THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Identification of vulnerable cell types in major brain disorders using single cell transcriptomes and expression weighted cell type enrichment

Citation for published version:
Skene, NG & Grant, SGN 2016, 'Identification of vulnerable cell types in major brain disorders using single cell transcriptomes and expression weighted cell type enrichment', *Frontiers in Neuroscience*, vol. 10, no. JAN, 16. https://doi.org/10.3389/fnins.2016.00016

Digital Object Identifier (DOI):
10.3389/fnins.2016.00016

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Frontiers in Neuroscience

Publisher Rights Statement:
Copyright © 2016 Skene and Grant. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

OPEN ACCESS

Download date: 24. Aug. 2022

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 27 January 2016
doi: 10.3389/fnins.2016.00016

Identification of Vulnerable Cell Types in Major Brain Disorders Using Single Cell Transcriptomes and Expression Weighted Cell Type Enrichment

*Nathan G. Skene and Seth G. N. Grant **

Centre for Clinical Brain Sciences, Edinburgh University, Edinburgh, UK

The cell types that trigger the primary pathology in many brain diseases remain largely unknown. One route to understanding the primary pathological cell type for a particular disease is to identify the cells expressing susceptibility genes. Although this is straightforward for monogenic conditions where the causative mutation may alter expression of a cell type specific marker, methods are required for the common polygenic disorders. We developed the Expression Weighted Cell Type Enrichment (EWCE) method that uses single cell transcriptomes to generate the probability distribution associated with a gene list having an average level of expression within a cell type. Following validation, we applied EWCE to human genetic data from cases of epilepsy, Schizophrenia, Autism, Intellectual Disability, Alzheimer's disease, Multiple Sclerosis and anxiety disorders. Genetic susceptibility primarily affected microglia in Alzheimer's and Multiple Sclerosis; was shared between interneurons and pyramidal neurons in Autism and Schizophrenia; while intellectual disabilities and epilepsy were attributable to a range of cell-types, with the strongest enrichment in interneurons. We hypothesized that the primary cell type pathology could trigger secondary changes in other cell types and these could be detected by applying EWCE to transcriptome data from diseased tissue. In Autism, Schizophrenia and Alzheimer's disease we find evidence of pathological changes in all of the major brain cell types. These findings give novel insight into the cellular origins and progression in common brain disorders. The methods can be applied to any tissue and disorder and have applications in validating mouse models.

Keywords: single cell genomics, transcriptome, RNA-seq, genetics, schizophrenia, autism, anxiety, Alzheimer's Disease

OPEN ACCESS

Edited by:
Mark Webber Miller,
Boston University School of Medicine, USA

Reviewed by:
Sulev Kõks,
University of Tartu, Estonia
Mark Logue,
Boston University School of Medicine, USA

***Correspondence:**
Seth G. N. Grant
seth.grant@ed.ac.uk

Specialty section:
This article was submitted to Neurogenomics, a section of the journal Frontiers in Neuroscience

Received: *25 November 2015*
Accepted: *12 January 2016*
Published: *27 January 2016*

Citation:
Skene NG and Grant SGN (2016) Identification of Vulnerable Cell Types in Major Brain Disorders Using Single Cell Transcriptomes and Expression Weighted Cell Type Enrichment. Front. Neurosci. 10:16. doi: 10.3389/fnins.2016.00016

INTRODUCTION

The brain has a highly complex cellular architecture characterized by a diverse set of cell types that are highly interconnected. Identifying the cell types involved with the pathogenesis of disease is particularly challenging in heterogeneous tissues where cell types are often poorly defined. In the majority of brain disorders evidence exists for changes affecting multiple cell types. It has proven problematic to determine which cells are associated with the primary disease pathology and which

are altered as secondary "reactive" responses. Genomic technologies have contributed important mechanistic insights into the primary genetic basis of pathogenesis through studies of mutations and variants that increase disease susceptibility. The recent availability of single cell transcriptomes (SCT) from brain tissue (Darmanis et al., 2015; Dueck et al., 2015; Saraiva et al., 2015; Shin et al., 2015; Usoskin et al., 2015; Zeisel et al., 2015) offer the opportunity to ask which cells express these susceptibility genes, and thereby identify the primary cellular pathology. Since changes in primary and secondary cell types will be reflected in transcriptomic studies of diseased tissue, SCTs could also be used to identify these two classes of cell types.

A method known as Population Specific Expression Analysis (PSEA) (Kuhn et al., 2011) was previously developed with the goal of extracting cell-type data from disease transcriptomes and has been successfully used to determine the cellular pathology in Huntington's and Parkinson's disease (Capurro et al., 2014). A limitation with this method was the dependence on predefined sets of cellular genes that are specific to particular cell types. As the method depended on linear modeling, it could not be applied to unquantified gene sets thereby limiting its applicability for use in gene expression studies. To overcome these limitations, we have developed the Expression Weighted Cell-type Enrichment (EWCE) method that statistically evaluates whether a set of genes has higher expression within a particular cell type than can be reasonably expected by chance. For the purposes of the current study, the cell type expression profiles are defined using data from single cell RNA-sequencing. The method can be applied to any experimental source that generates a list of genes, including but not limited to genetic, transcriptomic and proteomic studies. The workflow for two such experimental designs are shown in **Figure 1A** alongside data from an example gene set, which demonstrates the principles of the methods in **Figure 1B**.

The EWCE method has been enabled by single cell transcriptomes from brain tissue (Darmanis et al., 2015; Dueck et al., 2015; Saraiva et al., 2015; Shin et al., 2015; Usoskin et al., 2015; Zeisel et al., 2015). As many of these studies selected the cells through unbiased sampling, they provide data on all cells present in the sampled tissues, including less well studied cells such as vascular endothelial cells (Zeisel et al., 2015). Furthermore, transcriptomes obtained through RNA-Seq provide a broader dynamic range than those from microarrays, allowing more precise quantification of the degree to which a gene is expressed in a cell (Wang et al., 2009). The major methodological advance of EWCE lies in taking advantage of this improved dynamic range and data availability, in that we are able to use *all* of the genes expressed within a cell type to determine enrichment. This yields a significant improvement in capabilities relative to methods focusing on a small set of selected cell type markers.

Here, we have used EWCE with single cell mouse brain transcriptome data 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. These applications inform on the primary and secondary cell types involved with brain disorders. The EWCE method is a robust approach, detecting consistent cellular signatures across transcriptome datasets from 17 Alzheimer's brain regions and two independent Autism datasets. The EWCE method enables diverse sets of omic data to be integrated and can be applied to a wide range of biological problems in metazoan organisms.

METHODS

Summary of Expression Weighted Cell-Type Enrichment (EWCE) Method

The EWCE method takes two arguments: (1) a target gene list of length n denoted as T, and (2) a background set of genes, indexed by B. The objective of the method is to determine the probability that the genes in T have higher expression in a cell type than can be assigned to random chance. To find this we need the probability distribution of average expression in the cell of interest amongst gene lists of length n. Assume that one wanted to test the enrichment of astrocyte genes in T. From single-cell transcriptome data we know the expression level of every gene in astrocytes. First we calculate the average expression level in astrocytes of each gene in T. We then randomly sample 100000 gene lists from the background gene set, each with length n. The probability distribution is then is estimated from the average level of expression in astrocytes in each of these random gene lists. We have released the *EWCE* package through Bioconductor which can be used to run the method on novel datasets. The method is explained in greater detail below.

