Biological Psychiatry: Celebrating 50 Years

Archival Report

A Longitudinal Model of Human Neuronal Differentiation for Functional Investigation of Schizophrenia Polygenic Risk

Anil P.S. Ori, Merel H.M. Bot, Remco T. Molenhuis, Loes M. Olde Loohuis, and Roel A. Ophoff

ABSTRACT

BACKGROUND: Common psychiatric disorders are characterized by complex disease architectures with many small genetic effects that contribute and complicate biological understanding of their etiology. There is therefore a pressing need for in vitro experimental systems that allow for interrogation of polygenic psychiatric disease risk to study the underlying biological mechanisms.

METHODS: We have developed an analytical framework that integrates genome-wide disease risk from genome-wide association studies with longitudinal in vitro gene expression profiles of human neuronal differentiation.

RESULTS: We demonstrate that the cumulative impact of risk loci of specific psychiatric disorders is significantly associated with genes that are differentially expressed and upregulated during differentiation. We find the strongest evidence for schizophrenia, a finding that we replicate in an independent dataset. A longitudinal gene cluster involved in synaptic function primarily drives the association with schizophrenia risk.

CONCLUSIONS: These findings reveal that in vitro human neuronal differentiation can be used to translate the polygenic architecture of schizophrenia to biologically relevant pathways that can be modeled in an experimental system. Overall, this work emphasizes the use of longitudinal in vitro transcriptomic signatures as a cellular readout and the application to the genetics of complex traits.

Keywords: Genome-wide disease risk, Neuronal stem cells, Polygenicity, Psychiatric disorders, Schizophrenia, Synaptic function

https://doi.org/10.1016/j.biopsych.2018.08.019

Major psychiatric disorders feature a high heritability but have a largely unknown etiology (1,2). The increasing sample sizes of genome-wide association studies (GWASs) successfully result in identification of more susceptibility loci for these disorders (3). A major challenge is to understand and interpret the cumulative impact of many loci that collectively contribute to psychiatric disease risk, and how to translate this complex polygenic architecture to biological pathways that drive the underlying molecular and cellular disease processes. Lack of applicable in vitro model systems and a framework to study polygenic psychiatric risk hinders the translation of genetics findings to disease biology (4).

Early brain development has been implicated in psychiatric disorders such as schizophrenia (SCZ) (5–8), autism spectrum disorder (9,10), and self-reported depression (SRD) (11). Differentiation of human embryonic stem cells into neuronal lineages has been demonstrated to hold great promise for modeling early brain development (12–14) and may thus offer a unique opportunity to study psychiatric disease biology in vitro. However, it has remained unclear whether the molecular dynamics underlying in vitro human neuronal differentiation are associated with polygenic psychiatric disease susceptibility.

We set out to investigate in vitro human neuronal differentiation in the context of polygenic psychiatric disease risk. To accomplish this, we performed a densely sampled time series experiment and robustly detected transcriptome-wide changes across neuronal differentiation. To study the aggregate impact of risk loci, we integrated longitudinal in vitro gene expression signatures with GWAS summary statistics of major psychiatric disorders. We observed significant enrichment of genetic risk for multiple disorders in genes that are upregulated across differentiation. We further show that this effect is strongest for SCZ and primarily driven by a longitudinal gene cluster that is involved in synaptic functioning. These findings support the use of in vitro neuronal differentiation as a promising model system to study genetic psychiatric risk, particularly in the context of SCZ.

METHODS AND MATERIALS

Approval for Stem Cell Research

This study and all described work were approved by the University of California, Los Angeles, Embryonic Stem Cell Research Oversight Committee.

SEE COMMENTARY ON PAGE 532

 ISSN: 0006-3223

In Vitro Human Neuronal Differentiation

WA09(H9)-derived human neural stem cells were commercially obtained (Thermo Fisher Scientific, Waltham, MA) as neural progenitors and were subsequently expanded as adherent culture according to the manufacturer's guidelines. Low-passage human neural stem cells (<4 passage rounds) were plated in 12-well plates coated with poly-D-lysine (0.1 mg/mL) (VWR, Radnor, PA) and laminin (4.52 $\mu g/cm^2$) (Corning, Corning, NY) at 1.5×10^5 cells, which were equally distributed and subsequently cultured in expansion medium as described above. After 24 hours of proliferation, media were changed to neuronal differentiation media consisting of Neurobasal Medium (Thermo Fisher Scientific), 2% B-27 Serum-Free Supplement (Thermo Fisher Scientific), 2-mM GlutaMax-I Supplement, 0.05-mM β-mercaptoethanol (Thermo Fisher Scientific), and 1× penicillin-streptomycin. Media were changed every 2 or 3 days.

Experimental Design and Assessment of Gene Expression

Human neural stem cells were differentiated over a course of 30 days, and RNA were harvested at seven time points (days 0, 2, 5, 10, 15, 20, and 30) in triplicates or quadruplicates ($n = 24$). Genome-wide array-based transcriptome data were collected at the University of California, Los Angeles, Neuroscience Genomics Core using Illumina's HumanHT-12 v4 Expression BeadChip Kit (Illumina, San Diego, CA).

Data Preprocessing and Quality Control

Gene expression data were extracted using the Gene Expression Module in GenomeStudio Software 2011.1 (Illumina). Data were background corrected with subsequent variance-stabilizing transformation, and robust spline normalization was applied (15,16). We excluded low-quality probes and subsequently performed sample outlier detection by Euclidean distance and standardized connectivity. The FactoMineR package (v1.28) in R, version R-3.3.3 (R Foundation for Statistical Computing, Vienna, Austria) was used to perform principal component analysis. For subsequent downstream analyses, we used the normalized expression values of 19,012 high-quality filtered probes for all 24 samples.

Transcriptome-Based In Vitro Cellular Identity

To investigate in vitro cellular identity across differentiation, we used transcriptomic signatures of cell type–specific genes of seven main cell types identified in the mouse cerebral cortex (17). We extracted normalized gene expression values of these genes for each cell type from our own in vitro dataset and calculated mean standardized expression levels of cell type–specific genes for each of the seven cell types across days of differentiation.

Transition Mapping to a Spatiotemporal Atlas of Early Human Brain Development

To investigate global transcriptomic matching between in vitro gene expression profiles and in vivo gene expression profiles of neocortical brain regions, we applied transition mapping, which is implemented in the online CoNTExT bioinformatic pipeline (https://context.semel.ucla.edu) (14). Analyses were run for in vitro time points day 0 versus day 30, day 0 versus day 5, day 5 versus day 15, and day 15 versus day 30 across both temporal and spatial dimensions of human cortical development.

Time Series Differential Gene Expression and Cluster Analysis

Two multivariate empirical Bayes models were used to identify differentially expressed genes across differentiation. We computed the one-sample T^2 statistic and a probability of being differentially expressed using the mb.long() function in the Timecourse package (v 1.42) and the betr() function in the BETR package (version 1.26) in R, respectively (18,19). As both methods rank probes by their differential expression over time, differentially expressed genes were classified as the union of the set of probes with a probability of 1.0 using BETR and an equally sized set of top ranked probes using the T^2 statistic. We subsequently applied fuzzy c-means clustering to all differentially expressed probes and computed cluster membership values using the fclusList() and membership() function in the Mfuzz package in R (20,21). Clusters were annotated using Database for Annotation, Visualization, and Integrated Discovery (DAVID) (v6.8) (22) and probes with a membership >0.5.

