bioRxiv preprint doi: https://doi.org/10.1101/032680; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Expression weighted cell type enrichments reveal genetic and cellular nature of major brain disorders

Nathan G. Skene[1] and Seth G.N. Grant[1]

Affiliations:

[1] Centre for Clinical Brain Sciences, Chancellors Building, University of Edinburgh, 49 Little France Crescent, Edinburgh, EH16 4SB, UK.

*Correspondence to: Seth.Grant@ed.ac.uk

Abstract

The cell types which trigger the primary pathology in many major brain diseases remains unknown. Correspondingly, many open questions remain regarding which cells are phenotypically disturbed in the same disorders. We report a new method, Expression Weighted Cell Type Enrichment (EWCE), which utilizes single cell transcriptomes to generate robust answers to both questions. We find that three major disorders show evidence of a primary microglial origin. Through post-mortem transcriptomes we find that Autism, Schizophrenia and Alzheimer's are found to show robustly detectable phenotypes across most of the major brain cell types.

Cell enrichments in disease susceptibility genes

Recent studies have dramatically increased the availability of unbiased single cell transcriptomes from brain tissue[1]. We have developed a set of new methods utilizing this data and applied it to obtain significant interpretative advances with two important data sources: *(1)* human disease associated genes (2) transcriptomic datasets from post-mortem human brains of diseased and control patients.

The method works by first calculating the mean level of expression within a cell type of a set of genes. The expression levels are first normalized to represent proportions: such that an expression level of 55 would mean that 55% of transcripts detected are within astrocytes (assuming equal sampling of each cell type). The expected distribution of cell type expression is then calculated through bootstrapping with 10000 random gene lists having the same length as the target list.

bioRxiv preprint doi: https://doi.org/10.1101/032680; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

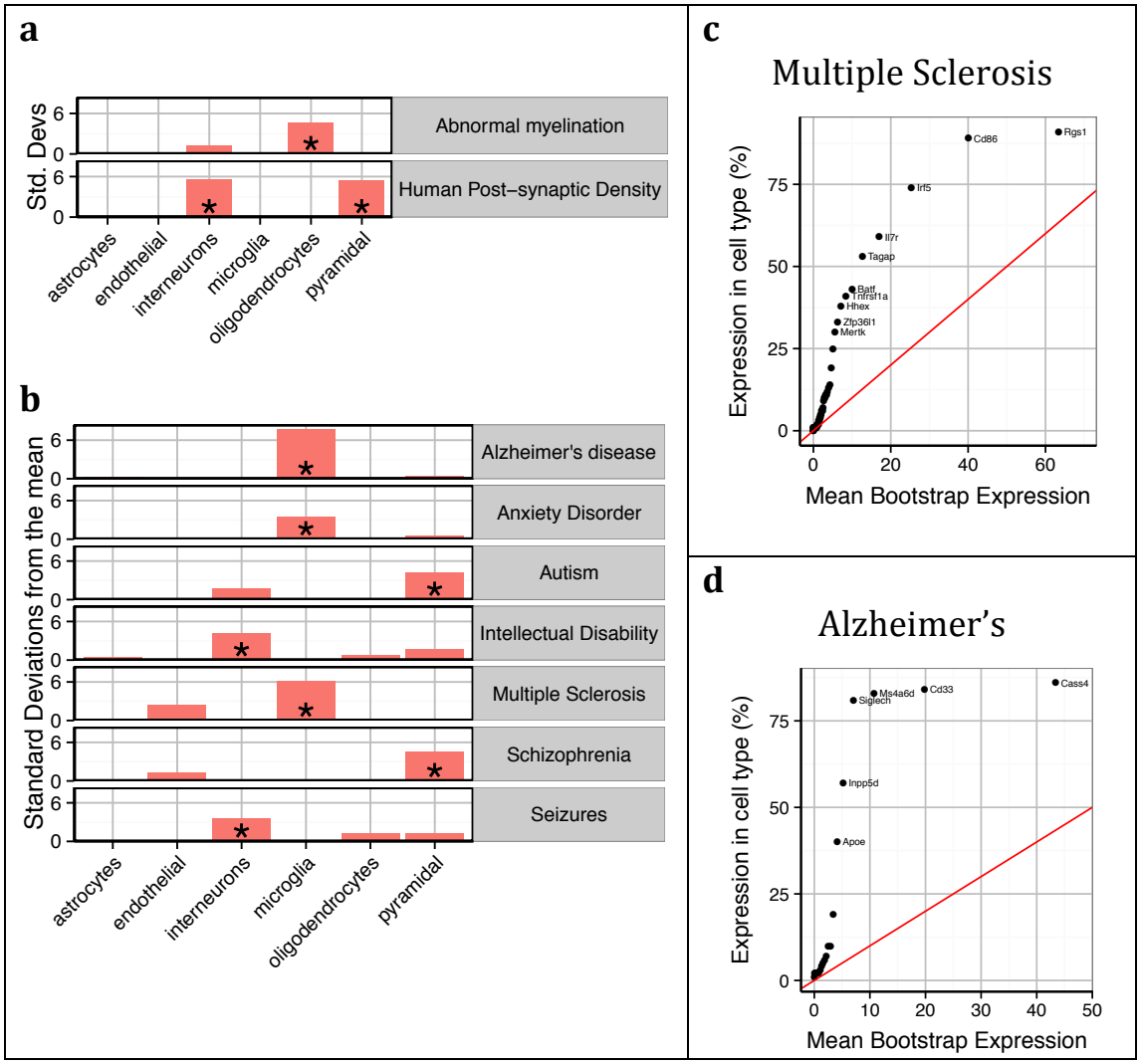
We first sought to confirm that the method detects expected cell type enrichments (Fig. 1a). We used two gene lists for this: the 1461 genes associated with the human post-synaptic density (hPSD)[2], and 185 genes which associated with abnormal myelination according to their Human Phenotype Ontology annotations. As expected, the hPSD was significantly associated with interneurons and pyramidal neurons ($p<0.0001$) and abnormal myelination phenotypes were enriched for mutations in oligodendrocyte enriched genes (Benjamini-Hochberg corrected $p=0.001$).

