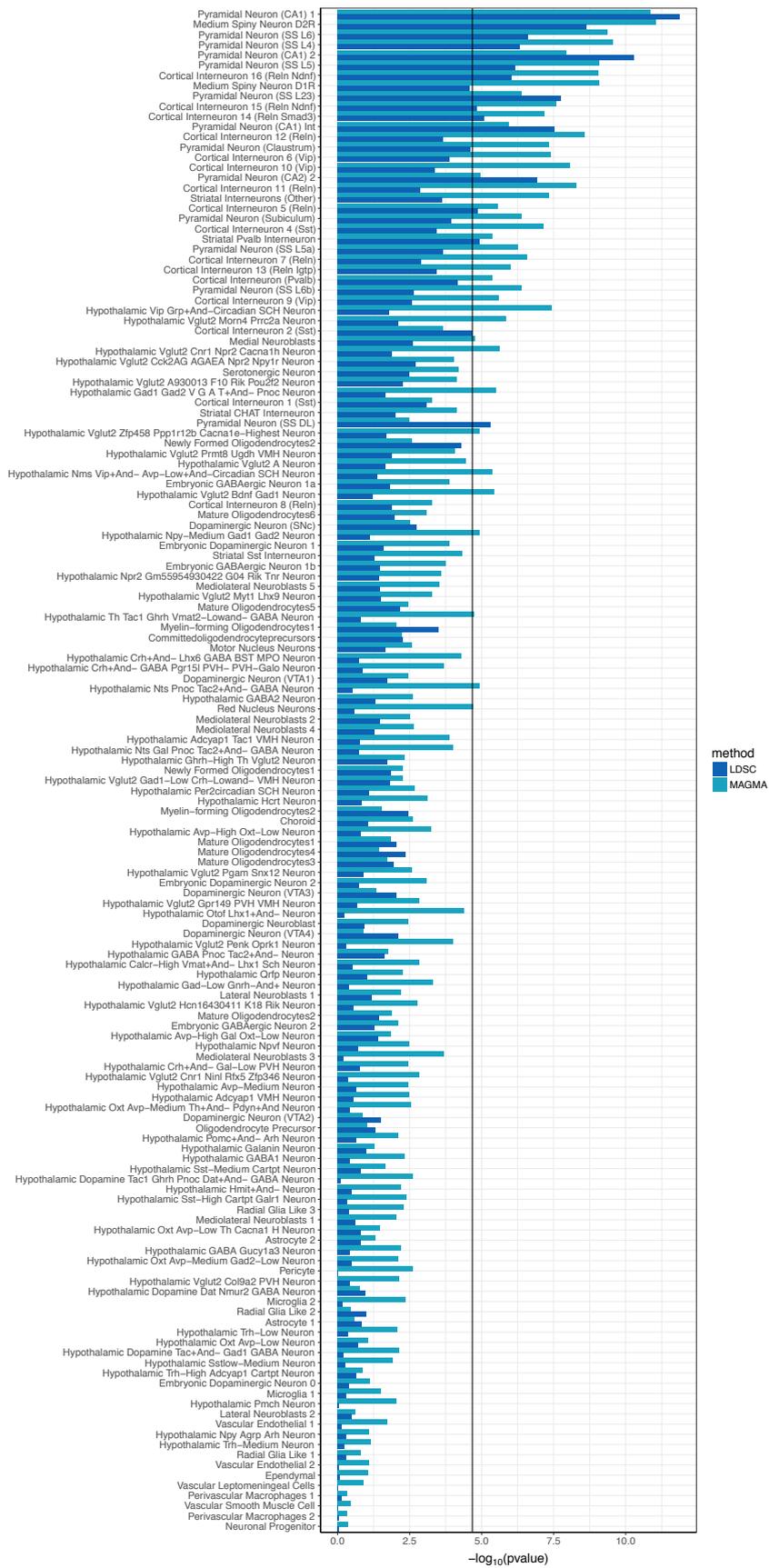
Supplementary Figure 3. MAGMA enrichment of CLOZUK schizophrenia GWA results in relation to human tissue-specific gene expression (from GTEx). Brain and pituitary are most associated, but there are significant associations with multiple non-brain tissues), and tissues not believed to be etiologically involved in schizophrenia (colon, heart, uterus). Black line shows Bonferroni correction.



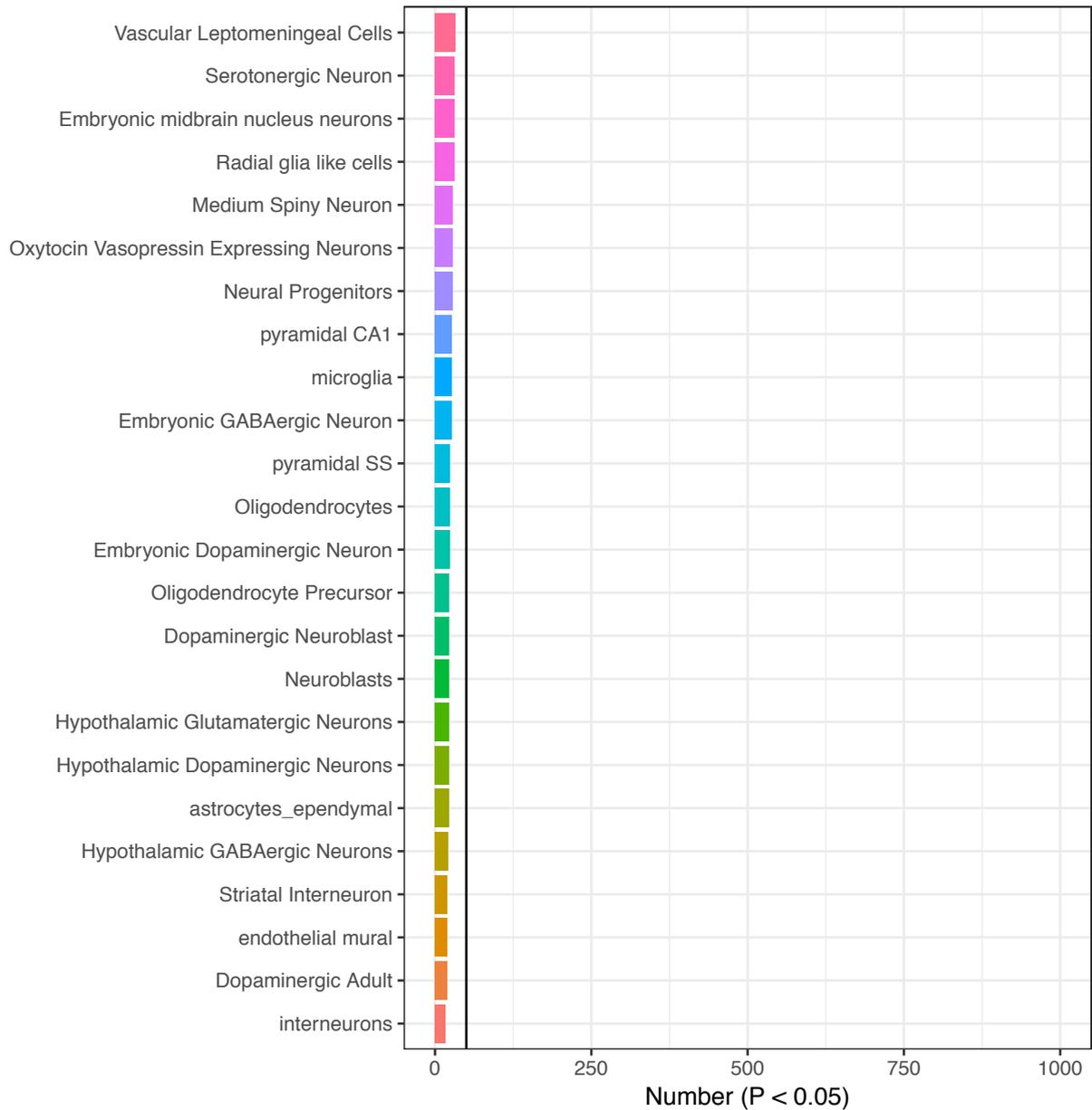
Supplementary Figure 4. LDSC schizophrenia (CLOZUK) enrichment values in each specificity decile for each of the KI level 1 cell types. Error bars indicate 95% confidence intervals. The rightmost point and its confidence intervals are marked in red as this is the decile used for reported LDSC p-values throughout this paper (rather than the p-value of the slope used for reporting MAGMA p-values). The leftmost point (marked 'N') represents all SNPs which map onto genes not expressed in MSNs. Blue line shows the linear regression slope fitted to the enrichment values. The grey boxing around the blue regression line depicts the confidence intervals of the regression line.



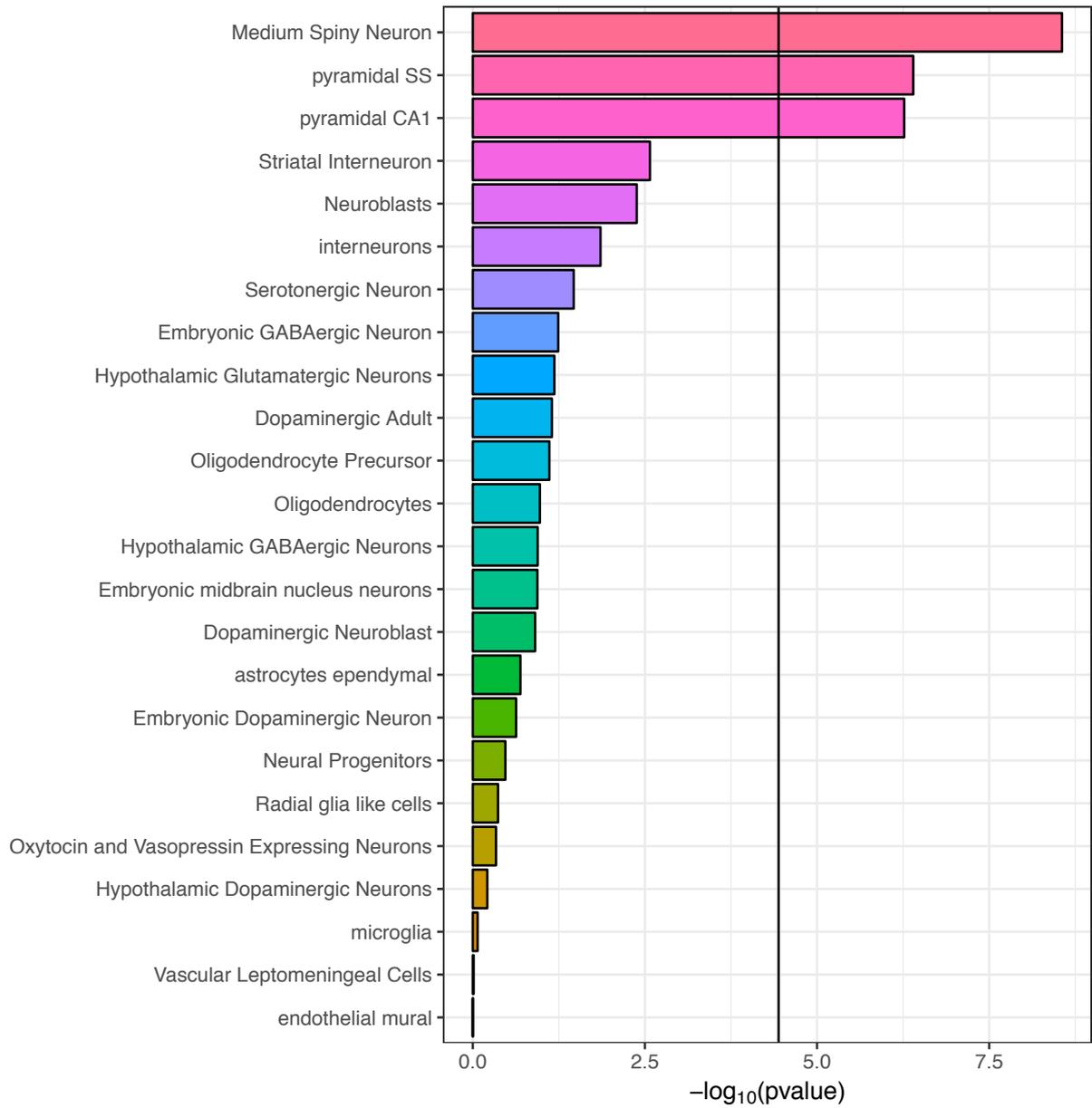
Supplementary Figure 5. MAGMA schizophrenia (CLOZUK) gene level model fit for each KI level 1 cell type. The Y-axis (residuals) was obtained by regressing the gene length, gene density and their logs from the gene-level z-score obtained from MAGMA using the CLOZUK schizophrenia GWAS. Negative residuals indicate that genes are less associated with schizophrenia, while positive residuals indicate that genes are more associated with schizophrenia. The x-axis shows the 41 binned tissue specificity measures (each bin represent a 2.5% quantile of the distribution of proportions in the cell type). The colored line shows the best non-linear fit to the data using a generalized additive model with its 95% confidence interval. The black line represents the linear regression of the residuals by the binned proportions.