Integration of GWAS Data With In Vitro Transcriptomic Signatures

Illumina probe IDs were mapped to Ensembl gene identifiers (IDs) using National Center for Biotechnology Information build 37.3, duplicate IDs were removed, and gene boundaries were extended symmetrically by 10 kb to include regulatory regions. Annotation files were then created mapping each gene ID or chromosomal position with in vitro gene parameters of interest, such as T^2 statistic and cluster membership values. These files were then used as input to MAGMA and stratified linkage disequilibrium (LD) score regression (sLDSR) to integrate in vitro signatures with GWAS data and study the cumulative impact across risk loci.

GWAS Summary Statistics and Ancestry-Matched Reference Panels. GWAS summary statistics were obtained for SCZ (23), major depressive disorder (MDD) (24), SRD (11), bipolar disorder (25), autism spectrum disorder (26), attention-deficit/hyperactivity disorder (ADHD) (27), cross disorder (28), Alzheimer's disease (AD) (29), and adult human height (30) (Table S2 in Supplement 1). For each trait we used the most recent GWAS summary statistics that was publicly available at the time of the analysis. The 1000 Genomes Project phase 3 release was used as reference panel to model ancestry-matched LD (31).

MAGMA Gene-Set Analysis. MAGMA (v1.06) (32) was used to perform gene-set analyses of GWAS data. MAGMA uses a multiple regression framework to associate a continuous or binary gene variable to GWAS gene-level p values. For each GWAS phenotype, we generated gene-level p values by computing the mean single nucleotide polymorphism (SNP) association using the default gene model (snp-wise=mean) with ±10-kb extensions of gene boundaries and SNPs with minor allele frequency >5%. For each annotation, we then regressed gene-level GWAS test statistics on the

corresponding gene annotation variable using the –gene-covar function while adjusting for gene size, SNP density, and LD-induced correlations (–model correct=all), which is estimated from an ancestry-matched 1000 Genomes Project phase 3 release reference panel. Testing only for a positive association, i.e., enrichment of GWAS signal, we report one-sided *p* values along with the corresponding regression coefficient.

Stratified LD Score Regression. We applied an extension to sLDSR, a statistical method that partitions SNP-based heritability (h^2) from GWAS summary statistics (8). This allows us to quantify the effects of continuous-valued annotations on the heritability (33). For each annotation, we first estimated partitioned LD scores using the ldsc.py –l2 function with minor allele frequency >5%, a 1-cM window, and an ancestry-match 1000 Genomes Project phase 3 release reference panel. We ran sLDSR (ldsc.py –h2) for each annotation of interest while accounting for the full baseline model, as recommended by the developers (8,33), and an extra annotation of all genes detected in our in vitro model (n = 12,414). As we only tested for a positive association, we report the contribution to the per-SNP h^2 (t) and the associated one-sided *p* value, which is

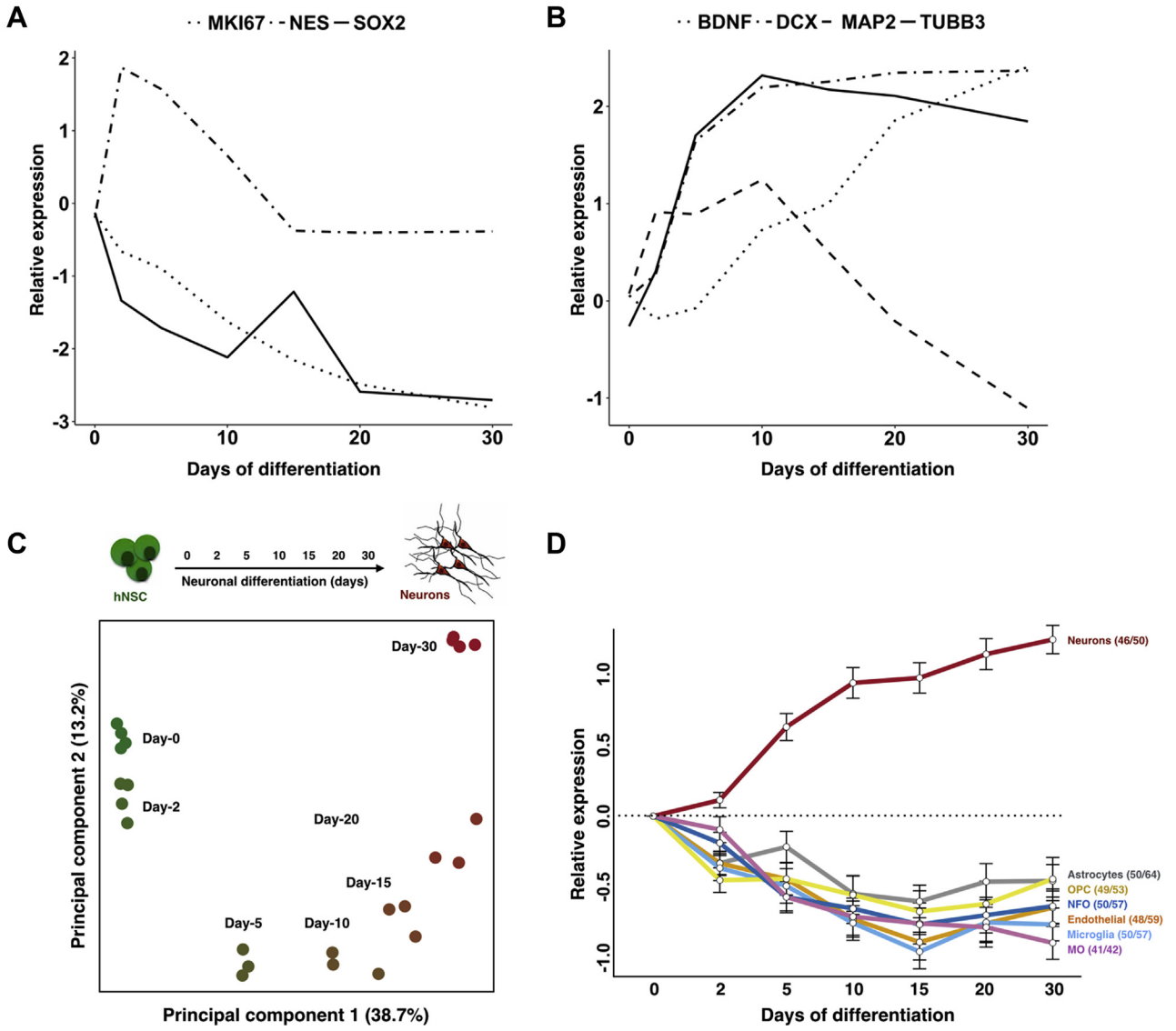

Figure 1. In vitro gene expression profiles confirm a neuron-specific differentiation process. Relative gene expression of **(A)** traditional stem cell and **(B)** neuronal markers plotted across days of differentiation. **(C)** Principal component analysis of in vitro transcriptomic data with principal component 1 (x-axis) and principal component 2 (y-axis) visualized. Variance explained per component is shown in parentheses. **(D)** Transcriptome-based cellular identity is shown by average expression of cell type–specific genes across days of differentiation. The first number in parentheses represents the number of genes for which the average expression is plotted. The second number represents the corresponding number of probes assayed. hNSC, human neural stem cell; MAP2, microtubule associated protein 2; MK167, marker of proliferation Ki-67; MO, myelinating oligodendrocytes; NFO, newly formed oligodendrocytes; OPC, oligodendrocyte precursor cells.

calculated using standard errors that are obtained via a block jackknife procedure (8,34).

