Schizophrenia Bulletin vol. 49 no. 4 pp. 914–922, 2023
https://doi.org/10.1093/schbul/sbad002
Advance Access publication 20 February 2023

Genetic Insights of Schizophrenia via Single Cell RNA-Sequencing Analyses

Yong Wu[1,2,6], Chu-Yi Zhang[3,4,6], Lu Wang[3], Yi Li[1,2,5,7], and Xiao Xiao[*,3,4,7]

[1]Research Center for Mental Health and Neuroscience, Wuhan Mental Health Center, Wuhan, Hubei, China; [2]Affiliated Wuhan Mental Health Center, Jianghan University, Wuhan, Hubei, China; [3]Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences and Yunnan Province, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China; [4]Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, Yunnan, China; [5]Research Center for Psychological and Health Sciences, China University of Geosciences, Wuhan, Hubei, China

[6]These authors contributed equally to the article.

[7]Xiao Xiao and Yi Li are co-corresponding authors who jointly directed this work.

[*]To whom correspondence should be addressed; Kunming Institute of Zoology, Chinese Academy of Sciences, NO 17 Long-Xin Lu, Kunming, Yunnan 650201, P.R. China. tel: 86-871-65189508; e-mail: xiaoxiao2@mail.kiz.ac.cn

***Background*: Schizophrenia is a complex and heterogeneous disorder involving multiple regions and types of cells in the brain. Despite rapid progress made by genome-wide association studies (GWAS) of schizophrenia, the mechanisms of the illness underlying the GWAS significant loci remain less clear. *Study Design*: We investigated schizophrenia risk genes using summary-data-based Mendelian randomization based on single-cell sequencing data, and explored the types of brain cells involved in schizophrenia through the expression weighted cell-type enrichment analysis. *Results*: We identified 54 schizophrenia risk genes (two-thirds of these genes were not identified using sequencing data of bulk tissues) using single-cell RNA-sequencing data. Further cell type enrichment analysis showed that schizophrenia risk genes were highly expressed in excitatory neurons and caudal ganglionic eminence interneurons, suggesting putative roles of these cells in the pathogenesis of schizophrenia. We also found that these risk genes identified using single-cell sequencing results could form a large protein-protein interaction network with genes affected by disease-causing rare variants. *Conclusions*: Through integrative analyses using expression data at single-cell levels, we identified 54 risk genes associated with schizophrenia. Notably, many of these genes were only identified using single-cell RNA-sequencing data, and their altered expression levels in particular types of cells, rather than in the bulk tissues, were related to the increased risk of schizophrenia. Our results provide novel insight into the biological mechanisms of schizophrenia, and future single-cell studies are necessary to further facilitate the understanding of the disorder.**

Key words: schizophrenia/risk genes/scRNA-seq/eQTL/cell types

Introduction

Schizophrenia is a highly heritable and polygenic disorder, and genome-wide association studies (GWAS) have reported more than 200 genomic loci harboring common variants that confer the risk of this illness.[1] Nonetheless, the majority of these schizophrenia GWAS risk loci are in noncoding regions of the genome, and it is of great interest to identify variants with functional impact in the pathogenesis of the disorder. As accumulating studies suggest that noncoding genetic variants may participate in complex diseases by modulating the mRNA expression of certain genes, expression quantitative trait loci (eQTL) analyses based on data from human brain tissues have been widely applied and made inspiring findings in illuminating the mechanisms of schizophrenia pathogenesis.[2–4] Nonetheless, systematic colocalization analysis based on tissue-level eQTL data[5] indicated that gene expression alterations detected by such investigations could only explain ~10% of schizophrenia heritability,[6] partly due to the limitation of data from bulk brain tissues, as the brain is a highly heterogenic organ with numerous regions showing specific characteristics in their structure, functionality and cellular composition.

While majority of previous eQTL analyses were conducted using data from bulk tissues due to inevitable technical and resource restrictions, recent studies are being conducted using data obtained from different types of cells, cell states, and developmental stages of the human brain,[7–11] to increase the "resolution" of studies for brain function and pathology of psychiatric illnesses. For example, eQTLs have been identified in primary human neural progenitors and their sorted neuronal progenies,[10]

in dopaminergic neurons derived from induced pluripotent stem cells (iPSCs),[12] and in 8 types of brain cells (using single nuclei RNA-sequencing).[13] Indeed, these studies confirmed that gene expression regulation occurs in a region-, temporal-, and cell-type-specific manner, and have identified numerous eQTLs that were masked in bulk tissues. It is hence necessary to perform further fine-mapping analyses of GWAS loci to characterize cell-type specific eQTLs to better understand the genetic and biological basis of schizophrenia. We herein carried out integrated analyses of schizophrenia GWAS data and brain cell-type eQTL data, and identified risk genes that were not prioritized in previous analyses of bulk tissues.

Materials and Methods

PGC3 Schizophrenia GWAS Data

The PGC3 GWAS was performed in 69 369 schizophrenia cases and 236 642 controls, and 270 genome-wide significant loci were identified. We downloaded the GWAS summary statistics data from the PGC website (https://www.med.unc.edu/pgc/). Detailed information regarding sample collection, analytical methods, and results can be found in the original publication.[1]

Alzheimer's Disease GWAS Data

Bellenguez et al performed a meta-analysis of 2-stage Alzheimer's disease (AD) GWAS containing a total of 111 326 clinically diagnosed/"proxy" AD cases and 677 663 controls and identified 75 risk loci.[14] We downloaded the stage I GWAS data (39 106 clinically diagnosed AD cases, 46 828 "proxy" AD cases, and 401 577 controls) from https://www.ebi.ac.uk/gwas/ under accession no. GCST90027158. Detailed information about sample collection, analytical methods, and results can be found in the original publication.[14]