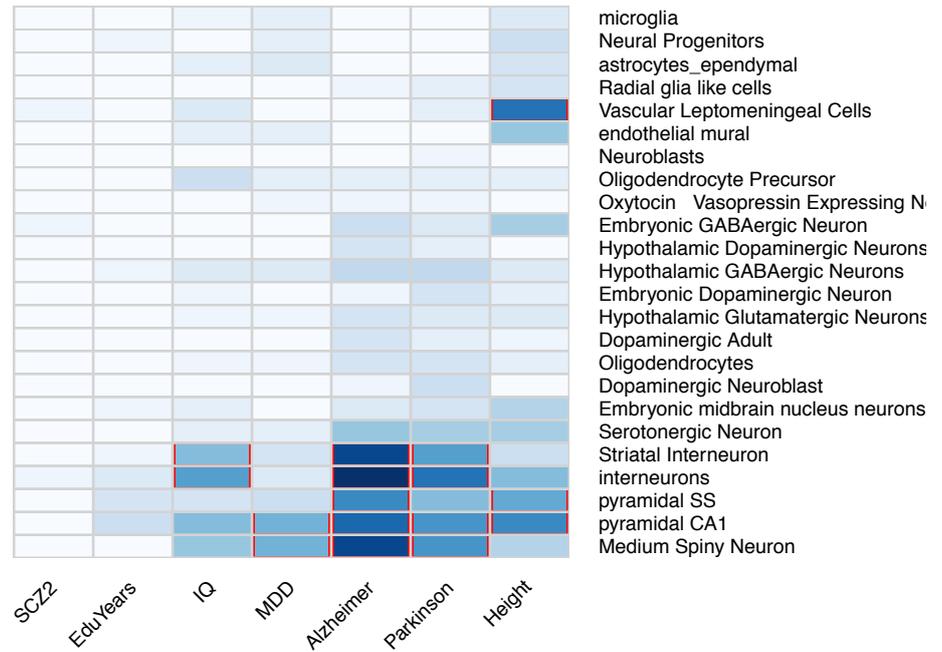
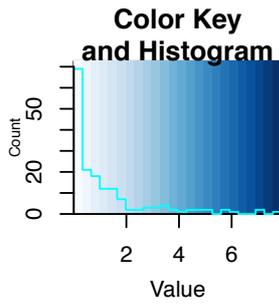
Supplementary Figure 6. Association between KI Level 2 brain cell types and schizophrenia. Cell types are ranked by the minimum average rank of LDSC and MAGMA (minimum average P-values for cell types with equal average rank). The black line represents the Bonferroni significance threshold.



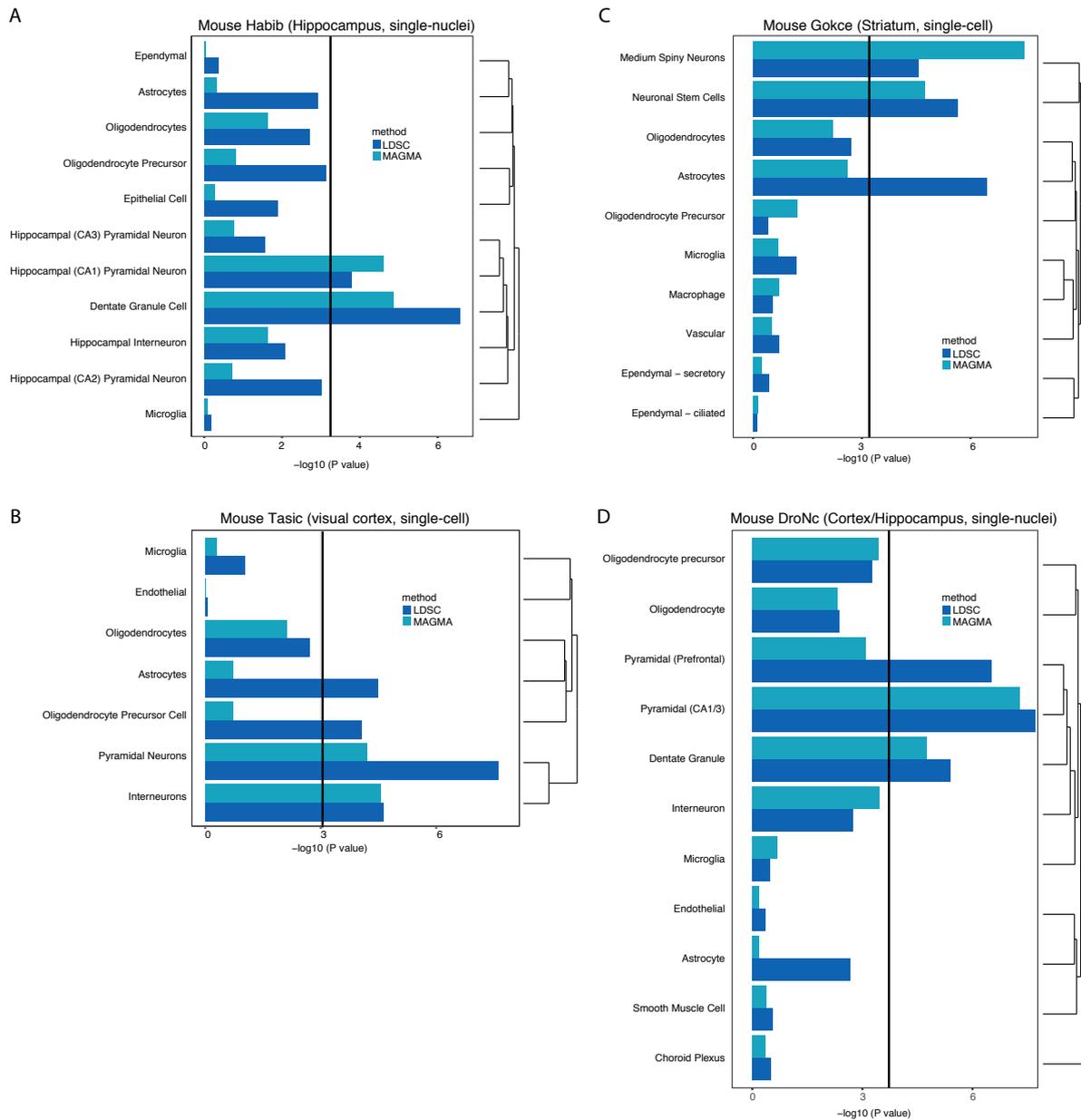
Supplementary Figure 7. The MAGMA approach is conservative. Number of associations with $P < 0.05$ after randomly permuting the gene label of the gene level association statistics a thousand times using the MAGMA approach. The black line represents the number of results expected by chance at $P < 0.05$ for 1000 permutations (50). We observed two times fewer significant results than expected by chance using MAGMA v1.04 indicating that MAGMA is conservative.



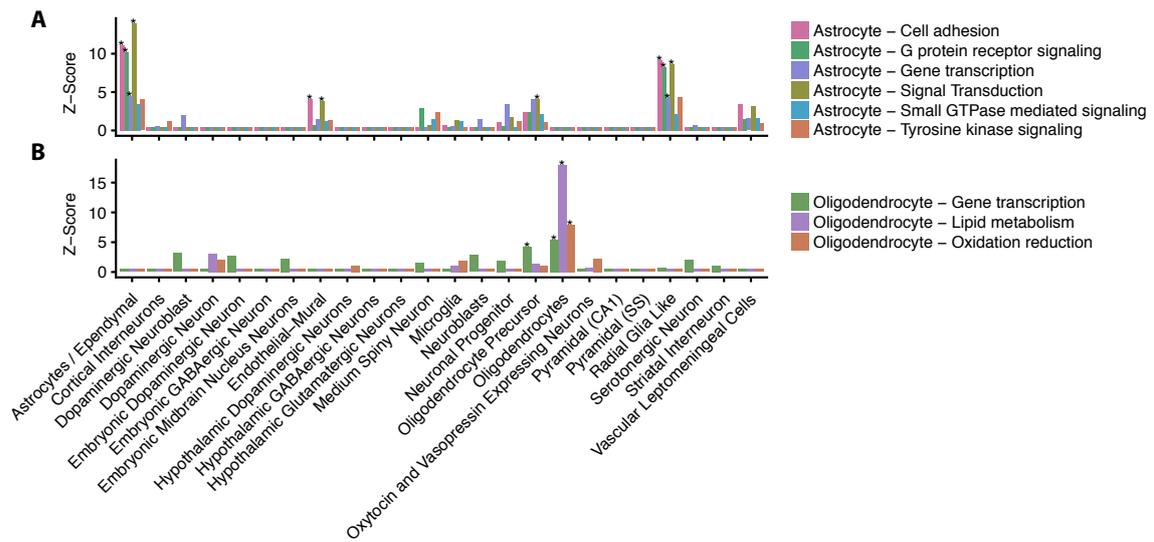
Supplementary Figure 8. Cell type schizophrenia association based on differential expression analysis. The top 10% most upregulated genes in each cell types with respect to the other 23 cell types were tested for association with schizophrenia using LDSC. The black bar represents the Bonferroni significance threshold [$=0.05/((24+149)*8)$]. Results are similar to our approach based on specificity measures except that interneurons are not significant using the differential expression approach.



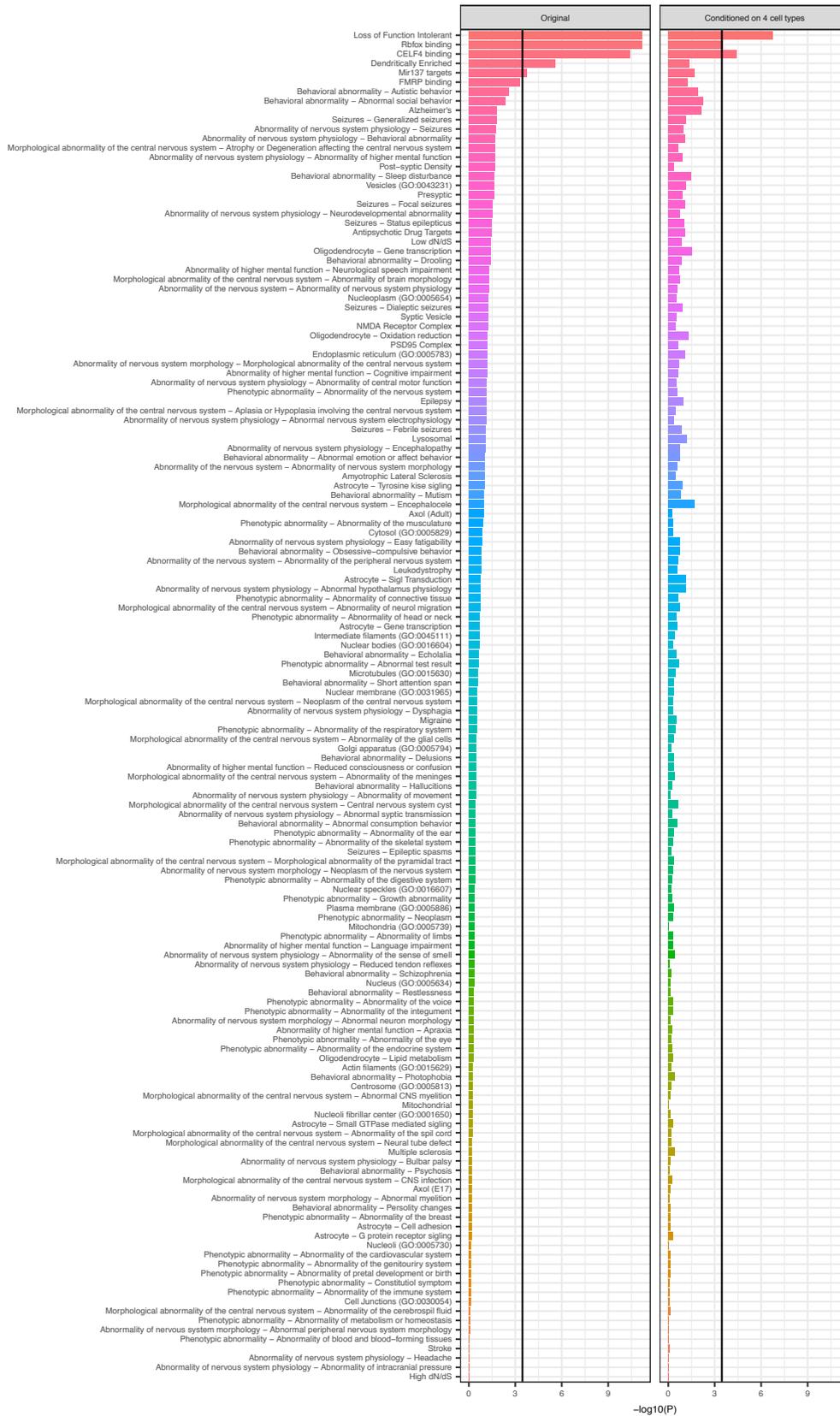
Supplementary Figure 9. Testing cell type specificity for schizophrenia and other GWA studies. Heat map representing the $-\log_{10}(P\text{-value})$ of the difference in MAGMA regression coefficient between CLOZUK and seven other traits for each cell type (height and six brain-related studies). The red box represent Bonferroni significant results [$=0.05/(24*7)$]. No cell types were differently associated between CLOZUK and the previous schizophrenia GWA (SCZ2).



