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

Authors:

Anil P. S. Ori[1], Merel H. M. Bot[1], Remco T. Molenhuis[1⊥], Loes M. Olde Loohuis[1], Roel A. Ophoff[1,2,3*]

Affiliations:

[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

[3]Brain Center Rudolf Magnus, Department of Psychiatry, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

⊥Current affiliation: Department of Translational Neuroscience, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

*Correspondence to Roel A. Ophoff, Ph.D.

695 Charles E. Young Drive South, CA 90095, Los Angeles, USA

Tel: (310) 794 9602

E-mail: Ophoff@ucla.edu

Abstract

There is a pressing need for *in vitro* experimental systems that allow for interrogation of polygenic psychiatric disease risk to study the underlying biological mechanisms. We developed an analytical framework that integrates genome-wide disease risk from GWAS with longitudinal *in vitro* gene expression profiles of human neuronal differentiation. We demonstrate that the cumulative impact of risk loci of specific psychiatric disorders is significantly associated with genes that are differentially expressed across neuronal differentiation. We find significant evidence for schizophrenia, which is driven by a longitudinal synaptic gene cluster that is upregulated during differentiation. Our findings reveal that *in vitro* 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.

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.

77 **Material and Methods**

78 ***Approval for stem cell research***

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

81

82 ***In vitro human neuronal differentiation***

83 WA09(H9)-derived hNSCs were commercially obtained (Gibco) as neural progenitors and
84 subsequently expanded as adherent culture according to the manufacturer's guidelines
85 (Supplementary Methods). Low passage hNSCs (< 4 passage rounds) were plated in 12-well
86 plates coated with poly-D-lysine (0.1 mg/mL, VWR) and laminin (4.52ug/cm^2, Corning^{TM}) at
87 1.5x10^5 cells, which were equally distributed and subsequently cultured in expansion medium as
88 described above. After 24h of proliferation, media was changed to neuronal differentiation
89 medium consisting of Neurobasal® Medium (Gibco), 2% B-27® Serum-Free Supplement (Gibco),
90 2mM GlutaMax™-I Supplement, 0.05 mM β-mercaptoethanol (Gibco), and 1x Pen Strep. Media
91 was changed every 2-3 days.

92

93 ***Experimental design and assessment of gene expression***

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

98

99 ***Data preprocessing and quality control***

100 Gene expression data was extracted using the Gene Expression Module in
101 GenomeStudio Software 2011.1. Data was background corrected with subsequent variance-
102 stabilizing transformation and robust spline normalization was applied[15, 16]. We excluded low

103 quality probes and subsequently performed sample outlier detection by Euclidean distance and
104 standardized connectivity (see Supplementary Methods). The FactoMineR package (v1.28) in R
105 was used to perform principal component analysis (PCA). For subsequent downstream analyses,
106 we used the normalized expression values of 19,012 high quality filtered probes for all 24
107 samples.
108
109 ***Transcriptome-based in vitro cellular identity***
110 We identified cell-type specific genes of neurons, astrocytes, oligodendrocyte precursor
111 cells (OPC), newly formed oligodendrocytes (NFO), myelinating oligodendrocytes (MO),
112 microglia, and endothelial cells from mouse cerebral cortex[17] (Supplementary Methods). Next, we
113 extracted normalized gene expression values of these genes for each cell type from our own *in*
114 *vitro* dataset. We then standardized expression values to time point zero and calculated mean
115 standardized expression levels of cell type-specific genes for these seven cell types across time
116 points. This allowed us to investigate *in vitro* cellular identity across differentiation using cell-type
117 specific gene expression information of many genes for these seven cell types.
118
119 ***Transition mapping to a spatiotemporal atlas of early human brain development***
120 To investigate global transcriptomic matching between *in vitro* gene expression profiles
121 and *in vivo* gene expression profiles of neocortical brain regions, we applied transition mapping
122 (TMAP), which is implemented in the online CoNTExT bioinformatic pipeline
123 (https://context.semel.ucla.edu)[14]. Analyses were run for *in vitro* time points day-0 vs day-30, day-
124 0 vs day-5, day-5 vs day-15, and day-15 vs day-30 across both temporal and spatial dimensions
125 of human cortical development (see Supplementary Methods).
126
127 ***Time-series differential gene expression analysis***

128 Two multivariate empirical Bayes models were used to identify differentially expressed
129 genes across *in vitro* neuronal differentiation. The first method was implemented in the
130 Timecourse package (v 1.42) in R. We used the mb.long() function to calculate the one-sample
131 T^2 statistic that ranks genes based on their log10 probability to have differential expression over
132 time[18]. Bayesian Estimation of Temporal Regulation (BETR), an extension of the first approach,
133 uses a flexible random-effect model that allows for correlations between the magnitude of
134 differential expression at different time points[19]. BETR is implemented in the betr package (v 1.26)
135 in R. Differentially expressed genes were classified as the union of the set of genes with a
136 probability of 1.0 using BETR and an equally-sized set of top ranked genes using the T^2-statistic.
137

138 ***Fuzzy c-means cluster analysis***

139 To identify probes with similar expression patterns across differentiation, we applied fuzzy
140 c-means clustering to all differentially expressed probes. We calculated cluster membership
141 values using the fclusList() and membership() function in the Mfuzz package in R[20, 21]. See
142 Supplementary Methods for more details. Each probe receives a membership value for each
143 cluster. Probe membership values represent gene affiliation to a cluster and highlights the extent
144 of similarity in expression between genes. These values were used for subsequent downstream
145 analyses. We annotated clusters using Database for Annotation, Visualization, and Integrated
146 Discovery (DAVID, v6.8) and probes with a membership > 0.5 (Supplementary Methods).
147

148 ***Integration of GWAS data with in vitro transcriptomic signatures***

149 We first mapped Illumina probe IDs to Ensembl gene IDs using NCBI build 37.3, removed
150 duplicate Ensemble IDs, and extended gene boundaries symmetrically by 10kb to include
151 regulatory regions. Probe T^2-statistic and cluster membership values were collapsed per gene ID
152 using the mean value across probes. The mean gene-level T^2-statistic was then log-transformed
153 and the mean cluster membership values rank-transformed. These mean gene values were then