Further details on experimental methods and statistical analyses are available in Supplemental Methods in Supplement 1.

RESULTS

Longitudinal In Vitro Gene Expression Profiling Confirms Neuron-Specific Differentiation and Matches In Vivo Human Cortical Development

To study the molecular dynamics underlying in vitro human neuronal differentiation, we differentiated a human neural stem cell line (WA09/H9) to a neuronal lineage across 30 days. Genome-wide gene expression profiles were assayed densely at seven time points in at least triplicates (n = 24 samples). To verify that the data were in agreement with the intended differentiation protocols, we investigated specific gene expression signatures over time. We first examined gene expression patterns of traditional gene markers (35,36) and observed that neural stem cell and proliferation markers (*MKI67*, *NES*, and *SOX2*) are downregulated, while early neuronal markers (*BDNF* and *DCX*) are upregulated as differentiation progresses (Figure 1A, B). *MAP2*, a more mature neuronal marker (35,37), is first upregulated and then subsequently downregulated at later time points, suggesting that the differentiated culture maintains a relatively immature neuronal identity. Next, we explored principal component analysis on normalized gene expression values using the full transcriptome and found a large proportion of the variance in expression to be explained by the differentiation process, with minimal effects of technical variation (Figure 1C, Figure S1 in Supplement 1). Investigation of transcriptome-based cell type–specific gene expression signatures of major classes of cell types in the cerebral cortex shows that relative neuronal gene expression increases as neuronal differentiation progresses over time (Figure 1D). There is no evidence of glial- or endothelial-specific gene expression, which confirms a broadly neuronal in vitro cellular identity.

Having established that the in vitro differentiation process is predominantly neuronal, we applied transition mapping to assess the correspondence of longitudinal in vitro transcriptome data to in vivo signatures of both brain developmental stages and laminae of the human neocortex. We find significant matching between the in vitro longitudinal differential gene expression profiles (day 0 vs. day 30) and in vivo developmental stage from 4 weeks postconception (PCW) to 24 PCW (Figure S2 in Supplement 1). This overlaps with the primary period of neurogenesis in the neocortex, which starts around 6 PCW (38,39). To gain more insight into this overlap, we partitioned the transition mapping analyses into three comparisons and examined how in vitro to in vivo matching

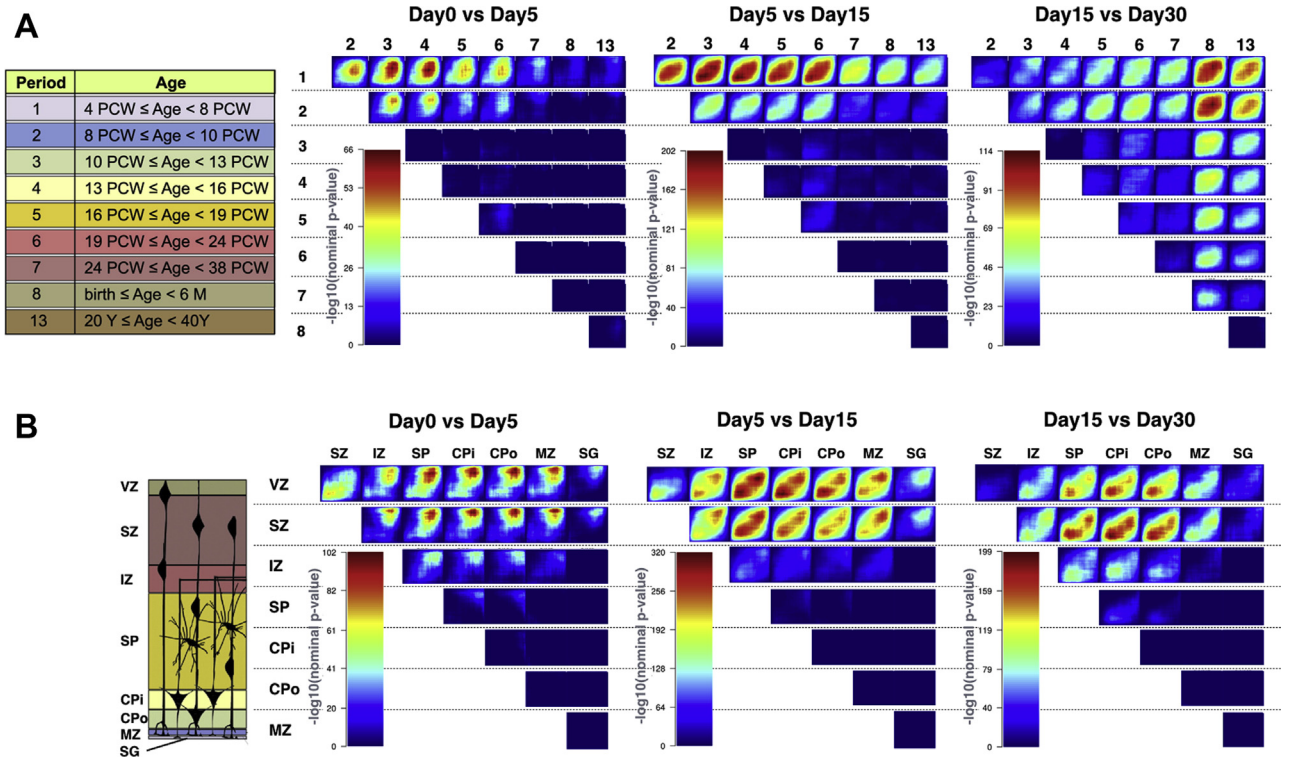

Figure 2. In vitro gene expression profiles match in vivo human cortical development. Transition mapping output visualizes the amount of overlap between in vitro and in vivo differential gene expression profiles colored by –log10 (p value) (see Figure S2 in Supplement 1 for more details on interpretation). Note that p values are shown on varying color scales between graphs. Abbreviations and numbering above maps correspond to schematic representations on the left [adapted with permission from Stein *et al.* (14)] of different **(A)** developmental stages and **(B)** laminae. CPi, inner cortical plate; CPo, outer cortical plate; IZ, intermediate zone; M, months; MZ, marginal zone; PCW, weeks postconception; Period, developmental stage; SG, subpial granular layer; SP, subplate zone; SZ, subventricular zone; VZ, ventricular zone; Y, years.

