Single nuclei RNA sequencing of 5 regions of the human prenatal brain implicates developing neuron populations in genetic risk for schizophrenia

Darren Cameron, Da Mi, Ngoc-Nga Vinh, Caleb Webber, Meng Li, Oscar Marín, Michael C. O’Donovan, Nicholas J. Bray

PII: S0006-3223(22)01404-4
DOI: https://doi.org/10.1016/j.biopsych.2022.06.033
Reference: BPS 14917

To appear in: Biological Psychiatry

Received Date: 22 March 2022
Revised Date: 17 June 2022
Accepted Date: 29 June 2022

Please cite this article as: Cameron D., Mi D., Vinh N.-N., Webber C., Li M., Marín O., O’Donovan M.C. & Bray N.J., Single nuclei RNA sequencing of 5 regions of the human prenatal brain implicates developing neuron populations in genetic risk for schizophrenia, Biological Psychiatry (2022), doi: https://doi.org/10.1016/j.biopsych.2022.06.033.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 Published by Elsevier Inc on behalf of Society of Biological Psychiatry.
Single nuclei RNA sequencing of 5 regions of the human prenatal brain implicates developing neuron populations in genetic risk for schizophrenia

Darren Cameron, Da Mi, Ngoc-Nga Vinh, Caleb Webber, Meng Li, Oscar Marin, Michael C. O’Donovan, Nicholas J. Bray

1 MRC Centre for Neuropsychiatric Genetics & Genomics, Division of Psychological Medicine & Clinical Neurosciences, Cardiff University, United Kingdom
2 Tsinghua-Peking Center for Life Sciences, IDG/McGovern Institute for Brain Research, School of Life Sciences, Tsinghua University, Beijing, China
3 UK Dementia Research Institute, Cardiff University, United Kingdom
4 Neuroscience and Mental Health Research Institute, Cardiff University, United Kingdom
5 Centre for Developmental Neurobiology and 6 MRC Centre for Neurodevelopmental Disorders, King’s College London, United Kingdom

Correspondence to: Nicholas Bray, MRC Centre for Neuropsychiatric Genetics & Genomics, Cardiff University School of Medicine, Hadyn Ellis Building, Maindy Road, Cardiff CF24 4HQ, United Kingdom. Email: BrayN3@Cardiff.ac.uk.

Running title: Fetal brain cells implicated in schizophrenia

Keywords: Schizophrenia, neurodevelopment, single nuclei RNA sequencing, genetic, gene expression, GWAS
Abstract

Background: While a variety of evidence supports a prenatal component to schizophrenia, there are few data regarding the cell populations involved. We sought to identify cells of the human prenatal brain mediating genetic risk for schizophrenia by integrating cell-specific gene expression measures generated through single nuclei RNA sequencing with recent large-scale GWAS and exome sequencing data for the condition.

Methods: Single nuclei RNA sequencing was performed on 5 brain regions (frontal cortex, ganglionic eminence, hippocampus, thalamus and cerebellum) from 3 fetuses from the second trimester of gestation. Enrichment of schizophrenia common variant genetic liability and rare damaging coding variation was assessed in relation to gene expression specificity within each identified cell population.

Results: Common risk variants were prominently enriched within genes with high expression specificity for developing neuron populations within the frontal cortex, ganglionic eminence and hippocampus. Enrichments were largely independent of genes expressed in neuronal populations of the adult brain that have been implicated in schizophrenia through the same methods. Genes containing an excess of rare damaging variants in schizophrenia had higher expression specificity for developing glutamatergic neurons of the frontal cortex and hippocampus that were also enriched for common variant liability.

Conclusions: We find evidence for a distinct contribution of prenatal neuronal development to genetic risk for schizophrenia, involving specific populations of developing neuron within the second trimester fetal brain. Our study significantly advances understanding of the neurodevelopmental origins of schizophrenia and provides a resource with which to investigate the prenatal antecedents of other psychiatric and neurological disorders.
Introduction

Schizophrenia is a severe psychiatric disorder characterized by profound disturbances of thought, perception and behaviour. Although the neurobiological mechanisms underlying the condition are poorly understood, it is known to have a substantial genetic component (1), of which common (>1% population frequency) risk alleles account for a sizeable fraction (2,3). One of the leading theories of schizophrenia pathogenesis holds that the condition, which is typically diagnosed in early adulthood, has origins in prenatal brain development (4,5). This hypothesis, founded partly on epidemiological evidence of environmental risk factors operating in the pre- or peri-natal period (6,7), has been supported in recent years by studies showing enrichment of schizophrenia risk variation within genes expressed in the prenatal human brain (8–10) and within common genetic variants associated with effects on gene regulation in fetal brain tissue (11–15). However, with few exceptions (16,17), these studies have been performed using ‘bulk’ samples of fetal brain tissue that encompass multiple neural cell types at different stages of development.