Processing of Single Cell Transcriptome Data

Raw cell type mRNA expression data was downloaded from the Linnarsson lab webpage (data annotated as being from 17th August 2014) (Zeisel et al., 2015). Throughout the methods section we refer to this as the Single-Cell Transcriptome (SCT) dataset. The data downloaded provides annotations of the cell type (e.g., "astrocyte," "interneuron," "oligodendrocyte") and sub-cell type annotations (e.g., Int1 ["interneuron type 1"], Int2 ["interneuron type 2"], Int15 ["interneuron type 15"], Peric ["Pericyte"], Vsmc ["Vascular Smooth Muscle Cell"]) which are expected to be analogous to either particular cell types (e.g., basket cells) or maturational stages (e.g., differentiating oligodendrocytes). We refer to these as "cell types" and "sub-cell type" annotations. The annotations were not known prior to sequencing and were instead determined using the backspin algorithm as described in the manuscript associated with the dataset. The cells were sampled from two brain regions, SS (somatosensory cortex) and hippocampus CA1: the pyramidal neurons are divided at a cell-type level into these groups. The cell types analyzed in this study are specified and described in Supplementary Table 1.

The dataset contains data from w cells associated with k sub-cell types. Each of the k sub-cell types can be associated with a numerical index from the set $\{1, .., k\}$. The cell type annotations for cell i are stored using a numerical index in m_i, while sub-cell type annotations are stored in l_i. For the EWCE algorithm,

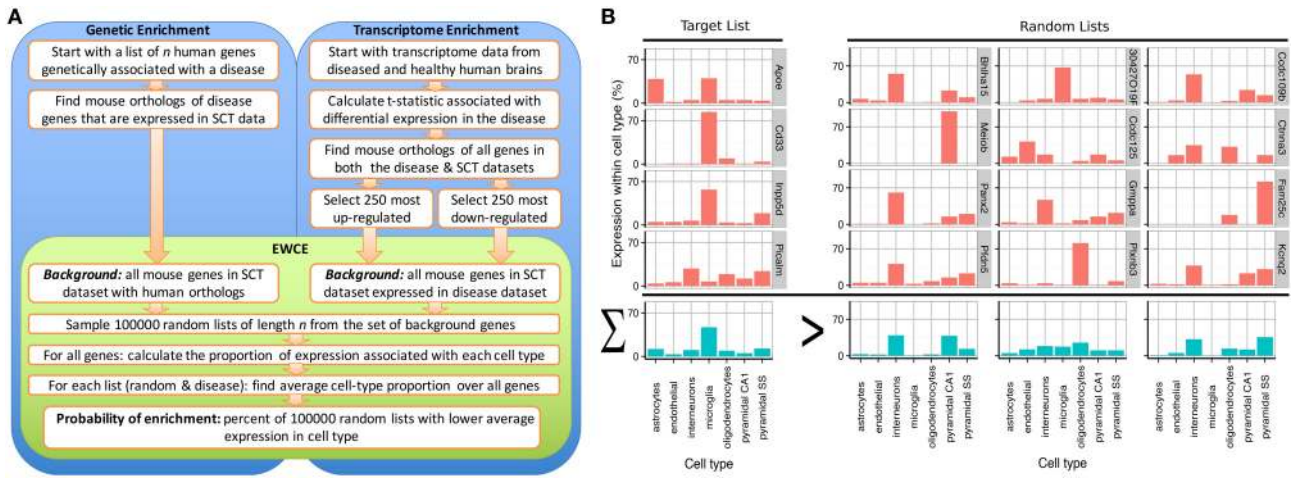

FIGURE 1 | Depiction of the expression weighted cell-type enrichment method. (A) Flow diagram showing the steps involved in going from either genetic or transcriptional data to a probability of enrichment. **(B)** Demonstration of the principles of EWCE method using four Alzheimer's genes (target list) and three randomly generated lists. In the left most column, the cell type expression proportions for the target list is shown. The row with blue bars shows the average over the genes shown above. All four of the target genes are not specific marker genes for microglia; Apoe for instance also has high expression in astrocytes. Nonetheless, when averaged together, there is a higher mean expression in microglia than in the averaged random list.

we need a single number to describe the relationship between a gene and a sub-cell type. When the number of cells from the cell-type indexed by c is given by N_c we calculate the mean level of expression for gene g as

$$e_{g,c} = \frac{\sum_{i=1}^{|L|} F(g,i,c)/N_c}{\sum_{r=1}^{k}\left(\sum_{i=1}^{|L|} F(g,i,r)/N_r\right)}$$

$$F(g,i,c) = \begin{cases} r_{g,i}, & l_i = c \\ 0, & l_i \neq c \end{cases}$$

where $r_{g,i}$ is the expression of gene "g" in cell "i" as described in the data file downloaded from the Linarsson lab website. Because $e_{g,c}$ is independent of the overall expression level of a gene, it is desirable to drop genes with very low expression levels, as a small number of reads in one cell can make the gene appear to be a highly specific cell marker. We thus drop all genes for which $\sum_{i=1}^{|L|} F(g,i,c)/N_c$ is not greater than 0.2 for at least one sub-cell type.

For some of the analyses (when testing enrichment in genetic susceptibility genes) the values for expression in cell types rather than sub-cell types are used (definition stated above). For this, the values for the expression matrix E are first calculated separated for sub-cell types (i.e., S1 Pyramidal cells from Layer 6b) then summed to get the values for grouped cell types (i.e., S1 Pyramidal cells). To get the expression level of gene g for the cell type indexed by e:

$$p_{g,e} = \sum_{i=1}^{k} G(g,i,e) \text{ where } G(g,i,e) = \begin{cases} e_{g,i} & m_i = e \\ 0 & m_i \neq e \end{cases}$$

For the remainder of the methods section all the formula's are provided using $e_{g,c}$ (the measure of expression in subcell types) but $p_{g,e}$ should be used instead if it is desired to apply the method to full cell types instead.

Bootstrapping Enrichment within Gene Lists

The background set used for EWCE analysis depends on the analysis being performed. For gene set (but not transcriptome) enrichment the background gene set is comprised of all genes which have orthologs between human and mice—including those in the target list—but excluding any which were not detected in the SCT dataset. For transcriptome enrichment analysis the background set has an additional restriction relative to simple gene set analysis, in that the background genes must also be expressed in the disease transcriptome dataset. Human genes are converted to mouse orthologs using Biomart.

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. Thus, for a gene list indexed by X we calculate the average expression in the sub-cell type indexed by c as:

$$\gamma(X,c) = \sum_{g \in X} e_{g,c}$$

This calculation is then repeated for 100,000 randomly generated gene lists, having the same length as the target gene list, with the genes randomly selected from the background gene set. Sampling of random gene sets is done without replacement. When the target gene list has length n, D_j denotes the jth set of n indices for bootstrapping genes, with $j \in \{1,\ldots,100000\}$. 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

$$P\left(X \textit{ enriched for cell type } c\right) = \frac{\sum_{j=1}^{100000} \begin{cases} 1 & \gamma(X,c) > \gamma\left(D_j, c\right) \\ 0 & \gamma(X,c) < \gamma\left(D_j, c\right) \end{cases}}{100000}$$

Where probabilities are stated for gene list enrichments, all p-values stated are adjusted for multiple testing. Multiple testing corrections are done separately for the results shown in **Figures 2**, **3A–C**. In the section on schizophrenia transcriptomics, for ease of reading we report p-values for cell-types though the analysis was done on the level of sub-cell types: the p-values reported are for the subcell-type with the most significant relevant enrichment. The fold enrichment $\varphi(X, c)$ is calculated as the expression in the target gene list divided by the mean level of expression in the bootstrap samples

$$\varphi(X, c) = \gamma(X, c) / \left(\sum_{j=1}^{100000} \gamma\left(D_j, c\right)/100000\right)$$

Throughout the paper the number of standard deviations which $\gamma(X, c)$ falls from the mean of $\gamma(D, c)$ is used as a measure of significance. We denote this value for gene list X as $d(\gamma_X, \overline{\gamma_D})$.