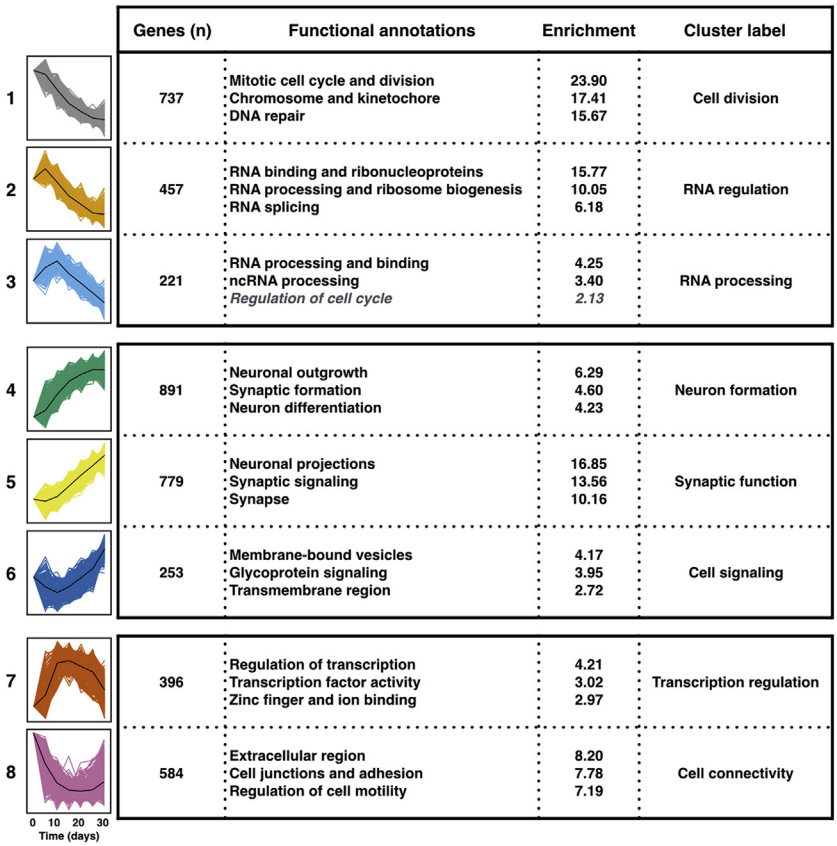

Figure 3. Identified gene clusters highlight biological pathways important for neuronal differentiation. Top significant functional annotations and corresponding enrichment score are shown for each gene cluster. Longitudinal gene expression is visualized for high member genes only (black line represents mean gene expression). Each cluster is color-coded with the number of genes at membership >0.5 denoted. See Table S1 in Supplement 2 for full annotation results. ncRNA, noncoding RNA.

progressed over time across differentiation. We see a clear progression in matching from early developmental stages to later stages (Figure 2A). For example, in vitro day 0 versus day 5 shows strong overlap with in vivo period 1 (4–8 PCW) versus period 4 (13–16 PCW), while in vitro day 15 versus day 30 shows stronger overlap with in vivo period 2 (8–10 PCW) versus period 8 (birth to 6 months of age). Similarly, in vitro longitudinal DGE shows progression from overlap of early time points with inner laminae, to overlap with more upper cortical layers as in vitro neuronal differentiation advances (Figure 2B, Figure S2 in Supplement 1).

In Vitro Neuronal Differentiation Reveals Specific Longitudinal Gene Clusters

To identify biological pathways associated with neuronal differentiation, we applied an analysis framework specifically tailored to time series gene expression data (see Methods and Materials as well as Supplemental Methods and Materials in Supplement 1). A total of 7734 probes, mapping to 5818 genes, were differentially expressed over time (Figure S3 in Supplement 1). We found that these genes are, on average, more constrained to genetic variation than nondifferentially expressed genes (section S2) are. Using only differentially expressed probes, we next applied fuzzy c-means clustering and identified eight distinct longitudinal gene clusters (Figure 3, Figure S4 in Supplement 1). For each probe, we generated a corresponding cluster membership value, representing the degree to which a gene belongs to a cluster. To identify the most informative biological interpretation of each cluster, we analyzed genes with high cluster membership for enrichment of functional annotations using DAVID (Supplemental Methods and Table S1 in Supplement 2).

First, we identified three clusters with decreasing gene expression over time that are significantly enriched for cell division and for RNA regulation and processing genes, reflective of stem cell proliferation and cell fate determination that is tightly controlled and regulated by RNA-dependent processes (40). Second, there are three clusters showing increased gene expression levels over time that are primarily enriched for neuronal processes, such as neuron formation and synaptic function. Another independent cluster shows an inverted U-shaped expression pattern during development, enriched for genes involved in transcriptional regulation. The final cluster is enriched for genes involved in extracellular region and cell adhesions. These processes are important for cell connectivity and have also been implicated in cell proliferation and neuronal migration (41,42). Together, these eight gene clusters reveal different biological mechanisms that are associated with neuronal differentiation and consistent with known biology of neurodevelopment. We hypothesize that the study of these longitudinal gene expression clusters can help decipher disease mechanisms involved in psychiatric phenotypes.

Differentially Expressed Genes Are Enriched for Polygenic Psychiatric Disease Risk

To examine how aggregate psychiatric disease risk is distributed across genes that are important for neuronal differentiation, we applied gene set analysis and partitioning of h^2 with MAGMA and sLDSR, respectively. We used GWAS summary statistics from major psychiatric disorders in addition to AD and adult human height, which served as nonpsychiatric control phenotypes that are heritable and polygenic. Using a two-step approach, we first investigated disease susceptibility on overall differential expression level and subsequently proceeded to deconstruct these associations across the longitudinal gene clusters. We found that genes that are differentially expressed are enriched for genetic risk of multiple psychiatric disorders. We found significant effects with MAGMA for SCZ (p = .001), ADHD (p = .002), and SRD (p = .003) (Table 1, Table S3 in Supplement 2). With sLDSR, we found nominally significant effects for SCZ (p = .01) and SRD (p = .02) and a suggestive association for ADHD (p = .06) (Table 1, Table S4 in Supplement 2). We observed a suggestive enrichment for bipolar disorder, and no enrichment for cross disorder, autism spectrum disorder, and MDD (24), or for adult height and AD.

We next investigated whether enrichment across differentially expressed genes was driven by up- or downregulation of genes during differentiation. For SCZ, we found that the effect is driven by genes that are upregulated (MAGMA $p = 5.0 \times 10^{-7}$, sLDSR $p = 6.1 \times 10^{-5}$) and not by genes that are downregulated (MAGMA p = .98, sLDSR p = .61) (Figure 4, Figure S6 in Supplement 1). For SRD, we found only a stronger enrichment in upregulated genes with MAGMA ($p = 3.5 \times 10^{-4}$), while ADHD shows no specific evidence for either up- or downregulated genes.

