UCLA
UCLA Previously Published Works

Title

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

Permalink

https://escholarship.org/uc/item/3t41d5zt

Journal

Biological Psychiatry, 85(7)

ISSN

0006-3223

Authors

Ori, Anil PS
Bot, Merel HM
Molenhuis, Remco T
et al.

Publication Date

2019-04-01

DOI

10.1016/j.biopsych.2018.08.019

Peer reviewed

A longitudinal model of human neuronal differentiation for functional investigation of schizophrenia polygenic risk

Anil P.S. Ori[1], **Merel H.M. Bot**[1], **Remco T. Molenhuis**[1], **Loes M. Olde Loohuis**[1], and **Roel A. Ophoff**[1,2,*]

[1]Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, California, USA

[2]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, California, USA

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 GWAS 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 SCZ 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.

*Correspondence to Roel A. Ophoff, Ph.D., 695 Charles E. Young Drive South, CA 90095, Los Angeles, USA, Tel: (310) 794 9602, Ophoff@ucla.edu.

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

Keywords

schizophrenia; psychiatric disorders; genome-wide disease risk; polygenicity; neuronal stem cells; synaptic function

Introduction

Major psychiatric disorders feature a high heritability but have a largely unknown etiology(1, 2). The increasing sample sizes of genome-wide association studies (GWAS) 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 (ASD)(9, 10), and self-reported depression (SRD) (11). Differentiation of human embryonic stem cells (hESCs) into neuronal lineages has been demonstrated to hold great promise to model 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 observe 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 to use of *in vitro* neuronal differentiation as a promising model system to study genetic psychiatric risk, particularly in the context of schizophrenia.

Methods and Materials

Approval for stem cell research

This study and all described work was approved by the University of California, Los Angeles Embryonic Stem Cell Research Oversight (ESCRO) committee.

In vitro human neuronal differentiation

WA09(H9)-derived hNSCs were commercially obtained (Gibco) as neural progenitors and subsequently expanded as adherent culture according to the manufacturer's guidelines. Low

passage hNSCs (< 4 passage rounds) were plated in 12-well plates coated with poly-Dlysine (0.1 mg/mL, VWR) and laminin (4.52ug/cm^2, Corning™) at 1.5×10^5 cells, which were equally distributed and subsequently cultured in expansion medium as described above. After 24h of proliferation, media was changed to neuronal differentiation medium consisting of Neurobasal® Medium (Gibco), 2% B-27® Serum-Free Supplement (Gibco), 2mM GlutaMax™-I Supplement, 0.05 mM β-mercaptoethanol (Gibco), and 1x Pen Strep. Media was changed every 2–3 days.

Experimental design and assessment of gene expression

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

Data preprocessing and quality control

Gene expression data was extracted using the Gene Expression Module in GenomeStudio Software 2011.1. Data was background corrected with subsequent variancestabilizing 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 was used to perform principal component analysis (PCA). 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 typespecific 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 (TMAP), 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 vs day-30, day-0 vs day-5, day-5 vs day-15, and day-15 vs 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 (v 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 IDs using NCBI build 37.3, duplicate IDs removed, and gene boundaries extended symmetrically by 10kb 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 Multi-marker Analysis of GenoMic Annotation (MAGMA) and stratified 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 (BPD)(25), ASD(26), attention deficit hyperactivity disorder (ADHD)(27), cross disorder(28), Alzheimer's disease (AD)(29), and adult human height(30) (Supplemental Table S2). For each trait we used the most recent GWAS summary statistics that was publically available at the time of the analysis. The 1000 Genomes Project Phase 3 release (1KG) 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 SNP association using the default gene model ('snp-wise=mean') with +/− 10kb extensions of gene boundaries and SNPs with minor allele frequency (MAF) > 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 ancestrymatched 1KG 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 stratified LD score regression (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 MAF > 5%, a 1 centimorgan (cm) window, and an ancestry-match 1KG 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 test for a positive association, we report the contribution to the per-SNP h^2 (t) and the associated one-sided p-value, which is 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.