Bootstrapping with Controls for Transcript Length and GC Content

Enrichments found in gene sets from genetic studies have been shown to be biased by gene characteristics including transcript length and GC content (Jia et al., 2010). To account for this, for the analyses associated with **Figure 2** we used a semi-random form of gene selection which controlled for these two properties. For each human gene list, we first obtained the ensembl ID's associated with the gene symbols through Biomart and then ran a second query to obtain the transcript lengths and GC content associated with each ensemble ID. Where multiple transcript lengths were associated with a single HGNC gene we took the mean value. The deciles of gene size and GC content were calculated over the set of genes expressed in the SCT dataset (after dropping those with low expression levels as described above). The two sets of decile values were used to define a grid, and each gene assigned to a position within the grid based on it's transcript lengths and GC content. To run a bootstrap analysis on a particular target list, 100,000 random lists were constructed with equal length to the target list. Gene i in each random list was selected from the same grid square as gene i in the target list.

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 2. References associated with the disease lists are listed here: Abnormal myelination (Groza et al., 2015), Human post-synaptic proteome (Bayés et al., 2011), Alzheimer's disease (Bertram et al., 2007; Lambert et al., 2013), Anxiety disorders (Le-Niculescu et al., 2011), Autism (Sanders et al., 2015), Intellectual Disability (Sanders et al., 2015), Multiple Sclerosis (Lill et al., 2011), Schizophrenia (Girard et al., 2011; Ayalew et al., 2012; Xu et al., 2012; Fromer et al., 2014; Schizophrenia Working Group of the Psychiatric Genomics Consortium., 2014). Genes associated with Intellectual Disabilities and Epilepsy were also sourced by finding all genes associated with "Intellectual Disability" and "Seizure" Human Phenotype Ontology terms using the HPO Browser (http://compbio.charite.de/hpoweb). Enrichment probabilities were corrected using the Bonferroni methods.

Application of EWCE to Human Disease Transcriptome Datasets

The transcriptome datasets used in the study were all obtained from publically available sources (Katsel et al., 2005; Narayan et al., 2008; Haroutunian et al., 2009; Maycox et al., 2009; Barnes et al., 2011; Voineagu et al., 2011; Roussos et al., 2012; Gupta et al., 2014; Reinhart et al., 2015) that are detailed in Supplementary Methods. The methods used to calculate differential expression for each study are also detailed in Supplementary Methods. The R package *Limma* (Ritchie et al., 2015) was used for all tests of differential expression. For all Autism and Schizophrenia studies differential expression was determined for a diagnosis of disease. For Alzheimer's disease, differential expression was calculated for Braak score. For each study, genes were ordered based on the t-statistic. The 250 genes with the largest positive t-statistics were taken as the upregulated gene set, and the 250 genes with the largest negative t-statistic were taken to be downregulated. Where multiple probes targeted the same gene these were dropped after selected the 250 genes, and the length of the random lists set to have the length of the number of unique genes.

For the Schizophrenia analysis, data from multiple independent studies was available for a number of brain regions. Superior temporal gyrus came from two studies (Katsel et al., 2005; Barnes et al., 2011). Middle temporal gyrus was based on two studies (Katsel et al., 2005; Roussos et al., 2012). Hippocampus was based on three studies, including one Bipolar cohort (Katsel et al., 2005; Reinhart et al., 2015). Frontal pole was based on two studies (Katsel et al., 2005; Maycox et al., 2009). Dorsolateral prefrontal cortex was based on five studies, including one Bipolar cohort (Katsel et al., 2005; Narayan et al., 2008; Roussos et al., 2012; Reinhart et al., 2015). Cingulate cortex was based on three studies, including samples from anterior and posterior areas (Katsel et al., 2005; Roussos et al., 2012).

To merge these schizophrenia datasets together the EWCE method was extended as follows. Standard cell type bootstrapping was done for each individual study and the cell type expression proportions for each bootstrap sample was stored as a matrix, with a row for each of the 100,000 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.

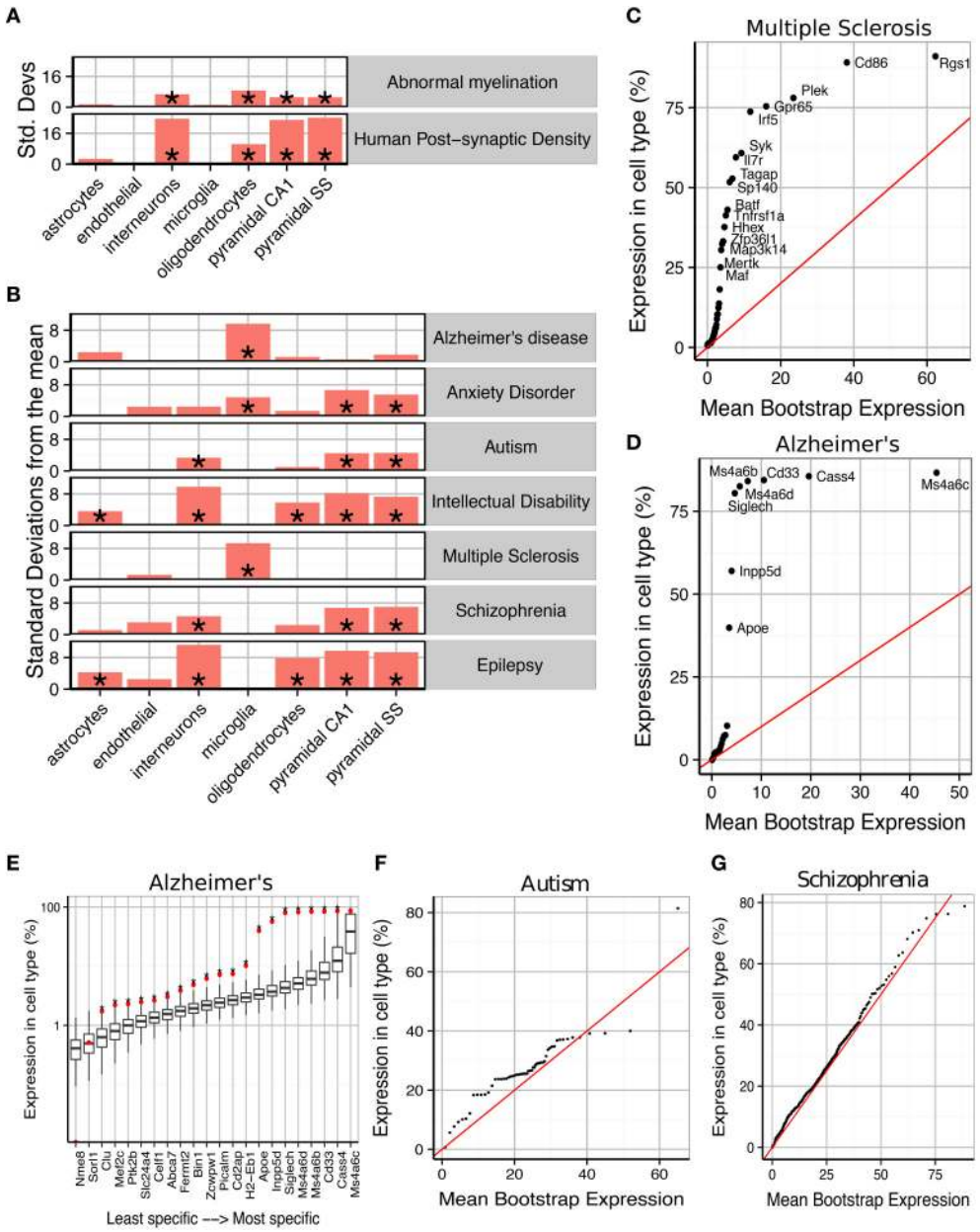

FIGURE 2 | Susceptibility genes for major human brain disorders show distinct cell type enrichments. (A) Two gene sets with a strong prior expectation for cell type enrichment—the human postsynaptic density, and genes with Human Phenotype Ontology annotations for abnormal myelination—were detected using 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 *i*th 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 disease susceptibility genes are more enriched for microglial expression than expected by chance. **(E)** Bootstrap distributions of expected microglial expression levels of Alzheimer's disease genes. Red dots mark the expression level of the susceptibility genes, while the associated boxplots denote the expected expression level of the *i*th 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)** Genes associated with Autism are found to show increased 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.