154 used to integrate *in vitro* signatures with GWAS data and to study the cumulative impact across
155 risk loci using Multi-marker Analysis of GenoMic Annotation (MAGMA) and stratified LD score
156 regression (sLDSR).
157
158 *GWAS summary statistics and ancestry matched reference panels*
159 GWAS summary statistics were obtained for SCZ[22], major depressive disorder (MDD)[23],
160 SRD[11], bipolar disorder (BPD)[24], ASD[25], attention deficit hyperactivity disorder (ADHD)[26], cross
161 disorder[27], Alzheimer's disease (AD)[28], and adult human height[29] (Supplementary Methods and
162 Table S2). For each trait we used the most recent GWAS summary statistics that was publically
163 available at the time of the analysis. The 1000 Genomes Project Phase 3 release (1KG) was used
164 as reference panel to model LD[30]. We used 503 individuals of European ancestry and 301
165 individuals of East Asian ancestry in analyses of GWAS data derived from target population of
166 Europeans and Han Chinese, respectively.
167
168 *MAGMA gene-set analysis*
169 Multi-marker Analysis of GenoMic Annotation (MAGMA v1.06)[31] was used to run "gene
170 property" analyses, which uses a multiple regression framework to associate a continuous gene
171 variable to GWAS gene level p-values. SNPs were mapped to genes using Ensembl gene IDs
172 and NCBI build 37.3 gene boundaries +/- 10kb extensions using the --annotate flag. For each
173 phenotype, we generated gene-level p-values by computing the mean SNP association using the
174 default gene model ('snp-wise=mean'). We only included SNP with minor allele frequency (MAF)
175 > 5% and dropped synonymous or duplicate SNPs after the first entry ('synonym-dup=drop-dup').
176 For each annotation, we then regressed gene-level GWAS test statistics on the corresponding
177 gene annotation variable using the '--gene-covar' function while adjusting for gene size, SNP
178 density, and LD-induced correlations ('--model correct=all'), which is estimated from an ancestry-
179 matched 1KG reference panel. In all analyses, we included only genes for which we had both the

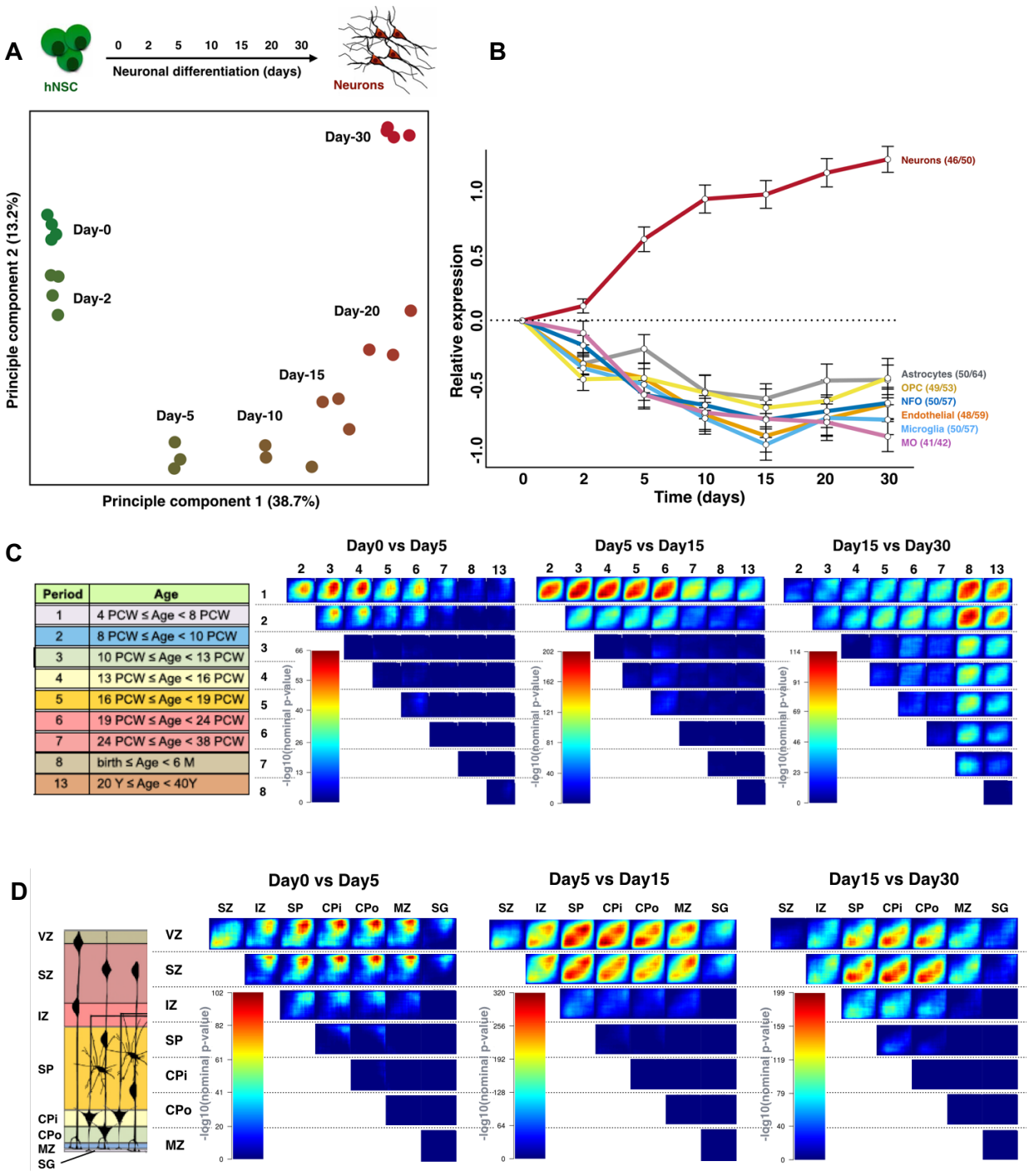
180 gene variable and GWAS gene level test statistic available. Testing only for a positive association,
181 i.e. enrichment of GWAS signal, we report one-sided p-values along with the corresponding
182 regression coefficient.
183
184 *Stratified LD Score Regression*
185 We applied a recent extension to stratified LD score regression (sLDSR), a statistical
186 method that partitions SNP-based heritability (h^2) from GWAS summary statistics[8]. This extension
187 allows us to quantify the effects of continuous-valued annotations on the heritability[32]. For each
188 annotation, we first estimated partitioned LD scores using the ldsc.py --l2 function with MAF >
189 5%, a 1 centimorgan (cm) window, and an ancestry-match 1KG reference panel (Supplementary
190 Methods). We ran sLDSR (ldsc.py --h2) for each annotation of interest while accounting for the
191 full baseline model, as recommended by the developers[8, 32], and an extra annotation of all genes
192 detected in our *in vitro* model (n = 12,414). That is, for each annotation we ran the following model;
193 1. Full baseline model with 53 annotations.
194 2. Annotation of all genes detected during *in vitro* neuronal differentiation.
195 3. Annotation of interest (e.g. cluster membership).
196 If an annotation of interest (3) is associated with increased h^2, LD to SNPs with large
197 values of that annotation will increase the χ^2 statistic of a SNP more than LD to SNPs with smaller
198 values. To determine if this effect is significant and specific to this annotation, it estimates the
199 contribution of that annotation to the per-SNP h^2 while accounting for the baseline and the all
200 genes detected annotation (1 + 2). As we only test for a positive association, we report the
201 contribution to the per-SNP h^2 (τ) and the associated one-sided p-value, which is calculated using
202 standard errors that are obtained via a block jackknife procedure[8, 33].
203
204 **Statistical analyses**