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 an hNSC 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 was in agreement with the intendend 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*, *Nestin*, 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 subsequently downregulated at later time points, suggesting that the differentiated culture maintains a relatively immature neuronal identity. Next, we explored PCA 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 & S1). 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 endothelialspecific 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 (TMAP) 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 DGE profiles (day-0 vs day-30) and *in vivo* developmental stage from 4 weeks post-conception (PCW) to 24 PCW (Figure S2). 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 TMAP analyses in three comparisons and examined how *in vitro* to *in vivo* matching 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 vs day-5 show strong overlap with *in vivo* period-1 (4–8 PCW) vs period-4 (13–16 PCW), while *in vitro* day-15 vs day-30 shows stronger overlap with *in vivo* period-2 (8–10 PCW) vs period-8 (birth-6M). 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 and S2).

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 Supplemental Methods). A total of 7,734 probes, mapping to 5,818 genes, were differentially expressed over time (Figure S3). We find that these genes are, on average, more constrained to genetic variation compared to non-differentially expressed genes (section S2). Using only differentially expressed probes, we next applied fuzzy c-means clustering and identified eight distinct longitudinal gene clusters (Figure 3 and S4). For each probe, we generated a corresponding cluster membership value, representing the degree to which a gene belongs to a cluster. To identify 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).

We identified three clusters with decreasing gene expression over time that are significantly enriched for cell division and 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 Alzheimer's disease (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 find that genes that are differentially expressed are enriched for genetic risk of multiple psychiatric disorders. We find significant effects with MAGMA for SCZ (P=0.001), ADHD (P=0.002), and SRD (P=0.003) (Table 1 and Table S3). With sLDSR, we find nominally significant effects for SCZ (P=0.01) and SRD (P=0.02) and a suggestive association for ADHD (P=0.06) (Table 1 and Table S4). We observed a suggestive enrichment for BPD, and no enrichment for the cross disorder, ASD, MDD CONVERGE 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 find that the effect is driven by genes that are upregulated (MAGMA P=5.0×10^{-7}, sLDSR P=6.1×10^{-5}) and not by genes that are downregulated (MAGMA P=0.98, sLDSR P=0.61) (Figure 4 and Figure S6). For SRD, we only find a stronger enrichment in upregulated genes with MAGMA (P=3.5×10^{-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 cluster level. For this analysis, we only included 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 (rho = 0.92, p<2.2×10^{-16}, n=40, see also Figure S7). After Bonferroni correction (n=40), we find five significant phenotype-cluster associations with MAGMA and three with sLDSR (Figure 5 and Table S5/S6).

We find that multiple upregulated clusters show enrichment for SCZ with the strongest evidence for the *synaptic function* cluster (MAGMA P=1.8×10^{-7}, sLDSR P=7.2×10^{-5}) (see Figure S8). For SRD, we find significant associations in the *transcription regulation* (P=2.5×10^{-5}) and the *neuron formation* (P=1.2×10^{-4}) gene cluster 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 is observed at the cluster level. The *cell connectivity* cluster (P=3.7×10^{-4}) is enriched for height, in addition to suggestive enrichments in the *cell division* and *RNA regulation* cluster, which are not present for any of the psychiatric phenotypes. Remarkably, across all 8 clusters the enrichments of SCZ and height are inversely correlated (rho=−0.85, P=0.011, n=8; see also section S3 and Figure S9–10).

Finally, in order to take into account the full spectrum of correlations and dependencies between clusters (Figure S11), we performed a conditional analysis for SCZ, the trait for which the strongest cluster enrichments are 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 find that the SCZ enrichment is driven by the *synaptic function* cluster (p=2.88×10^{-3}) only. The same conditional analysis for SRD, which only showed a significant enrichment with MAGMA, shows that this effect is primarily driven by the *transcription regulation* cluster (p=5.42×10^{-3}) (Table S7).

Replication in the CORTECON RNA-seq dataset shows strong concordance with discovery analyses