Psychiatric Disease Risk Aggregates to Specific Longitudinal Gene Clusters

Next, we explored the relationship between differentially expressed genes and disease risk on the cluster level. For this analysis, we included only traits that show significant disease enrichment across differentially expressed genes using MAGMA after correcting for multiple testing (SCZ, ADHD, SRD) and our control traits (AD, height). These disease traits showed at least a nominally significant effect with sLDSR as well. Using both MAGMA and sLDSR, we integrated cluster membership values with GWAS summary statistics (n = 5) and assessed whether genome-wide disease risk aggregates to any of the eight experimentally identified longitudinal gene clusters. Overall, MAGMA and sLDSR show a strong concordance across phenotypes and clusters (ρ = .92, $p < 2.2 \times 10^{-16}$, n = 40) (see also Figure S7 in Supplement 1). After Bonferroni correction (n = 40), we found five significant phenotype-cluster associations with MAGMA and three with sLDSR (Figure 5, Tables S5 and S6 in Supplement 2).

We found that multiple upregulated clusters show enrichment for SCZ with the strongest evidence for the synaptic function cluster (MAGMA $p = 1.8 \times 10^{-7}$, sLDSR $p = 7.2 \times 10^{-5}$) (see Figure S8 in Supplement 1). For SRD, we found significant associations in the transcription regulation ($p = 2.5 \times 10^{-5}$) and neuron formation ($p = 1.2 \times 10^{-4}$) gene clusters with MAGMA only. While the analysis of adult height using all differentially expressed genes did not yield any evidence for enrichment of genetic signal, enrichment was observed at the cluster level. The cell connectivity cluster ($p = 3.7 \times 10^{-4}$) was enriched for height, in addition to suggestive enrichments in the cell division and RNA regulation clusters, which were not present for any of the psychiatric phenotypes. Remarkably, across all eight clusters, the enrichments of SCZ and height were inversely correlated ($\rho = -.85$, p = .011, n = 8; see also Supplemental Results section S3, Figures S9 and S10 in Supplement 1).

Finally, to take into account the full spectrum of correlations and dependencies between clusters (Figure S11 in Supplement 1), we performed a conditional analysis for SCZ, the trait for which the strongest cluster enrichments were observed with both methods. Using the same MAGMA model, for each cluster, we conditioned on the highest gene members (membership >0.5) of the other seven clusters (Table 2). We

Table 1. Differentially Expressed Genes Are Enriched for Polygenic Risk of Multiple Psychiatric Disorders

	MAGMA			sLDSR	
Phenotype	β (SE)	β_SD	p	τ (SE)	p
Psychiatric					
Schizophrenia	.022 (.007)	.094	.001[a]	1.70×10^{-9} (7.45×10^{-10})	.01
ADHD	.014 (.005)	.059	.002[a]	1.92×10^{-9} (1.25×10^{-9})	.06
Self-reported depression	.013 (.005)	.057	.003[a]	4.34×10^{-10} (2.10×10^{-10})	.02
Bipolar disorder	.007 (.005)	.032	.06	6.16×10^{-9} (3.64×10^{-9})	.05
Cross disorder	.005 (.005)	.020	.16	1.19×10^{-9} (1.00×10^{-9})	.12
MDD	.000 (.004)	−.001	.51	6.07×10^{-9} (4.39×10^{-9})	.08
ASD	.000 (.004)	−.002	.54	2.97×10^{-9} (3.48×10^{-9})	.20
Neurodegenerative					
Alzheimer's disease	.003 (.004)	.015	.22	1.30×10^{-10} (1.02×10^{-9})	.45
Nonbrain					
Height	.009 (.011)	.037	.21	-1.62×10^{-9} (1.36×10^{-9})	.88

The table shows the results of MAGMA and stratified linkage disequilibrium score regression (sLDSR) for differentially expressed genes (see Tables S3 and S4 in Supplement 2 for more details). β_SD indicates change in Z-value given a change of 1 SD in log (T^2 statistic), and τ indicates the contribution to the per–single nucleotide polymorphism heritability.

ADHD, attention-deficit/hyperactivity disorder; ASD, autism spectrum disorder; MDD, major depressive disorder.

[a]Phenotypes that survive multiple testing correction (n = 9).

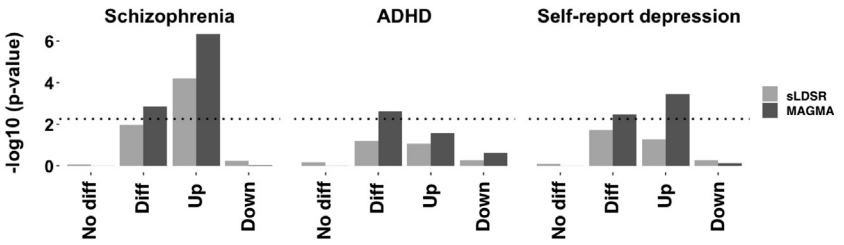

Figure 4. Schizophrenia polygenic risk lies in genes upregulated during neuronal differentiation. A more detailed investigation of the effect of differentially expressed genes on the heritability of schizophrenia, attention-deficit/hyperactivity disorder (ADHD), and self-report depression. The y-axis denotes the –log10 *p* value of the enrichment. No diff indicates genes that are not differentially expressed, Diff indicates log (T^2 statistic) (as shown in Table 1), Up indicates genes upregulated during differentiation, and Down indicates genes downregulated during differentiation. The dotted line represents the threshold for $p = .0056$ ($n = 9$ traits). sLDSR, stratified linkage disequilibrium score regression.

found that the SCZ enrichment was driven by the synaptic function cluster ($p = 2.88 \times 10^{-3}$) only. The same conditional analysis for SRD, which only showed a significant enrichment with MAGMA, showed that this effect was primarily driven by the transcription regulation cluster ($p = 5.42 \times 10^{-3}$) (Table S7 in Supplement 1).

Replication in the CORTECON RNA Sequencing Dataset Shows Strong Concordance With Discovery Analyses

To evaluate reproducibility of our findings, we performed a comprehensive replication analysis in the CORTECON RNA sequencing dataset of in vitro human cortical differentiation (13). While the CORTECON project was executed using widely different experimental procedures (Supplemental Results section S4.1 in Supplement 1), we detected largely overlapping transcriptomic patterns with the discovery dataset. Between datasets, we saw robust sample correlations across the differentiation trajectory (Supplemental Results section S4.2, Figure S12 in Supplement 1), including in stem cell and early neuronal gene marker expression patterns (Supplemental Results section S4.4, Figures S14 and S15 in Supplement 1). We observed a highly significant overlap in differentially expressed genes (Supplemental Results section S4.5 in