Cell-type eQTL Data From the Study by Jerber et al

The cell-type eQTL data of mature midbrain dopaminergic neurons (from 175 donors) and serotonergic-like neurons (from 161 donors) were downloaded from the study by Jerber et al[12] Briefly, the authors differentiated human iPSC lines from the Human Induced Pluripotent Stem Cell Initiative (http://www.hipsci.org) with a midbrain neural fate and performed scRNA-seq analyses on cells captured on particular days across differentiation. For example, sequencing was performed on day 52 when the dominant type of cells was mature midbrain dopaminergic neurons,[12,15,16] and the identity of dopaminergic neurons was also confirmed through satisfactory transcriptome-wide alignment (mapping rates between 85% and 99%) to existing single-cell atlases of human iPSC-derived dopaminergic neurons and fetal or adult human midbrain samples.[15,17] Considering the variability in neuronal differentiation efficiency between cell lines, the authors mapped the eQTLs using aggregated expression levels for each donor when calculating eQTLs in distinct cell populations (corresponding to the profiled cell type-condition contexts of the dominant cell types) and applied linear mixed models by incorporating additional variance components.[18] We downloaded the eQTL analysis results from Zenodo at https://zenodo.org/record/4333872.

Cell-Type eQTL Data From the Study by Bryois et al

We downloaded the cell-type eQTL data from the study by Bryois et al.[13] Different from the study by Jerber et al, which obtained eQTL information from iPSC-derived cells, Bryois et al directly performed eQTL analysis using scRNA-seq in the prefrontal cortex, temporal cortex, and deep white matter from 196 individuals. They performed scRNA-seq and genotyping in 123 independent individuals and then combined their data with 127 human brain samples, which were previously conducted using scRNA-seq analysis.[19–21] After quality control and data normalization, gene expression and genotype information from a total of 196 individuals were included in the final data processing. Eight major central neuron system (CNS) cell types were identified according to canonical markers of cell type identity. The cis-eQTLs were calculated by testing all SNPs (single nucleotide polymorphism) within a 1 MB window on the flanking transcription start site by adjusting for known covariates using FastQTL[22] in each cell type. In total, 6108 genes with a cis-eQTL at a 5% false discovery rate (FDR) across the 8 different CNS cell types were discovered. We downloaded the eQTL results from https://doi.org/10.5281/zenodo.5543734.

Summary-Data-Based Mendelian Randomization Analysis

Summary-data-based Mendelian randomization (SMR) is a widely used integrative analysis method to identify genes whose expression levels are associated with a complex trait.[23] Similar to the concept of Mendelian randomization, the SMR tests the causative effect of an exposure of gene expression on an outcome of phenotype using a genetic variant as an instrumental variable. Different from other colocalization and imputation methods that require individual genotype information from eQTL data, SMR integrates summary-level data from independent GWASs with data from summary-level eQTL studies. To rule out the linkage effect, the developers of SMR proposed the HEIDI (heterogeneity in dependent instruments) test to distinguish pleiotropy from the linkage. Contrary to other statistical analyses, the HEIDI test disregarded results with a small heterogeneity test *P*-value, which means that the smaller the *P*-value of HEIDI is, the larger the probability of the observations being derived by a linkage

model. Detailed information about the method theory, model design, and algorithm implementation can be found in the original publication.[23]

The eQTL summary data output by other QTL calculation methods were first transformed to BESD format, which is an efficient binary format to save storage space, using the—make-BESD option. Then, the reference genotype data (downloaded from the 1000 Genomes Project[24]), GWAS summary statistics data, and BESD formatted eQTL summary data were used as SMR inputs. Other settings were set as the default arguments. The significance for SMR was set at the Bonferroni-corrected threshold of $0.05/N$, in which N was the number of genes with significant eQTLs (eGenes) tested for a given cell type. An eQTL was considered significant when its P-value of the top associated locus was less than 5.00×10^{-8}. The numbers of eGenes for dopaminergic neurons, serotonergic-like neurons, astrocytes, endothelial cells, excitatory neurons, inhibitory neurons, microglia, oligodendrocytes, oligodendrocyte precursors, and pericytes were 703, 892, 366, 80, 867, 286, 212, 758, 243, and 23, respectively. Only genes with Bonferroni-corrected P-values of SMR < .05 and P-values of HEIDI test >.01 were considered to be schizophrenia-associated risk genes. Because of the complexity of linkage disequilibrium in the major histocompatibility complex region (26–34 Mb), we excluded genes in this genomic area.

Cell Type Enrichment Analysis

To further identify schizophrenia-associated cell types, we conducted cell type enrichment analysis with the single cell transcriptomic and Expression Weighted Cell-type Enrichment (EWCE) R package.[25] The EWCE evaluated whether a set of genes had higher expression within a particular cell type were selected by chance. The EWCE first calculated the average expression level of the background set of genes in every cell type in a selected single-cell transcriptomic dataset. A simulation test was then performed by randomly sampling a gene list with the same length as the target gene list for 10 000 times, and the probability distribution was estimated from the average level of expression in all cell types in each of these random gene lists. If the expression level of the target gene list was significantly higher than that of the randomly selected gene list in a given cell type, the cell type was thought to be associated with traits of interest. A given cell type with a P-value of EWCE less than .05 was considered a schizophrenia-associated cell type.