We then tested for cell type enrichments in susceptibility genes for seven major brain disorders: Alzheimer's disease, Anxiety disorders, Autism, Intellectual Disability, Multiple Sclerosis, Schizophrenia and Seizures. Three of the disorders were found to show strong evidence for being primary microglial disorders (Alzheimer's, $p=0$; Anxiety, $p=0.016$; Multiple Sclerosis, $p=0$). It has previously been noted that a number of Alzheimer's susceptibility genes have high levels of expression in microglia[3], and so we tested whether the enrichment seen here is a result only of expression in those few genes, or whether all susceptibility genes have higher levels of expression than expected by chance. We found that evidence strongly supports the latter hypothesis for both Alzheimer's and Multiple Sclerosis (Fig. 1c—e): for Alzheimer's, every gene except Cass4 was found to have higher expression in microglia than expected by chance (Fig. 1e).

Anxiety disorders covers a diverse set of syndromes, including *social anxiety* and *panic disorders*. As with the other major psychiatric disorders, such as Schizophrenia and Bipolar disorder, they generally exhibit teenage and early-adult onset. Surprisingly however, our results suggest that anxiety disorders have a microglial origin ($p= 0.0175$) while Schizophrenia genes have high expression in cortical pyramidal neurons ($p=0.0023$). Genetic association between anxiety disorders and the brain's primary immune cell could be related to why anxiety disorders are found to be more prevalent amongst those suffering from disease and immunological disorders[4]. One of the more intriguing findings is that intellectual disabilities are enriched for interneuron genes ($p<0.0001$) and are not correspondingly enriched for pyramidal neurons; with 750 genes associated with Intellectual disabilities this is not likely to be a result of transient bias in the set composition. Autism on the other hand is enriched for pyramidal neuron genes ($p=0.012$). It is striking that the two broad

bioRxiv preprint doi: https://doi.org/10.1101/032680; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

classes of infant onset mental disorders appear to affect distinct cell classes, particularly so as we have shown that synaptic genes—which are associated with both classes of disorder[5,6]---are, as expected, strongly enriched in both cell types. One candidate explanation is that classical autism symptoms may be specifically related to pyramidal cell pathology; genes then associated with non-syndromic intellectual disabilities, rather than the autistic spectrum, may thus be depleted for pyramidal neuron expressed genes. The enrichments for Autism and Schizophrenia were again found to affect all genes throughout the lists, rather than a subset of strong markers (Figs. 1f--g). Seizures were found to be the fourth neuron-associated condition, and also were specific to a particular type, interneurons[7,8], which have been found to be causally associated with several classes of seizures.



bioRxiv preprint doi: https://doi.org/10.1101/032680; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