To evaluate reproducibility of our findings, we performed a comprehensive replication analysis in the CORTECON RNA sequencing (RNA-seq) dataset of *in vitro* human cortical differentiation(13). While the CORTECON project was executed using widely different experimental procedures (section S4.1), we detect largely overlapping transcriptomic patterns with the discovery dataset. Between datasets, we see robust sample correlations across the differentiation trajectory (section S4.2, Figure S12), including in stem cell and early neuronal gene marker expression patterns (section S4.4, Figure S14–15). We observe a highly significant overlap in differentially expressed genes (section S4.5) and in identified gene clusters (section S4.6, Figure S16–17). We in addition find that genes differentially expressed during 37 days of differentiation in CORTECON, which closely maps to 30 days of differentiation in the discovery set, are significantly associated with SCZ risk (beta=0.047, P=0.007, section S4.7). As in the discovery dataset, this association is driven by genes that are upregulated over time (P=0.008) but not downregulated (P=0.74). While the identified gene clusters show significant overlap with the eight gene clusters from the discovery analysis (Figure S17), 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 find that the strength of the association between SCZ risk and day-37 upregulated genes decreases when we account for synaptic genes from the discovery analysis (beta=0.044, P=0.031, section S4.7). 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). Taken together this suggests that the same group of genes underlie 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 GWAS. 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), such as Calcium Voltage-Gated Channel Subunit Alpha 1C (*CACNA1C*), located at a genomewide significant SCZ locus(23), are suitable candidates for functional follow-up in this *in vitro* model. We find no evidence for AD, a late-onset non-psychiatric 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 show 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). 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 multi-sample design (e.g. cases vs controls)(19, 53), which makes it relevant for many diseasespecific experimental settings.

Our experimental procedure applied differentiation towards a broad neuronal phenotype. Our work does not exclude disease associations with specific subtypes of neuronal cells or other major brain cell types nor does it exclude cell non-autonomous 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(54, 55) 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(56–58).

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 GWAS 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 finemapping 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

References

1. Geschwind DH, Flint J (2015): Genetics and genomics of psychiatric disease. Science. 349: 1489–1494. [PubMed: 26404826]

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. [PubMed: 25985137]

3. Sullivan PF, Agrawal A, Bulik C, Andreassen OA, Borglum A, Breen G, et al. (2017): Psychiatric Genomics: An Update and an Agenda. doi: 10.1101/115600.

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. 1167–1179. [PubMed: 27240529]

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. [PubMed: 23911319]

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. [PubMed: 24463508]

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. [PubMed: 26158538]

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. [PubMed: 26414678]

9. Geschwind DH (2011): Genetics of autism spectrum disorders. Trends in Cognitive Sciences. 15.

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. [PubMed: 25363760]

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. Nature Publishing Group. 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: 477–86, S1. [PubMed: 22306606]

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. [PubMed: 24991954]

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. [PubMed: 24991955]

15. Du P, Kibbe W a., Lin SM (2008): lumi: a pipeline for processing Illumina microarray. Bioinformatics. 24: 1547–1548. [PubMed: 18467348]

16. Lin SM, Du P, Huber W, Kibbe W a. (2008): Model-based variance-stabilizing transformation for Illumina microarray data. Nucleic Acids Res. 36: 1–9. [PubMed: 17962301]

17. Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O'Keeffe S, et al. (2014): An RNASequencing Transcriptome and Splicing Database of Glia, Neurons, and Vascular Cells of the Cerebral Cortex. Journal of Neuroscience. 34: 11929–11947. [PubMed: 25186741]

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 timecourse data: BETR (Bayesian Estimation of Temporal Regulation). BMC Bioinformatics. 10: 409. [PubMed: 20003283]

20. Kumar L E Futschik M (2007): Mfuzz: a software package for soft clustering of microarray data. Bioinformation. 2: 5–7. [PubMed: 18084642]

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. [PubMed: 20880957]

22. Huang DW, Lempicki R a., Sherman BT (2009): Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 4: 44–57. [PubMed: 19131956]

23. Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014): Biological insights from 108 schizophrenia-associated genetic loci. Nature. 511.

24. CONVERGE Consortium (2015): Sparse whole-genome sequencing identifies two loci for major depressive disorder. Nature. 523: 588. [PubMed: 26176920]

25. Group PGCBDW (2011): Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat Genet. 43: 977–983. [PubMed: 21926972]

26. The 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. [PubMed: 28540026]

27. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. (2017): Discovery of the first genome-wide significant risk loci for ADHD. bioRxiv. doi: 10.1101/145581.

28. Consortium C-DG of TPG (2013): Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet. 381: 1371–1379. [PubMed: 23453885]

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. [PubMed: 24162737]

30. Wood AR Esko T Yang J Vedantam S Pers TH Gustafsson S Chu AY Estrada K Luan J Kutalik Z Amin N Buchkovich ML Croteau-Chonka DC Day FR Duan Y Fall T Fehrmann R Ferreira T Jackson AU Karjalainen J Lo KS Locke AEMägi R Mihailov E Por FTM (2014): Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet. 46: 1173–1186. [PubMed: 25282103]

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. [PubMed: 26432245]

32. de Leeuw CA, Mooij JM, Heskes T, Posthuma D (2015): MAGMA: Generalized Gene-Set Analysis of GWAS Data. PLoS Comput Biol. 11. doi: 10.1371/journal.pcbi.1004219.