RESULTS

We first sought to confirm that the method detects expected cell type enrichments (**Figure 2A**). We first gene list we used were the 1461 genes associated with the human cortical post-synaptic density (hPSD) (Bayés et al., 2011). As expected the hPSD was significantly associated with cortical pyramidal neurons (Bonferroni corrected $p < 0.00001$, $d(\gamma_X, \overline{\gamma_D}) = 24.1$), interneurons ($p < 0.00001$, $d(\gamma_X, \overline{\gamma_D}) = 23.6$) and hippocampal pyramidal neurons ($p < 0.00001$, $d(\gamma_X, \overline{\gamma_D}) = 23$). Enrichment was also found for oligodendrocytes ($p < 0.00001$, $d(\gamma_X, \overline{\gamma_D}) = 10.4$), which is likely explained by myelin contamination of the protein samples (amongst the most oligodendrocyte enriched genes in the hPSD list are many of the major protein constituents of myelin including Mbp, Mag, Mog, Pllp, and Cnp). The enrichment can also be partially attributed to precursor cells of oligodendrocyte have also been shown to develop post-synaptic densities which are expected to be comprised of many of the same proteins as in neurons (De Biase and Bergles, 2011).

We then examined 185 genes which have Human Phenotype Ontology annotations for abnormal myelination with the prior hypothesis that they would be enriched for oligodendrocyte genes. The majority of the genes are associated with rare neurological disorders in which patients/families have been shown to exhibit either demyelination or absence of myelinated fibers. EWCE analysis confirmed that oligodendrocytes are the cell type most enriched amongst these genes ($p < 0.00001$, $d(\gamma_X, \overline{\gamma_D}) = 8.7$). All classes of neurons were also found to be enriched with $p < 0.00001$ suggesting that in many cases the myelination deficit is secondary to a neuronal change.

Cell Enrichments in Human Disease Susceptibility Genes

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 epilepsy (gene lists used shown in Supplementary Table 2). Two of the disorders were found to show strong evidence for being primary microglial disorders (Alzheimer's, $p < 0.00001$, $d(\gamma_X, \overline{\gamma_D}) = 9.6$ and Multiple Sclerosis, $p < 0.00001$, $d(\gamma_X, \overline{\gamma_D}) = 9.3$). We denote the number of standard deviations from the bootstrapped mean as $d(\gamma_X, \overline{\gamma_D})$. It has previously been noted that a number of Alzheimer's disease susceptibility genes have high levels of expression in microglia (Mhatre et al., 2015), 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 (**Figures 2C–E**): for Alzheimer's, every gene except Cass4 was found to have higher expression in microglia than expected by chance (**Figure 2E**). No other cell type was found to be significantly enriched for either of these disorders.

Schizophrenia and Autism were found to be the only exclusively neuronal disorders, and both had their strongest enrichment in pyramidal rather than interneurons. Schizophrenia associated genes were enriched for all three classes of neurons with $p < 0.00001$ but for cortical pyramidal neurons $d(\gamma_X, \overline{\gamma_D}) = 7$ while for interneurons $d(\gamma_X, \overline{\gamma_D}) = 4.7$. For autism the enrichment for interneurons was close to the significance threshold ($p = 0.024$ and $d(\gamma_X, \overline{\gamma_D}) = 3.3$) indicating that it may be secondary to the enrichment of pyramidal neuron genes ($p = 0.00126$ and $d(\gamma_X, \overline{\gamma_D}) = 4.6$). The enrichments for both of these disorders were again found to affect all genes throughout the lists, rather than a subset of strongly cell-type specific genes (**Figures 2F,G**). Anxiety disorders were also found to be enriched for pyramidal neuron genes ($p < 0.00001$, $d(\gamma_X, \overline{\gamma_D}) = 6.7$) but also showed evidence for a microglial involvement ($p < 0.00001$, $d(\gamma_X, \overline{\gamma_D}) = 6.7$).

Interneurons were found to be the most enriched cell type for intellectual disabilities (ID) ($p < 0.00001$, $d(\gamma_X, \overline{\gamma_D}) = 9.8$) and epilepsy ($p < 0.00001$, $d(\gamma_X, \overline{\gamma_D}) = 11.2$). Both disorders also showed enrichment for pyramidal neurons but to a lesser degree (for cortical pyramidal neurons $d(\gamma_X, \overline{\gamma_D}) = 7.3$ for ID and $d(\gamma_X, \overline{\gamma_D}) = 9.4$ for epilepsy). Astrocytes ($p = 0.003$, epilepsy; $p = 0.034$, ID) and Oligodendrocytes (both $p < 0.00001$) were also found to contribute to both of these conditions. We note one gene in particular, Aass, which is markedly specific to astrocytes and is associated with both seizures and intellectual disabilities.

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.

The first step towards obtaining the gene set from transcriptomics was in each case a standard differential expression analysis. Genes were rank ordered by t-statistic and the 250 most upregulated as well as 250 most downregulated genes were taken for further analysis. We then perform an Expression Weighted Cell type Enrichment (EWCE) analysis, wherein the random samples are obtained by reordering the ranked list 100,000 times (see **Figure 1A** for workflow).

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 that examined changes in 14 cortical and three non-cortical regions, with between 51 and 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 **Figure 3A**). We calculated the 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