205 Statistical analyses were performed with R (https://www.r-project.org) or an otherwise specified
206 algorithm. Significance with MAGMA and sLDSR was determined at a one-sided α level of 0.05.
207 If applicable, Bonferroni correction for multiple comparison is applied and denoted in figures and
208 tables. For all analyses, further details on statistics are available in the supplementary results.
209

210 **Results**

211 ***Longitudinal in vitro gene expression profiling confirms neuron-specific differentiation***
212 ***and matches in vivo human cortical development***

213 To study the molecular dynamics underlying *in vitro* human neuronal differentiation, we
214 differentiated an hNSC line (WA09/H9) to a neuronal lineage across 30 days. Genome-wide gene
215 expression profiles were assayed densely at seven time points in at least triplicates (n=24
216 samples). Principal component analysis (PCA) on normalized gene expression values shows a
217 large proportion of the variance in expression to be explained by the differentiation process, with
218 minimal effects of technical variation (Figure 1A & S1). Investigation of transcriptome-based cell
219 type-specific gene expression signatures of major classes of cell types in the cerebral cortex
220 shows that relative neuronal gene expression increases as neuronal differentiation progresses
221 over time (Figure 1B). There is no evidence of glial- or endothelial-specific gene expression, which
222 confirms a broadly neuronal *in vitro* cellular identity.



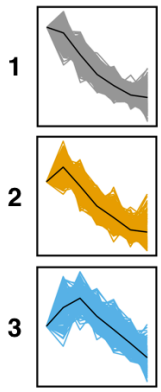
Period	Age
1	4 PCW ≤ Age < 8 PCW
2	8 PCW ≤ Age < 10 PCW
3	10 PCW ≤ Age < 13 PCW
4	13 PCW ≤ Age < 16 PCW
5	16 PCW ≤ Age < 19 PCW
6	19 PCW ≤ Age < 24 PCW
7	24 PCW ≤ Age < 38 PCW
8	birth ≤ Age < 6 M
13	20 Y ≤ Age < 40Y

Figure 1. *In vitro* gene expression profiles confirm a neuron-specific differentiation process and match *in vivo* human cortical development. (A) PCA of *in vitro* transcriptomic data with PC1 (x-axis) and PC2 (y-axis) visualized. Variance explained per component is shown in parentheses. Time points are color-coded and labeled by days across differentiation. (B) Transcriptome-based cellular identity is shown by average expression of cell type specific genes across days of

252 differentiation. Cell types are highlighted by their name and corresponding color. The first number in
253 the parentheses represents the number of genes for which the average expression is plotted. The
254 second number represents the corresponding number of probes assayed. OPC = oligodendrocyte
255 precursor cells, NFO = newly formed oligodendrocytes, MP = myelinating oligodendrocytes. (C+D)
256 TMAP output visualizes the amount of overlap between *in vitro* and *in vivo* DGE profiles colored by –
257 log10(p-value) (see figure S2 for more details on interpretation). Abbreviations and numbering above
258 maps correspond to schematic representations on the left (adopted from Stein et al., 2014) of different
259 developmental stages (C) and laminae (D). VZ = ventricular zone, SZ = subventricular zone, IZ =
260 intermediate zone, SP=subplate zone, CPi= inner cortical plate, CPo = outer cortical plate, MZ =
261 marginal zone, PCW = post conception weeks, M = months, Y = years, Period = developmental stage.
262

263 Having established that the *in vitro* differentiation process is predominantly neuronal, we
264 applied transition mapping (TMAP) to assess the correspondence of longitudinal *in vitro*
265 transcriptome data to *in vivo* signatures of human cortical development. TMAP uses a
266 spatiotemporal transcriptome atlas of the human neocortex and laminar expression data to
267 assess global overlap in differential gene expression (DGE) profiles between *in vitro* time points
268 and *in vivo* brain developmental stages or laminae of the human neocortex. We find significant
269 matching between the *in vitro* longitudinal DGE profiles (day-0 vs day-30) and *in vivo*
270 developmental stage from 4 weeks post-conception (PCW) to 24 PCW (Figure S2). This overlaps
271 with the primary period of neurogenesis in the neocortex, which starts around 6 PCW[34, 35]. To gain
272 more insight into this overlap, we partitioned the TMAP analyses in three comparisons and
273 examined how *in vitro* to *in vivo* matching progressed over time across *in vitro* neuronal
274 differentiation. We see a clear progression in matching from early developmental stages to later
275 stages (Figure 1C). For example, *in vitro* day-0 vs day-5 show strong overlap with *in vivo* period-
276 1 (4-8 PCW) vs period-4 (13-16 PCW), while *in vitro* day-15 vs day-30 shows stronger overlap
277 with *in vivo* period-2 (8-10 PCW) vs period-8 (birth-6M). Similarly, *in vitro* longitudinal DGE shows
278 progression from overlap of early time points with inner laminae, to overlap with more upper
279 cortical layers as *in vitro* neuronal differentiation advances (Figure 1D and S2).
280
281 ***In vitro neuronal differentiation reveals specific longitudinal gene clusters***
282 To identify biological pathways associated with neuronal differentiation, we applied an
283 analysis framework specifically tailored to time-series gene expression data (see Methods and
284 Supplementary Methods). A total of 7,734 probes, mapping to 5,818 genes, were differentially
285 expressed over time (Figure S3). We find that these genes are, on average, more constrained to
286 genetic variation compared to non-differentially expressed genes (Supplementary Results). Using
287 only differentially expressed probes, we next applied fuzzy c-means clustering and identified eight
288 distinct longitudinal gene clusters (Figure 2 and S4). For each probe, we generated a

289 corresponding cluster membership value, representing the degree to which a gene belongs to a
290 cluster. To identify most informative biological interpretation of each cluster, we analyzed genes
291 with high cluster membership for enrichment of functional annotations using DAVID
292 (Supplementary Methods and Table S1).