The target gene list was defined as the union of significant genes identified by SMR with all single cell types and genes with rare variants identified by exome sequencing from 24 248 cases and 97 322 controls and mutations from 3402 parent-proband trios with an FDR < 5%.[26] We selected single-cell transcriptome data from three studies. The first was a study by Polioudakis et al downloaded from http://solo.bmap.ucla.edu/shiny/webapp/. They collected developing human neocortices at gestation weeks 17 to 18 from 4 donors[27] and obtained ~40 000 cells belonging to 16 distinct cell types, including microglia, pericytes, interneurons, and excitatory neurons through Drop-seq. The second study was a part of the Brain Initiative Cell Census Network, which aimed to identify and catalog the diverse cell types in brains of the human, monkey, and mouse.[28] The dataset contained cells from 10 distinct major forebrain, midbrain, and hindbrain regions from 13 individuals at gestation weeks 14 to 25 and 6 neocortical areas from the same donors. In total, this dataset contained 698 820 high-quality cells of 10 major cell types including excitatory neurons, interneurons, microglia, and intermediate progenitor cells. Due to the large amount of data beyond the memory range of the workstation (128 G), we selected one individual at gestation week 18 for the follow-up study, which contained 84 730 cells. The cell count matrix meta cell types can be downloaded from https://data.nemoarchive.org/biccn/grant/u01_devhu/kriegstein/transcriptome/scell/10x_v2/human/processed/counts/. The third dataset was from the study by Trevino et al, who generated a single-cell RNA transcriptome of the developing human cerebral cortex using 10× Genomics from 4 samples at gestation weeks 16, 20, 21, and 24.[11] Overall, they obtained 31 304 single-cell transcriptomes and identified 23 main cell types. All the processed data could be downloaded at Gene Expression Omnibus with the accession number GSE162170.[11]

Protein-Protein Interaction Analysis

The schizophrenia risk genes identified by SMR and exome sequencing were further used to conduct protein-protein interaction (PPI) analysis on the STRING website (Version 11.5, https://cn.string-db.org/).[29]

Results

Integrative Analyses of Cell-Type eQTL Data Prioritized Risk Genes for Schizophrenia

Bryois et al conducted an eQTL analysis using single nuclei RNA-sequencing data of 8 types of brain cells from 196 individuals, and discovered 6108 cis-eQTL genes including 2620 genes showing cell-type specific effects.[13] Their data were included in the current study. In addition, Jerber et al performed an eQTL analysis using scRNA-seq data of human iPSCs across differentiation, and we retrieved their cell-type eQTL data in mature midbrain dopaminergic neurons (from 175 donors) and serotonergic-like neurons (from 161 donors).[12] Based on the GWAS summary statistics of 69 369 schizophrenia cases and 236 642 controls,[1] we conducted SMR to examine the contribution of these cell-type eQTLs to the genetic risk of schizophrenia. To correct for the impact

of varied numbers of eGenes in different cell types on the results, we used the Bonferroni method to correct the statistical *P*-value and analyzed the cell type-specific genes for AD, which is a neurodegenerative disease.

Briefly, integrating the eQTL data in excitatory neurons identified the largest number of schizophrenia risk genes (figure 1 and Supplementary table S1, $n = 15$) that passed the significance level of SMR, among which 9 genes (GLT8D1, PBRM1, CYP7B1, FAM114A2, GATAD2A, SFMBT1, C12orf65, TRANK1, and ACTR1B) were not identified in other cell types (figure 2B), suggesting that those genes may confer schizophrenia risk by exerting functions primarily in excitatory neurons. This result is also in line with the genetic enrichment of excitatory