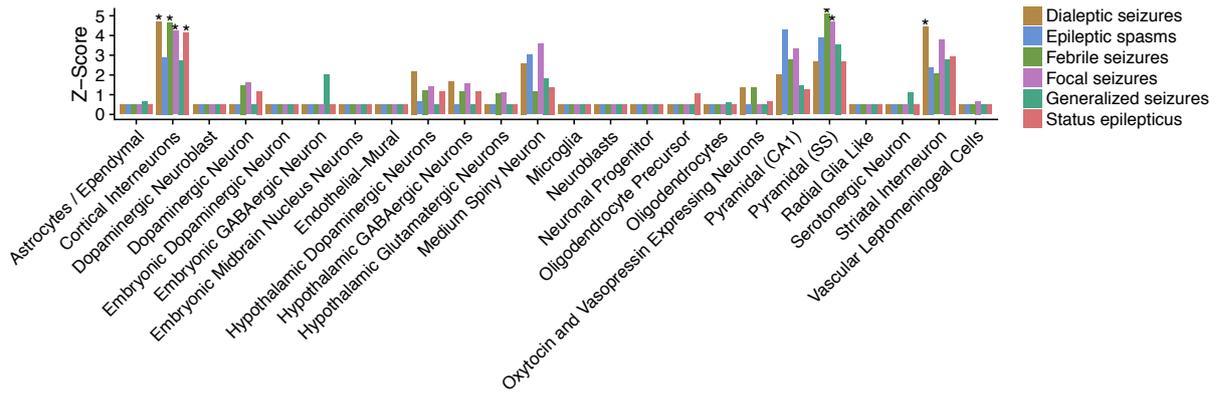
Supplementary Figure 10. Confirmation of enrichment of common variant CLOZUK schizophrenia GWA results from the KI mouse data in independent mouse brain studies. The KI Level 1 findings connect the CLOZUK schizophrenia results to hippocampal CA1 pyramidal neurons, cortical pyramidal neurons, cortical interneurons, and medium spiny neurons. (A) snRNAseq from mouse hippocampus²¹ showing enrichment of hippocampal CA1 pyramidal cells. (B) scRNAseq from mouse cortex⁵ demonstrating enrichment of cortical pyramidal neurons and cortical interneurons. (C) scRNAseq from mouse striatum with enrichment of “striatal neurons”, which were predominantly medium spiny neurons²⁰. There was also enrichment for hippocampal dentate granule cells in Figure 6a and migratory neural precursors in striatum in Figure 6c, but these were not included in the larger KI dataset. (D) snRNAseq data set from²⁴ showing enrichment in pyramidal and dentate granule cells from hippocampus but not cortical pyramidal or interneurons cells. This data set also show a near significant enrichment of OPCs.



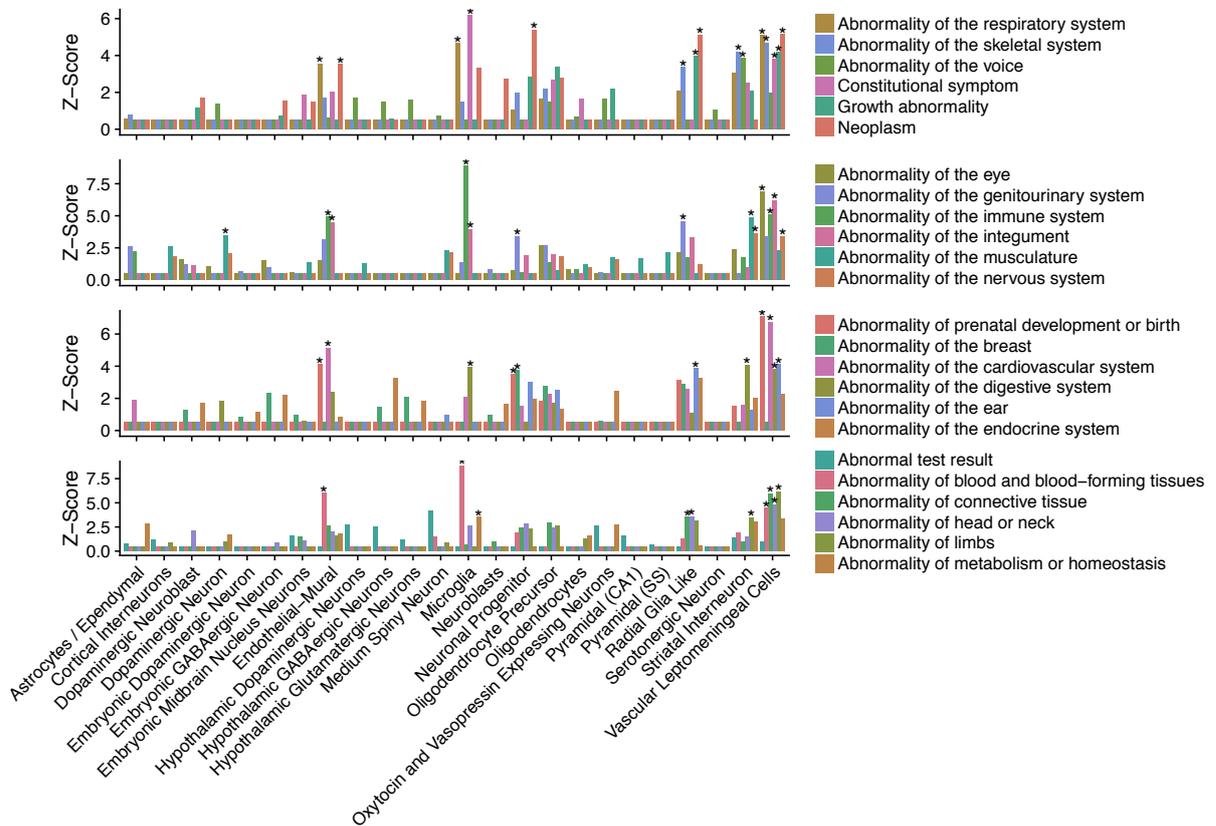
Supplementary Figure 12: Cell types associated with glial gene sets that had previously been shown to be associated with schizophrenia. (A) Expert curated astrocyte specific functional gene sets. (B) Expert curated oligodendrocyte specific functional gene sets. Enrichments were tested with EWCE. Stars mark significance at $p < 0.05$ after correcting for multiple testing with the Benjamini-Hochberg method over all gene sets tested with EWCE.



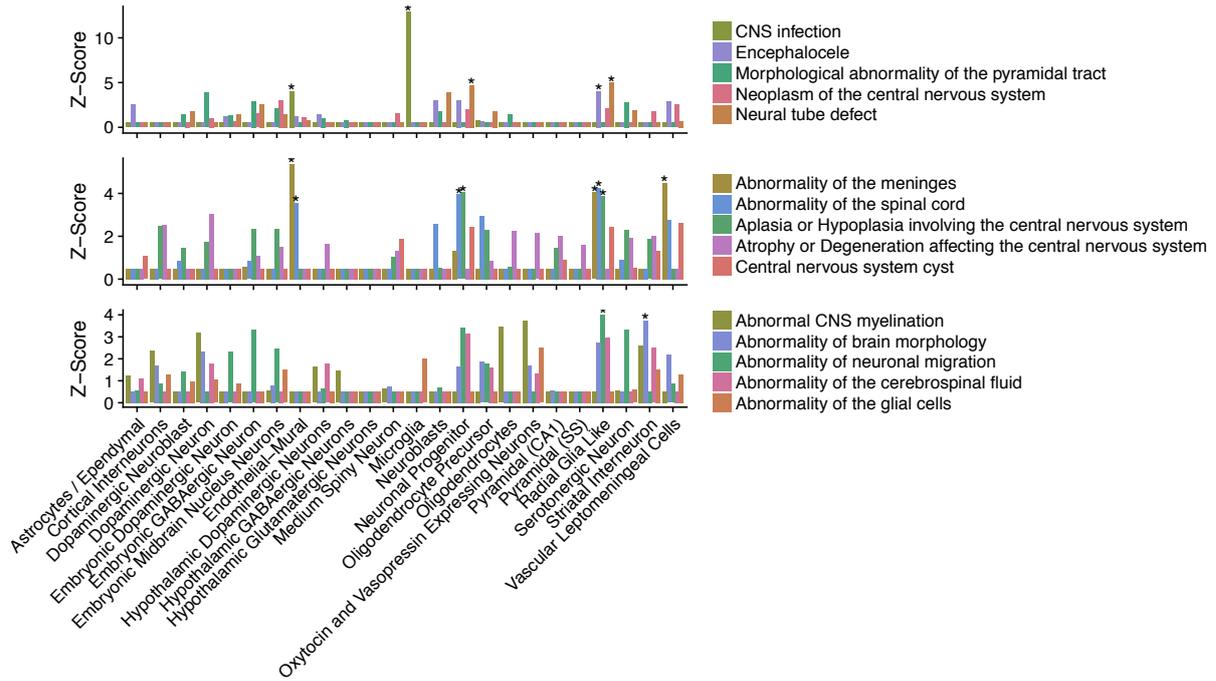
Supplementary Figure 13. Association of gene sets with schizophrenia (CLOZUK) using MAGMA. Schizophrenia (CLOZUK) association probabilities of all the gene sets analyzed. The left column ('original') shows the simple association probabilities, while the right column used MAGMA to condition on the four significantly associated cell types (MSNs, CA1 and SS pyramidal neurons, and interneurons). The black bar represents the Bonferroni significance threshold.



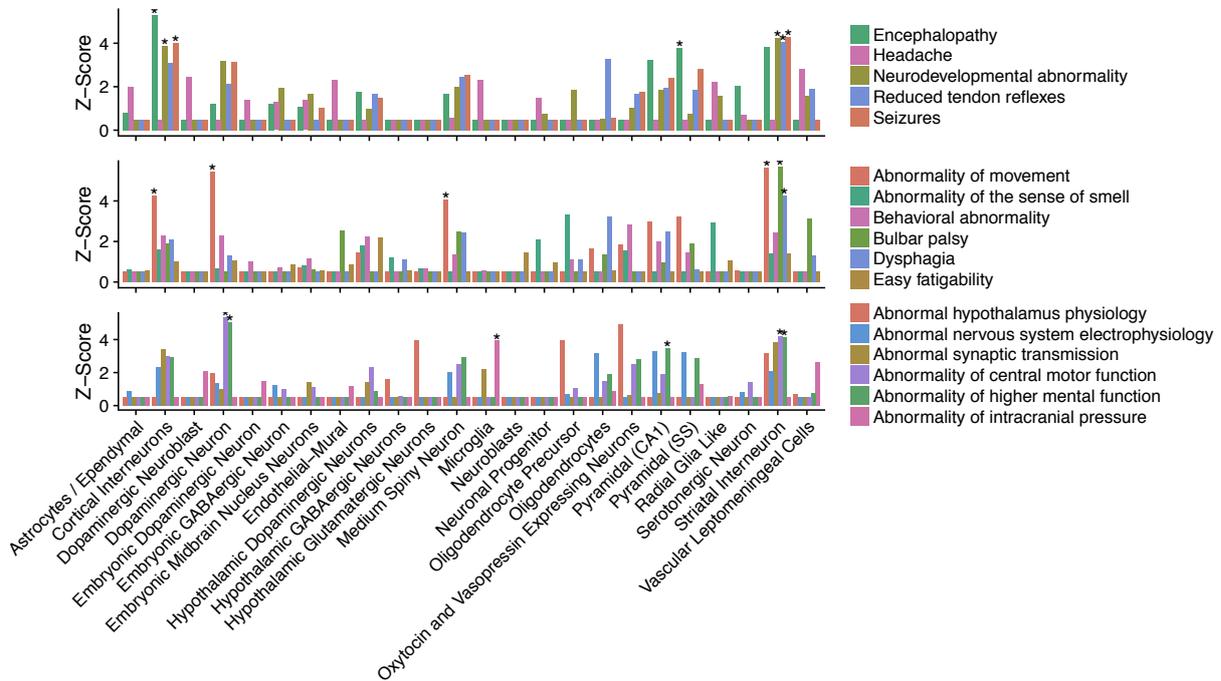
Supplementary Figure 14: Cell types associated with gene sets that are children of the Human Phenotype Ontology term "Seizures". Enrichments were tested with EWCE. Stars mark significance at $p < 0.05$ after correcting for multiple testing with the Benjamini-Hochberg method over all gene sets tested with EWCE.