Advances in single cell sequencing technologies have made it possible to profile the transcriptomes of thousands of individual cells / nuclei in parallel, allowing delineation and identification of cell populations of the brain based on global gene expression (16, 18–22). Thus defined, individual cell populations can be assessed for relevance to a given trait by testing for enrichment of trait-associated variation in the genes they express or in putative regulatory DNA elements mapped within them. This approach has provided evidence for microglial involvement in Alzheimer’s disease (22), implicated oligodendrocytes as well as monaminergic and cholinergic neurons in Parkinson’s disease (23,24) and linked common variant genetic risk for schizophrenia with gene expression in neuronal, rather than glial, cell populations in the adult brain (3,23,25).
To date, single cell sequencing studies of the human fetal brain have largely focused on individual brain regions (16-21), with the majority of data derived from the developing cerebral cortex (16–20, 26). The only published studies to our knowledge that have used single cell / nuclei sequencing data to explore cellular mediators of genetic risk for schizophrenia in the human prenatal brain have likewise focused on cortical cell populations (16,17). To provide a more extensive dataset, using uniform methodology, with which to investigate the developmental neurobiology of schizophrenia, we here report single nuclei RNA sequencing (snRNA-seq) in 5 regions of the human second trimester fetal brain of potential relevance to the disorder (frontal cortex, ganglionic eminence, hippocampus, thalamus and cerebellum). We combine these data with results from the largest available genome-wide association (GWAS) and exome sequencing studies of schizophrenia (3,27) to implicate specific cell populations of the human prenatal brain in genetic risk for the condition.

Methods and Materials

Samples

Single nuclei RNA sequencing was performed on brain tissue from 3 karyotypically-normal human fetuses from the second trimester of gestation (two of 14 and one of 15 post-conception weeks, all female). Tissue was acquired from the MRC-Wellcome Trust Human Developmental Biology Resource (HDBR) as fresh whole brain in Hibernate-E media (Thermo Fisher Scientific). All samples were obtained through elective terminations of pregnancy, with consent from female donors. Ethical approval for the collection and distribution of fetal material for scientific research was granted to the HDBR by the Royal Free Hospital research ethics committee and NRES Committee North East - Newcastle & North Tyneside. The frontal cortex, whole ganglionic eminence, hippocampus, thalamus and cerebellum were dissected from both hemispheres of each sample under a light microscope and individually dounce homogenized on ice to produce cell suspensions. Aliquots of 1-10 million cells from each dissected region of each fetus were stored at -80°C until required.
Single nuclei RNA Sequencing (snRNA-seq)

Cryopreserved cell suspensions were gently lysed and processed to obtain intact nuclei. snRNA-seq libraries were prepared from ~10,000 nuclei from each sample using Chromium Single Cell 3’ (v3) reagents (10X Genomics). Quality control of libraries was performed using the Agilent 5200 Fragment Analyzer before sequencing on an Illumina NovaSeq 6000 to a depth of at least 865 million (median = 1.01 billion) read pairs per library.

Raw sequencing data were converted into FASTQ files, aligned to the hg38 build of the human reference genome and quantified using cellranger count (10X Genomics). A stringent quality control procedure was carried out to ensure that only high-quality single nuclei data were processed. Cells were excluded if they expressed fewer than 1000, or greater than 5000, genes, if greater than 5% or 10% of their transcriptome mapped to the mitochondrial genome or ribosomal genes respectively, or if they were identified as a doublet. Genes from the mitochondrial genome or which were expressed in fewer than 3 cells were excluded. Normalization, dimensionality reduction, clustering and cluster visualization were carried out using Seurat version 4.0.3 (28). A supervised support vector machine classifier was used to assess cluster stability (29). We primarily assigned cell-type identity to each cluster based on expression of known cell markers (Supplementary Figures 1-5) and other differentially-expressed genes (Supplementary tables S2-S6). Where public scRNA-seq / snRNA-seq datasets where available from the human fetal brain region assayed (16,19,21), a cell label transfer approach was also implemented in which a reference dataset was created from the pre-existing data and the cell type labels in the reference data were projected onto cells in the query dataset based on the similarity of their gene expression profiles (28,30). A full description of snRNA-Seq analyses and assignment of cell-type identity are provided in Supplementary Methods.
MAGMA cell-typing for testing enrichment of common variant genetic association