	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 2. Identified gene clusters highlight biological pathways important for neuronal differentiation. 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.

300 We identified three clusters with decreasing gene expression over time that are
301 significantly enriched for cell division and RNA regulation and processing genes, reflective of stem
302 cell proliferation and cell fate determination that is tightly controlled and regulated by RNA
303 dependent processes[36]. Second, there are three clusters showing increased gene expression
304 levels over time that are primarily enriched for neuronal processes, such as neuron formation and
305 synaptic function. Another independent cluster shows an inverted U-shaped expression pattern
306 during development, enriched for genes involved in transcriptional regulation. The final cluster is
307 enriched for genes involved in extracellular region and cell adhesions. These processes are
308 important for cell connectivity and have also been implicated in cell proliferation and neuronal
309 migration[37, 38]. Together, these eight gene clusters reveal different biological mechanisms that are
310 associated with neuronal differentiation and consistent with known biology of neurodevelopment.
311 We hypothesize that the study of these longitudinal gene expression clusters can help decipher
312 disease mechanisms involved in psychiatric phenotypes.
313
314 ***Differentially expressed genes are enriched for polygenic psychiatric disease risk***
315 To examine how aggregate psychiatric disease risk is distributed across genes that are
316 important for neuronal differentiation, we applied gene-set analysis and partitioning of h^2 with
317 MAGMA and sLDSR, respectively. We investigated the cumulative impact of risk loci associated
318 with major psychiatric disorders using GWAS results from large-scale studies. See
319 Supplementary Table S2 for details on all included phenotypes. For MDD, we included GWAS
320 results from the China Oxford and VCU Experimental Research on Genetic Epidemiology
321 (CONVERGE) consortium[23] and 23andMe Inc., a personal genetics company[11]. The latter uses
322 SRD as a proxy for major depression. Alzheimer's disease (AD) and adult human height served
323 as non-psychiatric control phenotypes that are heritable and polygenic. We used a two-step
324 approach where we first investigated disease susceptibility on overall differential expression level
325 and subsequently proceeded to deconstruct these associations across the longitudinal gene

326 clusters. Together, these analyses allow us to integrate *in vitro* longitudinal transcriptomic
327 signatures with GWAS summary statistics and assess if our approach can capture the polygenic
328 nature of psychiatric disorders.

329 We find that genes that are differentially expressed across *in vitro* neuronal differentiation
330 are enriched for genetic risk of multiple psychiatric disorders. We find significant effects with
331 MAGMA for SCZ (P=0.001), ADHD (P=0.002), and SRD (P=0.003) (Table 1 and Table S3). With
332 sLDSR, we find nominally significant effects for SCZ (P=0.01) and SRD (P=0.02) and a
333 suggestive association for ADHD (P=0.06) (Table 1 and Table S4). We observed a suggestive
334 enrichment for BPD, and no enrichment for the cross disorder, ASD, MDD CONVERGE or for
335 adult height and AD.

	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} ($1.02 \times 10^{-9)}$	0.45
Non-brain					
Height	0.009 (0.011)	0.037	0.21	-1.62×10^{-9} ($1.36 \times 10^{-9)}$	0.88

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.

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 $P=5.0x10^{-7}$, sLDSR $P=6.1x10^{-5}$) and not by genes that are downregulated (MAGMA P=0.98, sLDSR P=0.61) (Figure 3 and Figure S5). For SRD, we only find a stronger enrichment in upregulated genes with MAGMA ($P=3.5x10^{-4}$), while ADHD shows no specific evidence for either up or downregulated genes.

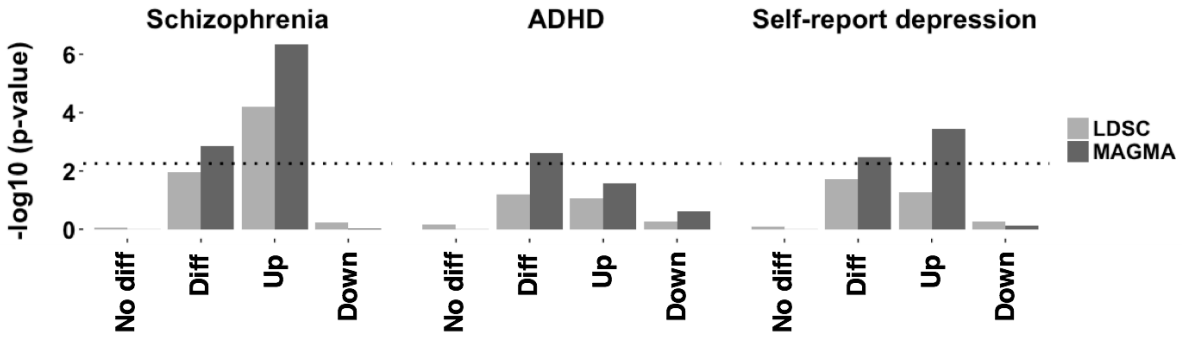


Figure 3. Schizophrenia polygenic risk lies in genes up-regulated during neuronal differentiation. A more detailed investigation of the effect of differentially expressed genes on the heritability of SCZ, ADHD, and SRD. The y-axis denotes the –log10 P-value of the enrichment. No diff = genes that are not differentially expressed; Diff = log (T^2-statistic) as shown in Table 1; Up = genes up-regulated during differentiation; Down = genes down-regulated during differentiation. The dotted line represents the threshold for P=0.0056 (n=9 traits).

367 ***Psychiatric disease risk aggregates to specific longitudinal gene clusters***

368 Next, we explored the relationship between differentially expressed genes and disease
369 risk on cluster level. For this analysis, we only included traits that show significant disease
370 enrichment across differentially expressed genes using MAGMA after correcting for multiple
371 testing (SCZ, ADHD, SRD) and our control traits (AD, height). These disease traits showed at
372 least a nominally significant effect with sLDSR as well. Using both MAGMA and sLDSR, we
373 integrated cluster membership values with GWAS summary statistics (n=5) and assessed
374 whether genome-wide disease risk aggregates to any of the eight experimentally identified
375 longitudinal gene clusters. Overall, MAGMA and sLDSR show a strong concordance across
376 phenotypes and clusters (rho = 0.92, $p<2.2x10^{-16}$, n=40, see also Figure S6). After Bonferroni
377 correction (n=40), we find five significant phenotype-cluster associations with MAGMA and three
378 with sLDSR (Figure 4 and Table S5/S6).