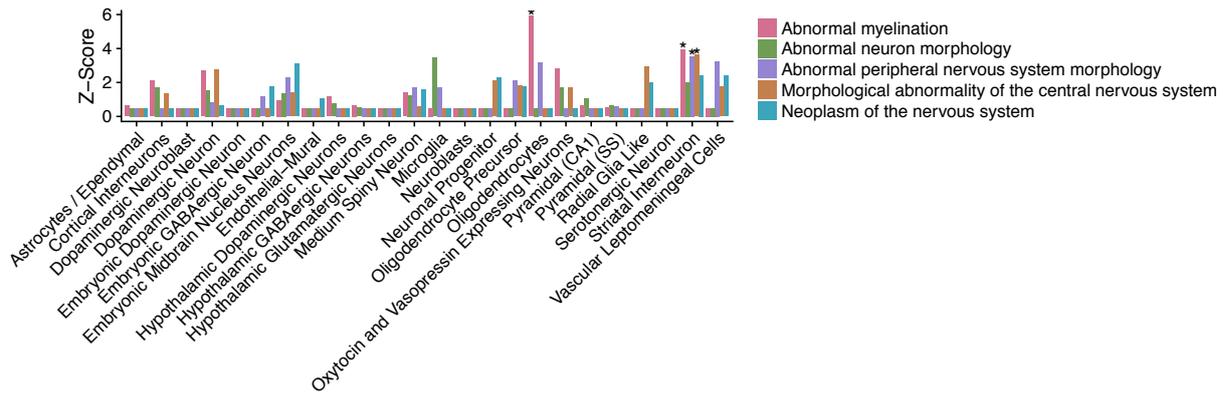
Supplementary Figure 15: Cell types associated with gene sets that are children of the Human Phenotype Ontology term "Phenotypic Abnormality". Enrichments were tested with EWCE. Stars mark significance at $p < 0.05$ after correcting for multiple testing with the Benjamini-Hochberg method over all gene sets tested with EWCE.



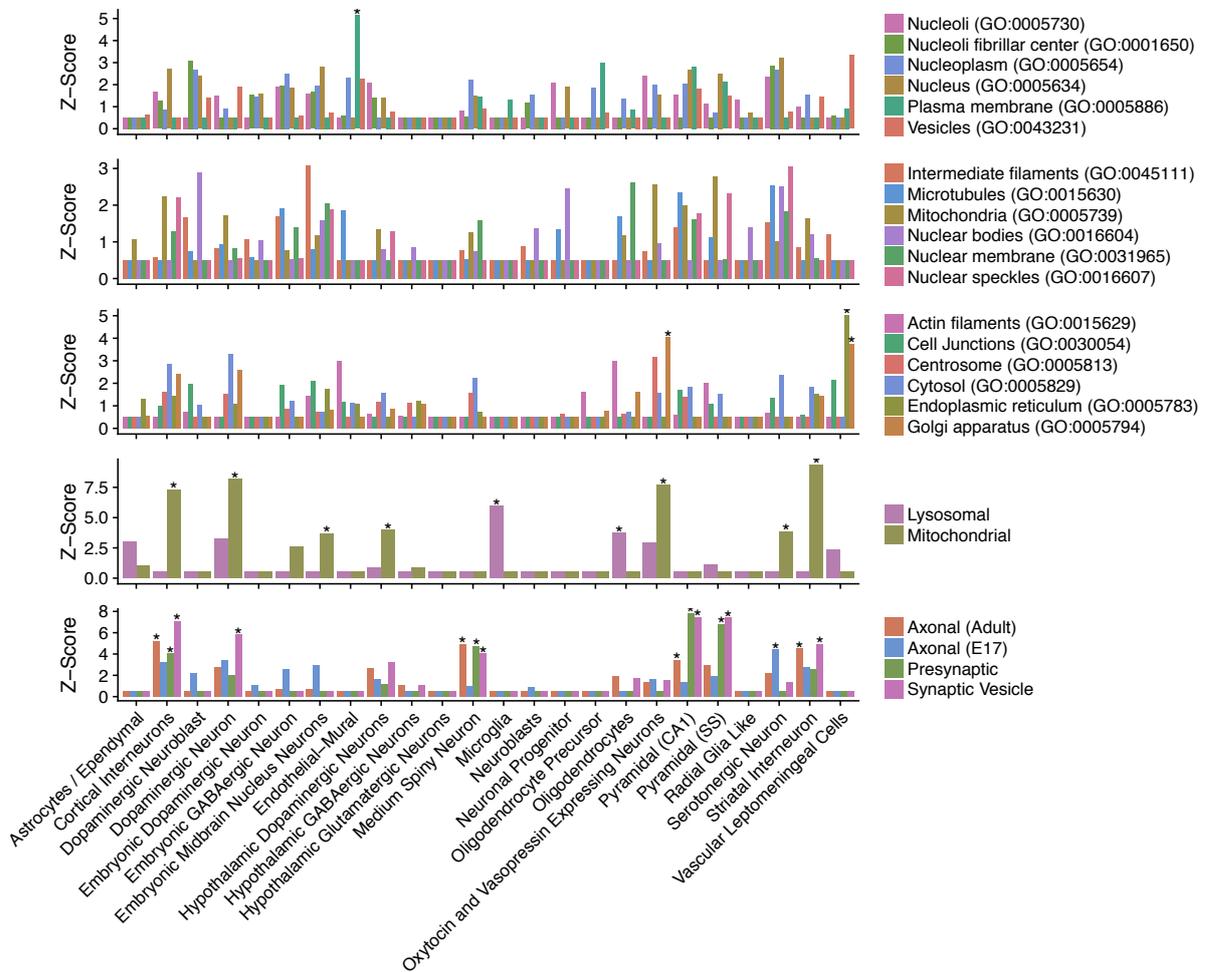
Supplementary Figure 16: Cell types associated with gene sets that are children of the Human Phenotype Ontology term "Morphological abnormality of the central nervous system". Enrichments were tested with EWCE. Stars mark significance at $p < 0.05$ after correcting for multiple testing with the Benjamini-Hochberg method over all gene sets tested with EWCE.



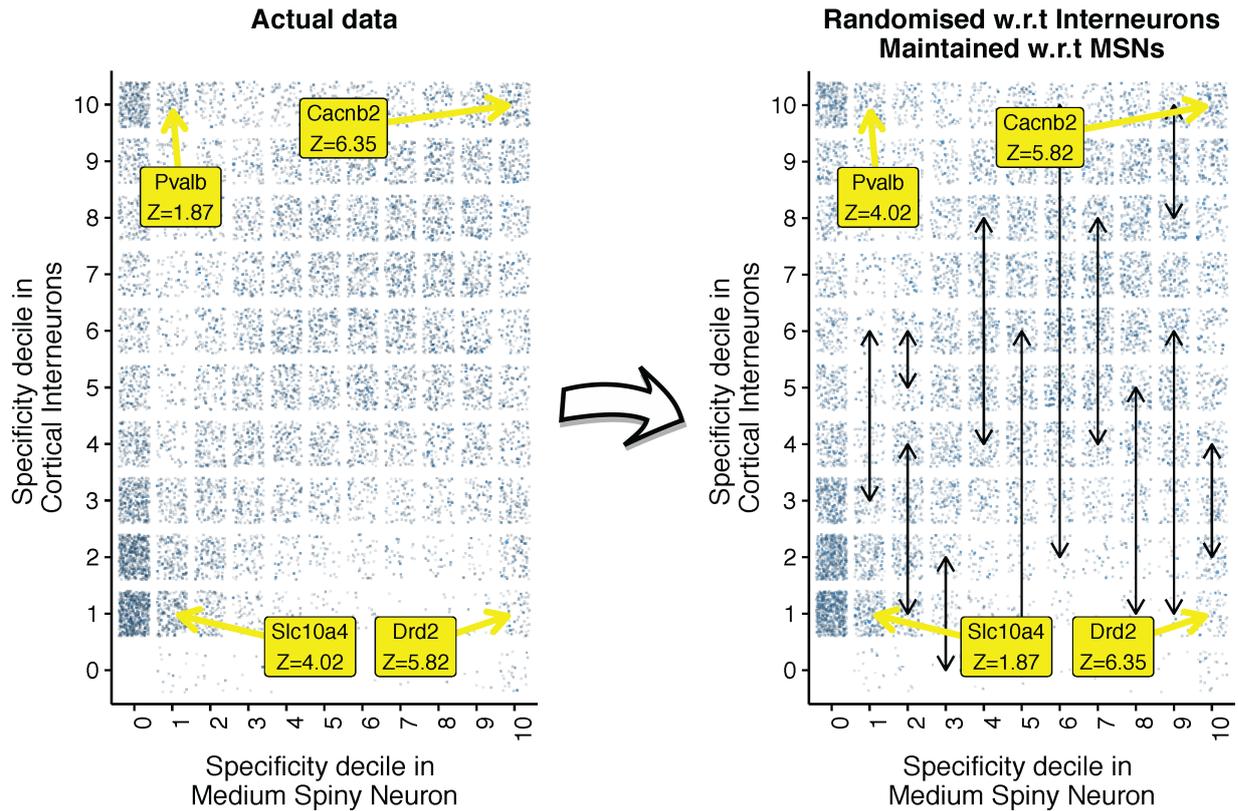
Supplementary Figure 17: Cell types associated with gene sets that are children of the Human Phenotype Ontology term "Abnormality of nervous system physiology". Enrichments were tested with EWCE. Stars mark significance at $p < 0.05$ after correcting for multiple testing with the Benjamini-Hochberg method over all gene sets tested with EWCE.



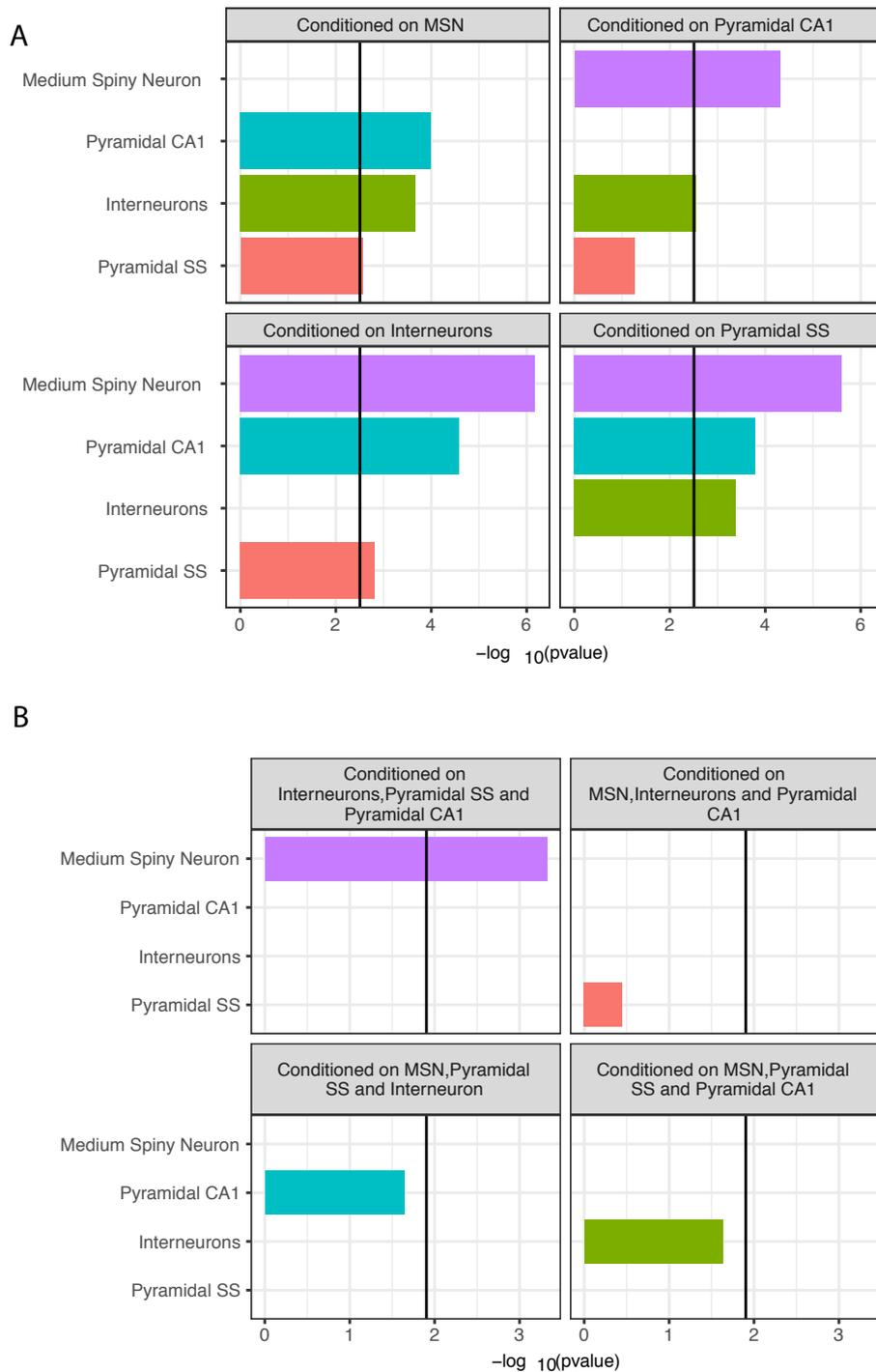
Supplementary Figure 18: Cell types associated with gene sets that are children of the Human Phenotype Ontology term “Abnormality of nervous system morphology”. Enrichments were tested with EWCE. Stars mark significance at $p < 0.05$ after correcting for multiple testing with the Benjamini-Hochberg method over all gene sets tested with EWCE.