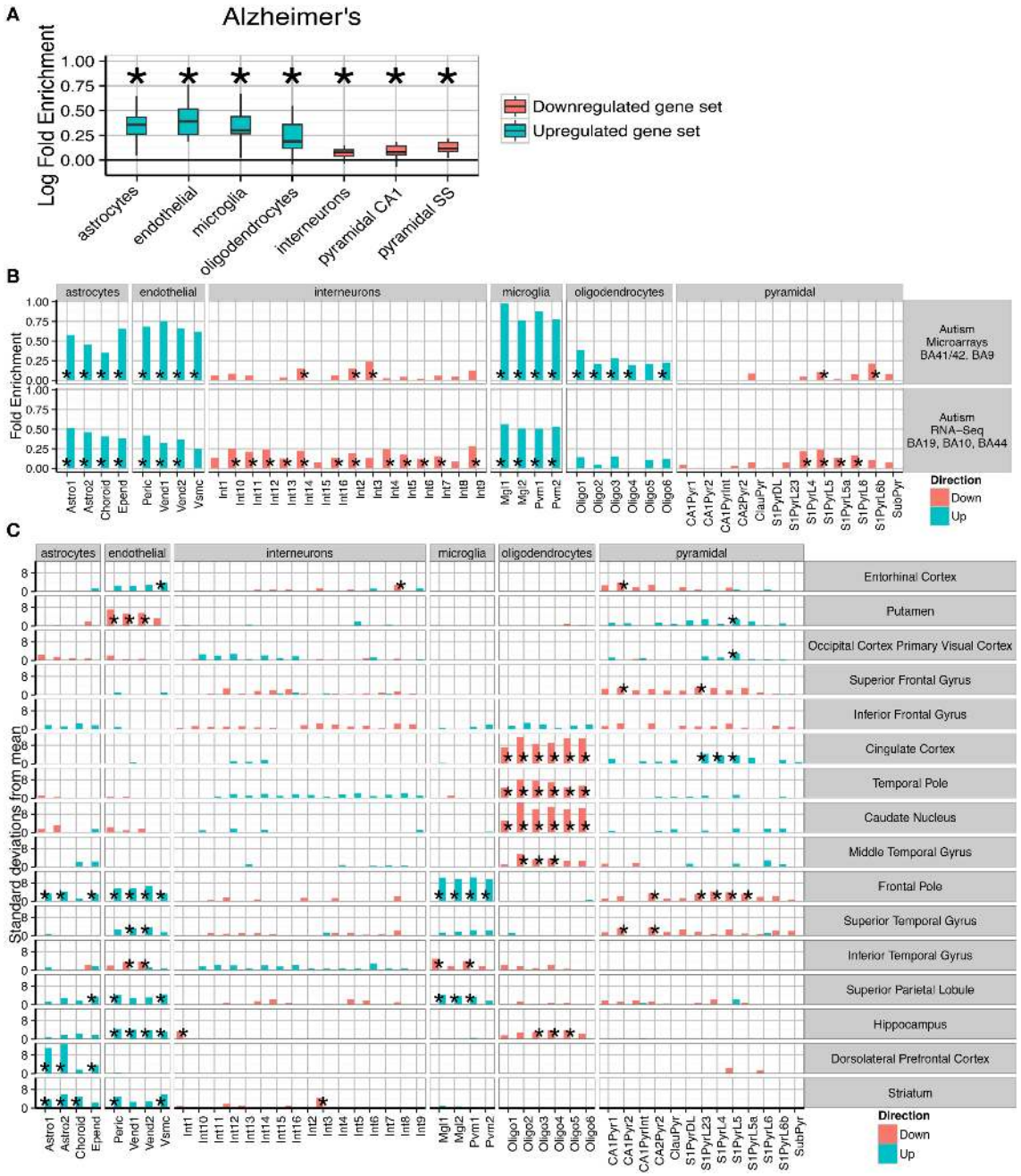

FIGURE 3 | Post-mortem transcriptomes from patients with Alzheimer's disease, Autism and Schizophrenia show distinctive cellular phenotypes. (A) Consistent fold enrichments were found for each cell type across **fourteen** cortical and **three** subcortical brain regions of Alzheimer's patients. The box plots mark the distribution of cellular fold enrichments across all the brain regions examined. Asterisks 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 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.

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 (Grammas, 2011; Wyss-Coray and Rogers, 2012), strongly supporting the notion that this is a powerful and robust technique for determining hidden variables in transcriptome data.

Having validated the method's ability to detect known pathological phenotypes, we sought to apply the method to two disorders that are less well characterized: Autism and Schizophrenia. Two publically available transcriptome datasets were used for Autism (Voineagu et al., 2011; Gupta et al., 2014), with one study using samples from areas BA41/42, BA9, BA19, BA10, and BA44. The two studies were analyzed 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 (**Figure 3B**). We emphasize that these were two independent studies, performed by different laboratories, on different cortical regions, with one study using RNA-Sequencing and the other Illumina microarrays.

Both Autism studies, like the Alzheimer's disease 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 (van Kooten et al., 2008; Courchesne et al., 2011; Casanova et al., 2013). 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 glial marker genes, increased cell densities and altered morphologies for both cell types (Laurence and Fatemi, 2005; Vargas et al., 2005; Morgan et al., 2012; Tetreault et al., 2012). The finding that endothelial genes are up-regulated could be related to the decrease in cerebral blood flow seen in temporal and frontal cortices of autistic patients (Ohnishi et al., 2000).

We next extended the study to Schizophrenia. We utilized 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 and upregulation of pyramidal neuron genes; (2) increased astrocyte and/or endothelial expression. All samples from the prefrontal cortex fell into the second cluster. Four regions fall outside of these clusters and show few/no significant changes: the primary visual cortex, putamen, superior and inferior frontal gyrus. While we showed in **Figure 2B** that schizophrenia is genetically associated only with neurons, the primary effects appear to be in astrocytes, endothelial cells and oligodendrocytes. A number of regions do show changes in neuronal transcripts, including downregulation of a Somatostatin and Neuropeptide-Y expressing interneuron in the hippocampus ($p < 0.0266$, $d(\gamma_X, \overline{\gamma_D}) = 3.5$), but these changes are less consistent across regions.

One of the most significant changes was in the dorsolateral prefrontal cortex (BA46), with an enrichment of astrocyte genes 12.7 standard deviations from the bootstrapped mean ($p < 0.00001$, $d(\gamma_X, \overline{\gamma_D}) = 12.7$). Significant up-regulation of astrocyte genes (after Benjamini Hochberg correction) was also seen in Striatum ($p = 0.00001$, $d(\gamma_X, \overline{\gamma_D}) = 5.8$), Frontal Pole ($p = 0.004$, $d(\gamma_X, \overline{\gamma_D}) = 4.2$) and the Superior Parietal Lobe ($p = 0.037$, $d(\gamma_X, \overline{\gamma_D}) = 3.5$). Six regions showed significant up-regulation of endothelial cells in Schizophrenia, including three of those regions which also had up-regulated astrocytes. The most significant effects were seen in the Frontal Pole ($p < 0.00001$, $d(\gamma_X, \overline{\gamma_D}) = 6.7$) and Striatum ($p<0.00001$, $d(\gamma_X, \overline{\gamma_D}) = 6.0$). Also affected was the Superior Parietal Lobule ($p = 0.00992$, $d(\gamma_X, \overline{\gamma_D}) = 4.5$), the Hippocampus ($p = 0.0031$, $d(\gamma_X, \overline{\gamma_D}) = 4.5$), the Entorhinal cortex ($p = 0.018$, $d(\gamma_X, \overline{\gamma_D}) = 4.0$) and Superior Temporal Gyrus ($p = 0.013$, $d(\gamma_X, \overline{\gamma_D}) = 3.7$). The Frontal Pole, also known as the Anterior Prefrontal Cortex, fits within this cluster but also showed a very strong upregulation of microglial genes ($p < 0.00001$, $d(\gamma_X, \overline{\gamma_D}) = 10.4$) and downregulation of pyramidal neurons ($p < 0.0009$, $d(\gamma_X, \overline{\gamma_D}) = 4.4$).

The second set of brain regions affected in schizophrenia 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 Caudate Nucleus ($p < 0.00001$, $0d(\gamma_X, \overline{\gamma_D}) = 13.1$), Cingulate Cortex ($p < 0.00001$, $d(\gamma_X, \overline{\gamma_D}) = 11.6$), Hippocampus ($p < 0.0057$, $d(\gamma_X, \overline{\gamma_D}) = 4$), Middle Temporal Gyrus ($p < 0.0001$, $d(\gamma_X, \overline{\gamma_D}) = 5.7$), and the Temporal Pole ($p < 0.0001$, $d(\gamma_X, \overline{\gamma_D}) = 8.2$). For four of the five regions the most significant change is in the pre-myelinating (type 2) oligodendrocytes, which are understood to be mid-way through the stages of oligodendrocyte maturation.

DISCUSSION

Using the EWCE method, we have shown that single cell transcriptome data can be integrated with genetic susceptibility data or tissue transcriptome data to identify cell types involved with disease. Using lists of genetic susceptibility data with single cell transcriptome data that define cell types, we could identify specific cell types that are the likely primary targets of the genetic susceptibility. In a separate analysis, using lists of genes from post-mortem transcriptomes, we found that a broader range of cell types were affected, indicating that the cellular pathology of the disease extends from the cells affected by the primary genetic susceptibility to a wider set of "secondary" or "reactive" cell types.