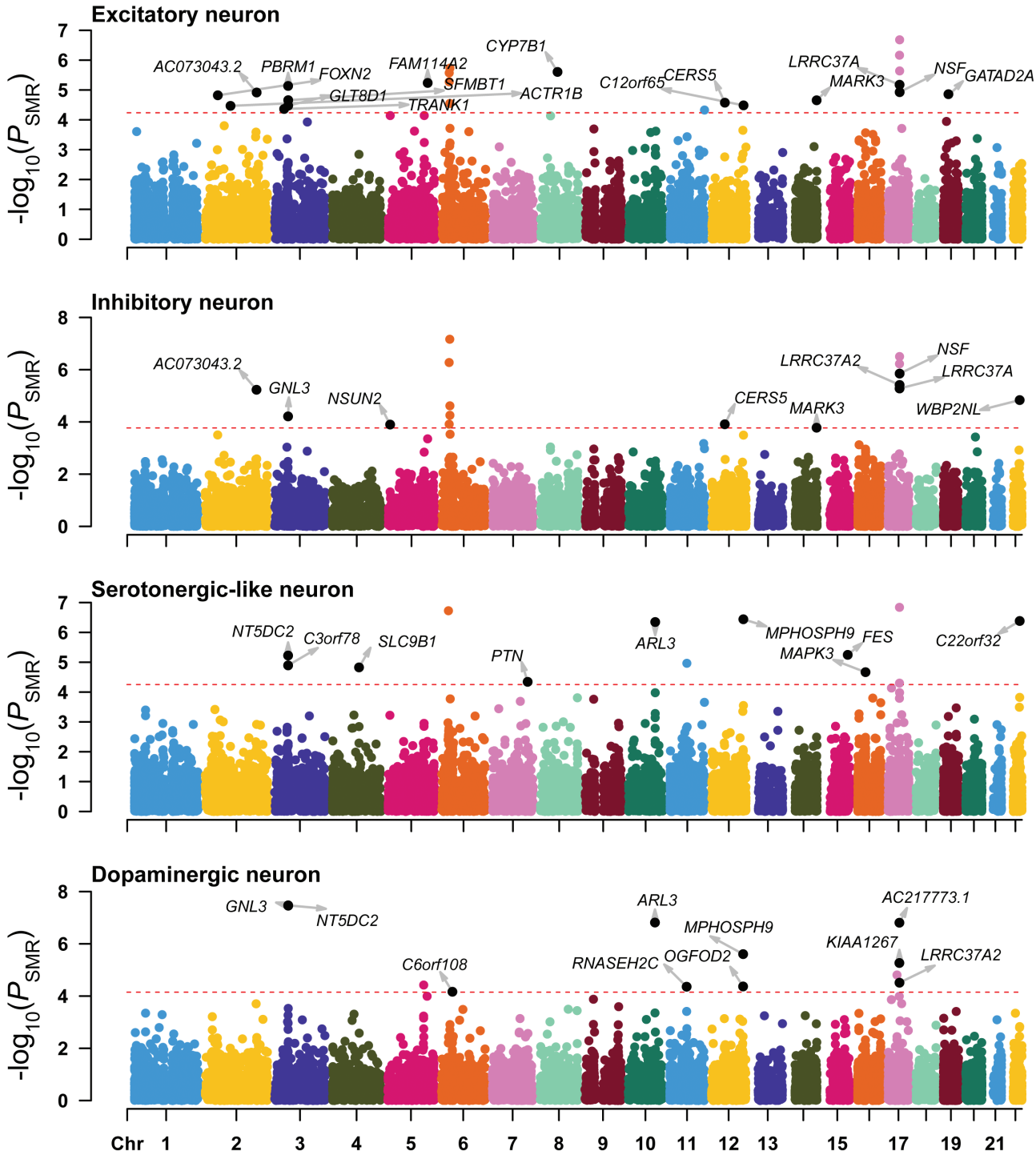

Fig. 1. Manhattan plots of SMR integrating PGC3 GWAS data with single-cell eQTL data in excitatory neuron, inhibitory neuron, serotonergic-like neuron, and dopaminergic neuron.

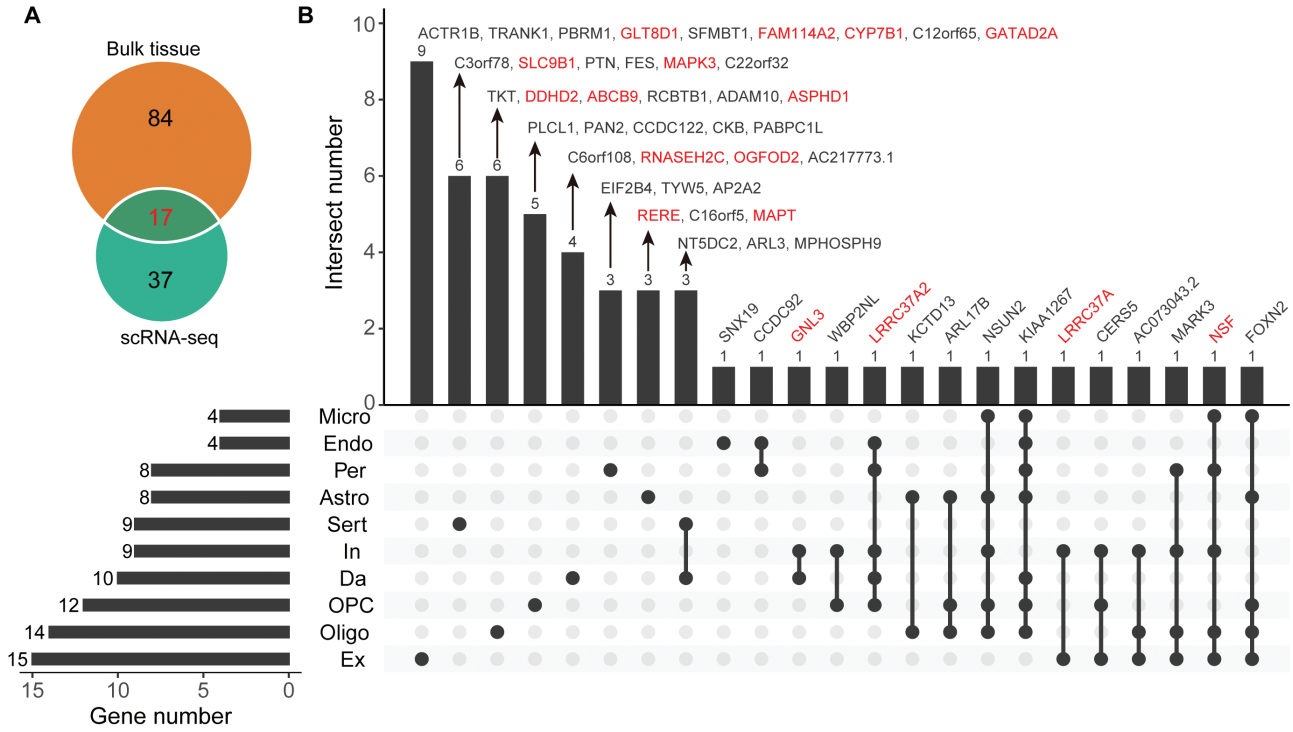

Fig. 2. (A) Venn plot of number of genes identified in bulk tissues level and single-cell level. (B) Upset plot of genes identified in different single cell types. Horizontal bar on left represents number of genes identified in each cell types. Dots and lines represent subsets of genes. Vertical histogram represents number of genes in each subset. Genes both identified in bulk tissue level and single-cell level were red marked. Micro, microglia; Endo, endothelial; Astro, astrocyte; Sert, serotonergic-like neuron; Da, dopaminergic neuron; Per, pericyte; In, inhibitory neuron; OPC, oligodendrocyte precursors; Oligo, oligodendrocyte; Ex, excitatory neuron.

neuron-specific genes in schizophrenia[30] and the previous finding that schizophrenia was primarily associated with variants in transcriptional enhancers and promoters in neurons.[31]

Integrating the eQTL data in inhibitory neurons identified 9 schizophrenia risk genes (figure 1 and Supplementary table S1), all of which were identified in at least one other cell type. Integrating the eQTL data in oligodendrocytes identified 14 schizophrenia risk genes (Supplementary figure S1 and Supplementary table S1), among which 6 genes (ASPHD1, RCBTB1, TKT, ABCB9, ADAM10, and DDHD2) were not identified in other cell types (figure 2B). Integrating the eQTL data in serotonergic-like neurons identified 9 schizophrenia risk genes (figure 1), among which 6 genes (FES, SLC9B1, C3orf78, MAPK3, PTN, and C22orf32) were not identified in other cell types. Integrating the eQTL data in mature midbrain dopaminergic neurons identified 10 schizophrenia risk genes (figure 1), among which 4 genes (C6orf108, RNASEH2C, OGFOD2, and RNASEH2C) were not identified in other cell types. Integrating the eQTL data in astrocytes identified 8 schizophrenia risk genes, among which 3 genes (MAPT, RERE, and C16orf5) were not identified in other cell types (Supplementary figure S1).

