ARTICLE

Transcriptional programs regulating neuronal differentiation are disrupted in DLG2 knockout human embryonic stem cells and enriched for schizophrenia and related disorders risk variants

Bret Sanders1, Daniel D’Andrea2, Mark O. Collins3, Elliott Rees2, Tom G. J. Steward4, Ying Zhu1, Gareth Chapman1, Sophie E. Legge2, Antonio F. Pardiñas2, Adrian J. Harwood1, William P. Gray1, Michael C. O’Donovan2, Michael J. Owen1,2, Adam C. Errington1, Derek J. Blake2, Daniel J. Whitcomb4, Andrew J. Pocklington2, & Eunju Shin1,5

Coordinated programs of gene expression drive brain development. It is unclear which transcriptional programs, in which cell-types, are affected in neuropsychiatric disorders such as schizophrenia. Here we integrate human genetics with transcriptomic data from differentiation of human embryonic stem cells into cortical excitatory neurons. We identify transcriptional programs expressed during early neurogenesis in vitro and in human foetal cortex that are down-regulated in \( DLG2^{-/-} \) lines. Down-regulation impacted neuronal differentiation and maturation, impairing migration, morphology and action potential generation. Genetic variation in these programs is associated with neuropsychiatric disorders and cognitive function, with associated variants predominantly concentrated in loss-of-function intolerant genes. Neurogenic programs also overlap schizophrenia GWAS enrichment previously identified in mature excitatory neurons, suggesting that pathways active during prenatal cortical development may also be associated with mature neuronal dysfunction. Our data from human embryonic stem cells, when combined with analysis of available foetal cortical gene expression data, de novo rare variants and GWAS statistics for neuropsychiatric disorders and cognition, reveal a convergence on transcriptional programs regulating excitatory cortical neurogenesis.

1Neuroscience and Mental Health Research Institute, Cardiff University, Cardiff CF24 4HQ, UK. 2MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff CF24 4HQ, UK. 3Department of Biomedical Science, University of Sheffield, Sheffield S10 2TN, UK. 4Bristol Medical School, University of Bristol, Bristol BS1 3NY, UK. 5School of Life Sciences, Keele University, Keele ST5 5BG, UK. *EBmail: pocklingtonaj@cardiff.ac.uk; e.shin@keele.ac.uk

https://doi.org/10.1038/s41467-021-27601-0

NATURE COMMUNICATIONS | (2022)13:27 | https://doi.org/10.1038/s41467-021-27601-0 | www.nature.com/naturecommunications
S
chizophrenia (SZ) is a highly heritable1–3 psychiatric disorder, with genetic variation ranging from common polymorphisms (SNPs) to rare mutations contributing to disease risk4–7. Rare variant studies consistently implicate disruption of postsynaptic signaling complexes in SZ etiology8–11, however the cellular pathways mediating common variant risk (an estimated 30–50% of the total genetic contribution to liability5) remain unclear. Genome-wide association studies (GWAS) have shown SZ common variant enrichment in broad, synapse-related gene sets12,13, but these sets only capture a modest proportion of the overall common variant association signal12. In contrast, nearly 50% of genic SNP-based heritability is captured by loss-of-function intolerant (LoFi) genes12. Being under extreme selective constraint, LoFi genes are likely to play important developmental roles. Indeed, LoFi genes are enriched for rare variants contributing to autism spectrum disorders (ASD) and intellectual disability/severe neurodevelopmental delay (ID/NDD)14, conditions that manifest early in life. Rare variation in LoFi genes, including many of those implicated in ASD and ID/NDD, also contributes to SZ2,14–16. We hypothesized that a significant proportion of SZ common variants may contribute to disease via the disruption of neurodevelopmental pathways harboring a concentration of LoFi genes.