The MAGMA celltyping package (https://github.com/neurogenomics/MAGMA_Celltyping) (25) was used to test for enrichment of schizophrenia common variant genetic associations (3) (available through: https://doi.org/10.6084/m9.figshare.14672178) within each fetal brain cell population. This package is a wrapper for the gene-set enrichment analysis MAGMA (31), and, following Bryois et al (23) and Trubetskoy et al (3), was used to test for enrichment of schizophrenia associations in genes with the top 10% highest expression specificity values for each fetal brain cell population. The function map.snps.to.genes was run to map SNPs in the schizophrenia GWAS (3) to genes and then to compute gene-wide association $P$-values. The 1000 genomes data (phase 3) (32) was used as the reference to account for linkage disequilibrium (LD) between SNPs. The boundaries of each gene’s transcribed region were extended at the default of 10kb upstream and 1.5kb downstream. Next, for each brain region, a specificity score was calculated for each gene in each cell type by dividing a gene’s normalized counts per million UMI count in one cell type by the sum of that gene’s expression in all cell types. Uninformative (i.e. sporadically expressed) genes were removed prior to this process using the function drop.uninformative.genes. Genes in the MHC region on chromosome 6 were also removed due to its complex LD structure. These scores were then scaled and ranked into deciles so that each cell type was comparable. A linear regression was then run to test for a one-sided association between the 10% most specific genes in each cell type and the gene-level genetic association with schizophrenia. Covariates for gene size, gene density, the inverse of the minor allele counts, per-gene sample size and the log of these measures were accounted for. A total of 91 tests were performed (one for each cell population) and, following Trubetskoy et al (3), we primarily report enrichment (1-tailed) $P$-values that exceed the Bonferroni threshold ($P < 5.5 \times 10^{-4}$) for these analyses. We also report enrichments with a false discovery rate (FDR) < 0.05. For comparison, MAGMA was also performed on GWAS data for autism (33) (available through:}
https://figshare.com/articles/dataset/asd2019/14671989) and human height (34) (available through:

Stratified Linkage Disequilibrium Score Regression

Following others (3,23,25), we also performed stratified linkage disequilibrium score regression (sLDSR) (35) to test enrichment of schizophrenia SNP heritability in genes in the top expression specificity decile of each cell population. We used HapMap Project phase 3 SNPs with a minor allele frequency >5% and extended the genomic coordinates for each gene by 100kb upstream and downstream of the transcribed region, as recommended by the sLDSR authors (35). Again, uninformative genes and genes overlapping the MHC region were excluded from the analysis. For each gene set, LD scores were then estimated for each SNP in relation to nearby SNPs within a 1 centimorgan window using the 1000 genomes phase 3 reference panel files to estimate LD. Finally, schizophrenia SNP heritability was stratified for each gene set using a joint fit model accounting for SNP heritability attributable to 53 genomic annotations, including genic, enhancer and conserved regions (baseline model v1.2), as performed previously (3,23,25). Significance was determined empirically by calculating a Z-score based on whether schizophrenia SNP-heritability was greater in each gene set compared to the baseline model annotations. As with MAGMA, we also performed sLDSR on GWAS data for autism (33) and human height (34).

Gene Ontology analysis

To explore biological processes mediating common variant genetic liability to schizophrenia in different populations of developing neuron, we first performed Gene Ontology (GO) enrichment analyses on genes in the top decile of expression specificity for each implicated cell type, with a
background list of all genes expressed in that cell, using the DAVID Bioinformatics Resource 6.8 (36). We then used MAGMA (31) to test enrichment of schizophrenia associations (3) within genes annotated to the most significantly over-represented GO terms for each implicated cell type to identify those processes most relevant to schizophrenia.

Expression analyses of genes containing an excess of rare damaging coding variation

Following Singh and colleagues (27), we tested for higher expression specificity of genes carrying an excess (FDR < 0.05) of rare damaging coding variation in schizophrenia (27) by performing a one-sided Wilcoxon rank sum test comparing the expression specificity score ranking of those 32 genes against the ranking of all remaining genes for each cell population. Expression specificity scores for each cell population are provided in Supplementary tables S8-S12.