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. doi: 10.1038/ng.3954.

34. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Consortium SWG of TPG, et al. (2015): LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 47: 291–295. [PubMed: 25642630]

35. Tanapat P (2013): Neuronal Cell Markers. Materials and Methods. 3. doi: 10.13070/mm.en.3.196.

36. Magavi SSP, Macklis JD (2002): Immunocytochemical analysis of neuronal differentiation. Methods Mol Biol. 198: 291–297. [PubMed: 11951631]

37. von Bohlen Und Halbach O(2007): Immunohistological markers for staging neurogenesis in adult hippocampus. Cell Tissue Res. 329: 409–420. [PubMed: 17541643]

38. Clancy B, Darlington RB, Finlay BL (2001): Translating developmental time across mammalian species. Neuroscience. 105: 7–17. [PubMed: 11483296]

39. Stiles J, Jernigan TL (2010): The basics of brain development. Neuropsychology Review. 20.

40. Hattori A, Buac K, Ito T (2016): Regulation of Stem Cell Self-Renewal and Oncogenesis by RNA-Binding Proteins. RNA Processing Disease and Genome-wide probing. pp 153–188.

41. Barros CS, Franco SJ, Muller U (2011): Extracellular Matrix: Functions in the nervous system. Cold Spring Harb Perspect Biol. 3: 1–24.

42. Bikbaev A, Frischknecht R, Heine M (2015): Brain extracellular matrix retains connectivity in neuronal networks. Sci Rep. 5: 14527. [PubMed: 26417723]

43. Skene NG, Bryois J, Bakken TE, Breen G, Crowley JJ, Gaspar H, et al. (2017): Genetic Identification Of Brain Cell Types Underlying Schizophrenia. doi: 10.1101/145466.

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. [PubMed: 27694994]

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. [PubMed: 28803920]

46. Finucane H, Reshef Y, Anttila V, Slowikowski K, Gusev A, Byrnes A, et al. (2017): Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. bioRxiv. doi: 10.1101/103069.

47. Hall J, Trent S, Thomas KL, O'Donovan MC, Owen MJ (2015): Genetic risk for schizophrenia: Convergence on synaptic pathways involved in plasticity. Biological Psychiatry. 77.

48. Lips ES, Cornelisse LN, Toonen RF, Min JL, Hultman CM, Holmans P a., et al. (2011): Functional gene group analysis identifies synaptic gene groups as risk factor for schizophrenia. Mol Psychiatry. 4: 1–11.

49. Pocklington AJ, O'Donovan M, Owen MJ (2014): The synapse in schizophrenia. Eur J Neurosci. 39: 1059–1067. [PubMed: 24712986]

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. [PubMed: 27056717]

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. [PubMed: 25599223]

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. [PubMed: 28002544]

53. Tai YC, Speed TP (2006): A multivariate empirical Bayes statistic for replicated microarray time course data. Ann Stat. 34: 2387–2412.

54. 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. [PubMed: 23759711]

55. Johnson MA, Weick JP, Pearce RA, Zhang S-C (2007): Functional neural development from human embryonic stem cells: accelerated synaptic activity via astrocyte coculture. J Neurosci. 27: 3069–3077. [PubMed: 17376968]

56. 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. [PubMed: 24874100]

57. 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.