For seven different brain disorders, the EWCE method was used to identify the putative primary cell types affected by genetic susceptibility. Consistent with current models, pyramidal neurons were the cell type most associated with schizophrenia and autism genes. A primary role of microglia in multiple sclerosis is also consistent with primary pathology in the immune

system (Hemmer et al., 2015). The association of Alzheimer's with microglia also builds on a body of existing work on the role of that cell type in the disease (Mhatre et al., 2015), but goes further in suggesting it is the primary focus of genetic risk. The identification of anxiety disorders as having a microglial/immune component was surprising and may be relevant to the reported prevalence of anxiety disorders amongst those suffering from immunological disorders (Roy-Byrne et al., 2008). For this later association we caution that unlike the gene sets used for all other disorders, the gene set for Anxiety disorder was generated through a convergent functional genomics approach, which was not solely informed by human genetic studies.

Intellectual disabilities and epilepsy were both found to be most strongly associated with interneurons. Past studies strongly support the finding that interneurons are the causative cell type underlying seizures (Ogiwara et al., 2007; Hammad et al., 2014). Other cell types, including astrocytes and oligodendrocytes were also found to contribute to the etiology of these disorders; this is supported by evidence showing that a human disorder featuring intellectual disabilities and seizures can be attributed to cell-specific mutations in astrocytes (Uhlmann et al., 2002). The shared cellular enrichment profiles between ID and epilepsy explains why there exists far high rates of comorbidity between individuals with the two disorders: around 26% of individuals with ID suffer from seizures compared to only 0.4–1% in the general population (McGrother et al., 2006). Noting the stronger enrichment for interneurons over pyramidal neurons in epilepsy and ID, the findings also suggest an explanation for why individuals with autism are significantly more likely to have seizures if they also have ID (Amiet et al., 2008): while a mutation affecting synapses in pyramidal neurons may cause Autism, our data suggests that if the mutation affects a gene that is also highly expressed in interneurons this would increase the chance of seizures and ID.

Once denser sequencing of interneurons has been performed from a wider set of brain regions it may be possible to identify specific interneuron populations associated with distinct types of seizures, as well as particular cognitive deficits. The ability to precisely distinguish affected subtypes of interneurons may however require a modification to the method: at present four of the five disorders which are stated as affecting one population of neurons, have significant enrichments for all three neuron types. As neurons share many genes, and vary only in limited subsets and graduated expression levels, it may be that one can only distinguish between them by considering the single most enriched category. The limitations of the method can be seen by considering the case of autism, which shows it's strongest enrichment in pyramidal cells but is also enriched in interneurons to a lesser degree. These results could support either of two hypotheses: that autism is a primary pyramidal neuron disorder, or that it is a disorder of broad neuronal dysfunction. An extension to the EWCE method that penalizes neuronal subtypes for absence of expression when a gene is expressed in similar cells could potentially resolve this issue.

Contrasting the results of EWCE from genetic susceptibility data with transcriptomes from diseased tissue suggested that the cellular pathology spreads from the primary affected cells to secondary cells. These putative secondary changes appear to extend between classes of cells. For example, the genetic susceptibility of Autism and Schizophrenia appears to primarily impact neurons, yet both show evidence for secondary endothelial disruption. Interestingly, both disorders have been shown to have decreased cerebral blood flow (Sabri et al., 1997; Ohnishi et al., 2000; Kindler et al., 2013) although it is unclear whether this is related to up-regulation of endothelial genes. Though mechanisms are well established for how synaptic activity alters blood flow across brief time scales (Attwell et al., 2010) we are unaware of any studies investigating how persistent mutational changes in the level of synaptic/neuronal activity alters the vascular system. Based on the results we have found here, we suggest that understanding which pyramidal neuronal properties need to be altered (through mutation) to induce secondary transcriptional enrichments of endothelial and/or glial genes could cast new light on the polygenic nature of Schizophrenia and Autism.

Implementation of EWCE in mouse models of human disease could underlie a new approach to studying brain disorders. Once the transcriptomic cell type enrichments are determined for a disease, conditional mouse models (carrying cell type specific mutations) could then be validated or rejected based on whether they recapitulate some or all of the disease associated secondary effects. One problem with this approach is that a range of distinct changes could result in the same transcriptomic alteration (for instance, down regulation of interneuron genes could be caused by either decreased cell density, altered cellular state, or decreased synaptic connectivity). The directionality and regional specificity of transcriptional phenotypes should however be able to act as a guide to narrow down the nature of the cellular changes. Even without more extensive follow-up EWCE could heighten confidence in the biological validity of disease models for which only behavioral similarities to human disease could otherwise be shown.

Three limitations with the current study are that the cells were obtained from mice, they were immature and only from CA1 and somatosensory cortex. Using human rather than mouse data may make a significant difference for disease enrichments: one study which compared expression profiles of cell types between humans and mice found that as few as 52% of genes identified as being astrocyte-enriched in mice, were also found to be astrocyte-enriched in humans (Zhang et al., 2015). Developmental age is likely to be important since the Braincloud dataset shows that significant transcriptional changes occur in human brain tissue between the pre- to post-natal period (Colantuoni et al., 2011). The few anatomical regions so far studied also limits the range of diseases which can currently be analyzed. For instance, while a suitable number of genome wide significant genes are known for Parkinson's disorder (Nalls et al., 2014), the known association of that disorder with the Striatum lead us to avoid testing for enrichment in those genes.

In the coming years the quantity of single cell data that is available will increase and the utility of the EWCE method is expected to expand commensurate with this. The single cell transcriptome dataset used for this study comprised 3005 cells from the cortex and CA1 of mice aged p21–31. With

greater depth of cellular sequencing, it may become possible to detect changes in more specific populations of cells. This is potentially of greatest importance for interneurons for which the data presently available is sparse—of the 1314 cells sequenced from CA1, only 126 were from interneurons. With over thirty types of interneuron estimated to exist just within CA1 based on morphology, electrophysiology and expression of classical molecular markers (Wheeler et al., 2015), the sparse sampling means there is likely to be substantial noise within the current cell type estimates. Presently it is unclear whether diseases are likely to specifically affect such particular neuron types, and an alternative approach could involve testing along branches of the cell lineage tree—i.e., interneurons derived from medial or caudal ganglionic eminences.

We also note that there is no reason why EWCE should be restricted to the study of brain disorders and as sufficient cellular data becomes available the same methodology could be applied to any other disease. Indeed the EWCE method can be applied to other omic gene lists for the purposes of interrogating the relevant cell types. For example, gene lists from mouse phenotyping studies, such as the International Mouse Phenotyping Consortium (Brown and Moore, 2012) could be used to identify the cell types underlying specific phenotypes.