Results

Single nuclei RNA sequencing of 5 regions of the human fetal brain

We performed snRNA-seq on 5 dissected brain regions from 3 human fetuses from the second trimester of gestation. After strict quality control, we retained snRNA-seq data from 13,597 nuclei from the frontal cortex, 11,310 nuclei from the ganglionic eminence, 12,594 nuclei from the hippocampus, 17,283 nuclei from the thalamus and 22,652 nuclei from the cerebellum. Nuclei were clustered according to gene expression profile using shared nearest neighbor modularity optimization-based clustering (26). We thus identified 17 transcriptionally distinct clusters of nuclei in the fetal frontal cortex, 11 in the ganglionic eminence, 19 in the hippocampus, 23 in the thalamus and 21 in the cerebellum (Figure 1). Defined clusters in all regions showed high stability (median F1 per region > 0.92; Supplementary table S1).
Clusters were annotated based on expression of known cell markers (Supplementary Figures S1 - S5) and other differentially expressed genes (Supplementary Tables S2 - S6). In addition to populations of radial glia (markers: GLI3, TNC), intermediate progenitors (marker: EOMES), oligodendrocyte precursor cells (marker: OLIG1), endothelial cells (marker: ITM2A) and microglia (marker: C3), we were able to define a variety of developing neuron populations appropriate for each region. Thus, in the frontal cortex, we found developing excitatory (glutamatergic) neuron populations differentially expressing markers of upper cortical layers, such as TLE3 and LHX2 (FC-ExN-1, FC-ExN-2 and FX-ExN-3), or markers of deep cortical layers, such as CRYM and FEZF2 (FC-ExN-4 and FC-ExN-5), as well as developing GABAergic interneuron populations expressing CALB2 (FC-InN-1) or SST (FC-InN-4). In the ganglionic eminence, a transitory structure of the prenatal brain that gives rise to cortical and subcortical inhibitory neurons (37), we observed interneuron populations expressing the medial ganglionic eminence marker LHX6 (GE-InN-1, GE-InN-2 and GE-InN-5) or the caudal ganglionic eminence marker PROX1 (GE-InN-3), in addition to a population (GE-InN-6) that we predict to be developing medium spiny neurons from the lateral ganglionic eminence based on expression of SIX3 and TSHZ1 (37). In the hippocampus, we identified two populations highly expressing the Cajal-Retzius cell marker RELN (Hipp-CR-1 and Hipp-CR-2), two excitatory neuron populations expressing the dentate gyrus granule cell marker PROX1 (Hipp-ExN-3 and Hipp-ExN-6), four excitatory neuron populations expressing the CA3 marker GRIK4 (Hipp-ExN-1, Hipp-ExN-2 and Hipp-ExN-4 and Hipp-ExN-5) and populations of interneuron predicted to derive from the medial (Hipp-InN-2) or caudal ganglionic eminence (Hipp-InN-1 and Hipp-InN-4). In the thalamus, we discerned three developing glutamatergic neuron populations expressing SLC17A6 (Thal-ExN-1, Thal-ExN-2 and Thal-ExN-3), along with 8 GABAergic populations expressing SLC6A1 and / or GAD2 (Thal-InN-1, Thal-InN-2, Thal-InN-3, Thal-InN-4, Thal-InN-5, Thal-InN-6, Thal-InN-7 and Thal-InN-8). In the cerebellum, we observed five populations expressing markers of (excitatory) granule neurons such as RBFOX3 and RELN (Cer-ExN-1, Cer-ExN-2, Cer-ExN-3, Cer-ExN-4 and Cer-ExN-5) as well as four populations of (inhibitory) Purkinje cells marked by expression of CA8 and ITPR1 (Cer-InN-1, Cer-InN-2, Cer-InN-3,
Fetal brain cells implicated in schizophrenia

Cer-InN-4). There was strong concordance between our annotations and predictions using data from previous scRNA-seq / snRNA-seq studies of the human fetal brain (16,19,21) (Supplementary Figures S6 - S8).

Common risk alleles for schizophrenia are enriched in genes with high expression specificity for developing neurons of the fetal brain

Single cell / nuclei data from mouse and human have indicated enrichment of common schizophrenia risk variation in genes with high expression specificity for populations of mature neuron of the adult brain (3,23,25). To investigate cells of the prenatal human brain mediating common variant genetic liability for schizophrenia, we similarly determined the top decile of gene expression specificity for each of the identified fetal brain cell populations (Supplementary Table S7) and used two different statistical methods (MAGMA [31] and sLDSR [35]) to assess enrichment of schizophrenia genetic risk (3) within those gene loci.

Both MAGMA and sLDSR showed significant enrichment (exceeding the Bonferroni threshold of $P < 5.5 \times 10^{-4}$) of schizophrenia genetic risk in genes with high expression specificity for 6 fetal cell populations, all of which were annotated as developing neurons (Figure 2). The most significant enrichment was observed for genes with high expression specificity for FC-ExN-2 ($P_{\text{MAGMA}} = 1.98 \times 10^{-11}$; $P_{\text{sLDSR}} = 3.33 \times 10^{-8}$), a maturing upper layer pyramidal neuron of the frontal cortex. Schizophrenia genetic risk was also prominently enriched in genes with high expression specificity for another upper layer glutamatergic neuron of the frontal cortex (FC-ExN-3), a deep layer excitatory neuron of the frontal cortex (FC-ExN-4), a developing interneuron population within the (medial) ganglionic eminence (GE-InN-2) and two developing glutamatergic neuron populations from the hippocampus (Hipp-ExN-3, expressing markers of dentate gyrus granule cells, and Hipp-ExN-5, expressing markers
of CA3 pyramidal cells). A further 8 fetal cell populations were enriched for schizophrenia genetic liability at the less conservative FDR < 0.05 threshold, including another deep layer excitatory neuron of the frontal cortex (FC-ExN-5), 4 additional developing inhibitory neuron populations of the ganglionic eminence (GE-InN-1, GE-InN-4, GE-InN-5 and GE-InN-7) and 3 neuronal populations of the thalamus (Thal-ExN-1, Thal-ExN-3 and Thal-InN-7). We found little evidence for enrichment of schizophrenia common variant liability in genes with high expression specificity for progenitor or glial cells in any region. This pattern of neuronal, rather than progenitor, enrichment was maintained even when large differences in gene boundary were applied (Supplementary Figures S9 and S10).