379 We find that multiple upregulated clusters show enrichment for SCZ with the strongest
380 evidence for the *synaptic function* cluster (MAGMA $P=1.8x10^{-7}$, sLDSR $P=7.2x10^{-5}$) (see Figure
381 S7). For SRD, we find significant associations in the *transcription regulation* ($P=2.5x10^{-5}$) and the
382 *neuron formation* ($P=1.2x10^{-4}$) gene cluster with MAGMA only. While the analysis of adult height
383 using all differentially expressed genes did not yield any evidence for enrichment of genetic signal,
384 enrichment is observed at the cluster level. The *cell connectivity* cluster ($P=3.7x10^{-4}$) is enriched
385 for height, in addition to suggestive enrichments in the *cell division* and *RNA regulation* cluster,
386 which are not present for any of the psychiatric phenotypes. Remarkably, across all 8 clusters the
387 enrichments of SCZ and height are inversely correlated (rho=-0.85, P=0.011, n=8, see also Figure
388 S8).

389

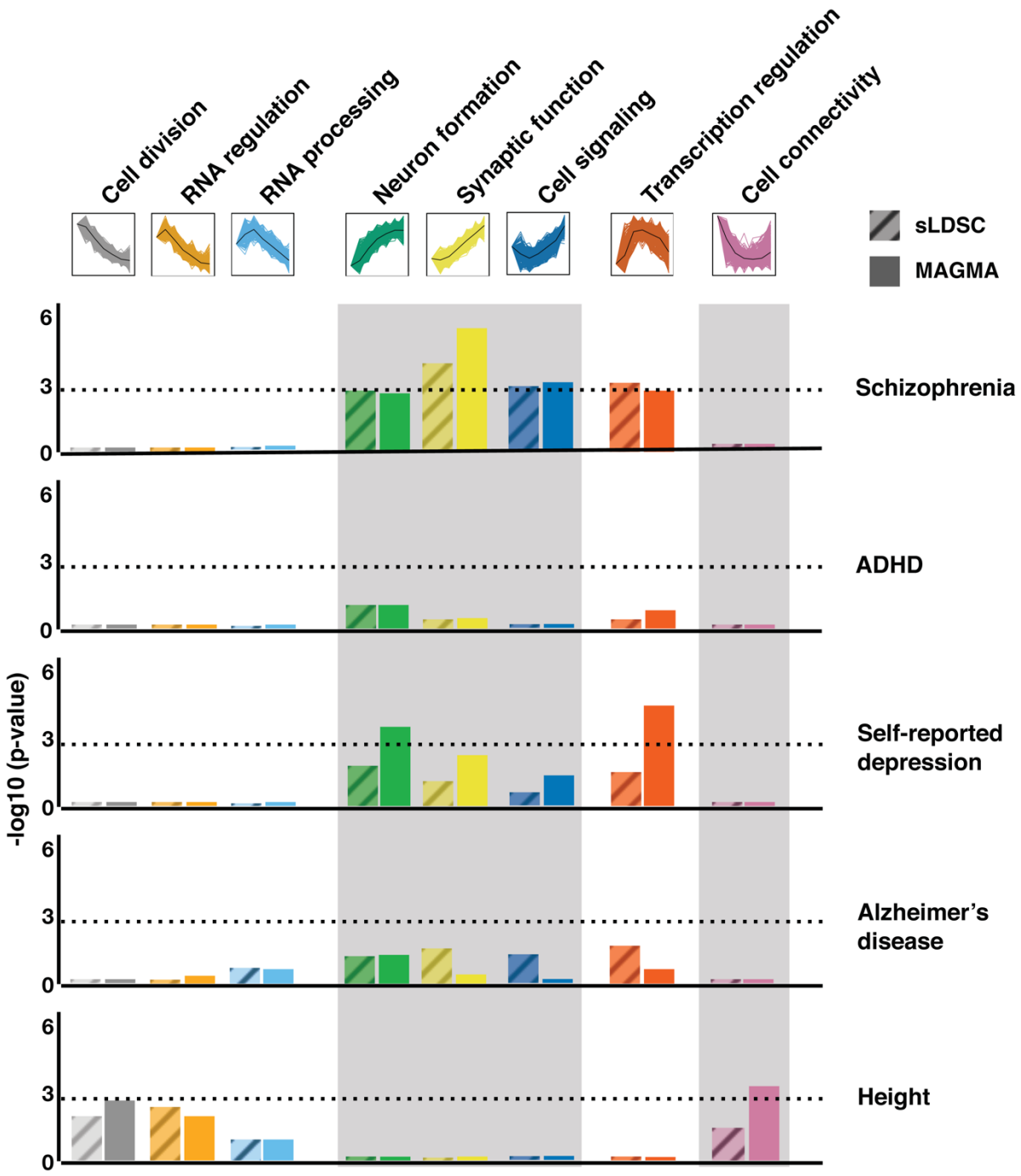


Figure 4. Polygenic psychiatric risk is distributed across specific longitudinal gene clusters. Results from sLDSC (diagonal pattern) and MAGMA (solid colors) are shown for each phenotype (labels on the right) colored by gene cluster. Gene cluster annotation and cluster expression pattern are shown on top. The y-axis states the –log10 (p-value). The dotted horizontal line represents the threshold for Bonferroni correction (p=0.05/40).