AUTHOR CONTRIBUTIONS

NS (Implementation, analysis, and author); SG (Interpretation, drafting, and revising)

ACKNOWLEDGMENTS

This work was supported by the Wellcome Trust and the European Union Seventh Framework Programme under grant agreements n° HEALTH-F2-2009-241995 ("Gencodys" project) and HEALTH-F2-2009-242167 ("Synsys" project) and W. Richardson (UCL).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fnins.2016.00016

REFERENCES

Amiet, C., Gourfinkel-An, I., Bouzamondo, A., Tordjman, S., Baulac, M., Lechat, P., et al. (2008). Epilepsy in autism is associated with intellectual disability and gender: evidence from a meta-analysis. *Biol. Psychiatry* 64, 577–582. doi: 10.1016/j.biopsych.2008.04.030

Attwell, D., Buchan, A. M., Charpak, S., Lauritzen, M., MacVicar, B. A., and Newman, E. A. (2010). Glial and neuronal control of brain blood flow. *Nature* 468, 232–243. doi: 10.1038/nature09613

Ayalew, M., Le-Niculescu, H., Levey, D., Jain, N., Changala, B., Patel, S., et al. (2012). Convergent functional genomics of schizophrenia: from comprehensive understanding to genetic risk prediction. *Mol. Psychiatry* 17, 887–905. doi: 10.1038/mp.2012.37

Barnes, M. R., Huxley-Jones, J., Maycox, P. R., Lennon, M., Thornber, A., Kelly, F., et al. (2011). Transcription and pathway analysis of the superior temporal cortex and anterior prefrontal cortex in schizophrenia. *J. Neurosci. Res.* 89, 1218–1227. doi: 10.1002/jnr.22647

Bayés, À., van de Lagemaat, L. N., Collins, M. O., Croning, M. D., Whittle, I. R., Choudhary, J. S., et al. (2011). Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nat. Neurosci.* 14, 19–21. doi: 10.1038/nn.2719

Bertram, L., McQueen, M. B., Mullin, K., Blacker, D., and Tanzi, R. E. (2007). Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat. Genet.* 39, 17–23. doi: 10.1038/ng1934

Brown, S. D., and Moore, M. W. (2012). Towards an encyclopaedia of mammalian gene function: the International Mouse Phenotyping Consortium. *Dis. Model. Mech.* 5, 289–292. doi: 10.1242/dmm.009878

Capurro, A., Bodea, L.-G., Schaefer, P., Luthi-Carter, R., and Perreau, V. M. (2014). Computational deconvolution of genome wide expression data from Parkinson's and Huntington's disease brain tissues using population-specific expression analysis. *Front. Neurosci.* 8:441. doi: 10.3389/fnins.2014.00441

Casanova, M. F., El-Baz, A. S., and Suri, J. S. (2013). *Imaging the Brain in Autism*. New York, NY: Springer. doi: 10.1007/978-1-4614-6843-1

Colantuoni, C., Lipska, B. K., Ye, T., Hyde, T. M., Tao, R., Leek, J. T., et al. (2011). Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* 478, 519–523. doi: 10.1038/nature10524

Courchesne, E., Mouton, P. R., Calhoun, M. E., Semendeferi, K., Ahrens-Barbeau, C., Hallet, M. J., et al. (2011). Neuron number and size in prefrontal cortex of children with autism. *JAMA* 306, 2001–2010. doi: 10.1001/jama.2011.1638

Darmanis, S., Sloan, S. A., Zhang, Y., Enge, M., Caneda, C., Shuer, L. M., et al. (2015). A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci. U.S.A.* 112, 7285–7290. doi: 10.1073/pnas.1507125112

De Biase, L. M., and Bergles, D. E. (2011). Same players, different game: AMPA receptor regulation in oligodendrocyte progenitors. *Nat. Neurosci.* 14, 1358. doi: 10.1038/nn.2965

Dueck, H., Khaladkar, M., Kim, T. K., Spaethling, J. M., Francis, C., Suresh, S., et al. (2015). Deep sequencing reveals cell-type-specific patterns of single-cell transcriptome variation. *Genome Biol.* 16, 122. doi: 10.1186/s13059-015-0683-4

Fromer, M., Pocklington, A. J., Kavanagh, D. H., Williams, H. J., Dwyer, S., Gormley, P., et al. (2014). De novo mutations in schizophrenia implicate synaptic networks. *Nature* 506, 179–184. doi: 10.1038/nature12929

Girard, S. L., Gauthier, J., Noreau, A., Xiong, L., Zhou, S., Jouan, L., et al. (2011). Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat. Genet.* 43, 860–863. doi: 10.1038/ng.886

Grammas, P. (2011). Neurovascular dysfunction, inflammation and endothelial activation: implications for the pathogenesis of Alzheimer's disease. *J. Neuroinflammation* 8, 26. doi: 10.1186/1742-2094-8-26

Groza, T., Köhler, S., Moldenhauer, D., Vasilevsky, N., Baynam, G., Zemojtel, T., et al. (2015). The human phenotype ontology: Semantic unification of common and rare disease. *Am. J. Hum. Genet.* 97, 111–124. doi: 10.1016/j.ajhg.2015.05.020

Gupta, S., Ellis, S. E., Ashar, F. N., Moes, A., Bader, J. S., Zhan, J., et al. (2014). Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nat. Commun.* 5 doi: 10.1038/ncomms6748

Hammad, M., Schmidt, S. L., Zhang, X., Bray, R., Frohlich, F., and Ghashghaei, H. T. (2014). Transplantation of GABAergic interneurons into the neonatal primary visual cortex reduces absence seizures in stargazer mice. *Cereb. Cortex* 25, 2970–2979. doi: 10.1093/cercor/bhu094

Haroutunian, V., Katsel, P., and Schmeidler, J. (2009). Transcriptional vulnerability of brain regions in Alzheimer's disease and dementia. *Neurobiol. Aging* 30, 561–573. doi: 10.1016/j.neurobiolaging.2007.07.021

Hemmer, B., Kerschensteiner, M., and Korn, T. (2015). Role of the innate and adaptive immune responses in the course of multiple sclerosis. *Lancet Neurol.* 14, 406–419. doi: 10.1016/S1474-4422(14)70305-9

Jia, P., Tian, J., and Zhao, Z. (2010). Assessing gene length biases in gene set analysis of Genome-Wide Association Studies. *Int. J. Comput. Biol. Drug Des.* 3, 297–310. doi: 10.1504/IJCBDD.2010.038394

Katsel, P., Davis, K. L., and Haroutunian, V. (2005). Variations in myelin and oligodendrocyte-related gene expression across multiple brain regions

in schizophrenia: a gene ontology study. *Schizophr. Res.* 79, 157–173. doi: 10.1016/j.schres.2005.06.007

Kindler, J., Jann, K., Homan, P., Hauf, M., Walther, S., Strik, W., et al. (2013). Static and dynamic characteristics of cerebral blood flow during the resting state in schizophrenia. *Schizophr. Bull.* 41, 163–170. doi: 10.1093/schbul/sbt180

Kuhn, A., Thu, D., Waldvogel, H. J., Faull, R. L., and Luthi-Carter, R. (2011). Population-specific expression analysis (PSEA) reveals molecular changes in diseased brain. *Nat. Methods* 8, 945–947. doi: 10.1038/nmeth.1710

Lambert, J.-C., Ibrahim-Verbaas, C. A., Harold, D., Naj, A. C., Sims, R., Bellenguez, C., et al. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* 45, 1452–1458. doi: 10.1038/ng.2802

Laurence, J., and Fatemi, S. (2005). Glial fibrillary acidic protein is elevated in superior frontal, parietal and cerebellar cortices of autistic subjects. *Cerebellum* 4, 206–210. doi: 10.1080/14734220500208846

Le-Niculescu, H., Balaraman, Y., Patel, S., Ayalew, M., Gupta, J., Kuczenski, R., et al. (2011). Convergent functional genomics of anxiety disorders: translational identification of genes, biomarkers, pathways and mechanisms. *Transl. Psychiatry* 1:e9. doi: 10.1038/tp.2011.9

Lill, C. M., Roehr, J. T., McQueen, M. B., Bagade, S., Schjeide, B. M., Zipp, F., et al. (2011). *The MSGene Database.* Alzheimer Research Forum. Available online at: http://www.msgene.org/ (Accessed May 20, 2014).