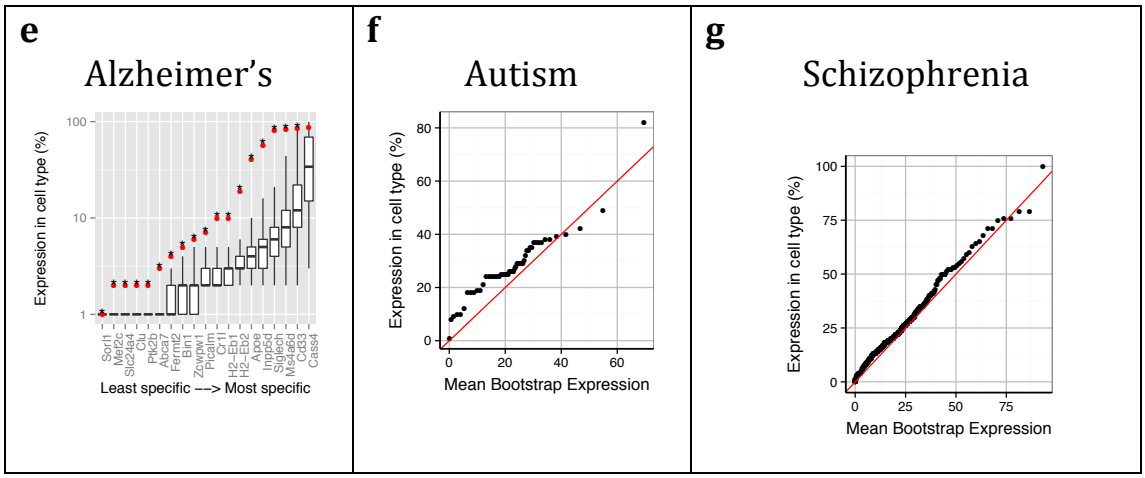


Figure 1: *Susceptibility genes for major human brain disorders show distinct cell type enrichments.*

(a) Two gene sets with a strong prior expectation for cell type enrichment---the human postsynaptic density, and genes with Human Phenotype Ontology annotations for abnormal myelination---were detected using bootstrapping to have higher expression in neurons and oligodendrocytes respectively.

(b) Bootstrapping tests performed using the EWCE method show that seven different classes of brain disorder show enrichment in particular cell types.

(c) Multiple Sclerosis associated genes are strongly enriched for microglial expression. This plot shows that this is not just a property of a few genes, but instead almost every single gene shows higher levels of expression in microglia than would be expected by chance. The plot shows the actual level of expression of the susceptibility gene, against the mean expression level of the ith most expressed gene in a bootstrapping analysis of lists of 19 genes. If microglial expression in MS genes was randomly distributed, the genes would be expected to fall along the red line.

(d) All Alzheimer's susceptibility genes are more enriched for microglial expression than expected by chance.

(e) Bootstrap distributions of expected microglial expression levels of Alzheimer's genes. Red dots mark the expression level of the susceptibility genes, while the associated boxplots denote the expected expression level of the ith most expressed gene, in a list of 19 genes, as determined using bootstrapping. Asterisks behind the red dots denote that the gene has higher expression in the cell type than expected by chance ($p<0.05$).

(f) Though hundreds of genes are associated with Autism, they are all found to show a moderate increase in expression in pyramidal neurons.

(g) The hundreds of genes associated with Schizophrenia are found to show a moderate, but highly significantly, increased expression in pyramidal neurons.

bioRxiv preprint doi: https://doi.org/10.1101/032680; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Robust cell enrichments in post-mortem disease transcriptomes

We next sought to ascertain whether the method could be used to describe the cellular nature of disease phenotypes found in post-mortem brain samples. Many transcriptome studies have been performed for major brain disorders, in an effort to cast light on the pathological basis of the conditions. We reasoned that because our method utilizes genome-wide data to define 'set membership' in a quantitative and brain specific manner, it may be more robust and relevant than GO enrichments at detecting the hidden variables underlying brain diseases.

To perform this analysis, we first perform a standard differential expression analysis and rank order the affected probes by t-statistic, then take the 250 most upregulated, and 250 most downregulated genes. We then perform an Expression Weighted Cell type Enrichment (EWCE) analysis, wherein the random samples are obtained by reordering the ranked list 10000 times.

The pathological characteristics of Alzheimer's disease are relatively well understood, with inflammatory gliosis and synapse loss becoming more acute as the disease progresses. We applied the method to an Alzheimer's dataset which examined changes in fourteen cortical and three non-cortical regions, with between 51—70 samples per region. Differential expression was calculated for genes whose expression is affected by increases in Braak score. Across all the brains tested, a consistent cell enrichment signature was detected (see Fig 2a). We calculated the mean fold-enrichment for each cell type in each region, for both up- and down- regulated genes, and applied Bayesian estimation to determine whether an actual fold enrichment of zero falls within the 95% Highest Density Interval (HDI): for each of the cell types, we found this to not be the case. Interneuron and pyramidal neuron genes were found to be enriched amongst those which were down-regulated, while oligodendrocytes, microglia, astrocytes and endothelial cells showed evidence of up-regulation. The enrichments detected are directionally consistent with the known pathology of the disease[9], strongly supporting the notion that this is a powerful and robust technique for determining hidden variables in transcriptome data.

bioRxiv preprint doi: https://doi.org/10.1101/032680; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