58. 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. [PubMed: 26814963]
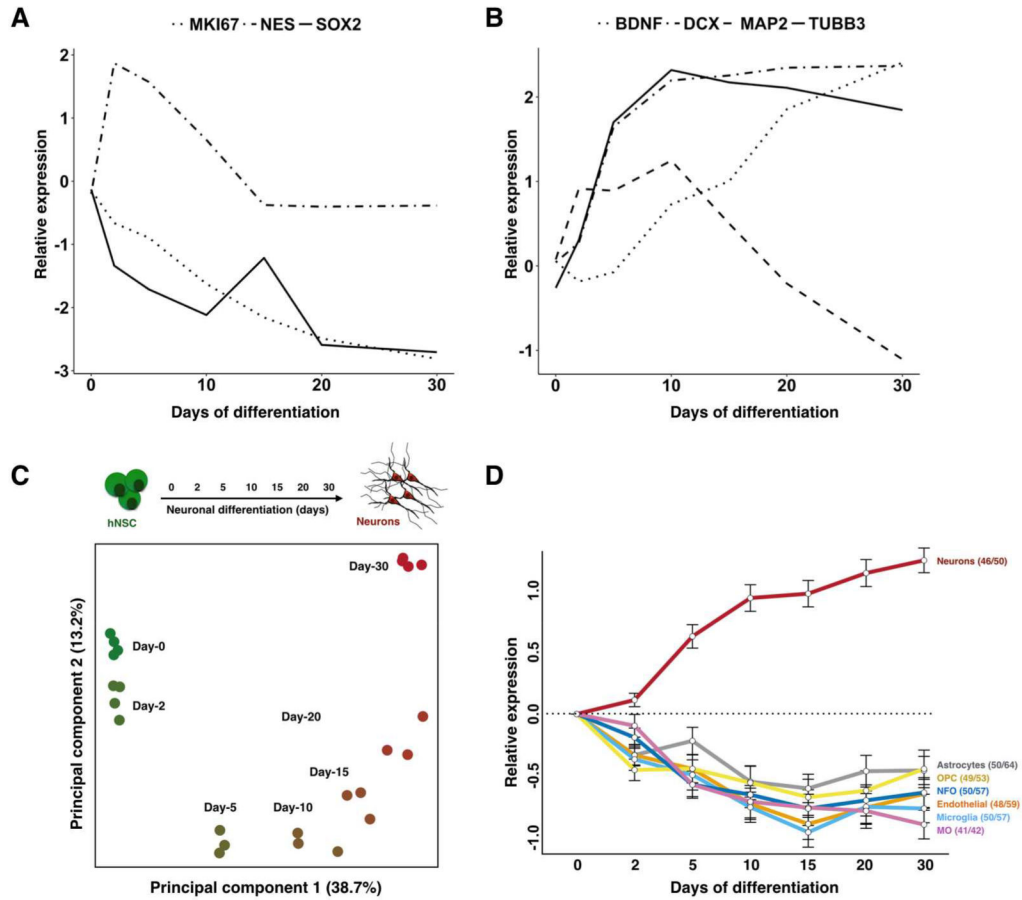
Figure 1. *In vitro* **gene expression profiles confirm a neuron-specific differentiation process.**
Relative gene expression of traditional stem cell (A) and neuronal (B) markers plotted across days of differentiation. (C) PCA of *in vitro* transcriptomic data with PC1 (x-axis) and PC2 (y-axis) visualized. 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 the parentheses represents the number of genes for which the average expression is plotted. The second number represents the corresponding number of probes assayed. OPC = oligodendrocyte precursor cells, NFO = newly formed oligodendrocytes, MP = myelinating oligodendrocytes.
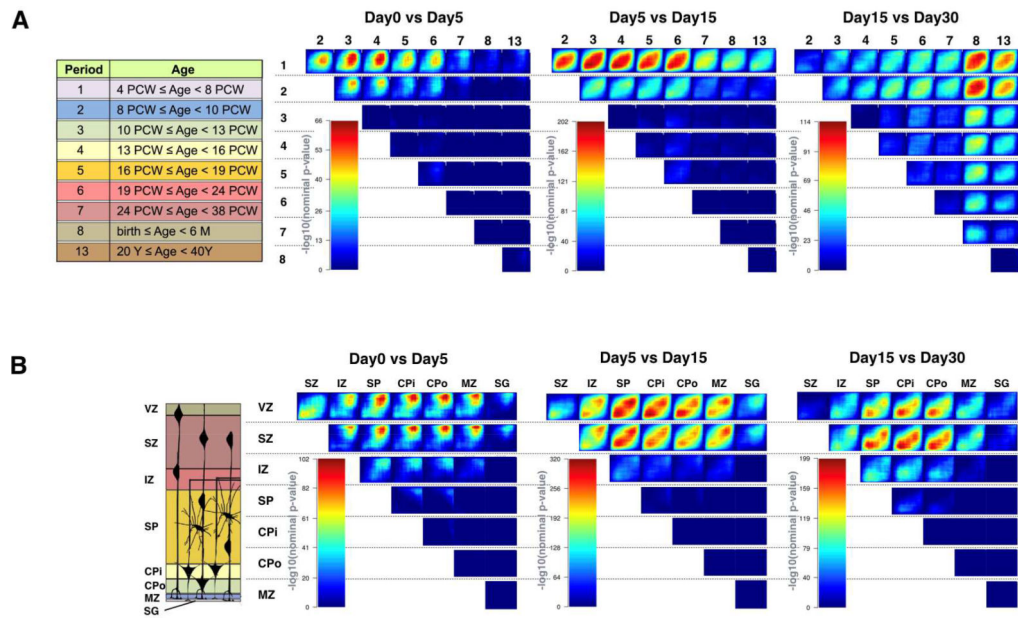
Figure 2. *In vitro* gene expression profiles match *in vivo* human cortical development.
TMAP output visualizes the amount of overlap between *in vitro* and *in vivo* DGE profiles colored by – log10(p-value) (see figure S2 for more details on interpretation). Note that p-values are shown on varying color scales between graphs. Abbreviations and numbering above maps correspond to schematic representations on the left (adopted from Stein et al., 2014) of different developmental stages (A) and laminae (B). VZ = ventricular zone, SZ = subventricular zone, IZ = intermediate zone, SP=subplate zone, CPi= inner cortical plate, CPo = outer cortical plate, MZ = marginal zone, PCW = post conception weeks, M = months, Y = years, Period = developmental stage.