For comparison, we tested enrichment of common genetic variation associated with autism (33) and human height (34), as neurodevelopmental and non-brain phenotypes respectively, within the same high expression specificity genes for each fetal cell population. Although autism GWAS currently provide limited statistical power for these analyses, we observed nominally significant ($P < 0.05$) enrichment of associations in deep layer excitatory neuron FC-ExN-4 of the frontal cortex and oligodendrocyte precursor Thal-OPC of the thalamus by both MAGMA and sLDSR (Supplementary Figure S11). In contrast, despite strong statistical power, variants influencing human height were not enriched in any neural population; instead, MAGMA indicated significant enrichment within genes in the top expression specificity decile for endothelial cells (Supplementary Figure S12).

**Independence of schizophrenia genetic association between implicated cell types of the fetal brain**

We next investigated the extent to which observed enrichments of schizophrenia risk alleles in different populations of developing neuron represent independent signals. Figure 3a shows the overlap of nominally significant schizophrenia-associated genes (MAGMA gene-wise $P < 0.05$) in the top expression specificity decile of the 6 cell populations of the fetal brain that we implicate in the
Fetal brain cells implicated in schizophrenia disorder at the Bonferroni significance threshold. Cell populations generally shared less than half of the schizophrenia-associated genes in their top specificity deciles with any other implicated cell population, suggesting partly independent enrichment signals. We formally tested for enrichment of schizophrenia genetic association independent of genes shared with the most strongly enriched cell type (FC-ExN-2) by repeating the MAGMA cell-specific expression analyses of the 5 other implicated cell types, conditioning on genes in the top expression specificity decile of FC-ExN-2 (Figure 3b). All implicated fetal cell populations remained significantly ($P < 0.05$) enriched for genetic association with schizophrenia when genes shared with FC-ExN-2 were accounted for.

Enrichment of schizophrenia genetic risk within developing neurons of the fetal brain is largely independent of genetic enrichments in adult neurons

We then assessed the extent to which enrichment of schizophrenia genetic risk in developing neuronal populations of the fetal brain can be explained by genes shared with neurons of the adult brain that have been implicated in the disorder through the same methods (3,23,25). Figure 4 shows the overlap of nominally significant schizophrenia-associated genes (MAGMA gene-wise $P < 0.05$) in the top expression specificity decile of the 6 fetal neuron populations that we implicate in the disorder at the Bonferroni threshold with those in the top expression specificity deciles of neuronal populations implicated in schizophrenia using snRNA-Seq data from adult human brain (frontal cortex pyramidal neurons, dentate gyrus granule neurons, hippocampal CA1 pyramidal neurons and hippocampal CA3 pyramidal neurons) and scRNA-Seq data from adolescent mouse brain (hippocampal CA1 pyramidal neurons, striatal medium spiny neurons, cortical interneurons and cortical somatosensory pyramidal neurons). None of the fetal cell populations shared more than 29% of schizophrenia-associated genes with any adult neuron, suggesting largely independent enrichments. We confirmed this by repeating the MAGMA cell-specific expression analyses, conditioning on genes in the top expression specificity decile of each implicated adult neuron.
population (Figure 5). All implicated fetal cell populations remained highly enriched (maximum $P < 0.007$) for schizophrenia genetic association, indicating that variants acting through genes expressed in developing neurons of the prenatal brain make a distinct contribution to risk for schizophrenia.

**Biological processes mediating common variant genetic association with schizophrenia in cells of the developing brain**

To better understand the biological processes mediating genetic association with schizophrenia in cells of the second trimester fetal brain, we first performed Gene Ontology (GO) enrichment analyses on genes in the top decile of expression specificity of the 6 cell populations implicated in the disorder by MAGMA and sLDSR at the Bonferroni threshold. As expected, cell-specific gene sets were significantly enriched (FDR < 0.05) for terms relating to neuronal development, although we note that genes with high expression specificity for implicated excitatory neuron populations were also enriched for mature neuronal functions such as synaptic signaling and regulation of membrane potential (Supplementary Figure S13). To determine which of these processes are relevant to schizophrenia genetic liability, we used MAGMA to test for enrichment of schizophrenia associations (3) within genes annotated to over-represented GO terms in the top expression specificity decile of each cell population (Figure 6). Consistent with an important role for neuronal development in schizophrenia susceptibility, risk alleles were significantly enriched (at a level exceeding the Bonferroni $P$-value threshold) within genes annotated to terms such as ‘neurogenesis’, ‘neuron differentiation’ and ‘neuron projection development’ belonging to all 6 cell populations. For genes with high expression specificity for FC-ExN-2 and FC-ExN-4, schizophrenia genetic associations were additionally enriched for synaptic signaling, which is likely to index a later, post-natal risk mechanism for the disorder. These analyses thus suggest several mechanisms by which common alleles confer risk to schizophrenia within developing neurons.
**Cellular expression specificity of genes implicated in schizophrenia through rare coding variation**