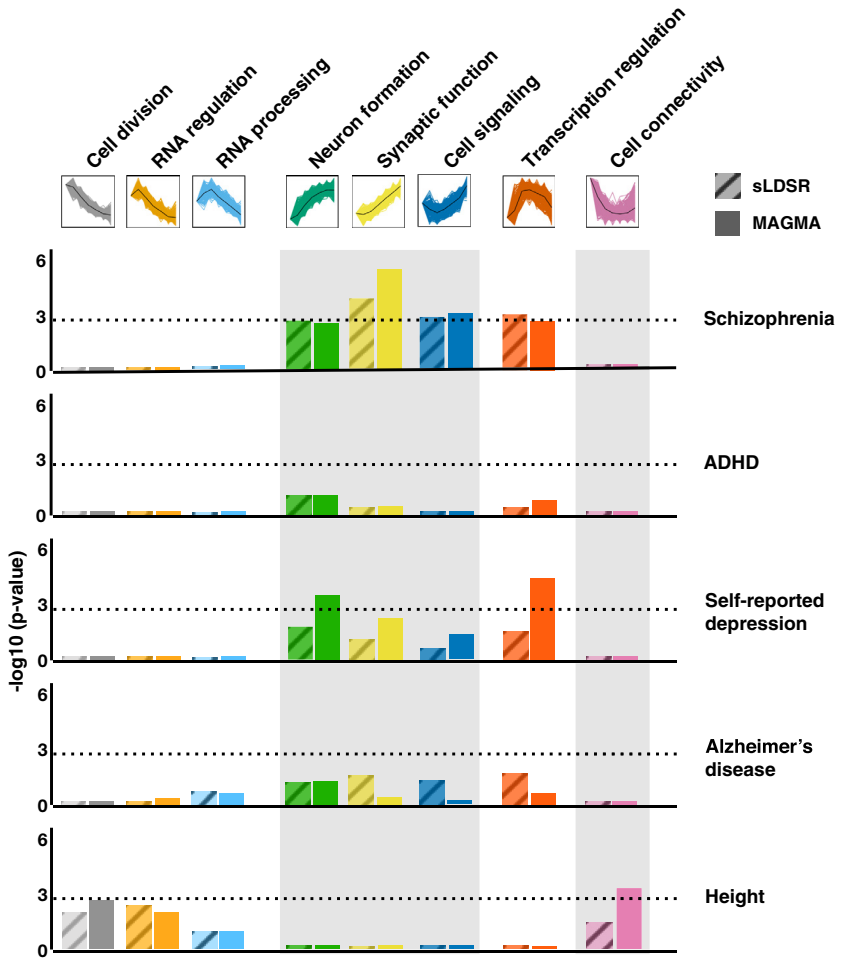

Figure 5. Polygenic psychiatric risk is distributed across specific longitudinal gene clusters. Results from stratified linkage disequilibrium score regression (sLDSR) (diagonal pattern) and MAGMA (solid colors) are shown for each phenotype (labels on the right) colored by gene cluster. Gene cluster annotation and cluster expression pattern are shown on top. The y-axis states the –log10 (*p* value). The dotted horizontal line represents the threshold for Bonferroni correction ($p = .05/40$). ADHD, attention-deficit/hyperactivity disorder.

Table 2. The Association With Schizophrenia Risk Is Driven by the Synaptic Function Gene Cluster

	MAGMA Primary		MAGMA Conditional	
Cluster	β (SE)	*p*	β (SE)	*p*
Cell Division	−.045 (.017)	1.00	−.047 (.027)	.96
RNA Regulation	−.040 (.017)	.99	−.044 (.027)	.95
RNA Processing	−.006 (.017)	.64	−.011 (.024)	.68
Neuron Formation	.048 (.017)	2.12×10^{-3}	.018 (.036)	.30
Synaptic Function[a]	.077 (.017)	1.82×10^{-6}	.070 (.026)	2.88×10^{-3}
Cell Signaling	.052 (.016)	6.88×10^{-4}	.032 (.023)	.08
Transcription Regulation	.048 (.016)	1.67×10^{-3}	.019 (.025)	.22
Cell Connectivity	−.061 (.017)	1.00	−.076 (.026)	1.00

Gene-level association signal is regressed on cluster membership while adjusting for high-membership genes of all other seven clusters.

[a]Gene cluster that remains significant after conditioning on other gene clusters.

Supplement 1) and in identified gene clusters (Supplemental Results section S4.6, Figures S16 and S17 in Supplement 1). In addition, we found that genes differentially expressed during 37 days of differentiation in the CORTECON dataset, which closely maps to 30 days of differentiation in the discovery set, were significantly associated with SCZ risk (β = .047, *p* = .007) (Supplemental Results section S4.7 in Supplement 1). As in the discovery dataset, this association is driven by genes that are upregulated over time (*p* = .008), but not downregulated (*p* = .74). While the identified gene clusters show significant overlap with the eight gene clusters from the discovery analysis (Figure S17 in Supplement 1), we do not observe the association with SCZ risk to be distributed to a single gene cluster. To investigate whether similar genes are driving the association with SCZ risk between our discovery analysis and the CORTECON dataset, we adjusted our analysis in the CORTECON dataset for the synaptic gene cluster (*n* = 779 genes) of the discovery analysis. We found that the strength of the association between SCZ risk and day 37 upregulated genes decreased when we accounted for synaptic genes from the discovery analysis (β = .044, *p* = .031) (Supplemental Results section S4.7 in Supplement 1). We have highlighted a set of genes that have high membership to the synaptic gene cluster, are differentially expressed in CORTECON, and are significantly associated to SCZ based on the GWAS (Figure S18 in Supplement 1). Taken together, this suggests that the same group of genes underlies the association between SCZ polygenic risk and transcriptomic signatures across differentiation, and further demonstrates the concordance between both datasets.

DISCUSSION

We investigated a longitudinal in vitro stem cell model of human neuronal differentiation to study psychiatric disease susceptibility based on evidence from GWASs. We confirmed that our in vitro model highlights transcriptomic profiles that are in line with an emerging neuronal identity that recapitulates signatures of in vivo cortical development across specific developmental time periods and laminae of the human neocortex. This is in line with previous findings (14), and highlights that longitudinal gene expression dynamics underlying our model of human neuronal differentiation can be informative to study genes and pathways involved in in vivo human cortical development. Importantly, neuronal cell types (43–45) and early brain development (7,23,46) have been postulated as integral components of SCZ disease susceptibility. Here, we observe that genes differentially expressed across neuronal differentiation are significantly associated with genome-wide disease risk of SCZ, a finding that we replicate in an independent dataset. Our findings suggest that SCZ risk aggregates to genes involved in synaptic functioning during development. Although not the only pathogenic process contributing to SCZ, synaptic dysfunction is most strongly supported by genetic data, postmortem expression studies, and animal models (44,47–51). We are the first to provide evidence for this hypothesis using a longitudinal in vitro cell-based model and aggregate polygenic disease risk. Our results suggest that high gene members of the synaptic function gene cluster enriched for SCZ (Figure S18 in Supplement 1), such as calcium voltage-gated channel subunit alpha 1C (*CACNA1C*), located at a genome-wide significant SCZ locus (23), are suitable candidates for functional follow-up in this in vitro model. We found evidence neither for AD, a late-onset nonpsychiatric brain disease, nor for adult human height in this neuronal cluster. Together, our findings demonstrate that longitudinal transcriptomic signatures important for neuronal differentiation recapitulate the in vivo context and align with the genetic basis of the disease. SCZ disease biology, and in particular synaptic functioning, can thus be studied through these molecular processes captured by this in vitro model.