Notably, TRANK1 was highlighted in excitatory neurons only, suggesting that this gene may confer a risk of schizophrenia primarily by exerting functions in excitatory neurons, which is also consistent with a previous expression analysis showing that the function of this gene might be related to dendritic spines, synaptic plasticity, and axon guidance.[32] Another previous study also suggested that the schizophrenia GWAS risk SNP could downregulate the expression of TRANK1 in iPSC-derived neural progenitor cells, and the decreased expression of TRANK1 further perturbed the expression of many genes involved in neural development and differentiation.[33] Moreover, the reduction in TRANK1 expression could be rescued by chronic treatment with therapeutic dosages of valproic acid.[33,34] All this evidence suggests that TRANK1 might be involved in the pathogenesis of schizophrenia by affecting the functions of neurons.

By integrating the single-cell eQTL data with AD GWAS data, we identified a total of 26 genes (Supplementary table S2 and figure S2A) in ten cell types, of which 4 genes (ADAM10, ARL17B, LRRC37A, and MAPT) were also associated with schizophrenia (Supplementary figure S2B and S2C). Notably, integrating microglial eQTLs with AD GWAS data identified the most risk genes, whereas the fewest genes were identified to be associated with schizophrenia in microglia. We also found that more genes were identified to be associated with AD in glial cell types, including microglia, oligodendrocytes, and astrocytes,

suggesting that these cells were more closely related to AD. These results further confirmed distinct genetic risk architectures between AD and schizophrenia.[35–37]

Comparison of Risk Genes Derived From Cell-Type and Tissue-Level eQTL Data

Single-cell eQTL-based SMR analysis identified 54 genes showing significant associations with the genetic risk of schizophrenia in at least one cell type (figure 2A). We then compared the SMR result obtained with cell-type eQTL data and that obtained with bulk tissue eQTL data. The SMR based on bulk tissue eQTL data identified 101 risk genes by SMR by integrating PGC3 GWAS data and eQTL data from fetal brain,[4] PsychENCODE,[38] and whole blood.[39] We noticed that 17 (figure 2B, ABCB9, ASPHD1, CYP7B1, DDHD2, FAM114A2, GATAD2A, GLT8D1, GNL3, LRRC37A, LRRC37A2, MAPK3, MAPT, NSF, OGFOD2, RERE, RNASEH2C, and SLC9B1) of the 54 risk genes identified using cell-type eQTL data were also identified in SMR based on bulk tissue eQTL data; and there were 38 genes (figure 2A) that were specifically significant in single-cell data. Therefore, analyses using single-cell sequencing data could effectively complement our knowledge regarding schizophrenia risk genes.

Specific Cell Types Underlying Schizophrenia

A critical step of translating GWAS data to molecular mechanisms of schizophrenia is to determine the cell types predominantly associated with the illness. Here, we used the EWCE R package to infer schizophrenia-associated cell types. The underlying rationale is that if a particular cell type is involved in the etiology of schizophrenia, more disease-associated genes will be highly expressed in this cell type. Single-cell transcriptomics from three different studies was used to construct the specific gene expression matrix. We added the genes identified by exome sequencing because they might play a synergistic role in the pathogenesis of schizophrenia with genes identified by common variants.[1,26] The genes identified by SMR of common variants and the genes affected by rare mutations (identified through exome sequencing with FDR < 5%) were examined to test whether their expression in a given cell type was significantly higher than randomly selected background genes. In total, 54 genes identified by single-cell eQTL-based SMR and 32 genes identified by exome sequencing were analyzed (Supplementary table S3). Using scRNA-seq data from the developing human cortex published by Polioudakis et al,[27] we found that the mRNA levels of schizophrenia risk genes in maturing excitatory neurons (ExM, $P = .008$), migrating excitatory neurons (ExN, $P = .03$), and caudal ganglionic eminence interneurons (CGEIn, $P = .04$) were significantly higher than the levels of randomly selected background genes (figure 3A), suggesting the involvement of these 3 types of cells in the pathogenesis of schizophrenia. Using another scRNA-seq dataset of the developing human cerebral cortex,[11] we observed that glutamatergic neurons, which are excitatory neurons, were significantly associated with schizophrenia (figure 3B). CGEIn did not reach the statistical significance level but was marginally detected ($P = .06$).

We further expanded the EWCE analysis using scRNA-seq data in an additional study.[28] Due to the largely different brain tissues (ie, allocortex, claustrum, ganglionic eminences, hypothalamus, neocortex, and thalamus) and analytical approaches utilized, the results were partly different from those obtained with the 2 aforementioned datasets.[11,27] Specifically, the expression of schizophrenia risk genes was primarily altered in intermediate progenitors in this third dataset ($P = .004$),[28] whereas this cell type was not identified in the other 2 datasets.[11,27] Nevertheless, altered expression of such genes was also seen in neurons ($P = .03$) and interneurons ($P = .04$) (figure 3C), further highlighting the putative role of these types of cells schizophrenia.

Protein-Protein Interaction

A total of 86 genes (54 genes identified by single-cell eQTL-based SMR and 32 genes identified by exome sequencing, Supplementary table S3) were used to construct the PPI network with STRING, 30 of which could form a large network. Among the 30 genes, XPO7, GRIN2A, GRIA3, DNM3, MAGI2, and EIF2S3 were affected by rare mutations (Supplementary figure S3). These results suggest synergistic effects of genes regulated by common and rare variants on the pathogenesis of schizophrenia.