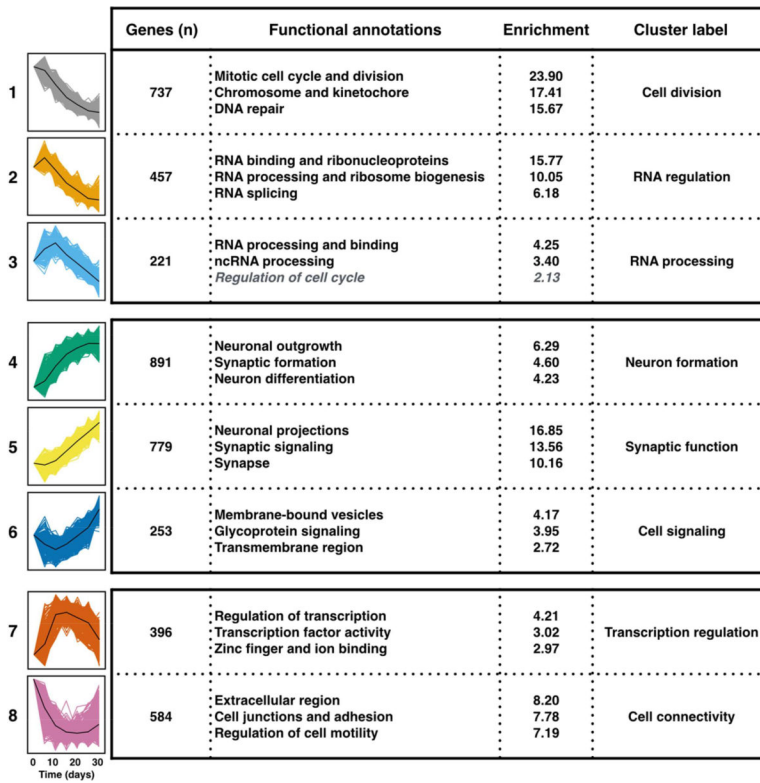|	Genes (n)	Functional annotations	Enrichment	Cluster label
1	737	Mitotic cell cycle and division Chromosome and kinetochore DNA repair	23.90 17.41 15.67	Cell division
2	457	RNA binding and ribonucleoproteins RNA processing and ribosome biogenesis RNA splicing	15.77 10.05 6.18	RNA regulation
3	221	RNA processing and binding ncRNA processing *Regulation of cell cycle*	4.25 3.40 *2.13*	RNA processing
4	891	Neuronal outgrowth Synaptic formation Neuron differentiation	6.29 4.60 4.23	Neuron formation
5	779	Neuronal projections Synaptic signaling Synapse	16.85 13.56 10.16	Synaptic function
6	253	Membrane-bound vesicles Glycoprotein signaling Transmembrane region	4.17 3.95 2.72	Cell signaling
7	396	Regulation of transcription Transcription factor activity Zinc finger and ion binding	4.21 3.02 2.97	Transcription regulation
8	584	Extracellular region Cell junctions and adhesion Regulation of cell motility	8.20 7.78 7.19	Cell connectivity