Supplementary Figure 19: Cell types associated with gene sets that are associated subcellular protein localisation. Gene lists associated with GO ID’s were obtained via the Human Protein Atlas. Enrichments were tested with EWCE. Stars mark significance at $p < 0.05$ after correcting for multiple testing with the Benjamini-Hochberg method over all gene sets tested with EWCE.



Supplementary Figure 20. Schematic of the resampling approach used for conditional cell type enrichments shown in Figure 4a. The purpose of the resampling was to enable testing of whether the schizophrenia enrichment detected for one cell type (here represented by cortical interneurons, INTs) results from significant enrichment in a second cell type (e.g., medium spiny neurons, MSNs). The left plot shows the correlation in specificity values between INTs and MSNs. Each point is a single gene. Darker points have higher schizophrenia CLOZUK association Z-scores (from MAGMA). The right plot shows how Z-scores are resampled within each MSN specificity decile such that while the specificity values of each gene in MSNs and INTs remains the same and the distribution of Z-scores remain constant relative to each MSN decile, the distribution of Z-scores is randomized relative to INTs. Yellow boxes and arrows mark the location and Z-scores of four genes. In the left plot, the Z-scores are from MAGMA, and the right plot shows resampled Z-scores (hence the Pvalb scores have been switched with those of Slc10a4 as both are in the same MSN specificity decile). For the analysis shown in Figure 4a, resampling of Z-scores is performed 10000 times over and the relative enrichment of INTs in the left plot compared to that in the resampled right plot.



Supplementary Figure 21. Univariate and multivariate conditional analysis using MAGMA. (A) Association of the 4 significant cell types with CLOZUK conditioning on one of the 3 other significant cell types. The black bar represents the Bonferroni significance threshold (0.05/12). All cell types are conditionally independent, except Pyramidal neurons from the somatosensory cortex which are not associated with CLOZUK when conditioning on Pyramidal neurons from the somato-sensory cortex. Note that the converse is not true. (B) Association of the 4 significant cell types with CLOZUK conditioning on the 3 other significant cell types. The black bar represents the Bonferroni significance threshold (0.05/4). Medium spiny neurons remain significantly associated with CLOZUK after conditioning on the three other cell types.

Supplementary Note

Rationale

The overall goal of this analysis was to attempt to connect human genomic findings to specific brain cell types defined by their gene expression profiles: to what specific brain cell types do the common variant genetic findings for schizophrenia best “fit”? Multiple studies have approached this issue¹⁻⁴, but using gene expression based on aggregates of millions of cells. We also evaluated whether gene sets previously implicated in schizophrenia mapped to similar or different brain cell types. We focused on the KI scRNAseq data from mouse (**Figure 1** and **Supplemental Table 2**). We did this because:

- a) These comprise the largest dataset currently available generated using identical procedures. As shown in **Figure 1**, the total cells with scRNAseq (9,970) and Level 2 cell types (149) exceed all other studies.
- b) The mouse data include more brain regions than in human. These regions include a better sampling of those believed to be important in schizophrenia (e.g., currently no data from human striatum or adult dopaminergic neurons).
- c) Due to the use of unique molecular identifiers in the KI data, the scRNAseq data reflect absolute counts, and are directly comparable across experiments (particularly for our goal of evaluating enrichment).
- d) The mouse data appear to have better signal quality. This could be due to better experimental control or the ability to isolate whole cells (excluding distal neurites) of good quality from mouse but only nuclei or lower quality cells from adult humans. For example, sampling 1,500-3,000 cells in cortical mouse data sets (KI and Tasic et al.⁵) allowed identification of 24 and 42 cortical neuronal subtypes. In contrast, sequencing over 3,000 human neuronal nuclei⁶ or 466 whole neurons⁷ allowed for the identification of only 16 and 7 subtypes. More types of inhibitory interneurons (16-23) have been identified in mouse but only 8 in human despite equal or greater sequencing depth but future work may improve the ability to discriminate cell types using single nuclei RNA-Seq data.
- e) Use of laboratory mice allow far greater experimental control of impactful perimortem and postmortem events. All mice are healthy without systemic illnesses and medication-free. All mice can be euthanized in the same way, and time from death to tissue processing is standardized and measured in minutes rather than hours. Causes of death in human are highly variable, and perimortem events can alter brain gene expression (e.g., systemic disease or prolonged hypoxia). Although human brain tissue can be obtained during certain neurosurgical procedures (e.g., resection of a seizure focus in refractory epilepsy), the individuals undergoing these procedures are atypical and subject to the effects of chronic brain disease and medication.
- f) As shown in **Figure 3a**, scRNAseq can be done in mouse brain cells but is far harder in human brain cells. scRNAseq provides better coverage of brain cell transcriptomes and avoids loss of transcripts important for schizophrenia.

Thus, from a practical perspective, more cell types have been identified in mouse, and the KI data comprise over half of currently available brain scRNAseq data. Key findings in the KI dataset were verified in other human and mouse datasets. We also applied independent statistical methods predicated on different assumptions and algorithms to evaluate the relation of brain cell types to GWA results for schizophrenia.

Limitations

Nonetheless, despite our use of multiple statistical methods and efforts to identify and resolve any spurious explanations for our findings, our work has to be considered in light of inevitable limitations. First, although we used the largest available schizophrenia GWA dataset, we still have an incomplete portrait of the genetic architecture of schizophrenia. This is an active area, and more informative results are sure to emerge in the next few years. Second, although the KI scRNAseq data cover a broad range of brain regions thought to be relevant to the neurobiology of schizophrenia, extensive coverage of cortical and striatal development is lacking at present (gestation, early postnatal, or adolescence).

Third, we focus principally on mouse scRNAseq data. Our reasons for doing so are explained above. A key part of our approach is replication of the main findings in human datasets. However, we would be remiss not to consider the comparability of mouse and human. Mice are widely used for modeling brain diseases. There are relatively high degrees of mouse-human conservation in genes expressed in brain. The most extensive study compared RNA-seq data from six organs (cortex, cerebellum, heart, kidney, liver, and testis) across ten species (human, chimpanzee, bonobo, gorilla, orangutan, macaque, mouse, opossum, platypus and chicken)⁸. Using principal components analysis, the largest amount of variation (PCs 1 and 2) explained differences between organs rather than between species. Gene expression in brain (including several key gene expression modules) was more conserved between species than any of the other tissues. These observations were broadly replicated using scRNAseq in ventral midbrain⁹. Furthermore, 75% of genes show similar laminar patterning in mouse and human cortex¹⁰.

Fourth, whatever the general similarities, there are certainly differences between mouse and human brain⁹, and there are even cortical cells present in human but not mouse (e.g., spindle or von Economo neurons). **We therefore evaluated mouse-human gene conservation.** Using empirical measures of gene conservation (Ensembl, URLs), we determined that the mouse genes in the KI Level 1 and Level 2 gene expression dataset that we analyzed were 89% identical (median, interquartile range 80-95%) to human 1:1 homologues. For these genes, the ratio of non-synonymous to synonymous amino acid changes (dN/dS) was 0.094 (median, interquartile range 0.045-0.173): mutations in these genes are thus subject to strong negative selection (dN/dS = 1 is consistent with neutrality). Pathway analysis of the 400 genes with the largest dN/dS values revealed enrichments in genes involved in defense responses, inflammation, cytokines, and immunoglobulin production. The 400 genes with extremely low dN/dS ratios were involved in neuron differentiation, RNA splicing, and mRNA processing.

In conclusion, for most brain cell types, use of KI mouse scRNAseq data was defensible and reasonable (particularly given verification in human transcriptomic data). The major caution is with respect to cells with prominent immune function (e.g., microglia). (See also the section on mouse-human gene mapping below.)

We can implicate a particular cell type (i.e., present consistent positive evidence) but it is premature to exclude cell types for which we do not have data, or those with dissimilar function or under selection pressure between mouse and human.

Single-cell transcriptome data

Supplemental Table 2 shows the scRNAseq or snRNAseq data from mouse or human brain. These include published and unpublished data (using the same protocols as in peer-reviewed papers). To the best of our knowledge, these comprise all or nearly all of the available adult brain single nuclei or scRNAseq data.

We focused on a superset of brain scRNAseq data from KI generated using identical methods from the same labs with the use of unique molecular identifiers that allow for direct comparison of transcription

data across regions (see above for full rationale). The KI mouse superset of 9,970 cells and 149 Level 2 cell types is more extensive than any other single nuclei or scRNAseq dataset now available, and includes most brain regions thought to be salient to schizophrenia. The papers contain full method details. Briefly, the KI scRNAseq data were generated using the same methods (Fluidigm C1 with Illumina 50 bp single end sequencing) with the use of unique molecular identifiers to enable absolute molecular counts. In the first paper describing the method it was estimated that an average of 1.2 million mapped reads per cells was sequenced (28). Level 1 and 2 clustering was done using the BACKSPIN algorithm ¹¹. All cells lacking annotations were excluded. For non-neuronal populations, except cells from oligodendrocyte lineage and VLMCs, we only included cells from Zeisel et al 2015 in the KI data set. The level 2 CA1 pyramidal cell contain a small number of cells from CA2 and Subiculum resulting from dissection inaccuracies, these are represented as separate level 2 classes. The resulting data have been shown to be insensitive to linear variation in total reads per cell. If a gene was detected in one dataset and not in another, it was considered to have zero reads in all cells where it had not been detected.

To confirm that no batch effects exist across KI regional subdatasets that may influence the merged results, we plotted three cell types using tSNE which were expected to show little real regional variation: endothelial cells, vascular smooth muscle cells and microglia (**Supplemental Figure 2**). The tSNE plots were generated in R using the Rtsne and Scater packages using 500 of the most variable features. Only the embryonic midbrain cells clustered separately, as was expected due to the difference between the embryonic and adult brain.

We include unpublished data generated by the Hjerling-Leffler and Linnarsson labs at KI using the same methods as in Zeisel et al. ¹¹. Cells were isolated from dorsolateral striatum from p21-p30 transgenic mice, the same age span as in Zeisel et al. ¹¹. Coverage of rare interneuron populations was enhanced by FACS sorting cells from either 5HT3a-EGFP or a Lhx6cre::TdTomato line. The cortical parvalbuminergic cells and striatal neurons were captured and prepared for sequencing as described in Zeisel et al. ¹¹.

The largest human dataset is an unpublished data set from the Allen Institute for Brain Science which consisted of 4401 cells from middle temporal gyrus of 3 post-mortem brains from healthy, adult subjects. Nuclei were dissociated from cortical tissue and FACS isolated based on NeuN staining, resulting in approximately 90% NeuN+ and 10% NeuN- nuclei. Single nucleus cDNA libraries were generated using SMARTerV4 and Nextera XT and sequenced to a depth of approximately 2 million reads per sample. Reads were aligned with Bowtie and gene expression quantified with RSEM plus intronic reads and normalized to counts per million. Clustering was performed with iterative PCA and tSNE with cluster robustness assessed with 100 bootstrap replicates. Level 1 clusters were characterized based on expression of known marker genes and included two broad classes of neurons – GABAergic interneurons and glutamatergic projection neurons – and 4 non-neuronal types: astrocytes, oligodendrocyte precursors, mature oligodendrocytes, and microglia.