A recent large-scale exome sequencing study identified 10 genes containing an exome-wide significant excess of rare damaging coding variants in schizophrenia, with 32 such genes at FDR < 0.05 (27). Using scRNA-Seq data from the adolescent mouse nervous system (38), the authors reported that the 32 FDR < 0.05 genes had significantly higher expression specificity for (brain-derived) neuronal populations. Applying similar methodology, we found higher expression specificity of these 32 genes in 4 cell populations of the human fetal brain at FDR < 0.05 (Figure 7); namely, FC-ExN-2, FC-InN-4, FC-OPC and Hipp-ExN-5. We note that FC-ExN-2 and Hipp-ExN-5 also display enrichment of common variant genetic liability for schizophrenia, suggesting points of biological convergence. We performed an identical analysis of 78 genes implicated in autism through exome sequencing (39) at FDR < 0.05, finding expression enrichment in a variety of developing cell populations (Supplementary Figure S14), consistent with prior literature (16,39).

**Discussion**

By combining single nuclei RNA sequencing of 5 regions of the human fetal brain with recent large-scale genomic data for schizophrenia (3,27), we implicate specific neuronal populations within the developing frontal cortex, ganglionic eminence and hippocampus in genetic liability to the disorder.

To our knowledge, only two other published studies have used human single cell sequencing data to explore cellular mediators of genetic risk for schizophrenia in the prenatal brain, both of which focused on the developing cerebral cortex. The first, by Polioudakis and colleagues (16), used sLDSR to test for enrichment of schizophrenia common variant liability in regions of open chromatin (an index of regulatory genomic regions) associated with genes enriched in specific cell types of the second trimester fetal neocortex (gestation week 17 to 18). Consistent with our findings, the authors
found enrichment of schizophrenia genetic risk in regulatory genomic sites associated with developing glutamatergic populations, although significant enrichments were also observed in sites associated with a variety of other cell types, including radial glia, microglia and oligodendrocyte precursors (the latter also enriched for genes containing an excess of rare damaging coding mutations in schizophrenia in the present study). More recently, Ziffra and colleagues (17) used sLDSR to test enrichment of schizophrenia common variant genetic risk in predicted enhancers among cell-specific open chromatin regions of the mid-gestation human cerebral cortex. While implicating no progenitor or glial cell population, the authors found an enrichment of schizophrenia-associated variation in enhancers attributed to upper layer excitatory neurons, consistent with the present study, as well as cortical interneurons predicted to derive from the caudal ganglionic eminence.

Although studies based on single cell sequencing data have implicated neuronal populations of the adult hippocampus in genetic risk for schizophrenia (3,23,25), our study is the first to implicate cell populations of the prenatal hippocampus in genetic liability to the disorder. Our finding that schizophrenia genetic risk is enriched in developing glutamatergic neurons of the frontal cortex and hippocampal formation supports the view that the disorder has origins in the initial formation of neuronal connectivity within those regions (40-43) and may partly explain reductions in synaptic markers observed in those areas in post-mortem studies of schizophrenia (43,44). Our observation that common variant genetic risk for schizophrenia is independently enriched within genes with high expression specificity for at least one interneuron population of the ganglionic eminence provides evidence that early GABAergic neuron development is also perturbed in the disorder (45,46). At a more relaxed significance threshold of FDR < 0.05, we additionally implicate neuronal populations of the developing thalamus in common variant liability to schizophrenia.
Limitations of this study are that we focused on one developmental timepoint, from only female samples. Future studies could also extend the range of brain regions to include others of potential relevance to schizophrenia. Although we were able to utilize GWAS and exome sequencing data from many thousands of participants (3,27), even larger sample sizes are likely to bring greater resolution of relevant cell types, as illustrated in gains observed through adult brain cell-specific enrichment analyses performed using successive schizophrenia GWAS of the Psychiatric Genomics Consortium (3). We therefore provide our cellular expression specificity data for future studies exploring the prenatal antecedents of schizophrenia and other brain-related traits.