Discussion

GWASs of schizophrenia have reported hundreds of genomic regions significantly associated with the illness, yet the molecular mechanisms underlying the majority of the risk loci are relatively unclear. Translating the genetic associations with clinical diagnosis into risk genes and etiological mechanisms is an urgent task in this post-GWAS era.[40] Multiple studies have been conducted by integrating schizophrenia GWAS statistics and eQTL data in bulk tissues of certain brain regions relevant to disease pathogenesis, such as the DLPFC, hippocampus, and caudate,[2,3,41,42] and a number of risk genes have been identified. However, it is acknowledged that the regulation of many genes is cell-type and developmental-stage specific, and eQTL effects in particular cell types might be therefore shadowed in analyses based on data of bulk tissues. Specifically, when a gene is associated with the genetic risk of schizophrenia only in certain cell types, it cannot be detected through eQTL analyses based on data of bulk tissues.

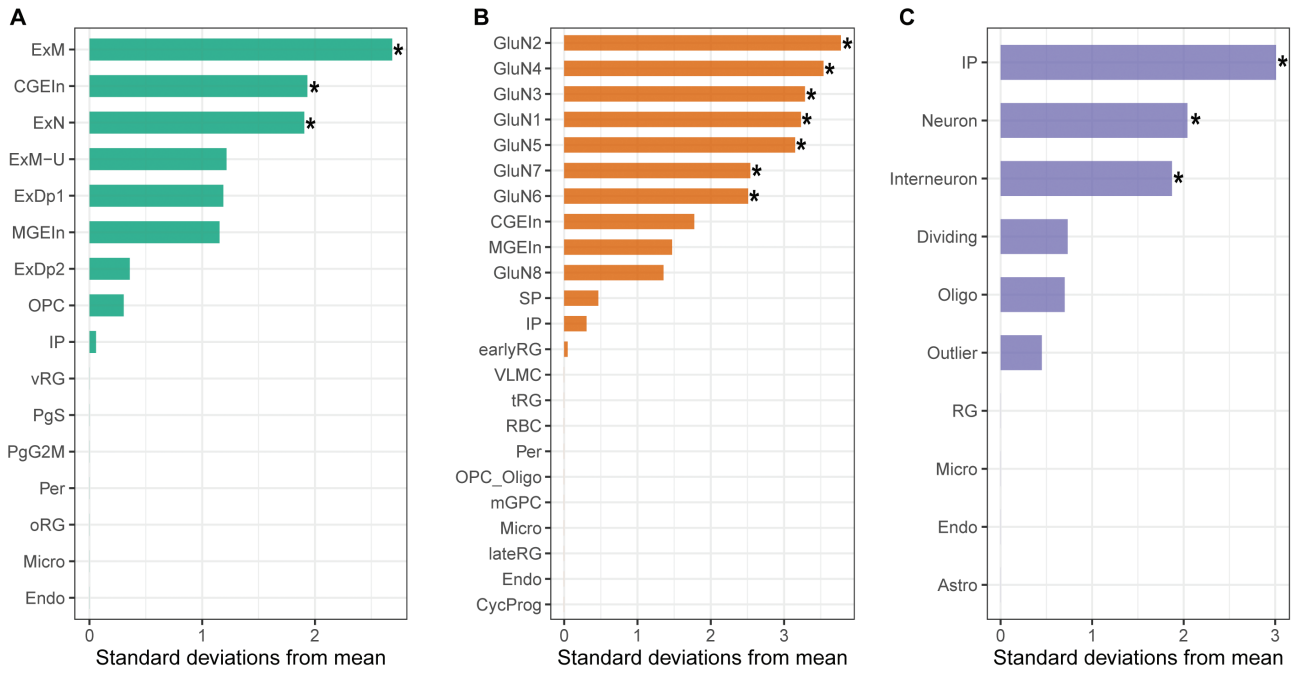

Fig. 3. Bootstrapping tests performed using the EWCE method show that schizophrenia risk genes enriched expressed in several cell types using data from (A) Polioudakis et al, (B) Trevino et al, (C) Bhaduri et al ExM, Maturing excitatory; CGEIn, caudal ganglionic eminence interneurons; ExN, Migrating excitatory; ExM-U, Maturing excitatory upper enriched; ExDp1, Excitatory deep layer 1; MGEIn, medial ganglionic eminence interneurons; ExDp2, Excitatory deep layer 2; IP, intermediate progenitors; vRG, ventricular radial glia; PgS, cycling progenitors (S phase); PgG2M, cycling progenitors (G2/M phase); oRG, outer radial glia; GluN, glutamatergic neuron; SP, subplate; early RG, early radial glia; VLMC, leptomeningeal cells; tRG, truncated radial glia; RBC, red blood cells; OPC_oligo, OPC and oligodendrocyte; mGPC, multipotent glial progenitor; lateRG, late radial glia; CycProg, cycling progenitors.

Thanks to the improvement of scRNA-seq analysis, several studies have been published in recent years to reveal cell-type specific eQTL effects in brains, which have greatly promoted our knowledge regarding gene expression profiles in particular brain cells. Notably, previous single-cell transcriptomic analyses revealed that schizophrenia genomic risk loci were primarily mapped to pyramidal cells, medium spiny neurons, and certain interneurons, whereas far fewer were mapped to embryonic, progenitor, or glial cells.[30,35] A more recent study showed that schizophrenia GWAS loci were enriched in deep-layer excitatory neurons, cortico-cortical projection neurons, and Parvalbumin-expressing basket interneurons, suggesting potentially pivotal roles of these cells in the genetic effects in the illness.[43] In addition, a study found that open chromatin regions in glutamatergic neurons were enriched for schizophrenia risk loci.[44] Based on these findings, we herein investigated whether schizophrenia genetic risk influenced gene expression in particular brain cells, such as excitatory neurons, inhibitory neurons, and oligodendrocytes.

Our data have implications in guiding future mechanistic studies. For example, genetic variants near TRANK1 have shown genome-wide significant associations with the risk of schizophrenia, but previous integrative analyses in bulk brain tissues have not identified this gene. In the current study, we found that the expression of TRANK1 was only associated with the genetic risk of schizophrenia in excitatory neurons. This result highlights that this gene likely exert its function only in certain cell types, and provides clues that future functional studies on this gene may focus on excitatory neurons and relevant neural circuits.