0 10 20 30
Time (days)

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 for full annotation results.
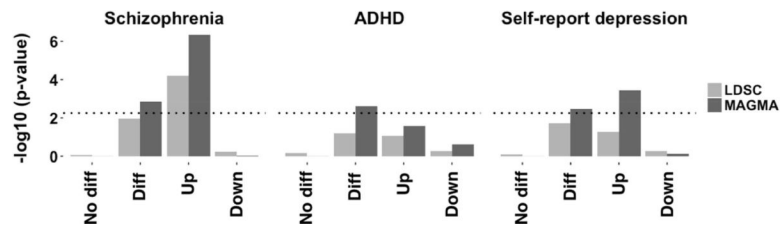
Figure 4. Schizophrenia polygenic risk lies in genes up-regulated during neuronal differentiation.
A more detailed investigation of the effect of differentially expressed genes on the heritability of SCZ, ADHD, and SRD. The y-axis denotes the –log10 P-value of the enrichment. No diff = genes that are not differentially expressed; Diff = log (T^2-statistic) as shown in Table 1; Up = genes upregulated during differentiation; Down = genes downregulated during differentiation. The dotted line represents the threshold for P=0.0056 (n=9 traits).
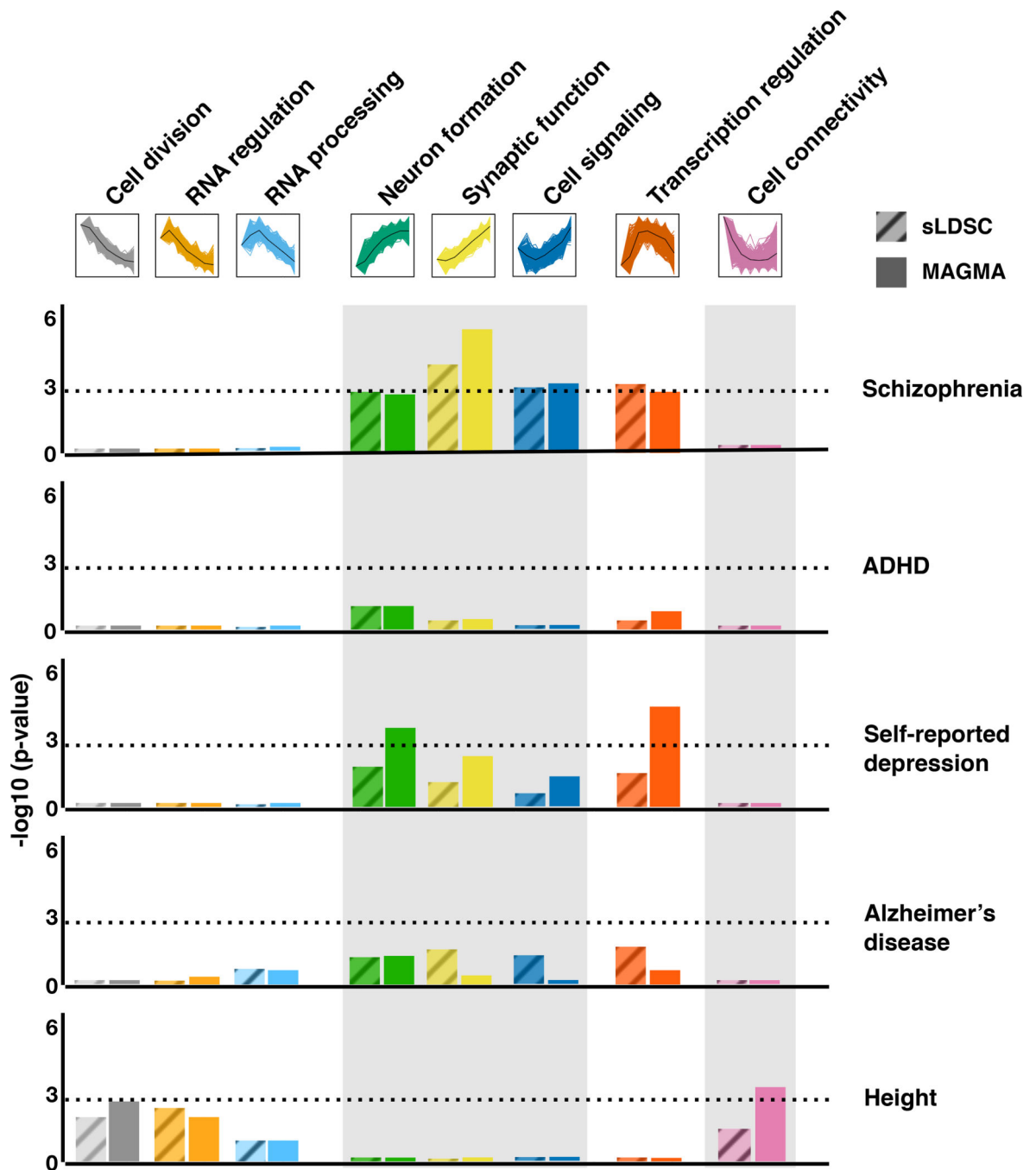
Figure 5. Polygenic psychiatric risk is distributed across specific longitudinal gene clusters.
Results from sLDSC (diagonal pattern) and MAGMA (solid colors) are shown for each
phenotype (labels on the right) colored by gene cluster. Gene cluster annotation and cluster
expression pattern are shown on top. The y-axis states the –log10 (p-value). The dotted
horizontal line represents the threshold for Bonferroni correction (p=0.05/40).