Acknowledgments

This work was supported by a project grant (MR/T002379/1) and Centre grant (MR/L010305/1) from the Medical Research Council (UK). We thank Dr Joanne Morgan for assistance with sequencing. The human fetal material was provided by the Joint MRC/Wellcome Trust (grant #099175/Z/12/Z) Human Developmental Biology Resource (www.hdbr.org).

Disclosures

NJB and MCO’D have received a research award from Takeda Pharmaceutical Company Ltd, unrelated to this work. All other authors report no biomedical financial interests or potential conflicts of interest.

References


Fetal brain cells implicated in schizophrenia


Figure legends

**Figure 1.** snRNA-seq clusters in 5 regions of the second trimester human fetal brain. Nuclei were clustered according to gene expression profile using shared nearest neighbor (SNN) modularity optimization-based clustering in Seurat 4.0 (28) and visualized in 2D space using Uniform Manifold Approximation and Projection (UMAP) (47). Clusters in all regions showed high stability (median F1 per region > 0.92). Clusters were annotated based on expression of known cell markers (Supplementary Figures S1 - S5) and other differentially expressed genes (Supplementary Tables S2 - S6). ExN = developing excitatory neuron; InN = developing inhibitory neuron; CR = Cajal-Retzius cell; N-undef = neuron of undefined class; IP = intermediate progenitor; RG = radial glia; CycPro = cycling progenitor cells; OPC = oligodendrocyte precursor cell; Endo = endothelial cell; MG = microglia.

**Figure 2.** -log10 P-values for enrichment of schizophrenia genetic associations (3) in genes in the top expression specificity decile for each identified fetal brain cell population using MAGMA and sLDSR. The dotted vertical line indicates nominal (P < 0.05) significance and the dashed vertical line indicates the Bonferroni-corrected P-value threshold for 91 tested cell populations (P < 5.5 X 10^-4). *Cell populations where enrichments in both MAGMA and sLDSR are FDR < 0.05. ExN = developing excitatory neuron; InN = developing inhibitory neuron; CR = Cajal-Retzius cell; N-undef = neuron of undefined class; IP = intermediate progenitor; RG = radial glia; CycPro = cycling progenitor cells; OPC = oligodendrocyte precursor cell; Endo = endothelial cell; MG = microglia.

**Figure 3.** Independence of schizophrenia genetic association between implicated cell types of the fetal brain. A) Number of genes exhibiting nominally significant genetic association with schizophrenia (MAGMA gene-wise P < 0.05) shared between the top expression specificity deciles of
each fetal cell population implicated in the disorder at a level exceeding the Bonferroni significance
threshold in both the MAGMA and sLDSR analysis. The total number of schizophrenia-associated
genes (MAGMA gene-wise $P < 0.05$) in the top expression specificity decile of each cell type is shown
in boxes with bold borders. B) MAGMA -log$_{10}$ $P$-values for enrichment of schizophrenia genetic
associations (3) in genes in the top expression specificity deciles of 5 fetal brain cell populations
implicated in the disorder, conditioning on genes in the top expression specificity decile of the most
significantly enriched cell type (FC-ExN-2). The dotted vertical line indicates $P = 0.05$. ExN =
developing excitatory (glutamatergic) neuron; InN = developing inhibitory (GABAergic) neuron; FC =
frontal cortex; GE = ganglionic eminence; Hipp = hippocampus.

Figure 4. Overlap of schizophrenia-associated genes in the top expression specificity decile of
implicated neuron populations of the fetal and adult human brain. A) Number of genes exhibiting
nominally significant genetic association with schizophrenia (MAGMA gene-wise $P < 0.05$) shared
between the top expression specificity deciles of each fetal cell population (implicated in the
disorder at a level exceeding the Bonferroni significance threshold in both the MAGMA and sLDSR
analysis) and each similarly implicated adult human cell population (3). Adult-hum-exCA1 =
hippocampal CA1 pyramidal neurons; Adult-hum-exCA3 = hippocampal CA1 pyramidal neurons;
Adult-hum-exDG = dentate gyrus granule neurons; Adult-hum-exPFC1 = frontal cortex pyramidal
neurons. B) Number of genes exhibiting nominally significant genetic association with schizophrenia
(MAGMA gene-wise $P < 0.05$) shared between the top expression specificity deciles of each fetal cell
population (implicated in the disorder at a level exceeding the Bonferroni significance threshold in
both the MAGMA and sLDSR analysis) and each similarly implicated adult mouse cell population.
Adult-mus-SS = cortical somatosensory pyramidal neurons; Adult-mus-InN = cortical interneurons;
Adult-mus-MSN = striatal medium spiny neurons; Adult-mus-CA1 = hippocampal CA1 pyramidal
neurons.
Figure 5. MAGMA -log10 P-values for enrichment of common variant risk for schizophrenia (3) in implicated fetal neuron populations, conditioning on genes in the top expression specificity decile of each implicated adult neuron population. The dotted vertical line indicates $P = 0.05$. ExN = developing excitatory (glutamatergic) neuron; InN = developing inhibitory (GABAergic) neuron; FC = frontal cortex; GE = ganglionic eminence; Hipp = hippocampus. A) MAGMA -log10 P-values for enrichment of schizophrenia common variant liability in fetal neuron populations (implicated in the disorder at a level exceeding the Bonferroni significance threshold in both the MAGMA and sLDSR analysis), conditioning on genes in the top expression specificity decile of each similarly implicated adult human cell population (3). Adult-hum-exCA1 = hippocampal CA1 pyramidal neurons; Adult-hum-exCA3 = hippocampal CA1 pyramidal neurons; Adult-hum-exDG = dentate gyrus granule neurons; Adult-hum-exPFC1 = frontal cortex pyramidal neurons. B) MAGMA -log10 P-values for enrichment of schizophrenia common variant liability in fetal neuron populations (implicated in the disorder at a level exceeding the Bonferroni significance threshold in both the MAGMA and sLDSR analysis), conditioning on genes in the top expression specificity decile of each similarly implicated adult mouse cell population (3). Adult-mus-SS = cortical somatosensory pyramidal neurons; Adult-mus-InN = cortical interneurons; Adult-mus-MSN = striatal medium spiny neurons; Adult-mus-CA1 = hippocampal CA1 pyramidal neurons.