This study has certain limitations. First, although we performed stringent Bonferroni correction on the SMR results according to the number of eGenes in each cell type, we did not further correct for the number of cell types. Despite we have applied the same correction criteria following previous studies[1,45] and our intention was to identify risk genes that were associated with schizophrenia in either single cells or multiple cell types, we acknowledge that this is a potential limitation, which might lead to false positive results, and further replications using independent datasets are warranted. Second, a total of 5 scRNA-seq datasets were used in this study,[11–13,27,28] and we assumed that the single-cell clustering and annotation were relatively consistent and correct across different studies. Nonetheless, different studies might apply different clustering parameters or methods, and even the markers used to annotate cell types were also different. The heterogeneity and batch effects between different studies should be therefore considered in interpreting the results. In addition, we

should not ignore the bias due to single-cell acquisition from different developmental stages and different tissue origins, as brain development is a highly dynamic process in the early stage.[46,47] Third, unlike bulk tissue RNA-seq, which could detect the expression of tens of thousands of genes in a single tissue, only thousands of genes, which was just a drop in the bucket, could be captured and detected at the single-cell level. Therefore, SMR analysis at the single-cell level might not be able to detect some risk genes with lower levels of expression, which might (at least partially) explain the fact that large numbers of genes detected at the bulk tissue level with SMR could not be replicated at the single-cell level. Fourth, limited by the current scRNA-seq technologies, data analysis methods and sample acquisition, SMR analysis and EWCE cell type analysis were performed in only a few cell types. While the current studies are crucial steps toward a better understanding of schizophrenia pathogenesis, we believe that further development of scRNA-seq technology and data analysis methods will be necessary to obtain more information regarding the cell types involved in this illness.

In summary, by integrating scRNA-seq data in multiple brain cells, we revealed many novel schizophrenia risk genes that were not identified in previous analyses using bulk tissues. Our study also provides evidence supporting the putative roles of excitatory neurons and CGEIn in the pathogenesis of schizophrenia. These data suggest the necessity and importance of utilizing expression profiling in particular cell types to elucidate the genetic mechanism of schizophrenia, and further fine-scale mapping analyses are necessary.

Supplementary Material

Supplementary material is available at https://academic.oup.com/schizophreniabulletin/.

Funding

This work was supported by grants from National Natural Science Foundation of China (82222024 and 82071275 to X.X., 82201661 to Y.W., 32100482 to L.W.), the Yunnan Fundamental Research Projects (202101AW070020 and 202101AT070348 to X.X., 202101AT070283 and 202101AT070359 to L.W.), Health Commission of Hubei Province scientific research project (WJ2021M025 to Y.W.), Wuhan Municipal Health Commission Scientific research project (WX20Q02 to Y.W.), Spring City Plan: the High-level Talent Promotion and Training Project of Kunming (2022SCP001), and the Open Research Fund (AMHD-2021-1 and AMHD-2022-1). X.X. was also supported by the Chinese Academy of Sciences Western Light Program, and Youth Innovation Promotion Association, CAS.

Conflict of Interest

The authors have no conflicts of interest to declare.

Author Contributions

Y.W., X.X., and Y.L. oversaw the project, conceived and designed the study, drafted the first version of the manuscript. Y.W. and C.-Y.Z. performed the primary analysis, L.W. provided assistance during the analysis. All authors revised the manuscript critically and approved the final version.

References

1. Trubetskoy V, Pardinas AF, Qi T, *et al*. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature.* 2022;604(7906):502–508.
2. Collado-Torres L, Burke EE, Peterson A, *et al*. Regional heterogeneity in gene expression, regulation, and coherence in the frontal cortex and hippocampus across development and schizophrenia. *Neuron.* 2019;103(2):203–216.e8.
3. Jaffe AE, Straub RE, Shin JH, *et al*; Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. *Nat Neurosci.* 2018;21(8):1117–1125.
4. O'Brien HE, Hannon E, Hill MJ, *et al*. Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders. *Genome Biol.* 2018;19(1):194.
5. GTEx C. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science.* 2020;369(6509):1318–1330.
6. Yao DW, O'Connor LJ, Price AL, Gusev A. Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nat Genet.* 2020;52(6):626–633.
7. Armand EJ, Li J, Xie F, Luo C, Mukamel EA. Single-cell sequencing of brain cell transcriptomes and epigenomes. *Neuron.* 2021;109(1):11–26.
8. Lake BB, Ai R, Kaeser GE, *et al*. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science.* 2016;352(6293):1586–1590.
9. Nowakowski TJ, Bhaduri A, Pollen AA, *et al*. Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science.* 2017;358(6368):1318–1323.
10. Aygun N, Elwell AL, Liang D, *et al*. Brain-trait-associated variants impact cell-type-specific gene regulation during neurogenesis. *Am J Hum Genet.* 2021;108(9):1647–1668.
11. Trevino AE, Muller F, Andersen J, *et al*. Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. *Cell.* 2021;184(19):5053–5069.e23.
12. Jerber J, Seaton DD, Cuomo ASE, *et al*. Population-scale single-cell RNA-seq profiling across dopaminergic neuron differentiation. *Nat Genet.* 2021;53(3):304–312.
13. Bryois J, Calini D, Macnair W, *et al*. Cell-type-specific cis-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders. *Nat Neurosci.* 2022;25(8):1104–1112.
14. Bellenguez C, Kucukali F, Jansen IE, *et al*. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet.* 2022;54(4):412–436.