Technical issues

The process of procuring cells for scRNAseq entails dissecting the tissue and preparing a single-cell suspension. This entails severing cellular processes such as dendrites and axons. It is probably safe to assume that different cell types are differentially sensitive to this drastic process. Accordingly, in early studies certain cell types, although present in the data sets, was clearly underrepresented –most likely due to lower rates of survival. Although the cell procurement methods have improved the field, in general, do not rely on scRNASeq data to make statements on relative abundance between cell types. Such statements should ultimately be done in tissue samples and this is one of the major reasons that multiplexed in situ hybridization techniques is a fast-growing field.

When sampling tissue for scRNASeq there are two general approaches used to increase the chances to sample underrepresented or low abundant cell types at enough high numbers: 1) if one has prior knowledge of what cell type is underrepresented these can be enriched using FACS isolation of genetically labeled cells, 2) to massively increase the sampling rate of cells (usually also entails decreasing the sequencing depth of each cell). ScRNAseq is a fast-moving field and future datasets will help fine-tune our knowledge on cell types but with the data sets used in this study we have captured a majority of cellular complexity. Even with *in situ* techniques emerging, the question “are we missing any cell types” cannot be answered in an absolute fashion –we can only estimate what are the minimum size of populations we can detect. For well characterized brain areas like cortex and hippocampus we can make these estimates more accurately but with structures that has high cellular complexity and is not so well characterized (i.e. hypothalamus) the risk is considerably higher that we are missing some populations.

We are not aware of how these kind of effects affects snRNAseq data but presumably the problem should be much less pronounced since as a first step nuclei are isolated by killing all cells simultaneously. The main scRNAseq data sets used in this paper (KI data set) have relied mainly on enrichment of low abundant/low surviving cells via FACS isolation in some of the sub-sets (e.g. cortex, hippocampus and striatum). This is also the case for the Tasic data set. The other main data sets used (AIBS and DroNc-seq) are snRNAseq data sets.

Evaluation of genomic biases

The algorithms used by LDSC and MAGMA both account for the non-independence introduced by linkage disequilibrium (LD), or the tendency for genomic findings to “cluster” due to strong intercorrelations. LD block size (discrete regions of high correlations between nearby genetic markers) average 15-20 kb in samples of European ancestry, but there are nearly 100 genomic regions with high LD extending over 1 mb (the extended MHC region on human chromosome 6 is the largest and has very high LD over 8 mb). Gene size is an additional consideration for MAGMA (accounting for gene size is a component of the algorithm), particularly as brain-expressed genes are considerably larger than genes not expressed in brain (mean of 80.7 kb vs 31.2 kb). The algorithms used by LDSC and MAGMA have been well-tested, and are widely used. However, it is conceivable that certain edge cases could defeat algorithms that work well for the vast majority of scenarios. An example might be if a large fraction of the genes that influence a brain cell type were located in a region of very high LD. First, brain-expressed genes were slightly more likely to be in a large LD block ($\geq 99^{\text{th}}$ percentile in size across the genome), 12.4% vs 10.2%. In discussion with the developers of LDSC and MAGMA, this should not yield an insuperable bias. Second, by counting the numbers of genes and brain-expressed genes per mb, we found that brain-expressed genes in the human genome were reasonably evenly scattered across the genome (R^2 0.85), and only 10 of 2,534 1-mb intervals were outliers. Most of these were gene clusters with fewer than expected brain-expressed genes (e.g., a late cornified envelope gene cluster on chr1:152-153 mb, an olfactory gene cluster on chr1:248-249 mb, and a keratin gene cluster on chr17:39-40 mb). Third, in a similar manner, we evaluated the locations of the human 1:1 mapped genes influential to the KI Level 1 classifications and found these to be relatively evenly scattered in the genome. Thus, these potential genomic biases did not appear to present difficulties for our key analyses (that used two independent methods in any event).

Mouse-to-human gene mapping

Because most of the scRNAseq data were from mouse brain and the schizophrenia genomic results are from human, it was necessary to map 1:1 homologs between *M. musculus* and *H. sapiens*. To accomplish this, used a best-practice approach in consultation with a senior mouse geneticist (UNC

Prof Fernando Pardo-Manuel de Villena de L'Epine, personal communication). We used the expert curated human-mouse homolog list (Mouse Genome Informatics, Jackson Laboratory, URLs, version of 11/22/2016). Only genes with a high-confidence, 1:1 mapping were retained. A large fraction of non-matches are reasonable given evolutionary differences between human and mouse (e.g., the distinctiveness of olfactory or vomeronasal receptor genes given the greater importance of smell in mouse). Nonetheless, we evaluated the quality and coherence of the mapping.

- The mouse brain cell expression levels for the KI Level 1 cell types were similar for mouse genes with and without a 1:1 human homologue. This is inconsistent with a strong bias due to the success/failure of identifying a human homologue.
- A high fraction (93%) of the KI genes detected in mouse brain samples that mapped to a human gene were expressed in human brain (CommonMind DLPFC RNA-seq) or 27 samples with RNA-seq from the Sullivan lab (unpublished, DLPFC from 9 schizophrenia cases and 9 controls plus 9 fetal frontal cortex samples). The ones that did not (7%) were expressed at considerably lower levels in mouse brain or in cell types not prevalent in cortex.
- Of genes with evidence of expression in human brain (via frontal cortex RNA-seq as noted above), human homologues of KI mouse genes accounted for 93.2% of intellectual disability genes, 93.7% of developmental delay genes, 93.8% of genes with a CHD8 binding site, 94.4% of post-synaptic genes, 95.0% of proteins involved in the ciliary proteome, 95.1% of genes intolerant to loss-of-function variation (ExAC pLI > 0.9), 95.6% of pre-synaptic genes, and 96.4% of FMRP interactors.

We evaluated the mapping carefully and the results above suggest the coherence of the mouse-human mapping. All key findings from the KI mouse scRNAseq data were evaluated in other mouse and human brain scRNAseq datasets.

Calculation of cell type expression specificity

We denote this specificity metric as $s_{g,c}$ for gene g and cell type c . Values of $s_{g,c}$ were calculated for the brain scRNAseq datasets in **Supplemental Table 2**.

Each dataset contains scRNAseq results from w cells associated with k cell types. Each of the k cell types is associated with a numerical index from the set $\{1, \dots, k\}$. The cell type annotations for cell i are stored using a numerical index in L , such that $l_{1005}=5$ indicates that the 1005th cell is of the 5th cell type. We denote N_c as the number of cells from the cell-type indexed by c . The expression proportion for gene g and cell type c (where $r_{g,i}$ is the expression of gene g in cell i) is given by:

$$s_{g,c} = \frac{\sum_{i=1}^w F(g,i,c)/N_c}{\sum_{r=1}^k (\sum_{i=1}^w F(g,i,r)/N_r)} \quad F(g,i,c) = \begin{cases} r_{g,i}, & l_i = c \\ 0, & l_i \neq c \end{cases}$$

This metric for cell specificity is closely related to other measures¹². For instance, the maximum value of s per gene yields similar results to τ such that $s_{max} > 0.5$ is equivalent to $\tau > 0.94$.

The size of a cell is generally correlated with the amount of mRNA detected upon isolation. This is reflected in the "Total Average Molecules" column of **Supplemental Table 3** which is representative of the real amount of mRNA detected in the cells. The values used in the KI data set are from Unique Molecular Identifiers (UMIs; each individual mRNA molecule is uniquely barcoded upon detection and an absolute number thus can be calculated even after PCR amplification). In order to not ignore meaningful biological differences we did not scale the expression levels between cell types. However, we tested the effect of scaling gene expression to 10'000 molecules for each cell type before calculating the specificity index and observed the same Bonferroni significant cell types in the same order, indicating that our approach is robust to differences in the total number of molecules detected

per cell type. Perhaps we should point out that MSNs are, as the name implies, medium-sized cells and were found to be significantly enriched in every test in spite of being the third smallest neuronal cell type based on mRNA counts.

Choice of MAGMA window size boundaries

In regard to choice of window size/bin boundaries, MAGMA by default combines *P*-values of SNPs located within gene boundaries (± 0 kb). We decided to extend the default window size as a large fraction of trait-associated SNPs are located just outside genes in regions likely to regulate gene expression^{1,13}. One of the authors of MAGMA (Christiaan de Leeuw) advised us to expand the window size by a limited amount in order to keep the ability to distinguish the genetic contribution of genes located in close proximity. Therefore, we set expanded gene boundaries to 10 kb upstream–1.5 kb downstream. We evaluated the effect of different choices of bin size including 35 kb upstream–10 kb downstream (as often used by the PGC¹⁴), 150 kb upstream–10 kb downstream, and 150 kb upstream–150 kb downstream (GTEx¹⁵ Supplemental Figure 9 from). The results were not substantially altered by window size as the ranking of cell types (KI level 1) were very similar for these different window sizes; if anything ours was a slightly conservative choice.

Members and their affiliations of the Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium

Naomi R Wray^{1,2†}, Stephan Ripke^{3,4,5†}, Manuel Mattheisen^{6,7,8,9†}, *Maciej Trzaskowski^{1†}, Enda M Byrne¹, Abdel Abdellaoui¹⁰, Mark J Adams¹¹, Esben Agerbo^{8,12,13}, Tracy M Air¹⁴, Till F M Andlauer^{15,16}, Silviu-Alin Bacanu¹⁷, Marie Bækvad-Hansen^{8,18}, Aartjan T F Beekman¹⁹, Tim B Bigdeli^{17,20}, Elisabeth B Binder^{15,21}, Douglas H R Blackwood¹¹, Julien Bryois²², Henriette N Buttenschøn^{7,8,23}, Jonas Bybjerg-Grauholm^{8,18}, Na Cai^{24,25}, Enrique Castelao²⁶, Jane Hvarregaard Christensen^{6,7,8}, Toni-Kim Clarke¹¹, Jonathan R I Coleman²⁷, Lucía Colodro-Conde²⁸, Baptiste Couvy-Duchesne^{29,30}, Nick Craddock³¹, Gregory E Crawford^{32,33}, Cheyenne A Crowley³⁴, Hassan S Dashti^{3,35}, Gail Davies³⁶, Ian J Deary³⁶, Franziska Degenhardt^{37,38}, Eske M Derks²⁸, Nese Direk^{39,40}, Conor V Dolan¹⁰, Erin C Dunn^{41,42,43}, Thalia C Eley²⁷, Nicholas Eriksson⁴⁴, Valentina Escott-Price⁴⁵, Farnush Farhadi Hassan Kiadeh⁴⁶, Hilary K Finucane^{47,48}, Andreas J Forstner^{37,38,49,50}, Josef Frank⁵¹, Héléna A Gaspar²⁷, Michael Gill⁵², Paola Giusti-Rodríguez⁵³, Fernando S Goes⁵⁴, Scott D Gordon⁵⁵, Jakob Grove^{6,7,8,56}, Lynsey S Hall^{11,57}, Eilis Hannon⁵⁸, Christine Sørholm Hansen^{8,18}, Thomas F Hansen^{59,60,61}, Stefan Herms^{37,38,50}, Ian B Hickie⁶², Per Hoffmann^{37,38,50}, Georg Homuth⁶³, Carsten Horn⁶⁴, Jouke-Jan Hottenga¹⁰, David M Hougaard^{8,18}, Ming Hu⁶⁵, Craig L Hyde⁶⁶, Marcus Ising⁶⁷, Rick Jansen^{19,19}, Fulai Jin^{68,69}, Eric Jorgenson⁷⁰, James A Knowles⁷¹, Isaac S Kohane^{72,73,74}, Julia Kraft⁵, Warren W. Kretschmar⁷⁵, Jesper Krogh⁷⁶, Zoltán Kutalik^{77,78}, Jacqueline M Lane^{3,35,79}, Yihan Li⁷⁵, Yun Li^{34,53}, Penelope A Lind²⁸, Xiaoxiao Liu⁶⁹, Leina Lu⁶⁹, Donald J MacIntyre^{80,81}, Dean F MacKinnon⁵⁴, Robert M Maier², Wolfgang Maier⁸², Jonathan Marchini⁸³, Hamdi Mbarek¹⁰, Patrick McGrath⁸⁴, Peter McGuffin²⁷, Sarah E Medland²⁸, Divya Mehta^{2,85}, Christel M Middeldorp^{10,86,87}, Evelin Mihailov⁸⁸, Yuri Milaneschi^{19,19}, Lili Milani⁸⁸, Jonathan Mill⁵⁸, Francis M Mondimore⁵⁴, Grant W Montgomery¹, Sara Mostafavi^{89,90}, Niamh Mullins²⁷, Matthias Nauck^{91,92}, Bernard Ng⁹⁰, Michel G Nivard¹⁰, Dale R Nyholt⁹³, Paul F O'Reilly²⁷, Hogni Oskarsson⁹⁴, Michael J Owen⁹⁵, Jodie N Painter²⁸, Carsten Bøcker Pedersen^{8,12,13}, Marianne Giørtz Pedersen^{8,12,13}, Roseann E. Peterson^{17,96}, Erik Pettersson²², Wouter J Peyrot¹⁹, Giorgio Pistis²⁶, Danielle Posthuma^{97,98}, Shaun M Purcell⁹⁹, Jorge A Quiroz¹⁰⁰, Per Qvist^{6,7,8}, John P Rice¹⁰¹, Brien P. Riley¹⁷, Margarita Rivera^{27,102}, Saira Saeed Mirza⁴⁰, Richa Saxena^{3,35,79}, Robert Schoevers¹⁰³, Eva C Schulte^{104,105}, Ling Shen⁷⁰, Jianxin Shi¹⁰⁶, Stanley I Shyn¹⁰⁷, Engilbert Sigurdsson¹⁰⁸, Grant C B Sinnamon¹⁰⁹, Johannes H Smit¹⁹, Daniel J Smith¹¹⁰, Hreinn Stefansson¹¹¹, Stacy Steinberg¹¹¹, Craig A Stockmeier¹¹², Fabian Streit⁵¹, Jana Strohmaier⁵¹, Katherine E Tansey¹¹³, Henning Teismann¹¹⁴, Alexander Teumer¹¹⁵, Wesley Thompson^{8,60,116,117}, Pippa A Thomson¹¹⁸, Thorgeir E Thorgeirsson¹¹¹, Chao Tian⁴⁴, Matthew Traylor¹¹⁹, Jens Treutlein⁵¹, Vassily Trubetskoy⁵, André G Uitterlinden¹²⁰, Daniel Umbricht¹²¹, Sandra