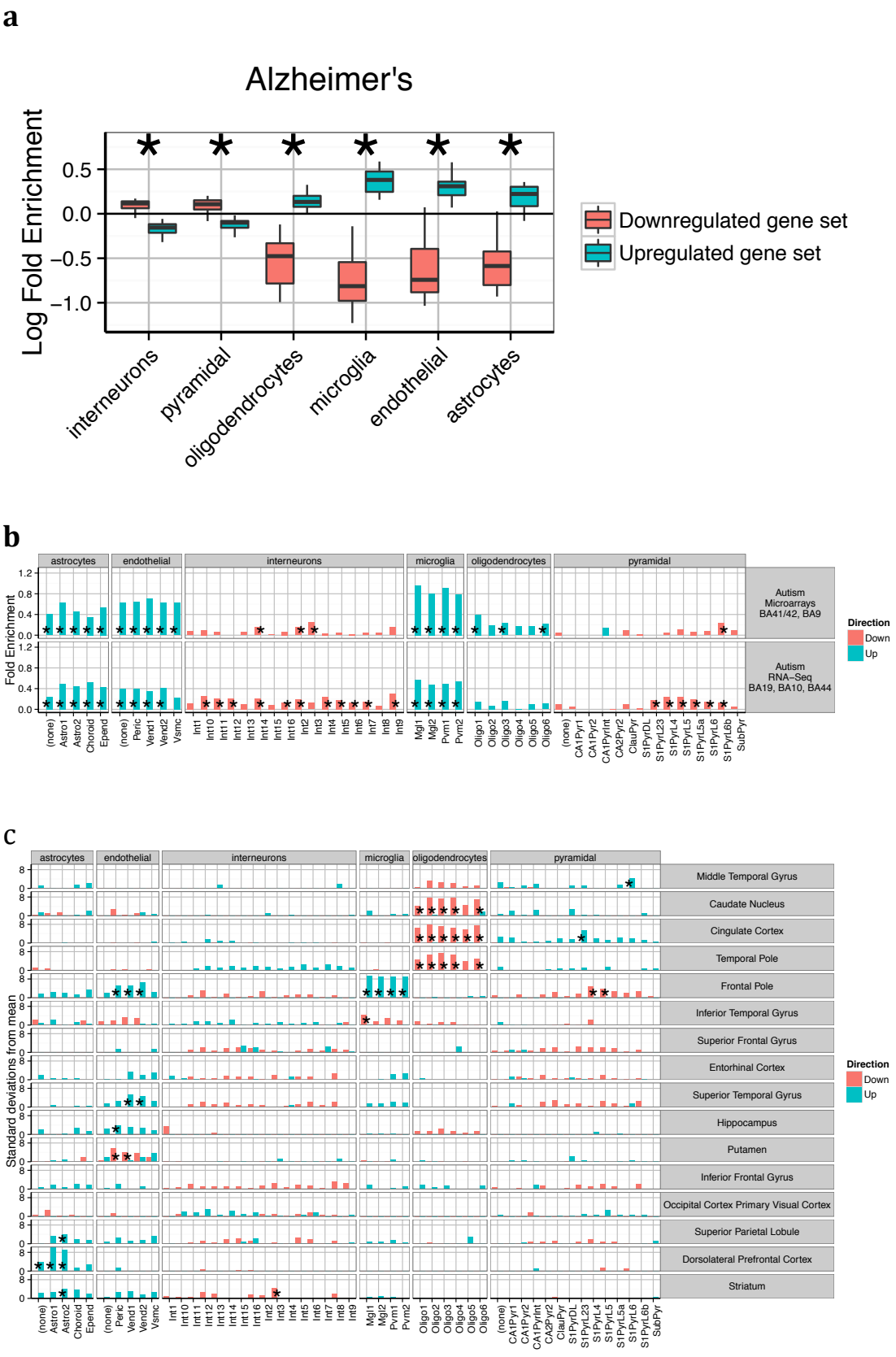


Figure 2: *Post-mortem transcriptomes from patients with Alzheimer's, Autism and Schizophrenia show distinctive cellular phenotypes.*

bioRxiv preprint doi: https://doi.org/10.1101/032680; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(a) Consistent fold enrichments were found for each cell type across fourteen cortical and three subcortical brain regions of Alzheimer's patients. The boxplots mark the distribution of fold enrichments for that cell types seen across all the brain regions examined. Asterixes mark that the fold enrichment for each cell type that was found to be significantly non-zero with p<0.05.

(b) Two independent autism studies show the same cellular phenotypes, including upregulation of glial cells and downregulation of neurons. Asterisks mark those cell types found to be significantly differential with p<0.05 after BH correction across over all groups.

(c) Cellular phenotypes in Schizophrenia are regionally dependent but cluster into groups, with a number of regions including the cingulate cortex and temporal pole showing downregulation of oligodendrocyte genes while the prefrontal cortex exhibits upregulation of endothelial and astrocyte genes as well as downregulation of deep pyramidal neurons in the anterior region. The analyses shown are based on an integrative analysis of six independent studies, though not all brain regions featured in all studies.

Having validated the method's ability to detect known pathological phenotypes, we sought to apply the method to two disorders which are less well characterized: Autism and Schizophrenia. Two publically available transcriptome datasets were used for Autism[10,11], with one study using samples from areas BA41/42, BA9, BA19, BA10 and BA44. The two studies were analysed separately, and within each study differential expression was tested over all cortical regions. Remarkably, we again found a consistent cellular enrichment signature across the two studies (Fig 2b). We emphasise that these were two independent studies, performed by different labs, on different cortical regions, with one study using RNA-Sequencing and the other Illumina microarrays.