Maycox, P. R., Kelly, F., Taylor, A., Bates, S., Reid, J., Logendra, R., et al. (2009). Analysis of gene expression in two large schizophrenia cohorts identifies multiple changes associated with nerve terminal function. *Mol. Psychiatry* 14, 1083–1094. doi: 10.1038/mp.2009.18

McGrother, C. W., Bhaumik, S., Thorp, C. F., Hauck, A., Branford, D., and Watson, J. M. (2006). Epilepsy in adults with intellectual disabilities: prevalence, associations and service implications. *Seizure* 15, 376–386. doi: 10.1016/j.seizure.2006.04.002

Mhatre, S. D., Tsai, C. A., Rubin, A. J., James, M. L., and Andreasson, K. I. (2015). Microglial malfunction: the third rail in the development of Alzheimer's disease. *Trends Neurosci.* 38, 621–636. doi: 10.1016/j.tins.2015.08.006

Morgan, J. T., Chana, G., Abramson, I., Semendeferi, K., Courchesne, E., and Everall, I. P. (2012). Abnormal microglial–neuronal spatial organization in the dorsolateral prefrontal cortex in autism. *Brain Res.* 1456, 72–81. doi: 10.1016/j.brainres.2012.03.036

Nalls, M. A., Pankratz, N., Lill, C. M., Do, C. B., Hernandez, D. G., Saad, M., et al. (2014). Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet*. 46, 989–993. doi: 10.1038/ng.3043

Narayan, S., Tang, B., Head, S. R., Gilmartin, T. J., Sutcliffe, J. G., Dean, B., et al. (2008). Molecular profiles of schizophrenia in the CNS at different stages of illness. *Brain Res.* 1239, 235–248. doi: 10.1016/j.brainres.2008.08.023

Ogiwara, I., Miyamoto, H., Morita, N., Atapour, N., Mazaki, E., Inoue, I., Takeuchi, T., et al. (2007). Nav1. 1 localizes to axons of parvalbumin-positive inhibitory interneurons: a circuit basis for epileptic seizures in mice carrying an Scn1a gene mutation. *J. Neurosci.* 27, 5903–5914. doi: 10.1523/JNEUROSCI.5270-06.2007

Ohnishi, T., Matsuda, H., Hashimoto, T., Kunihiro, T., Nishikawa, M., Uema, T., et al. (2000). Abnormal regional cerebral blood flow in childhood autism. *Brain* 123, 1838–1844. doi: 10.1093/brain/123.9.1838

Reinhart, V., Bove, S. E., Volfson, D., Lewis, D. A., Kleiman, R. J., and Lanz, T. A. (2015). Evaluation of TrkB and BDNF transcripts in prefrontal cortex, hippocampus, and striatum from subjects with schizophrenia, bipolar disorder, and major depressive disorder. *Neurobiol. Dis.* 77, 220–227. doi: 10.1016/j.nbd.2015.03.011

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007

Roussos, P., Katsel, P., Davis, K. L., Siever, L. J., and Haroutunian, V. (2012). A system-level transcriptomic analysis of schizophrenia using postmortem brain tissue samples. *Arch. Gen. Psychiatry* 69, 1205–1213. doi: 10.1001/archgenpsychiatry.2012.704

Roy-Byrne, P. P., Davidson, K. W., Kessler, R. C., Asmundson, G. J., Goodwin, R. D., Kubzansky, L., et al. (2008). Anxiety disorders and comorbid medical illness. *Gen. Hosp. Psychiatry* 30, 208–225. doi: 10.1016/j.genhosppsych.2007.12.006

Sabri, O., Erkwoh, R., Schreckenberger, M., Owega, A., Sass, H., and Buell, U. (1997). Correlation of positive symptoms exclusively to hyperperfusion or hypoperfusion of cerebral cortex in never-treated schizophrenics. *Lancet* 349, 1735–1739. doi: 10.1016/S0140-6736(96)08380-8

Sanders, S. J., He, X., Willsey, A. J., Ercan-Sencicek, A. G., Samocha, K. E., Cicek, A. E., et al. (2015). Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron* 87, 1215–1233. doi: 10.1016/j.neuron.2015.09.016

Saraiva, L. R., Ibarra-Soria, X., Khan, M., Omura, M., Scialdone, A., Mombaerts, P., et al. (2015). Hierarchical deconstruction of mouse olfactory sensory neurons: from whole mucosa to single-cell RNA-seq. *Sci. Rep.* 5, 18178. doi: 10.1038/srep18178

Shin, J., Berg, D. A., Zhu, Y., Shin, J. Y., Song, J., Bonaguidi, M. A., et al. (2015). Single-cell RNA-seq with waterfall reveals molecular cascades underlying adult neurogenesis. *Cell Stem Cell* 17, 360–372. doi: 10.1016/j.stem.2015.07.013

Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511, 421–427. doi: 10.1038/nature13595

Tetreault, N. A., Hakeem, A. Y., Jiang, S., Williams, B. A., Allman, E., Wold, B. J., et al. (2012). Microglia in the cerebral cortex in autism. *J. Autism Dev. Disord.* 42, 2569–2584. doi: 10.1007/s10803-012-1513-0

Uhlmann, E. J., Wong, M., Baldwin, R. L., Bajenaru, M. L., Onda, H., Kwiatkowski, D. J., et al. (2002). Astrocyte-specific TSC1 conditional knockout mice exhibit abnormal neuronal organization and seizures. *Ann. Neurol.* 52, 285–296. doi: 10.1002/ana.10283

Usoskin, D., Furlan, A., Islam, S., Abdo, H., Lönnerberg, P., Lou, D., et al. (2015). Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nat. Neurosci.* 18, 145–153. doi: 10.1038/nn.3881

van Kooten, I. A., Palmen, S. J., von Cappeln, P., Steinbusch, H. W., Korr, H., Heinsen, H., et al. (2008). Neurons in the fusiform gyrus are fewer and smaller in autism. *Brain* 131, 987–999. doi: 10.1093/brain/awn033

Vargas, D. L., Nascimbene, C., Krishnan, C., Zimmerman, A. W., and Pardo, C. A. (2005). Neuroglial activation and neuroinflammation in the brain of patients with autism. *Ann. Neurol.* 57, 67–81. doi: 10.1002/ana.20315

Voineagu, I., Wang, X., Johnston, P., Lowe, J. K., Tian, Y., Horvath, S., et al. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* 474, 380–384. doi: 10.1038/nature10110

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi: 10.1038/nrg2484

Wheeler, D. W., White, C. M., Rees, C. L., Komendantov, A. O., Hamilton, D. J., and Ascoli, G. A. (2015). Hippocampome. org: a knowledge base of neuron types in the rodent hippocampus. *Elife* 4:e09960. doi: 10.7554/eLife.09960

Wyss-Coray, T., and Rogers, J. (2012). Inflammation in Alzheimer disease—a brief review of the basic science and clinical literature. *Cold Spring Harb. Perspect. Med.* 2:a006346. doi: 10.1101/cshperspect.a006346

Xu, B., Ionita-Laza, I., Roos, J. L., Boone, B., Woodrick, S., Sun, Y., et al. (2012). *De novo* gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat. Genet.* 44, 1365–1369. doi: 10.1038/ng.2446

Zeisel, A., Muñoz-Manchado, A. B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., et al. (2015). Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* 347, 1138–1142. doi: 10.1126/science.aaa1934

Zhang, Y., Sloan, S. A., Clarke, L. E., Caneda, C., Plaza, C. A., Blumenthal, P. D., et al. (2015). Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. *Neuron* 89, 37–53. doi: 10.1016/j.neuron.2015.11.013

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer Dr. Logue and handling Editor, Dr. Miller declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2016 Skene and Grant. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.