Van der Auwera ¹²², Albert M van Hemert ¹²³, Alexander Viktorin ²², Peter M Visscher ^{1,2}, Yunpeng Wang ^{8,60,116}, Bradley T. Webb ¹²⁴, Shantel Marie Weinsheimer ^{8,60}, Jürgen Wellmann ¹¹⁴, Gonneke Willemssen ¹⁰, Stephanie H Witt ⁵¹, Yang Wu ¹, Hualin S Xi ¹²⁵, Jian Yang ^{2,126}, Futao Zhang ¹, eQTLGen Consortium ¹²⁷, 23andMe Research Team ⁴⁴, Volker Arolt ¹²⁸, Bernhard T Baune ¹⁴, Klaus Berger ¹¹⁴, Dorret I Boomsma ¹⁰, Sven Cichon ^{37,50,129,130}, Udo Dannlowski ¹²⁸, EJC de Geus ^{10,131}, J Raymond DePaulo ⁵⁴, Enrico Domenici ¹³², Katharina Domschke ¹³³, Tõnu Esko ^{3,88}, Hans J Grabe ¹²², Steven P Hamilton ¹³⁴, Caroline Hayward ¹³⁵, Andrew C Heath ¹⁰¹, David A Hinds ⁴⁴, Kenneth S Kendler ¹⁷, Stefan Kloiber ^{67,136,137}, Glyn Lewis ¹³⁸, Qingqin S Li ¹³⁹, Susanne Lucae ⁶⁷, Pamela AF Madden ¹⁰¹, Patrik K Magnusson ²², Nicholas G Martin ⁵⁵, Andrew M McIntosh ^{11,36}, Andres Metspalu ^{88,140}, Ole Mors ^{8,141}, Preben Bo Mortensen ^{7,8,12,13}, Bertram Müller-Myhsok ^{15,16,142}, Merete Nordentoft ^{8,143}, Markus M Nöthen ^{37,38}, Michael C O'Donovan ⁹⁵, Sara A Paciga ¹⁴⁴, Nancy L Pedersen ²², Brenda WJH Penninx ¹⁹, Roy H Perlis ^{42,145}, David J Porteous ¹¹⁸, James B Potash ¹⁴⁶, Martin Preisig ²⁶, Marcella Rietschel ⁵¹, Catherine Schaefer ⁷⁰, Thomas G Schulze ^{51,105,147,148,149}, Jordan W Smoller ^{41,42,43}, Kari Stefansson ^{111,150}, Henning Tiemeier ^{40,151,152}, Rudolf Uher ¹⁵³, Henry Völzke ¹¹⁵, Myrna M Weissman ^{84,154}, Thomas Werge ^{8,60,155}, Ashley R Winslow ^{156,157}, Cathryn M Lewis ^{27,158}*, Douglas F Levinson ¹⁵⁹*, Gerome Breen ^{27,160}*, Anders D Børglum ^{6,7,8}*, Patrick F Sullivan ^{22,53,161}*, for the Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium.

Author Affiliations:

- 1, Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, AU
- 2, Queensland Brain Institute, The University of Queensland, Brisbane, QLD, AU
- 3, Medical and Population Genetics, Broad Institute, Cambridge, MA, US
- 4, Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, US
- 5, Department of Psychiatry and Psychotherapy, Universitätsmedizin Berlin Campus Charité Mitte, Berlin, DE
- 6, Department of Biomedicine, Aarhus University, Aarhus, DK
- 7, iSEQ, Centre for Integrative Sequencing, Aarhus University, Aarhus, DK
- 8, iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, DK
- 9, Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, SE
- 10, Dept of Biological Psychology & EMGO+ Institute for Health and Care Research, Vrije Universiteit Amsterdam, Amsterdam, NL
- 11, Division of Psychiatry, University of Edinburgh, Edinburgh, GB
- 12, Centre for Integrated Register-based Research, Aarhus University, Aarhus, DK
- 13, National Centre for Register-Based Research, Aarhus University, Aarhus, DK
- 14, Discipline of Psychiatry, University of Adelaide, Adelaide, SA, AU
- 15, Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, DE
- 16, Munich Cluster for Systems Neurology (SyNergy), Munich, DE
- 17, Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, US
- 18, Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, DK
- 19, Department of Psychiatry, Vrije Universiteit Medical Center and GGZ inGeest, Amsterdam, NL
- 20, Virginia Institute for Psychiatric and Behavior Genetics, Richmond, VA, US
- 21, Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, US
- 22, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, SE
- 23, Department of Clinical Medicine, Translational Neuropsychiatry Unit, Aarhus University, Aarhus, DK
- 24, Statistical genomics and systems genetics, European Bioinformatics Institute (EMBL-EBI), Cambridge, GB
- 25, Human Genetics, Wellcome Trust Sanger Institute, Cambridge, GB
- 26, Department of Psychiatry, University Hospital of Lausanne, Prilly, Vaud, CH
- 27, MRC Social Genetic and Developmental Psychiatry Centre, King's College London, London, GB
- 28, Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Herston, QLD, AU
- 29, Centre for Advanced Imaging, The University of Queensland, Saint Lucia, QLD, AU
- 30, Queensland Brain Institute, The University of Queensland, Saint Lucia, QLD, AU
- 31, Psychological Medicine, Cardiff University, Cardiff, GB
- 32, Center for Genomic and Computational Biology, Duke University, Durham, NC, US
- 33, Department of Pediatrics, Division of Medical Genetics, Duke University, Durham, NC, US
- 34, Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, US
- 35, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA
- 36, Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, GB
- 37, Institute of Human Genetics, University of Bonn, Bonn, DE
- 38, Life&Brain Center, Department of Genomics, University of Bonn, Bonn, DE
- 39, Psychiatry, Dokuz Eylül University School of Medicine, Izmir, TR
- 40, Epidemiology, Erasmus MC, Rotterdam, Zuid-Holland, NL
- 41, Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, US
- 42, Department of Psychiatry, Massachusetts General Hospital, Boston, MA, US
- 43, Psychiatric and Neurodevelopmental Genetics Unit (PNGU), Massachusetts General Hospital, Boston, MA, US

44, Research, 23andMe, Inc., Mountain View, CA, US
45, Neuroscience and Mental Health, Cardiff University, Cardiff, GB
46, Bioinformatics, University of British Columbia, Vancouver, BC, CA
47, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, US
48, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, US
49, Department of Psychiatry (UPK), University of Basel, Basel, CH
50, Human Genomics Research Group, Department of Biomedicine, University of Basel, Basel, CH
51, Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Baden-Württemberg, DE
52, Department of Psychiatry, Trinity College Dublin, Dublin, IE
53, Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, US
54, Psychiatry & Behavioral Sciences, Johns Hopkins University, Baltimore, MD, US
55, Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, AU
56, Bioinformatics Research Centre, Aarhus University, Aarhus, DK
57, Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, GB
58, University of Exeter Medical School, Exeter, UK
59, Danish Headache Centre, Department of Neurology, Rigshospitalet, Glostrup, DK
60, Institute of Biological Psychiatry, Mental Health Center Sct. Hans, Mental Health Services Capital Region of Denmark, Copenhagen, DK
61, iPSYCH, The Lundbeck Foundation Initiative for Psychiatric Research, Copenhagen, DK
62, Brain and Mind Centre, University of Sydney, Sydney, NSW, AU
63, Interfaculty Institute for Genetics and Functional Genomics, Department of Functional Genomics, University Medicine and Ernst Moritz Arndt University Greifswald, Greifswald, Mecklenburg-Vorpommern, DE
64, Roche Pharmaceutical Research and Early Development, Pharmaceutical Sciences, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd, Basel, CH
65, Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH, US
66, Statistics, Pfizer Global Research and Development, Groton, CT, US
67, Max Planck Institute of Psychiatry, Munich, DE
68, Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH, US
69, Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH, US
70, Division of Research, Kaiser Permanente Northern California, Oakland, CA, US
71, Psychiatry & The Behavioral Sciences, University of Southern California, Los Angeles, CA, US
72, Informatics Program, Boston Children's Hospital, Boston, MA, US
73, Department of Medicine, Brigham and Women's Hospital, Boston, MA, US
74, Department of Biomedical Informatics, Harvard Medical School, Boston, MA, US
75, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, GB
76, Department of Endocrinology at Herlev University Hospital, University of Copenhagen, Copenhagen, DK
77, Swiss Institute of Bioinformatics, Lausanne, VD, CH
78, Institute of Social and Preventive Medicine (IUMSP), University Hospital of Lausanne, Lausanne, VD, CH
79, Dept of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston, MA, USA
80, Mental Health, NHS 24, Glasgow, GB
81, Division of Psychiatry, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, GB
82, Department of Psychiatry and Psychotherapy, University of Bonn, Bonn, DE
83, Statistics, University of Oxford, Oxford, GB
84, Psychiatry, Columbia University College of Physicians and Surgeons, New York, NY, US
85, School of Psychology and Counseling, Queensland University of Technology, Brisbane, QLD, AU
86, Child and Youth Mental Health Service, Children's Health Queensland Hospital and Health Service, South Brisbane, QLD, AU
87, Child Health Research Centre, University of Queensland, Brisbane, QLD, AU
88, Estonian Genome Center, University of Tartu, Tartu, EE
89, Medical Genetics, University of British Columbia, Vancouver, BC, CA
90, Statistics, University of British Columbia, Vancouver, BC, CA
91, DZHK (German Centre for Cardiovascular Research), Partner Site Greifswald, University Medicine, University Medicine Greifswald, Greifswald, Mecklenburg-Vorpommern, DE
92, Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Mecklenburg-Vorpommern, DE
93, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, QLD, AU
94, Humus, Reykjavik, IS
95, MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, GB
96, Virginia Institute for Psychiatric & Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, US
97, Complex Trait Genetics, Vrije Universiteit Amsterdam, Amsterdam, NL
98, Clinical Genetics, Vrije Universiteit Medical Center, Amsterdam, NL
99, Department of Psychiatry, Brigham and Women's Hospital, Boston, MA, US
100, Solid Biosciences, Boston, MA, US
101, Department of Psychiatry, Washington University in Saint Louis School of Medicine, Saint Louis, MO, US
102, Department of Biochemistry and Molecular Biology II, Institute of Neurosciences, Center for Biomedical Research, University of Granada, Granada, ES
103, Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, NL
104, Department of Psychiatry and Psychotherapy, Medical Center of the University of Munich, Campus Innenstadt, Munich, DE
105, Institute of Psychiatric Phenomics and Genomics (IPPG), Medical Center of the University of Munich, Campus Innenstadt, Munich, DE
106, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, US
107, Behavioral Health Services, Kaiser Permanente Washington, Seattle, WA, US
108, Faculty of Medicine, Department of Psychiatry, University of Iceland, Reykjavik, IS
109, School of Medicine and Dentistry, James Cook University, Townsville, QLD, AU