Both Autism studies, like the Alzheimer's studies, indicated that the disease processes affect every major cell type in the brain. As in Alzheimer's disease, interneuron and pyramidal enrichments were found in down-regulated genes, while enrichments of glial and endothelial cells were present in the up-regulated gene sets. The cellular changes expected to correspond to decreased expression of neuronal transcripts is unclear, with no consensus in the literature about how neurons are affected in autistic patients[12-14]. Up-regulation of astro- and micro- glial genes, which indicates activation in those cell types, is broadly supported by past studies which have shown elevated expression of markers, increased cell densities and altered morphologies for both cell types[15-18]. The finding that

bioRxiv preprint doi: https://doi.org/10.1101/032680; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

endothelial genes are up-regulated could be related to the decrease in cerebral blood flow seen in temporal and frontal cortices of autistic patients[19].

We next extended the study to Schizophrenia. We utilised data from six independent transcriptomic studies, providing data for many brain areas. Four of the studies included samples from the dorsolateral prefrontal cortex, while other brain regions including hippocampus and cingulate cortex were covered by at least two of the datasets. To maximize the utility of these replicate studies, the cell type bootstrap data was summed across each independent cohort allowing pooled estimates for cellular changes. As was expected based on the disease literature, regional changes were found to be divergent. Some regions (e.g. the primary visual cortex) were found to show no significant changes, while the most pronounced enrichments were seen in the prefrontal and cingulate cortices.

Those regions showing alterations were found to cluster into two groups: (1) those with decreased oligodendrocyte expression; (2) increased astrocyte and/or endothelial expression. All samples from the prefrontal cortex fell into the second cluster. The dorsolateral prefrontal cortex (BA46) showed an enrichment of astrocyte genes 10.3 standard deviations from the bootstrapped mean ($p<0.0001$, $dist(x,\bar{b})=10.3\sigma$). Significant up-regulation of astrocyte genes (after Benjamini Hochberg correction) was also seen in Striatum ($p<0.04$, $dist(x,\bar{b})=4.1\sigma$) and Superior Parietal Lobe ($p<0.03$, $dist(x,\bar{b})=4\sigma$). Three regions showed significant evidence for up-regulation of endothelial cells in Schizophrenics: the hippocampus ($p=0.03$, $dist(x,\bar{b})=3.8\sigma$), superior temporal gyrus ($p=0.0001$, $dist(x,\bar{b})=5.5\sigma$) and the frontal pole ($p<0.0007$, $dist(x,\bar{b})=5.3\sigma$). Whilst fitting within this cluster, the anterior prefrontal cortex was found to have a number of distinguishing features including a highly significant up-regulation of microglial genes ($p<0.0001$, $dist(x,\bar{b})=9.3\sigma$) and downregulation of layer 4/5 pyramidal neuron genes ($p<0.01$, $dist(x,\bar{b})=4.7\sigma$).

The second set of brain regions show totally distinct phenotypes from those described above. Astrocyte and endothelial expression appears normal, while highly significant down-regulation of Oligodendrocyte genes was found in the Cingulate Cortex ($p<0.0001$, $dist(x,\bar{b})=9.3\sigma$), Caudate Nucleus ($p<0.0001$, $dist(x,\bar{b})=7.4\sigma$), and

bioRxiv preprint doi: https://doi.org/10.1101/032680; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Temporal Pole ($p<0.0001$, $dist(x,\bar{b})=7.6\sigma$). There does not appear to be a bias towards a particular cell-type within the oligodendrocyte lineage suggesting that all stages, including precursor cells are affected, though sets shared across the lineage could mask a more specific effect.

Conclusions

The phenotypes discovered here open new avenues for modeling brain disorders in animals. Though on a genetic level each of the disorders could be pinpointed as primarily affecting a particular cell type, the post-mortem transcriptomes were found to show changes across a much broader set of cells, indicating that they are cell non-autonomous effects. There is little in the literature that can explain how these cell non-autonomous effects could occur: for instance, both Autism and Schizophrenia appear to be genetically neuronal disorders, and yet both show evidence for endothelial disruption. Though both disorders have been shown to have decreased cerebral blood flow, it is unclear whether this might be related to up-regulation of endothelial genes, or neuronal mutations should have any effect on vascular cells. Both of these are suitable problems for mouse models and defining disease models in terms of whether they recapitulate these cellular expression changes across multiple brain regions could open the path towards validation and community acceptance of particular models. We note that while this study has focused on brain disorders, the same methodology could be applied to other anatomical regions once suitable cell type expression data were available.