We also observed a significant enrichment of genetic signal with MAGMA for SRD in genes upregulated during differentiation, and showed that this enrichment is predominantly driven by genes in the transcription regulation gene cluster. Interestingly, the SRD GWAS reported that the top SNPs were enriched for transcription regulation related to neurodevelopment (11), which is in line with our in vitro findings. We observed no enrichment of the GWAS of recurrent and severe MDD in Han-Chinese women (24). The latter sample represents the most genetically and phenotypically homogeneous GWAS of MDD. The fact that for these results no enrichment for any of our gene sets was observed may suggest that neurodevelopmental processes play a lesser role in MDD (52). Alternatively, larger sample sizes are needed to better capture the genome-wide genetic risk associated with MDD (Figure S19 in Supplement 1). Self-reported depression is a much broader phenotype that may include other psychiatric traits, which could drive the observed neurodevelopment and transcription findings. Although it remains unclear how these results and the application of the model extrapolate to the MDD phenotype, our approach does highlight enrichment in distinct clusters for SRD and SCZ and could help shed light on how these two complex traits differ in their etiology.

A strength of our approach is the longitudinal analysis framework that we developed. We implemented an experimental design across a dense and repeatedly sampled time series and integrated longitudinal transcriptomic signatures with genome-wide disease risk using available GWAS summary statistics. This increases statistical power to directly

investigate the cumulative impact of risk loci on genes important to our model system. While we specifically chose to perform our experiments across an isogenic background to minimize variation and maximize statistical power to identify transcriptomic signatures, our framework can easily be extended to a multisample design (e.g., case vs. control) (18,19), which makes it relevant for many disease-specific experimental settings.

Our experimental procedure applied differentiation toward a broad neuronal phenotype. Our work excludes neither disease associations with specific subtypes of neuronal cells or other major brain cell types nor cell nonautonomous changes that may contribute. We provide a proof of concept of an in vitro model of neuronal cells for studying complex diseases, such as SCZ, and present an analytical framework that includes longitudinal assessment of gene expression profiles. This approach can readily be extended to study in vitro differentiation of other major brain cell types, such as astrocytes or oligodendrocytes. In addition, co-culture with astrocyte may facilitate a more mature neuronal culture (53,54) and provide further insights into the temporal specificity of SCZ genetic risk. Although we show strong evidence for SCZ risk in early prenatal neurodevelopment, our findings do not preclude an additional contribution of postnatal neurodevelopment to the etiology of the disease (55–57).

In summary, as GWAS risk loci have small effect sizes and are abundantly distributed across the genome, new approaches are needed that allow for functional investigation of polygenic disease architectures. Embracing the polygenic nature of psychiatric disorders is an important step forward in translating findings from GWASs to disease biology (52). Our approach allowed us to narrow down on potential core disease processes and opens up new avenues to study disease in the context of polygenicity. Future work may, for example, incorporate model perturbations to study aggregate disease risk in finer detail or use the model for functional fine mapping of specific SCZ GWAS loci across an isogenic background in a controlled environment. Overall, this work contributes to understand the functional mechanisms that underlie psychiatric disease heritability and polygenicity in the post GWAS era.

ACKNOWLEDGMENTS AND DISCLOSURES

This research was supported by National Institutes of Health/National Institute of Mental Health Grant Nos. R01 MH090553 (RAO, principal investigator) and U01MH105578 (to RAO).

The project was led by RAO. Experiments were designed and conceived by APSO and RAO. Experiments were optimized and conducted and samples were processed by APSO, MHMB, and RTM. Analysis of the data was performed by APSO and MHMB, and feedback was provided by LMOL and RAO. The main findings were interpreted by APSO, MHMB, RTM, LMOL, and RAO. Primary drafting of the manuscript was performed by APSO and main feedback was provided by RAO and LMOL. All authors contributed to the production and approval of the final manuscript.

We thank all research participants and researchers involved in making each genome-wide association study summary statistic available and this work possible, including the 23andMe Research Team. We thank C. de Leeuw for his helpful input and troubleshooting with MAGMA analyses and thank the linkage disequilibrium score regression team for their input and helpful troubleshooting with stratified linkage disequilibrium score regression.

The authors report no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From the Center for Neurobehavioral Genetics (APSO, MHMB, RTM, LMOL, RAO), Semel Institute for Neuroscience and Human Behavior; and Department of Human Genetics (RAO), David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California.

Address correspondence to Roel A. Ophoff, Ph.D., University of California, Center for Neurobehavioral Genetics, 695 Charles E. Young Drive South, Los Angeles, CA 90095; E-mail: ophoff@ucla.edu.

Received Feb 23, 2018; revised Jul 18, 2018; accepted Aug 9, 2018.

Supplementary material cited in this article is available online at https://doi.org/10.1016/j.biopsych.2018.08.019.

REFERENCES

1. Geschwind DH, Flint J (2015): Genetics and genomics of psychiatric disease. Science 349:1489–1494.
2. Polderman TJC, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, Posthuma D (2015): Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nat Genet 47:702–709.
3. Sullivan PF, Agrawal A, Bulik C, Andreassen OA, Borglum A, Breen G, *et al.* (2017): Psychiatric genomics: An update and an agenda. Am J Psychiatry 175:15–27.
4. Falk A, Heine VM, Harwood AJ, Sullivan PF, Peitz M, Brüstle O, *et al.* (2016): Modeling psychiatric disorders: From genomic findings to cellular phenotypes. Mol Psychiatry 21:1167–1179.
5. Gulsuner S, Walsh T, Watts AC, Lee MK, Thornton AM, Casadei S, *et al.* (2013): Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. Cell 154:518–529.
6. Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P, *et al.* (2014): A polygenic burden of rare disruptive mutations in schizophrenia. Nature 506:185–190.
7. Olde Loohuis LMO, Vorstman JAS, Ori AP, Staats KA, Wang T, Richards AL, *et al.* (2015): Genome-wide burden of deleterious coding variants increased in schizophrenia. Nat Commun 6:7501.
8. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, *et al.* (2015): Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet 47:1228–1235.
9. Geschwind DH (2011): Genetics of autism spectrum disorders. Trends Cogn Sci 15:409–416.
10. Rubeis SD, He X, Goldberg AP, Poultney CS, Samocha K (2014): Synaptic, transcriptional, and chromatin genes disrupted in autism A. Nature 515:209–215.
11. Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, *et al.* (2016): Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. Nat Genet 48:1031–1036.
12. Shi Y, Kirwan P, Smith J, Robinson HPC, Livesey FJ (2012): Human cerebral cortex development from pluripotent stem cells to functional excitatory synapses. Nat Neurosci 15:477–486.
13. van de Leemput J, Boles NC, Kiehl TR, Corneo B, Lederman P, Menon V, *et al.* (2014): CORTECON: A temporal transcriptome analysis of in vitro human cerebral cortex development from human embryonic stem cells. Neuron 83:51–68.
14. Stein JL, de la Torre-Ubieta L, Tian Y, Parikshak NN, Hernández IA, Marchetto MC, *et al.* (2014): A quantitative framework to evaluate modeling of cortical development by neural stem cells. Neuron 83:69–86.
15. Du P, Kibbe WA, Lin SM (2008): lumi: A pipeline for processing Illumina microarray. Bioinformatics 24:1547–1548.
16. Lin SM, Du P, Huber W, Kibbe WA (2008): Model-based variance-stabilizing transformation for Illumina microarray data. Nucleic Acids Res 36:e11.
17. Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O'Keeffe S, *et al.* (2014): An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. J Neurosci 34:11929–11947.