110, Institute of Health and Wellbeing, University of Glasgow, Glasgow, GB
111, deCODE Genetics / Amgen, Reykjavik, IS
112, Psychiatry & Human Behavior, University of Mississippi Medical Center, Jackson, MS, US
113, College of Biomedical and Life Sciences, Cardiff University, Cardiff, GB
114, Institute of Epidemiology and Social Medicine, University of Münster, Münster, Nordrhein-Westfalen, DE
115, Institute for Community Medicine, University Medicine Greifswald, Greifswald, Mecklenburg-Vorpommern, DE
116, KG Jepsen Centre for Psychosis Research, Norway Division of Mental Health and Addiction, Oslo University Hospital, Oslo, NO
117, Department of Psychiatry, University of California, San Diego, San Diego, CA, US
118, Medical Genetics Section, CGEM, IGMM, University of Edinburgh, Edinburgh, GB
119, Clinical Neurosciences, University of Cambridge, Cambridge, GB
120, Internal Medicine, Erasmus MC, Rotterdam, Zuid-Holland, NL
121, Roche Pharmaceutical Research and Early Development, Neuroscience, Ophthalmology and Rare Diseases Discovery & Translational Medicine Area, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd, Basel, CH
122, Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Mecklenburg-Vorpommern, DE
123, Department of Psychiatry, Leiden University Medical Center, Leiden, NL
124, Virginia Institute of Psychiatric & Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, US
125, Computational Sciences Center of Emphasis, Pfizer Global Research and Development, Cambridge, MA, US
126, Institute for Molecular Bioscience; Queensland Brain Institute, The University of Queensland, Brisbane, QLD, AU
127, Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, NL
128, Department of Psychiatry, University of Münster, Münster, Nordrhein-Westfalen, DE
129, Institute of Neuroscience and Medicine (INM-1), Research Center Juelich, Juelich, DE
130, Institute of Medical Genetics and Pathology, University Hospital Basel, University of Basel, Basel, CH
131, Amsterdam Public Health Institute, Vrije Universiteit Medical Center, Amsterdam, NL
132, Centre for Integrative Biology, Università degli Studi di Trento, Trento, Trentino-Alto Adige, IT
133, Department of Psychiatry and Psychotherapy, Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Rheinland-Pfalz, DE
134, Psychiatry, Kaiser Permanente Northern California, San Francisco, CA, US
135, Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, GB
136, Centre for Addiction and Mental Health, Toronto, ON, CA
137, Department of Psychiatry, University of Toronto, Toronto, ON, CA
138, Division of Psychiatry, University College London, London, GB
139, Neuroscience Therapeutic Area, Janssen Research and Development, LLC, Titusville, NJ, US
140, Institute of Molecular and Cell Biology, University of Tartu, Tartu, EE
141, Psychosis Research Unit, Aarhus University Hospital, Risskov, Aarhus, DK
142, University of Liverpool, Liverpool, GB
143, Mental Health Center Copenhagen, Copenhagen University Hospital, Copenhagen, DK
144, Human Genetics and Computational Biomedicine, Pfizer Global Research and Development, Groton, CT, US
145, Psychiatry, Harvard Medical School, Boston, MA, US
146, Psychiatry, University of Iowa, Iowa City, IA, US
147, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University, Baltimore, MD, US
148, Human Genetics Branch, NIMH Division of Intramural Research Programs, Bethesda, MD, US
149, Department of Psychiatry and Psychotherapy, University Medical Center Göttingen, Göttingen, Niedersachsen, DE
150, Faculty of Medicine, University of Iceland, Reykjavik, IS
151, Child and Adolescent Psychiatry, Erasmus MC, Rotterdam, Zuid-Holland, NL
152, Psychiatry, Erasmus MC, Rotterdam, Zuid-Holland, NL
153, Psychiatry, Dalhousie University, Halifax, NS, CA
154, Division of Epidemiology, New York State Psychiatric Institute, New York, NY, US
155, Department of Clinical Medicine, University of Copenhagen, Copenhagen, DK
156, Human Genetics and Computational Biomedicine, Pfizer Global Research and Development, Cambridge, MA, US
157, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, US
158, Department of Medical & Molecular Genetics, King's College London, London, GB
159, Psychiatry & Behavioral Sciences, Stanford University, Stanford, CA, US
160, NIHR BRC for Mental Health, King's College London, London, GB
161, Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, US

Supplementary Tables

Report	Genomic data	Convergence	Notable pathway/gene sets
Schizophrenia ¹⁶	dURV, WES (RVAS)	a b c d e f g h	Neurons (not astrocytes or oligodendrocytes), excitatory and inhibitory neurons similar Genes intolerant to loss-of-function variation Genes whose mRNAs bind to FMRP or CELF4 Genes with bindings sites for RBFOX1, 2, or 3 Synaptic genes Post-synaptic density Activity-related cytoskeleton complex NMDA receptor components Genes with miR-137 binding sites
Schizophrenia ¹⁷	dURV, WES (RVAS)	a b c f g	Genes intolerant to loss-of-function variation Genes whose mRNAs bind to FMRP Genes with bindings sites for RBFOX1, 2, or 3 Genes with bindings sites for CHD8 Activity-related cytoskeleton complex NMDA receptor components Overlap with autism genes Overlap with developmental delay genes
Schizophrenia ¹⁸	GWA	a b d	Genes intolerant to loss-of-function variation Genes whose mRNAs bind to FMRP Serotonin 2C receptor complex Calcium ion import Membrane depolarization during action potential Synaptic transmission Abnormal behavior Abnormal nervous system electrophysiology Abnormal long term potentiation
Schizophrenia ¹	GWA	d e f g h	Synaptic genes Post-synaptic density Activity-related cytoskeleton complex NMDA receptor components Genes with miR-137 binding sites
Schizophrenia ¹⁹	GWA		Oligodendrocyte – gene transcription, lipid metabolism, and oxidation reduction Astrocyte – cell adhesion, G protein receptor signaling, gene transcription, signal transduction, small GTPase mediated signaling, and tyrosine kinase signaling

Supplementary Table 1. Gene sets or biological pathways previously implicated in schizophrenia. These analyses ask whether schizophrenia case/control genetic association results are “enriched” in the genes comprising a gene set. At least 20 gene sets have been implicated, and many are implicated by different types of genetic studies. This convergence is highly notable. However, these connect genetic risk for schizophrenia to a highly diverse and puzzling set of genes and biological pathways. dURV=disruptive or damaging ultra-rare variants. WES=whole exome sequencing. GWA=genome-wide, common variant association study. RVAS=rare variant association study.

Species	Label	Citation	Source	RNAseq tech.	Stage	Region	Cells	L1	L2	
Mouse	Gokce	²⁰	GEO GSE82187	Single cell	Adult	Striatum	1208	10	10	
	Habib	²¹	GEO GSE84371	Single nuclei	Adult	Hippocampus	1287	11	29	
	KI	Unpublished	Pending	Single cell	Adult	Cortex (Pvalb interneurons)	89	1	1	
	KI	¹¹	Linnarsson lab (URLs)	Single cell	Adult	Cortex + hippocampus	1996	6	41	
	KI	²²	GEO GSE74672	Single cell	Adult	Hypothalamus	772	4	62	
	KI	⁹	GEO GSE76381	Single cell	Adult	Midbrain	243	1	5	
	KI	⁹	GEO GSE76381	Single cell	Fetal	Midbrain	1290	8	22	
	KI	²³	GEO GSE75330	Single cell	Adult	Oligodendrocytes	5051	3	13	
	KI	Unpublished	Pending	Single cell	Adult	Striatum	529	2	6	
	Tasic	⁵	GEO GSE71585	Single cell	Adult	Cortex	1679	7	49	
	DroNc human	²⁴	Broad Single Cell Portal	Single nuclei	Adult	Cortex + Hippocampus	13313	11	22	
	Human	AIBS	Unpublished	Pending	Single nuclei	Adult	Cortex	4401	6	6
		Darmanis	⁷	GEO GSE67835	Single nuclei	Adult, fetal	Cortex	420	8	8
Lake		⁶	dbGaP phs000833.v3.p1	Single nuclei	Adult	Cortex (N=1)	3042	2	16	
DroNc mouse		²⁴	Broad Single Cell Portal	Single nuclei	Adult	Cortex + Hippocampus	14137	10	15	

Supplementary Table 2. Single nuclei or scRNAseq data from mouse or human brain. Source column indicates provenance of the data (URLs). KI=Karolinska Institutet. AIBS=Allen Institute for Brain Science. L1=number of Level 1 cell type categories. L2=number of Level 2 cell types (subdivisions of L1 types). All KI datasets were merged into a superset; all other datasets were used separately. These data are depicted in **Figure 1**.

KI Level 1 cell type	Level 2 subdivisions	Cells	Total Average Molecules
Hypothalamic dopaminergic neurons	4	41	14220
Serotonergic neuron	1	59	10583
Oxytocin & vasopressin expressing neurons	7	62	12156
Dopaminergic neuroblast	1	71	4902
Vascular leptomeningeal cells	1	76	3256
Embryonic dopaminergic neuron	3	93	6147
Microglia	4	98	4567
Embryonic midbrain nucleus neurons	2	105	11181
Neuronal progenitor	1	149	7172
Astrocytes / ependymal	4	159	6130
Radial glia like	3	166	7762
Medium spiny neuron	2	218	8610
Endothelial-mural	4	220	4319
Embryonic GABAergic neuron	3	221	10129
Dopaminergic neuron	5	243	10912
Cortical interneurons	16	379	15979
Pyramidal (SS)	8	290	17131
Hypothalamic glutamatergic neurons	32	305	8503
Oligodendrocyte precursor	1	308	3384
Striatal interneuron	4	311	10796
Hypothalamic GABAergic neurons	19	364	9602
Neuroblasts	8	426	6156
Pyramidal (CA1)	5	939	16066
Oligodendrocytes	11	4667	7581

Supplementary Table 3. Detail on the KI scRNAseq Level 1 and 2 dataset. There are 24 Level 1 cell type categories and 149 Level 2 subdivisions. The median number of cells in the Level 1 categories was 218 (interquartile range 91-306), and the Level 2 subdivisions range from 1-32. The numbers of single cells contributing to the Level 1 classification are shown. We found no relation between the number of cells and the cell type found to “fit” schizophrenia. For example, there are large numbers of oligodendrocytes and neuroblasts (which were not enriched for schizophrenia genomic findings), and the number of cells for medium spiny neurons (which were associated) were at the median. Likewise, the total average number of molecules detected in each cell type (as determined using Unique Molecular Identifiers) does not explain the enrichments found (note that medium spiny neurons have almost half as many molecules as hypothalamus dopaminergic neurons).

Supplementary References

1. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-7 (2014).
2. Finucane, H.K. *et al.* Partitioning heritability by functional category using GWAS summary statistics. *Nature Genetics* **47**, 1228-35 (2015).
3. Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nature Neuroscience* **19**, 1442-1453 (2016).
4. Skene, N.G. & Grant, S.G. Identification of Vulnerable Cell Types in Major Brain Disorders Using Single Cell Transcriptomes and Expression Weighted Cell Type Enrichment. *Front Neurosci* **10**, 16 (2016).
5. Tasic, B. *et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci* **19**, 335-46 (2016).
6. Lake, B.B. *et al.* Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586-90 (2016).
7. Darmanis, S. *et al.* A survey of human brain transcriptome diversity at the single cell level. *Proc Natl Acad Sci U S A* **112**, 7285-90 (2015).
8. Brawand, D. *et al.* The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343-8 (2011).
9. La Manno, G. *et al.* Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. *Cell* **167**, 566-580 e19 (2016).
10. Zeng, H. *et al.* Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures. *Cell* **149**, 483-96 (2012).
11. Zeisel, A. *et al.* Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138-42 (2015).
12. Kryuchkova-Mostacci, N. & Robinson-Rechavi, M. A benchmark of gene expression tissue-specificity metrics. *Brief Bioinform* (2016).
13. Nicolae, D.L. *et al.* Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS genetics* **6**, e1000888 (2010).
14. Pathway Analysis Subgroup of the Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat Neurosci* **18**, 199-209 (2015).
15. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-60 (2015).
16. Genovese, G. *et al.* Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nature Neuroscience* **19**, 1433-1441 (2016).
17. Singh, T. *et al.* The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. *Nat Genet* **49**, 1167-1173 (2017).
18. Pardiñas, A.F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and maintained by background selection. *Nature Genetics* (In press).
19. Goudriaan, A. *et al.* Specific glial functions contribute to schizophrenia susceptibility. *Schizophrenia bulletin* **40**, 925-935 (2013).
20. Gokce, O. *et al.* Cellular Taxonomy of the Mouse Striatum as Revealed by Single-Cell RNA-Seq. *Cell Rep* **16**, 1126-37 (2016).
21. Habib, N. *et al.* Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science* **353**, 925-8 (2016).
22. Romanov, R.A. *et al.* Molecular interrogation of hypothalamic organization reveals distinct dopamine neuronal subtypes. *Nat Neurosci* **20**, 176-188 (2016).
23. Marques, S. *et al.* Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Science* **352**, 1326-9 (2016).
24. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat Methods* **14**, 955-958 (2017).