Supporting a neurodevelopmental role for SZ common variants, there is growing evidence that many such risk factors impact gene expression in the fetal brain17–20 and are enriched in cell-types at multiple stages of cortical excitatory neuron development21. This raises the question: do SZ common variants converge on specific gene expression programs that are normally activated or repressed during fetal cortical excitatory neuron development? Mutations disrupting regulators of such programs would be expected to have a larger effect size and lower allele frequency than risk variants impacting individual genes within the program. We therefore sought rare, single-gene mutations linked to SZ where the affected gene is expressed in human fetal brain and has the potential to regulate developmental processes. This led us to DLG2. Firstly, multiple independent deletions have been identified at the DLG2 locus in SZ and ASD patients8,22. Secondly, DLG2 mRNA is present from 8 weeks post-conception in humans23 and throughout all stages of in vitro differentiation from human embryonic stem cells (hESCs) to cortical projection neurons24. Thirdly, suggesting a potential regulatory role for DLG2 during early development, the single invertebrate orthologue of DLG1–4 (Dig) is a core component of the Scrib signaling module, which regulates cell polarity, differentiation and migration during development25. Primarily studied as a post-synaptic scaffold protein, DLG2 is required for the formation of NMDA receptor complexes26. These complexes regulate the induction of several forms of synaptic plasticity27 and are enriched for rare mutations in SZ cases4,5,8–11. This raises the possibility that DLG2 may be required for the normal operation of both adult and developmental signaling pathways relevant to SZ pathophysiology.

To explore the role of DLG2 in neurodevelopment we engineered hESCs with homozygous loss-of-function DLG2 mutations (DLG2−/−) using the CRISPR-CAS9 system. DLG2−/− knockout (KO) and isogenic wild-type (WT) hESC lines were differentiated into cortical excitatory neurons and cells were characterized at multiple developmental timepoints to identify phenotypes and gene expression changes in KO lines (Fig. 1a). Neurodevelopmental gene expression programs dysregulated in DLG2−/− lines were identified and analyzed for risk variant enrichment, first for SZ and then for related disorders. We explored the biological function of disease-associated programs, both computationally and experimentally, and evaluated the contribution of LoFi genes to common and rare variant associations. Returning to SZ, we investigated the relationship between developmental and mature neuronal pathways enriched for common variant association. Finally, we explored whether disease-associated neurogenic programs identified in vitro possessed a similar profile of expression across neurodevelopmental cell-types in human fetal cortex.

Results

**Knockout generation and validation.** DLG2 contains three PDZ domains, an SH3 and a GK domain, all involved in protein binding. Two DLG2−/− lines were created from H7 hESCs using the CRISPR/Cas9-D10A nickase system targeting the first PDZ domain, generating a frameshift and premature stop codon in both alleles (Supplementary Fig. 1). Sequencing of predicted off-target sites revealed no mutations (Methods, Supplementary Fig. 2, Supplementary Data 1). All subsequent analyses compared these lines to an isogenic WT sister line that underwent the same procedure but remained genetically unaltered.

DLG2−/− and WT lines were differentiated into cortical excitatory neurons using a modified dual SMAD inhibition protocol28,29; RNA was extracted in triplicate from each line at 4 timepoints spanning cortical excitatory neuron development and gene expression quantified (Fig. 1b, Supplementary Fig. 3). A significant decrease in DLG2 mRNA was observed for exons and transcripts spanning the first PDZ domain, indicating degradation via nonsense-mediated decay in DLG2−/− lines (Supplementary Fig. 4). Quantitative mass spectrometry of peptide-affinity pulldowns using the NMDA receptor NR2 subunit PDZ peptide ligand30 identified DLG2 in WT only, confirming the absence of DLG2 in KO lines (Fig. 1c–f, Supplementary Data 2). Genotyping revealed no CNVs in either DLG2−/− line relative to WT (Supplementary Fig. 5a). Both DLG2−/− lines expressed pluripotency markers OCT4, SOX2 and NANOG at 100% of WT levels (Supplementary Fig. 5b). Cells were extensively characterized for their cortical identity using western blotting and immunocytochemistry from days 20–60. Over 90% of day 20 cells were positive for FOXG1, PAX6 and SOX2 and <1% cells expressed ventral genes such as DLX1, GBX2, NKX2.1 and OLIG3 (Supplementary Fig. 6), confirming dorsal forebrain fate. In addition, staining of markers expressed in ventral forebrain-derived neurons from striatal, thalamic and hypothalamic nuclei confirmed no or trace expression (Supplementary Fig. 6).