15. La Manno G, Gyllborg D, Codeluppi S, *et al*. Molecular diversity of midbrain development in mouse, human, and stem cells. *Cell.* 2016;167(2):566–580.e19.
16. Arenas E, Denham M, Villaescusa JC. How to make a midbrain dopaminergic neuron. *Development.* 2015;142(11):1918–1936.
17. Welch JD, Kozareva V, Ferreira A, Vanderburg C, Martin C, Macosko EZ. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell.* 2019;177(7):1873–1887.e17.
18. Cuomo ASE, Seaton DD, McCarthy DJ, *et al*. Single-cell RNA-sequencing of differentiating iPS cells reveals dynamic genetic effects on gene expression. *Nat Commun.* 2020;11(1):810.
19. Cain A, Taga M, McCabe C, *et al*. Multi-cellular communities are perturbed in the aging human brain and with Alzheimer's disease. *bioRxiv [Preprint]*. 2020. doi:10.1101/2020.12.22.424084.
20. Zhou Y, Song WM, Andhey PS, *et al*. Human and mouse single-nucleus transcriptomics reveal TREM2-dependent and TREM2-independent cellular responses in Alzheimer's disease. *Nat Med.* 2020;26(1):131–142.
21. Mathys H, Davila-Velderrain J, Peng Z, *et al*. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature.* 2019;570(7761):332–337.
22. Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics.* 2016;32(10):1479–1485.
23. Zhu Z, Zhang F, Hu H, *et al*. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet.* 2016;48(5):481–487.
24. 1000 Genomes Project Consortium; Auton A, Brooks LD, Durbin RM, *et al*. A global reference for human genetic variation. *Nature.* 2015;526(7571):68–74.
25. Skene NG, Grant SGN. Identification of vulnerable cell types in major brain disorders using single cell transcriptomes and expression weighted cell type enrichment. *Front Neurosci.* 2016;10:16.
26. Singh T, Poterba T, Curtis D, *et al*. Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature.* 2022;604(7906):509–516.
27. Polioudakis D, de la Torre-Ubieta L, Langerman J, *et al*. A single-cell transcriptomic atlas of human neocortical development during mid-gestation. *Neuron.* 2019;103(5):785–801.e8.
28. Bhaduri A, Sandoval-Espinosa C, Otero-Garcia M, *et al*. An atlas of cortical arealization identifies dynamic molecular signatures. *Nature.* 2021;598(7879):200–204.
29. Szklarczyk D, Gable AL, Nastou KC, *et al*. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 2021;49(D1):D605–D612.
30. Skene NG, Bryois J, Bakken TE, *et al*. Genetic identification of brain cell types underlying schizophrenia. *Nat Genet.* 2018;50(6):825–833.
31. Nott A, Holtman IR, Coufal NG, *et al*. Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science.* 2019;366(6469):1134–1139.
32. Li W, Cai X, Li HJ, *et al*. Independent replications and integrative analyses confirm TRANK1 as a susceptibility gene for bipolar disorder. *Neuropsychopharmacology.* 2021;46(6):1103–1112.
33. Jiang X, Detera-Wadleigh SD, Akula N, *et al*. Sodium valproate rescues expression of TRANK1 in iPSC-derived neural cells that carry a genetic variant associated with serious mental illness. *Mol Psychiatry.* 2019;24(4):613–624.
34. Chen DT, Jiang X, Akula N, *et al*. Genome-wide association study meta-analysis of European and Asian-ancestry samples identifies three novel loci associated with bipolar disorder. *Mol Psychiatry.* 2013;18(2):195–205.
35. Olislagers M, Rademaker K, Adan RAH, Lin BD, Luykx JJ. Comprehensive analyses of RNA-seq and genome-wide data point to enrichment of neuronal cell type subsets in neuropsychiatric disorders. *Mol Psychiatry.* 2022;27(2):947–955.
36. Brainstorm C, Anttila V, Bulik-Sullivan B, *et al*. Analysis of shared heritability in common disorders of the brain. *Science.* 2018;360(6395):eaap8757.
37. Bryois J, Skene NG, Hansen TF, *et al*. Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson's disease. *Nat Genet.* 2020;52(5):482–493.
38. Gandal MJ, Zhang P, Hadjimichael E, *et al*. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science.* 2018;362(6420):eaat8127.
39. Vosa U, Claringbould A, Westra HJ, *et al*. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet.* 2021;53(9):1300–1310.
40. Guan F, Ni T, Zhu W, *et al*. Integrative omics of schizophrenia: from genetic determinants to clinical classification and risk prediction. *Mol Psychiatry.* 2022;27(1):113–126.
41. Benjamin KJM, Chen Q, Jaffe AE, *et al*. Analysis of the caudate nucleus transcriptome in individuals with schizophrenia highlights effects of antipsychotics and new risk genes. *Nat Neurosci.* 2022;25(11):1559–1568.
42. Jaffe AE, Hoeppner DJ, Saito T, *et al*. Profiling gene expression in the human dentate gyrus granule cell layer reveals insights into schizophrenia and its genetic risk. *Nat Neurosci.* 2020;23(4):510–519.
43. Ruzicka WB, Mohammadi S, Davila-Velderrain J, Subburaju S, Tso DR, Hourihan M, Kellis M. Single-cell dissection of schizophrenia reveals neurodevelopmental-synaptic axis and transcriptional resilience. *medRxiv [Preprint].* 2020. doi:10.1101/2020.11.06.20225342.
44. Hauberg ME, Creus-Muncunill J, Bendl J, *et al*. Common schizophrenia risk variants are enriched in open chromatin regions of human glutamatergic neurons. *Nat Commun.* 2020;11(1):5581.
45. Pavlides JM, Zhu Z, Gratten J, McRae AF, Wray NR, Yang J. Predicting gene targets from integrative analyses of summary data from GWAS and eQTL studies for 28 human complex traits. *Genome Med.* 2016;8(1):84.
46. Di Lullo E, Kriegstein AR. The use of brain organoids to investigate neural development and disease. *Nat Rev Neurosci.* 2017;18(10):573–584.
47. Gilmore JH, Knickmeyer RC, Gao W. Imaging structural and functional brain development in early childhood. *Nat Rev Neurosci.* 2018;19(3):123–137.