Methods

Cell type expression proportions

Raw cell type mRNA expression data was downloaded from the Linnarsson lab webpage (data annotated as being from 17th August 2014)[1]. Proportions are calculated separated for subcell types (I.e. S1 Pyramidal cell from Layer 6b) and grouped cell types (i.e. Astrocytes). The mean level of expression is calculated for each gene, for each cell type. For each gene, the expression level within each cell is divided by the summed expression over all cell types. The proportion calculated is thus relative to the number of other cell types considered, and does not take into account the relative number of cells of a given cell type that are found in the brain.

Bootstrapping enrichment within gene lists

bioRxiv preprint doi: https://doi.org/10.1101/032680; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Human genes were first converted to mouse orthologs based on MGI and HGNC symbols, using Biomart. For gene set (but not transcriptome) enrichment the background gene set is comprised of all genes which have orthologs between the two species, including those in the target list. The proportion of expression in each cell type is calculated as a matrix for each gene, then summed to get total expression in each cell type across the whole gene list. This calculation is then repeated for 10,000 randomly generated gene lists, having the same length as the target gene list, with the genes randomly selected from the background gene set. The probability of cellular enrichment is then calculated based on the number of bootstrapped gene lists with higher cell type specific expression than the target list. The fold enrichment is calculated as the expression in the target gene list ($x_{cell\ type}$) divided by the mean level of expression in the bootstrap samples ($\bar{b}_{cell\ type}$). The standard deviations from the mean (denoted throughout the paper as dist(x,$\bar{b}$)) is calculated as $\left(x_{cell\ type} - \bar{b}_{cell\ type}\right)/\sigma_{cell\ type}$ where $\sigma_{cell\ type}$ is the standard deviation of the bootstrapped expression levels. Where probabilities are stated for gene list enrichments, all p-values were adjusted for multiple testing using the Benjamini-Hochberg (BH) method, with correction performed over all tests involved in the associated figure.

Disease gene association lists

The disease gene associations were curated from the literature, being largely based on the most recent and authoritative studies. The sources are shown in and the genes comprising each list are in Supplementary Table 01. Most the disease lists are associated with publications, are appropriate references are listed as follows: Abnormal myelination[20], Human post-synaptic proteome[2], Alzheimer's disease[21,22], Anxiety disorder[23], Autism[24], Intellectual Disability[24], Multiple Sclerosis[25], Schizophrenia[26-30].

Sources of transcriptomic datasets used

The transcriptome datasets used in the study were all obtained from publically available sources. These are detailed below:

Alzheimer's:

Publication: PMID17845826[31]

Source: Author maintains on a private server. Details provided by email.

Brain regions: *frontal pole, occipital cortex primary visual cortex, inferior temporal gyrus, middle temporal gyrus, superior temporal gyrus, posterior cingulate cortex, anterior cingulate, entorhinal cortex, temporal pole, precentral gyrus, inferior frontal gyrus. dorsolateral refrontal cortex, superior parietal lobule, superior frontal gyrus, caudate nucleus, hippocampus, putamen*

Technology platform: Affymetrix Human Genome U133A + U133B Arrays

Analysis notes:

Raw data was read into R from .cel files using the affy package[32] and normalized using the RMA package. Probe annotations were obtained

bioRxiv preprint doi: https://doi.org/10.1101/032680; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

through Biomart. Linear models were fitted to the expression values for each probe using limma[33], with Braak stage as the independent variables and sex, age, and ancestry taken as extraneous variables. The U133A and U133B arrays were analysed separately and then merged by combining the rows from both studies into one table.

Autism:

Publication: PMID25494366[10]
Source: http://www.arkinglab.org/resources/
Brain regions: associative visual cortex, frontal pole, inferior frontal gyrus
Technology platform: RNA-Seq
Analysis notes:

The preprocessed data was downloaded from the lab webpage as 'Samples104BGenes-EDASeqFull.txt'. A linear model was fitted to the expression values for each probe using limma[33], with disease status as the independent variables and brain region, age, and processing site and sex taken as extraneous variables. Age was treated within the model as a spline with three degrees of freedom.

Publication: PMID21614001[11]
Source: GSE28521
Brain regions: superior temporal gyrus, dorsolateral prefrontal cortex
Technology platform: Illumina HumanRef-8 v3.0
Analysis notes:

Normalised data was obtained from GEO. Probe annotations were obtained through Biomart. A linear model was fitted to the expression values for each probe using limma[33], with disease status as the independent variables and brain region as an extraneous variable.

Schizophrenia:

Publication: PMID25796564[34]
Source: GSE53987
Brain regions: prefrontal cortex, hippocampus, striatum
Technology platform: Affymetrix Human Genome U133 Plus 2.0
Analysis notes:

Raw data was read into R from .cel files using the affy package[32] and normalized using the RMA package. Probe annotations were obtained through Biomart. Linear models were fitted to the expression values for each probe using limma[33], with disease status as the independent variables with sex and age as extraneous variables. Schizophrenia and Bipolar Disorder samples were both studied and the later treated as schizophrenia samples in the later analysis.