Figure 6. MAGMA -log10 P-values for enrichment of schizophrenia genetic associations (3) within over-represented gene ontology (GO) terms belonging to the top expression specificity decile of each implicated cell population. The dashed line indicates the Bonferroni-corrected $P$-value threshold for 121 tests ($P < 4.1 \times 10^{-4}$). InN = developing inhibitory (GABAergic) neuron; FC = frontal cortex; GE = ganglionic eminence; Hipp = hippocampus.
Fetal brain cells implicated in schizophrenia

Figure 7. Wilcoxon rank sum test -log_{10} P-values for higher gene expression specificity of 32 genes carrying an excess (FDR < 0.05) of rare damaging coding variation in schizophrenia (27) in cell populations of the human fetal brain. The dotted vertical line indicates nominal (P < 0.05) significance, the dashed vertical line indicates the Bonferroni-corrected P-value threshold for 91 tested cell populations (P < 5.5 X 10^{-4}) and asterisks indicate cell populations in which the 32 genes have higher expression specificity at FDR < 0.05. ExN = developing excitatory neuron; InN = developing inhibitory neuron; CR = Cajal-Retzius cell; N-undef = neuron of undefined class; IP = intermediate progenitor; RG = radial glia; CycPro = cycling progenitor cells; OPC = oligodendrocyte precursor cell; Endo = endothelial cell; MG = microglia.
<table>
<thead>
<tr>
<th></th>
<th>Adult.humanCA1 (499)</th>
<th>Adult.humanCA3 (511)</th>
<th>Adult.humanDG (504)</th>
<th>Adult.humanPFC1 (516)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC-ExN-2 (538)</td>
<td>137</td>
<td>136</td>
<td>134</td>
<td>148</td>
</tr>
<tr>
<td>FC-ExN-3 (476)</td>
<td>94</td>
<td>94</td>
<td>88</td>
<td>96</td>
</tr>
<tr>
<td>FC-ExN-4 (480)</td>
<td>114</td>
<td>121</td>
<td>103</td>
<td>133</td>
</tr>
<tr>
<td>GE-InN-2 (439)</td>
<td>54</td>
<td>52</td>
<td>45</td>
<td>53</td>
</tr>
<tr>
<td>Hipp-ExN-3 (432)</td>
<td>82</td>
<td>81</td>
<td>99</td>
<td>68</td>
</tr>
<tr>
<td>Hipp-ExN-5 (398)</td>
<td>113</td>
<td>97</td>
<td>104</td>
<td>86</td>
</tr>
<tr>
<td><strong>B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC-ExN-2 (538)</td>
<td>131</td>
<td>106</td>
<td>106</td>
<td>138</td>
</tr>
<tr>
<td>FC-ExN-3 (476)</td>
<td>91</td>
<td>76</td>
<td>72</td>
<td>89</td>
</tr>
<tr>
<td>FC-ExN-4 (480)</td>
<td>126</td>
<td>97</td>
<td>117</td>
<td>127</td>
</tr>
<tr>
<td>GE-InN-2 (439)</td>
<td>51</td>
<td>60</td>
<td>36</td>
<td>54</td>
</tr>
<tr>
<td>Hipp-ExN-3 (432)</td>
<td>99</td>
<td>65</td>
<td>86</td>
<td>86</td>
</tr>
<tr>
<td>Hipp-ExN-5 (398)</td>
<td>114</td>
<td>77</td>
<td>81</td>
<td>91</td>
</tr>
</tbody>
</table>