18. Tai YC, Speed TP (2006): A multivariate empirical Bayes statistic for replicated microarray time course data. Ann Stat 34:2387–2412.
19. Aryee MJ, Gutiérrez-Pabello JA, Kramnik I, Maiti T, Quackenbush J (2009): An improved empirical bayes approach to estimating differential gene expression in microarray time-course data: BETR (Bayesian Estimation of Temporal Regulation). BMC Bioinformatics 10:409.
20. Kumar L, E Futschik M (2007): Mfuzz: A software package for soft clustering of microarray data. Bioinformation 2:5–7.
21. Schwämmle V, Jensen ON (2010): A simple and fast method to determine the parameters for fuzzy c-means cluster analysis. Bioinformatics 26:2841–2848.
22. Huang DW, Sherman BT, Lempicki RA (2009): Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44–57.
23. Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014): Biological insights from 108 schizophrenia-associated genetic loci. Nature 511:421–427.
24. CONVERGE Consortium (2015): Sparse whole-genome sequencing identifies two loci for major depressive disorder. Nature 523:588–591.
25. Psychiatric GWAS Consortium Bipolar Disorder Working Group (2011): Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat Genet 43:977–983.
26. Autism Spectrum Disorders Working Group of the Psychiatric Genomics Consortium (2017): Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. Mol Autism 8:21.
27. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, *et al.* (2019): Discovery of the first genome-wide significant risk loci for ADHD. Nat Genet 51:63–75.
28. Cross-Disorder Group of the Psychiatric Genomics Consortium (2013): Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. Lancet 381:1371–1379.
29. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, 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.
30. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, *et al.* (2014): Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet 46:1173–1186.
31. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, *et al.* (2015): A global reference for human genetic variation. Nature 526:68–74.
32. de Leeuw CA, Mooij JM, Heskes T, Posthuma D (2015): MAGMA: Generalized gene-set analysis of GWAS data. PLoS Comput Biol 11:e1004219.
33. Gazal S, Finucane HK, Furlotte NA, Loh P-R, Palamara PF, Liu X, *et al.* (2017): Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection. Nat Genet 49:1421–1427.
34. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, *et al.* (2015): LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 47:291–295.
35. Tanapat P (2013): Neuronal cell markers. Mater Methods 3:196.
36. Magavi SSP, Macklis JD (2002): Immunocytochemical analysis of neuronal differentiation. Methods Mol Biol 198:291–297.
37. von Bohlen Und Halbach O (2007): Immunohistological markers for staging neurogenesis in adult hippocampus. Cell Tissue Res 329:409–420.
38. Clancy B, Darlington RB, Finlay BL (2001): Translating developmental time across mammalian species. Neuroscience 105:7–17.
39. Stiles J, Jernigan TL (2010): The basics of brain development. Neuropsychology Review 20:327–348.
40. Hattori A, Buac K, Ito T (2016): Regulation of stem cell self-renewal and oncogenesis by RNA-binding proteins. Adv Exp Med Biol 907:153–188.
41. Barros CS, Franco SJ, Muller U (2011): Extracellular matrix: Functions in the nervous system. Cold Spring Harb Perspect Biol 3:A005108.
42. Bikbaev A, Frischknecht R, Heine M (2015): Brain extracellular matrix retains connectivity in neuronal networks. Sci Rep 5:14527.
43. Skene NG, Bryois J, Bakken TE, Breen G, Crowley JJ, Gaspar H, *et al.* (2017): Genetic identification of brain cell types underlying schizophrenia. Nat Genet 50:825–833.
44. Genovese G, Fromer M, Stahl EA, Ruderfer DM, Chambert K, Landén M, *et al.* (2016): Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. Nat Neurosci 19:1433–1441.
45. Forrest MP, Zhang H, Moy W, McGowan H, Leites C, Dionisio LE, *et al.* (2017): Open chromatin profiling in hiPSC-derived neurons prioritizes functional noncoding psychiatric risk variants and highlights neurodevelopmental loci. Cell Stem Cell 21:305–318.e8.
46. Finucane H, Reshef Y, Anttila V, Slowikowski K, Gusev A, Byrnes A, *et al.* (2018): Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. Nat Genet 50:621–629.
47. Hall J, Trent S, Thomas KL, O'Donovan MC, Owen MJ (2015): Genetic risk for schizophrenia: Convergence on synaptic pathways involved in plasticity. Biol Psychiatry 77:52–58.
48. Lips ES, Cornelisse LN, Toonen RF, Min JL, Hultman CM, Holmans PA, *et al.* (2011): Functional gene group analysis identifies synaptic gene groups as risk factor for schizophrenia. Mol Psychiatry 17:996–1006.
49. Pocklington AJ, O'Donovan M, Owen MJ (2014): The synapse in schizophrenia. Eur J Neurosci 39:1059–1067.
50. Schwarz E, Izmailov R, Lio P, Meyer-Lindenberg A (2016): Protein interaction networks link schizophrenia risk loci to synaptic function. Schizophr Bull 42:1334–1342.
51. O'Dushlaine C, Rossin L, Lee PH, Duncan L, Parikshak NN, Newhouse S, *et al.* (2015): Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. Nat Neurosci 18:199–209.
52. Peterson RE, Cai N, Bigdeli TB, Li Y, Reimers M, Nikulova A, *et al.* (2017): The genetic architecture of major depressive disorder in Han Chinese women. JAMA Psychiatry 74:162–168.
53. Tang X, Zhou L, Wagner AM, Marchetto MCN, Muotri AR, Gage FH, Chen G (2013): Astroglial cells regulate the developmental timeline of human neurons differentiated from induced pluripotent stem cells. Stem Cell Res 11:743–757.
54. Johnson MA, Weick JP, Pearce RA, Zhang SC (2007): Functional neural development from human embryonic stem cells: Accelerated synaptic activity via astrocyte coculture. J Neurosci 27:3069–3077.
55. Birnbaum R, Jaffe AE, Hyde TM, Kleinman JE, Weinberger DR (2014): Prenatal expression patterns of genes associated with neuropsychiatric disorders. Am J Psychiatry 171:758–767.
56. Pers TH, Timshel P, Ripke S, Lent S, Sullivan PF, O'Donovan MC, *et al.* (2015): Comprehensive analysis of schizophrenia-associated loci highlights ion channel pathways and biologically plausible candidate causal genes. Hum Mol Genet 25:1247–1254.
57. Sekar A, Bialas AR, de Rivera H, Davis A, Hammond TR, Kamitaki N, *et al.* (2016): Schizophrenia risk from complex variation of complement component 4. Nature 530:177–183.