Publication: PMID16139990[35]
Source: Author maintains on a private server. Details provided by email.
Brain regions: *frontal pole, occipital cortex primary visual cortex, inferior temporal gyrus, middle temporal gyrus, superior temporal gyrus,*

bioRxiv preprint doi: https://doi.org/10.1101/032680; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

posterior cingulate cortex, anterior cingulate, entorhinal cortex, inferior frontal gyrus, dorsolateral prefrontal cortex, superior parietal lobule, superior frontal gyrus, caudate nucleus, hippocampus, putamen

Technology platform: Affymetrix Human Genome U133A + U133B Arrays

Analysis notes:

Raw data was read into R from .cel files using the affy package[32] and normalized using the RMA package. Probe annotations were obtained through Biomart. Linear models were fitted to the expression values for each probe using limma[33], with disease status as the independent variables and sex and age taken as extraneous variables. The U133A and U133B arrays were analysed separately and then merged by combining the rows from both studies into one table.

Publication: PMID22868662

Source: Author maintains on a private server. Details provided by email.

Brain regions: *middle temporal gyrus, anterior cingulate, temporal pole, dorsolateral prefrontal cortex*

Technology platform: Affymetrix Human Genome U133 Plus 2.0

Analysis notes: Raw data was read into R from .cel files using the affy package[32] and normalized using the RMA package. Probe annotations were obtained through Biomart. Linear models were fitted to the expression values for each probe using limma[33], with disease status as the independent variables with sex and age as extraneous variables.

Publication: PMID21538462[36]

Source: GSE21935

Brain regions: *superior temporal gyrus*

Technology platform: Affymetrix Human Genome U133 Plus 2.0

Analysis notes: Raw data was read into R from .cel files using the affy package[32] and normalized using the RMA package. Probe annotations were obtained through Biomart. Linear models were fitted to the expression values for each probe using limma[33], with disease status as the independent variables with sex and age as extraneous variables.

Publication: PMC2783475[37]

Source: GEO21138

Brain regions: *dorsolateral prefrontal cortex*

Technology platform: Affymetrix Human Genome U133 Plus 2.0 Array

Analysis notes: Processed data was downloaded from GEO. Probe annotations were obtained through Biomart. Linear models were fitted to the expression values for each probe using limma[33], with disease status as the independent variables. Actual ages were not provided with the dataset, however three catagories of how long the disease had been diagnosed for were provided. These three catagories were used as a proxy for age as an extraneous variable.

bioRxiv preprint doi: https://doi.org/10.1101/032680; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Publication: PMID19255580[38]
Source: GEO17612
Brain regions: *frontal pole*
Technology platform:
Analysis notes: Raw data was read into R from .cel files using the affy package[32] and normalized using the RMA package. Probe annotations were obtained through Biomart. Linear models were fitted to the expression values for each probe using limma[33], with disease status as the independent variables with sex and age as extraneous variables.

Merging of multiple datasets for Schizophrenia transcriptome analysis

For the Schizophrenia analysis, data from multiple independent studies was available for a number of brain regions. Superior temporal gyrus came from two studies (PMID1613999 and PMID21538462). Middle temporal gyrus was based on two studies (PMID22868662 and PMID16139990). Hippocampus was based on three studies, including one Bipolar cohort (PMID16139990 and PMID25796564). Frontal pole was based on two studies (PMID19255580 and PMID16139990). Dorsolateral prefrontal cortex was based on five studies, including one Bipolar cohort (PMID22868662, PMID18778695, PMID25796564 and PMID16139990). Cingulate cortex was based on three studies, including samples from anterior and posterior areas (PMID16139990 and PMID22868662).

Standard cell type bootstrapping was done for each individual study as reported above, and the cell type expression proportions for each bootstrap sample was stored as a matrix, with a row for each of the 10000 bootstrap replicates and a column for each cell type. For each individual study being merged, the bootstrap output matrices were summed to form a consensus estimated distribution of random cell type expression proportions. The cell type proportions for the target gene list in each individual study were summed. Calculation of p-values and fold enrichment was then performed as for an individual study, as described above.

References:

1 Zeisel, A. *et al.* Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138-1142 (2015).
2 Bayés, À. *et al.* Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nature neuroscience* **14**, 19-21 (2011).
3 Mhatre, S. D., Tsai, C. A., Rubin, A. J., James, M. L. & Andreasson, K. I. Microglial Malfunction: The Third Rail in the Development of Alzheimer's Disease. *Trends in neurosciences* **38**, 621-636 (2015).
4 Roy-Byrne, P. P. *et al.* Anxiety disorders and comorbid medical illness. *General hospital psychiatry* **30**, 208-225 (2008).
5 Hamdan, F. F. *et al.* Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. *The American Journal of Human Genetics* **88**, 306-316 (2011).

bioRxiv preprint doi: https://doi.org/10.1101/032680; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

6 Pinto, D. *et al.* Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *The American Journal of Human Genetics* **94**, 677-694 (2014).