398 Finally, in order to take into account the full spectrum of correlations and dependencies
399 between clusters (Figure S9), we performed a conditional analysis for SCZ, the trait for which the
400 strongest cluster enrichments are observed with both methods. Using the same MAGMA model,
401 for each cluster, we conditioned on the highest gene members (membership > 0.5) of the other
402 seven clusters (Table 2). We find that the SCZ enrichment is driven by the *synaptic function*
403 cluster ($p=2.88x10^{-3}$) only. The same conditional analysis for SRD, which only showed a
404 significant enrichment with MAGMA, shows that this effect is primarily driven by the *transcription*
405 *regulation* cluster ($p=5.42x10^{-3}$) (Table S7).
406
407

	MAGMA Primary		**MAGMA Conditional**	
Schizophrenia - clusters	**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)	$2.12x10^{-3}$	0.018 (0.036)	0.30
Synaptic function	**0.077 (0.017)**	**$1.82x10^{-6}$**	**0.070 (0.026)**	**$2.88x10^{-3}$**
Cell signaling	0.052 (0.016)	$6.88x10^{-4}$	0.032 (0.023)	0.08
Transcription regulation	0.048 (0.016)	$1.67x10^{-3}$	0.019 (0.025)	0.22
Cell connectivity	-0.061 (0.017)	1.00	-0.076 (0.026)	1.00

408
409 **Table 2. The association with SCZ risk is driven by the synaptic function gene cluster.** Gene
410 level association signal is regressed on cluster membership while adjusting for high membership
411 genes of all other seven clusters. Shown are the results of the primary analysis (not adjusted for other
412 clusters) and the conditional analysis with MAGMA. Beta = regression coefficient, SE = standard error.
413

414 **Discussion**

415 We investigated a longitudinal *in vitro* stem cell model of human neuronal differentiation to
416 study psychiatric disease susceptibility based on evidence from GWAS. Among five major
417 psychiatric disorders, we observe that SCZ disease susceptibility is significantly enriched in a set
418 of genes relevant to *synaptic functioning* that are upregulated during differentiation. We therefore
419 propose *in vitro* human neuronal differentiation as an experimental system to further understand
420 and decipher the polygenic architecture of psychiatric disease biology, in particular that of SCZ.

421 We confirmed that our *in vitro* model recapitulates neuronal signatures of *in vivo* cortical
422 development across specific developmental time periods and laminae of the human neocortex.
423 This is in line with previous findings[14] and highlights that longitudinal gene expression dynamics
424 underlying our model of human neuronal differentiation can be informative to study genes and
425 pathways involved in *in vivo* human cortical development. Importantly, neuronal cell types[39, 40]
426 and early brain development[7, 22, 41] have been postulated as integral components of SCZ disease
427 susceptibility. Here, we observe that genes differentially expressed across neuronal differentiation
428 are significantly associated with genome-wide disease risk of SCZ and that this risk aggregates
429 in genes involved in synaptic functioning during development. Although not the only pathogenic
430 process contributing to SCZ, synaptic dysfunction is most strongly supported by genetic data,
431 postmortem expression studies, and animal models[40, 42–45]. We are the first to provide evidence
432 for this hypothesis using a longitudinal *in vitro* cell-based model and aggregate polygenic disease
433 risk. Two of the highest gene members of the *synaptic function* gene cluster enriched for SCZ
434 include Calcium Voltage-Gated Channel Subunit Alpha 1C (*CACNA1C*) and Solute Carrier Family
435 45 Member 1 (*SLC45A1*), both located at a genome-wide significant SCZ locus[22]. We find no
436 evidence for AD, a late-onset non-psychiatric brain disease, nor for adult human height in this
437 neuronal cluster. Together, our findings demonstrate that longitudinal transcriptomic signatures
438 important for neuronal differentiation recapitulate the *in vivo* context *and* align with the genetic

439 basis of the disease. SCZ disease biology, and in particular synaptic functioning, can thus be
440 studied through these molecular processes captured by this *in vitro* model.

441 We also observed a significant enrichment of genetic signal with MAGMA for SRD in
442 genes upregulated during differentiation, and show that this enrichment is predominantly driven
443 by genes in the *transcription regulation* gene cluster. Interestingly, the SRD GWAS reported that
444 the top SNPs were enriched for transcription regulation related to neurodevelopment[11], which is
445 in line with our *in vitro* findings. We observed no enrichment of the GWAS of recurrent and severe
446 MDD in Han-Chinese women[23]. The latter sample represents the most genetically and
447 phenotypically homogeneous GWAS of MDD. The fact that for these results no enrichment for
448 any of our gene sets was observed may suggest that neurodevelopmental processes play a lesser
449 role in MDD[46]. Alternatively, larger sample sizes are needed to better capture the genome-wide
450 genetic risk associated with MDD. Self-reported depression is a much broader phenotype that
451 may include other psychiatric traits, which could drive the observed neurodevelopment and
452 transcription findings. Although it remains unclear how these results and the application of the
453 model extrapolate to the MDD phenotype, our approach does highlight enrichment in distinct
454 clusters for SRD and SCZ and could help shed light on how these two complex traits differ in their
455 etiology.

456 We did not find any evidence of significant association in the neuronal clusters for ADHD.
457 This could be due to the smaller sample size in the GWAS studies and thereby lack of power to
458 find a significant association with our transcriptomic signatures (Figure S10). As GWAS sample
459 sizes are expected to increase, these gene cluster associations should be revisited.

460 For height, we found effects in opposite direction of psychiatric traits in the downregulated
461 gene clusters. Strikingly, we observe an inverse correlation between SCZ and height enrichment
462 stratified across gene clusters (Supplementary Results, Figure S8 and S11), despite the absence
463 of any evidence of a genetic correlation across the whole-genome[47]. These observations not only

464 illustrate the added value of individual longitudinal gene clusters, but also highlight a complex
465 genetic relationship between these two phenotypes.

466 A strength of our approach is the longitudinal analysis framework that we developed. We
467 implemented an experimental design across a dense and repeatedly sampled time-series and
468 integrated longitudinal transcriptomic signatures with genome-wide disease risk using available
469 GWAS summary statistics. This increases statistical power to directly investigate the cumulative
470 impact of risk loci on genes important to our model system. While we specifically chose to perform
471 our experiments across an isogenic background to minimize variation and maximize statistical
472 power to identify transcriptomic signatures, our framework can easily be extended to a multi-
473 sample design (e.g. cases vs controls)[19, 48], which makes it relevant for many disease-specific
474 experimental settings.

475 Our experimental procedure applied differentiation towards a broad neuronal phenotype.
476 Our work does not exclude disease associations with specific subtypes of neuronal cells nor with
477 other major brain cell types. We provide a proof-of-concept of an *in vitro* model of neuronal cells
478 for studying complex diseases, such as SCZ, and present an analytical framework that includes
479 longitudinal assessment of gene expression profiles. This approach can readily be extended to
480 study *in vitro* differentiation of other major brain cell types, such as astrocytes or oligodendrocytes.
481 Although we show strong evidence for SCZ risk in early prenatal neurodevelopment, our findings
482 do not preclude an additional contribution of postnatal neurodevelopment to the etiology of the
483 disease[49–51].