Table 1.

Differentially expressed genes are enriched for polygenic risk of multiple psychiatric disorders.

Shown are results of MAGMA and sLDSR for differentially expressed genes. P-values highlighted in bold show phenotypes that survive multiple testing correction (n=9). See Table S3 and S4 for more details. Beta = regression coefficient, SE = standard error, Beta_std = change in Z-value given a change of one standard deviation in log T2 statistic, τ (tau) = the contribution to the per-SNP h^2.

	MAGMA			sLDSC	
Phenotype	**Beta (SE)**	**Beta_std**	**P-value**	**τ (SE)**	**P-value**
Psychiatric					
Schizophrenia	0.022 (0.007)	0.094	**0.001**	1.70×10^{-9} (7.45×10^{-10})	0.01
ADHD	0.014 (0.005)	0.059	**0.002**	1.92×10^{-9} (1.25×10^{-9})	0.06
Self-reported depression	0.013 (0.005)	0.057	**0.003**	4.34×10^{-10} (2.10×10^{-10})	0.02
Bipolar disorder	0.007 (0.005)	0.032	0.06	6.16×10^{-9} (3.64×10^{-9})	0.05
Cross disorder	0.005 (0.005)	0.020	0.16	1.19×10^{-9} (1.00×10^{-9})	0.12
MDD CONVERGE	0.000 (0.004)	-0.001	0.51	6.07×10^{-9} (4.39×10^{-9})	0.08
ASD	0.000 (0.004)	-0.002	0.54	2.97×10^{-9} (3.48×10^{-9})	0.20
Neurodegenerative					
Alzheimer's disease	0.003 (0.004)	0.015	0.22	1.30×10^{-10} (1.02×10^{-9})	0.45
Non-brain					
Height	0.009 (0.011)	0.037	0.21	-1.62×10^{-9} (1.36×10^{-9})	0.88

Table 2.

The association with SCZ risk is driven by the synaptic function gene cluster.

Gene level association signal is regressed on cluster membership while adjusting for high membership genes of all other seven clusters. Shown are the results of the primary analysis (not adjusted for other clusters) and the conditional analysis with MAGMA. Beta = regression coefficient, SE = standard error.

Schizophrenia - clusters	MAGMA Primary		MAGMA Conditional	
	Beta (SE)	P-value	Beta (SE)	P-value
Cell division	-0.045 (0.017)	1.00	-0.047 (0.027)	0.96
RNA regulation	-0.040 (0.017)	0.99	-0.044 (0.027)	0.95
RNA processing	-0.006 (0.017)	0.64	-0.011 (0.024)	0.68
Neuron formation	0.048 (0.017)	2.12×10^{-3}	0.018 (0.036)	0.30
Synaptic function	**0.077 (0.017)**	$\mathbf{1.82\times10^{-6}}$	**0.070 (0.026)**	$\mathbf{2.88\times10^{-3}}$
Cell signaling	0.052 (0.016)	6.88×10^{-4}	0.032 (0.023)	0.08
Transcription regulation	0.048 (0.016)	1.67×10^{-3}	0.019 (0.025)	0.22
Cell connectivity	-0.061 (0.017)	1.00	-0.076 (0.026)	1.00