7 Ogiwara, I. *et al.* Nav1. 1 localizes to axons of parvalbumin-positive inhibitory interneurons: a circuit basis for epileptic seizures in mice carrying an Scn1a gene mutation. *The Journal of neuroscience* **27**, 5903-5914 (2007).

8 Hammad, M. *et al.* Transplantation of GABAergic Interneurons into the Neonatal Primary Visual Cortex Reduces Absence Seizures in Stargazer Mice. *Cerebral Cortex*, bhu094 (2014).

9 Grammas, P. Neurovascular dysfunction, inflammation and endothelial activation: implications for the pathogenesis of Alzheimer's disease. *J Neuroinflammation* **8**, 26 (2011).

10 Gupta, S. *et al.* Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nature communications* **5** (2014).

11 Voineagu, I. *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380-384 (2011).

12 Courchesne, E. *et al.* Neuron number and size in prefrontal cortex of children with autism. *Jama* **306**, 2001-2010 (2011).

13 van Kooten, I. A. *et al.* Neurons in the fusiform gyrus are fewer and smaller in autism. *Brain* **131**, 987-999 (2008).

14 Casanova, M. F., El-Baz, A. S. & Suri, J. S. *Imaging the Brain in Autism*. (Springer, 2013).

15 Laurence, J. & Fatemi, S. Glial fibrillary acidic protein is elevated in superior frontal, parietal and cerebellar cortices of autistic subjects. *The Cerebellum* **4**, 206-210 (2005).

16 Tetreault, N. A. *et al.* Microglia in the cerebral cortex in autism. *Journal of autism and developmental disorders* **42**, 2569-2584 (2012).

17 Morgan, J. T. *et al.* Abnormal microglial–neuronal spatial organization in the dorsolateral prefrontal cortex in autism. *Brain research* **1456**, 72-81 (2012).

18 Vargas, D. L., Nascimbene, C., Krishnan, C., Zimmerman, A. W. & Pardo, C. A. Neuroglial activation and neuroinflammation in the brain of patients with autism. *Annals of neurology* **57**, 67-81 (2005).

19 Ohnishi, T. *et al.* Abnormal regional cerebral blood flow in childhood autism. *Brain* **123**, 1838-1844 (2000).

20 Groza, T. *et al.* The human phenotype ontology: Semantic unification of common and rare disease. *The American Journal of Human Genetics* **97**, 111-124 (2015).

21 Bertram, L., McQueen, M. B., Mullin, K., Blacker, D. & Tanzi, R. E. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nature genetics* **39**, 17-23 (2007).

22 Lambert, J.-C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature genetics* **45**, 1452-1458 (2013).

23 Le-Niculescu, H. *et al.* Convergent functional genomics of anxiety disorders: translational identification of genes, biomarkers, pathways and mechanisms. *Translational psychiatry* **1**, e9 (2011).

bioRxiv preprint doi: https://doi.org/10.1101/032680; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

24 Sanders, S. J. *et al.* Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87**, 1215-1233 (2015).
25 Lill, C. *et al.* in *Alzheimer Research Forum.*
26 Consortium, S. W. G. o. t. P. G. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427 (2014).
27 Xu, B. *et al.* De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nature genetics* **44**, 1365-1369 (2012).
28 Girard, S. L. *et al.* Increased exonic de novo mutation rate in individuals with schizophrenia. *Nature genetics* **43**, 860-863 (2011).
29 Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506**, 179-184 (2014).
30 Ayalew, M. *et al.* Convergent functional genomics of schizophrenia: from comprehensive understanding to genetic risk prediction. *Molecular psychiatry* **17**, 887-905 (2012).
31 Haroutunian, V., Katsel, P. & Schmeidler, J. Transcriptional vulnerability of brain regions in Alzheimer's disease and dementia. *Neurobiology of aging* **30**, 561-573 (2009).
32 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).
33 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*, gkv007 (2015).
34 Reinhart, V. *et al.* Evaluation of TrkB and BDNF transcripts in prefrontal cortex, hippocampus, and striatum from subjects with schizophrenia, bipolar disorder, and major depressive disorder. *Neurobiology of disease* **77**, 220-227 (2015).
35 Katsel, P., Davis, K. L. & Haroutunian, V. Variations in myelin and oligodendrocyte-related gene expression across multiple brain regions in schizophrenia: a gene ontology study. *Schizophrenia research* **79**, 157-173 (2005).
36 Barnes, M. R. *et al.* Transcription and pathway analysis of the superior temporal cortex and anterior prefrontal cortex in schizophrenia. *Journal of neuroscience research* **89**, 1218-1227 (2011).
37 Narayan, S. *et al.* Molecular profiles of schizophrenia in the CNS at different stages of illness. *Brain research* **1239**, 235-248 (2008).
38 Maycox, P. R. *et al.* Analysis of gene expression in two large schizophrenia cohorts identifies multiple changes associated with nerve terminal function. *Molecular psychiatry* **14**, 1083-1094 (2009).