484 As GWAS risk loci have small effect sizes and are abundantly distributed across the
485 genome, new approaches are needed that allow for functional investigation of polygenic disease
486 architectures. Embracing the polygenic nature of psychiatric disorders is an important step
487 forward in translating findings from GWAS to disease biology[52]. Our approach uses the cumulative
488 impact of risk loci to identify gene clusters most strongly associated with SCZ while accounting
489 for all other pathways identified in our model. This strategy allowed us to narrow down on potential

490 core disease processes and opens up new avenues to study disease in the context of
491 polygenicity. Future work can for example incorporate model perturbations to study aggregate
492 disease risk in finer detail or use the model for functional fine-mapping of specific SCZ GWAS loci
493 across an isogenic background in a controlled environment. Alternatively, cluster-specific
494 polygenic risk scores (cPRS) can be used to study patients with a high genetic load versus those
495 with a low genetic load to study *in vivo* or *in vitro* molecular and cellular differences. This will not
496 only help elucidate the underlying disease mechanisms but can also provide insights into
497 phenotypic domains within the heterogeneous schizophrenia patient population.

498 In summary, the current study establishes WA09 neuronal differentiation as an *in vitro*
499 genomic tool to study psychiatric disease susceptibility. We demonstrate the value of an analytical
500 framework that integrates longitudinal *in vitro* transcriptomic signatures with GWAS summary
501 statistics to study the cumulative impact of risk loci on biological pathways. Overall, this work
502 contributes to understand the functional mechanisms that underlie psychiatric disease heritability
503 and polygenicity in the post GWAS era.

504

505 **Data Availability**

506 The Illumina HT-12 v4 gene expression data is available through the Gene Expression
507 Omnibus (GEO) archive (*Accession number is available for review*). This dataset has the raw and
508 normalized gene expression values on all samples. Supplementary table 8 furthermore has
509 specific probe annotations, such as probabilities of differential expression and probe membership
510 values for all identified clusters.

511

512 **Author Contribution:**

513 The project was led by R.O. Experiments were designed and conceived by A.O. and R.O.
514 Experiments were optimized, conducted and samples processed by A.O., M.B., and R.M.
515 Analysis of the data was performed by A.O and M.B and feedback provided by L.O. and R.O. The

main findings were interpreted by A.O., M.B., R.M., L.O., and R.O. Primary drafting of the manuscript was performed by A.O. and main feedback provided by R.O and L.O. All authors contributed to the production and approval of the final manuscript.

Conflict of Interest

The authors declare no competing interests.

Acknowledgement

We thank all research participants and researchers involved in making each GWAS 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 LD score regression team for their input and helpful troubleshooting with stratified LDSR. This research was supported by NIH/NIMH R01 MH090553.

542 **References**

543 1. Geschwind DH, Flint J. Genetics and genomics of psychiatric disease. *Science* 2015; **349**:
544 1489–1494.

545 2. Polderman TJC, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM *et*
546 *al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat*
547 *Genet* 2015; **47**: 702–709.

548 3. Sullivan PF, Agrawal A, Bulik C, Andreassen OA, Borglum A, Breen G *et al.* Psychiatric
549 Genomics: An Update and an Agenda. Am J Psychiatry 2017 (ahead of print).

550 4. Falk A, Heine VM, Harwood AJ, Sullivan PF, Peitz M, Brüstle O *et al.* Modeling psychiatric
551 disorders: from genomic findings to cellular phenotypes. *Mol Psychiatry* 2016; 1167–1179.

552 5. Gulsuner S, Walsh T, Watts AC, Lee MK, Thornton AM, Casadei S *et al.* Spatial and
553 temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical
554 network. *Cell* 2013; **154**: 518–529.

555 6. Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P *et al.* A polygenic
556 burden of rare disruptive mutations in schizophrenia. *Nature* 2014; **506**: 185–190.

557 7. Olde Loohuis LMO, Vorstman JAS, Ori AP, Staats KA, Wang T, Richards AL *et al.*
558 Genome-wide burden of deleterious coding variants increased in schizophrenia. *Nat*
559 *Commun* 2015; **6**: 7501.

560 8. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R *et al.* Partitioning
561 heritability by functional annotation using genome-wide association summary statistics. *Nat*
562 *Genet* 2015; **47**: 1228–1235.

563 9. Geschwind DH. Genetics of autism spectrum disorders. *Trends in Cognitive Sciences*.
564 2011; **15**: 409–416.

565 10. Rubeis SD, He X, Goldberg AP, Poultney CS, Samocha K. Synaptic, transcriptional, and
566 chromatin genes disrupted in autism A. *Nature* 2014; **515**: 209–215.

567 11. Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR *et al.* Identification of 15
568 genetic loci associated with risk of major depression in individuals of European descent.
569 *Nature Publishing Group* 2016; **48**: 1031–1036.

570 12. Shi Y, Kirwan P, Smith J, Robinson HPC, Livesey FJ. Human cerebral cortex development
571 from pluripotent stem cells to functional excitatory synapses. *Nat Neurosci* 2012; **15**: 477–
572 486.

573 13. van de Leemput J, Boles NC, Kiehl TR, Corneo B, Lederman P, Menon V *et al.*
574 CORTECON: A temporal transcriptome analysis of in vitro human cerebral cortex
575 development from human embryonic stem cells. *Neuron* 2014; **83**: 51–68.

576 14. Stein JL, de la Torre-Ubieta L, Tian Y, Parikshak NN, Hernández IA, Marchetto MC *et al.* A
577 quantitative framework to evaluate modeling of cortical development by neural stem cells.
578 *Neuron* 2014; **83**: 69–86.

579 15. Du P, Kibbe W a., Lin SM. lumi: a pipeline for processing Illumina microarray.
580 *Bioinformatics* 2008; **24**: 1547–1548.

581 16. Lin SM, Du P, Huber W, Kibbe W a. Model-based variance-stabilizing transformation for
582 Illumina microarray data. *Nucleic Acids Res* 2008; **36**: 1–9.

583 17. Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O'Keeffe S *et al.* An RNA-
584 Sequencing Transcriptome and Splicing Database of Glia, Neurons, and Vascular Cells of
585 the Cerebral Cortex. *Journal of Neuroscience* 2014; **34**: 11929–11947.