**DLG2−/− alters gene expression during cortical differentiation.** To robustly identify genes dysregulated by DLG2 knockout, expression data from the two DLG2−/− lines was pooled and compared to WT at each timepoint (Methods). Disruption of DLG2 had a profound effect: of the >13,000 protein-coding genes regulated at each timepoint displayed SZ association above that seen for neurodevelopmentally expressed genes in general, ~7% displayed altered expression at day 15, rising to 40–60% between days 20 and 30 then decreasing to ~25% by day 60 (Fig. 1g, Supplementary Data 3).

Common risk variants implicate disruption of neurogenesis in SZ. We next tested whether genes differentially expressed in DLG2−/− lines at each timepoint were enriched for SZ common risk variants. Taking summary statistics from the most recent SZ GWAS available31, we utilized the competitive gene-set enrichment test implemented in MAGMA32. As expected for cells of neural lineage, the set of all genes expressed at one or more timepoint in DLG2−/− or WT lines (allWT+KO) was enriched for common variant association (P = 8.03 × 10−21, Ngene = 14,274). To investigate whether genes up-/downregulated at each timepoint displayed SZ association above that seen for neurodevelopmentally expressed genes in general, we tested them for association conditioning on allWT+KO. This revealed enrichment solely for genes down-regulated at day 30.
$P_{\text{corrected}} = 9.5 \times 10^{-8}$ (Fig. 2a), coinciding with active neurogenesis (Fig. 1b). Conditioning on timepoint-specific expressed genes gave the same result (Supplementary Data 4).

Compared to all WT + KO, $30_{\text{down}^-}$ genes were over-represented in Gene Ontology (GO) terms related to neuronal development, function and migration (Methods, Supplementary Data 5). Iterative refinement via conditional analyses identified 23 terms with independent evidence for over-representation (Fig. 2b, Methods). This suggests that loss of DLG2 dysregulates transcriptional programs underlying neurogenesis (neuronal differentiation, migration and maturation) and implicates these processes in SZ etiology.
Fig. 1 Study design and number of differentially expressed genes. a Study summary. DLG2−/− and wild-type (WT) hESCs were differentiated into cortical excitatory neurons and RNA collected at multiple timepoints: predominant cell types shown for each timepoint (NSCs, neural stem cells; NPCs, neural precursor cells). Genetic analysis of differentially expressed genes (DEG) revealed SZ GWAS enrichment in genes down-regulated at day 30 in DLG2−/− lines, coinciding with early neurogenesis: corresponding phenotypes predicted via GO term analysis were validated experimentally. Transcriptional programs active during neurogenesis were identified based on differential gene expression between successive developmental timepoints. Schizophrenia (SZ) common variant risk was concentrated in two early neurogenic programs down-regulated in DLG2−/− cells. Loss of function intolerant (LoFi) genes were over-represented in early neurogenic (but not later) programs. LoFi genes in early neurogenic programs were enriched for common/rare variants contributing to mental disorders (ASD Autism spectrum disorder; ADHD attention-deficit hyperactivity disorder; BP Bipolar disorder; MDD major depression disorder) and cognition (IQ), but not unaffected siblings of ASD cases (SIB) or Alzheimer’s disease (AD). Overlap with early and late neurogenic programs captures SZ GWAS association previously reported for genes with high expression in CA1 pyramidal neurons relative to other brain cell-types (pyramidal68). The expression profile seen for each disease-associated neurogenic program in vitro was recapitulated across neurodevelopmental cell-types from human fetal cortex. b Overview of cortical differentiation protocol with approximate timings of key developmental processes and predominant cell types present in culture. Asterisks indicate timepoints selected for RNA sequencing. c-f Label free quantification (LFQ) of DLG2 protein levels in PDZ-ligand (NR2 C-terminus) affinity pulldowns (n = 4 biological repeats each) in day 30 and 60 WT and DLG2−/− cells using LC-MS/MS analysis. One-way ANOVA with Bonferroni multiple comparison correction applied to p = 0.0061. d, e, f corrected for this set (H(2) = 10.46, p = 0.0061) and f (H(2) = 11.00, p = 0.0061). *p < 0.05; **p < 0.01; ****p < 0.0001 vs. WT. g Number of protein coding genes differentially expressed in DLG2−/− cells relative to WT at each timepoint. Source data are provided as a Source Data file.