586 18. Tai YC, Speed TP. A multivariate empirical Bayes statistic for replicated microarray time
587 course data. *Ann Stat* 2006; **34**: 2387–2412.

588 19. Aryee MJ, Gutiérrez-Pabello JA, Kramnik I, Maiti T, Quackenbush J. An improved empirical
589 bayes approach to estimating differential gene expression in microarray time-course data:
590 BETR (Bayesian Estimation of Temporal Regulation). *BMC Bioinformatics* 2009; **10**: 409.

591 20. Kumar L, E Futschik M. Mfuzz: a software package for soft clustering of microarray data.
592 *Bioinformation* 2007; **2**: 5–7.

593 21. Schwämmle V, Jensen ON. A simple and fast method to determine the parameters for
594 fuzzy c-means cluster analysis. *Bioinformatics* 2010; **26**: 2841–2848.

595 22. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights
596 from 108 schizophrenia-associated genetic loci. *Nature*. 2014; **511**: 421–427

597 23. CONVERGE Consortium. Sparse whole-genome sequencing identifies two loci for major
598 depressive disorder. *Nature* 2015; **523**: 588.

599 24. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide
600 association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat*
601 *Genet* 2011; **43**: 977–983.

602 25. The Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium.
603 Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights
604 a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol Autism* 2017; **8**:
605 21.

606 26. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E *et al.* Discovery of the
607 first genome-wide significant risk loci for ADHD. *bioRxiv* 2017;
608 https://doi.org/10.1101/145581

609 27. Consortium C-DG of TPG. Identification of risk loci with shared effects on five major
610 psychiatric disorders: a genome-wide analysis. *Lancet* 2013; **381**: 1371–1379.

611 28. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C *et al.* Meta-
612 analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease.
613 *Nat Genet* 2013; **45**: 1452–1458.

614 29. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S et al. Defining the role of
615 common variation in the genomic and biological architecture of adult human height. *Nat*
616 *Genet* 2014; **46**: 1173–1186.

617 30. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang
618 HM *et al.* A global reference for human genetic variation. *Nature* 2015; **526**: 68–74.

619 31. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis
620 of GWAS Data. *PLoS Comput Biol* 2015; **11**.

621 32. Gazal S, Finucane HK, Furlotte NA, Loh P-R, Palamara PF, Liu X *et al.* Linkage
622 disequilibrium–dependent architecture of human complex traits shows action of negative
623 selection. *Nat Genet* 2017.

624 33. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Consortium SWG of TPG *et al.*
625 LD Score regression distinguishes confounding from polygenicity in genome-wide
626 association studies. *Nat Genet* 2015; **47**: 291–295.

627 34. Clancy B, Darlington RB, Finlay BL. Translating developmental time across mammalian
628 species. *Neuroscience* 2001; **105**: 7–17.

629 35. Stiles J, Jernigan TL. The basics of brain development. *Neuropsychology Review*. 2010;
630 **20**: 327–348.

631 36. Hattori A, Buac K, Ito T. Regulation of Stem Cell Self-Renewal and Oncogenesis by RNA-
632 Binding Proteins. In: RNA Processing Disease and Genome-wide probing. 2016. p. 153–
633 188.

634 37. Barros CS, Franco SJ, Muller U. Extracellular Matrix: Functions in the nervous system.
635 *Cold Spring Harb Perspect Biol* 2011; **3**: 1–24.

636 38. Bikbaev A, Frischknecht R, Heine M. Brain extracellular matrix retains connectivity in
637 neuronal networks. *Sci Rep* 2015; **5**: 14527.

638 39. Skene NG, Bryois J, Bakken TE, Breen G, Crowley JJ, Gaspar H *et al.* Genetic
639 Identification Of Brain Cell Types Underlying Schizophrenia. bioRxiv 2017;
640 http://dx.doi.org/10.1101/145466

641 40. Genovese G, Fromer M, Stahl EA, Ruderfer DM, Chambert K, Landén M *et al.* Increased
642 burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia.
643 *Nat Neurosci* 2016; **19**: 1433–1441.

644 41. Finucane H, Reshef Y, Anttila V, Slowikowski K, Gusev A, Byrnes A *et al.* Heritability
645 enrichment of specifically expressed genes identifies disease-relevant tissues and cell
646 types. *bioRxiv* 2017;https://doi.org/10.1101/103069

647 42. Hall J, Trent S, Thomas KL, O'Donovan MC, Owen MJ. Genetic risk for schizophrenia:
648 Convergence on synaptic pathways involved in plasticity. *Biological Psychiatry*. 2015; **77**:
649 52–58.

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

653 44. Pocklington AJ, O'Donovan M, Owen MJ. The synapse in schizophrenia. *Eur J Neurosci*
654 2014; **39**: 1059–1067.

655 45. Schwarz E, Izmailov R, Lio P, Meyer-Lindenberg A. Protein Interaction Networks Link
656 Schizophrenia Risk Loci to Synaptic Function. *Schizophr Bull* 2016; **42**: 1334–1342.

657 46. Peterson RE, Cai N, Bigdeli TB, Li Y, Reimers M, Nikulova A *et al.* The Genetic
658 Architecture of Major Depressive Disorder in Han Chinese Women. *JAMA Psychiatry* 2017;
659 **74**: 162–168.

660 47. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Consortium R *et al.* An Atlas of
661 Genetic Correlations across Human Diseases and Traits. *Nat Genet* 2015; **47**: 1237–1241.

662 48. Tai YC, Speed TP. A multivariate empirical Bayes statistic for replicated microarray time
663 course data. *Ann Stat* 2006; **34**: 2387–2412.

664 49. Birnbaum R, Jaffe AE, Hyde TM, Kleinman JE, Weinberger DR. Prenatal expression
665 patterns of genes associated with neuropsychiatric disorders. *Am J Psychiatry* 2014; **171**:
666 758–767.

667 50. Pers TH, Timshel P, Ripke S, Lent S, Sullivan PF, O'Donovan MC *et al.* Comprehensive
668 analysis of schizophrenia-associated loci highlights ion channel pathways and biologically
669 plausible candidate causal genes. *Hum Mol Genet* 2015; **25**: 1247–1254.

670 51. Sekar A, Bialas AR, de Rivera H, Davis A, Hammond TR, Kamitaki N *et al.* Schizophrenia
671 risk from complex variation of complement component 4. *Nature* 2016; **530**: 177–183.

672 52. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to
673 Omnigenic. Cell 2017; 169: 1177–1186.