**DLG2−/− delays cortical cell-fate expression in newborn neurons.** To validate disruption of neurogenesis in DLG2−/− lines and investigate whether this leads to differences in the number or type of neurons produced, we compared the expression of cell-type specific markers in DLG2−/− and WT lines from days 30–60 via immunocytochemistry (ICC) and Western blotting (Fig. 2c–i). From ICC it was clear that DLG2−/− cells are able to differentiate and produce postmitotic neurons expressing characteristic neuronal markers such as NEUN and TUJ1 plus cortical deep layer markers TBR1 and BCL11B (CTIP2) (Supplementary Fig. 7). Western blot of NEUN (Fig. 2c) and MAP2 (Supplementary Fig. 7) and quantification of NEUN+ cells following ICC (Fig. 2f) revealed no difference in the percentage of neurons produced by DLG2−/− cultures. This is in line with the comparable percentage of cells in the cell cycle/neural progenitors at days 30–60 in DLG2−/− and WT cultures indicated by a similar proportion of Ki67+ and SOX2+ cells (Supplementary Fig. 7). At these early timepoints we would not expect to see the generation of upper layer neurons, which express markers such as SATB2. Although we could identify a small percentage of SATB2+ cells in both WT and KO lines, all co-expressed CTIP2 (Supplementary Fig. 7) indicating their deep layer identity. An analysis of deep layer markers TBR1 and CTIP2 revealed a significant decrease in CTIP2+ cells but a comparable proportion of TBR1+ neurons for all timepoints investigated (Fig. 2d, e, g–i). On average the proportion of CTIP2+ cells recovered from 15% of the WT level on day 30 to 50% by day 60, although there was notable variation between DLG2−/− lines (Supplementary Fig. 8); total CTIP2 protein levels also recovered to some extent, but at a slower rate (Supplementary Fig. 8). Thus, DLG2−/− does not affect the rate at which neurons are produced but delays the expression of subtype identity in newborn deep layer neurons.

**DLG2−/− lines display deficits in neuron morphology & migration.** Given the over-representation of 30down−/− genes in terms related to neuron morphogenesis and migration (Fig. 2b), we sought to experimentally validate these phenotypes. Immature (day 30) and mature (day 70) neurons were traced and their morphology quantified (Fig. 3). At both timepoints DLG2−/− neurons displayed a simpler structure than WT, characterized by a similar number of primary neurites projecting from the soma (Fig. 3a) but with greatly reduced secondary neurite branching (Fig. 3b). Total neurite length did not differ (Fig. 3c), leading to a clear DLG2−/− phenotype of longer, relatively unbranched primary neurites (Fig. 3e). There was no significant difference in soma area (Fig. 3d). Day 40 DLG2−/− neurons had a slower speed of migration (Fig. 3f) and reduced displacement from their origin after 70 h (Fig. 3g, h). In summary, DLG2−/− neurons show clear abnormalities in both morphology and migration, validating the GO term analysis.

**DLG2-regulated transcriptional programs enriched for SZ genetic risk.** We postulated that loss of DLG2 inhibits the activation of transcriptional programs driving neurogenesis, which starts between days 20 and 30 and steadily increases thereafter. If this is the case, SZ association in 30down−/− should be captured by genes normally upregulated between days 20 and 30 in WT cultures (20–30up WT). Analyzing differential expression between WT samples at successive timepoints, we found risk variant enrichment only in 20–30up WT (conditioning on all WT-expressed genes, Fig. 4a). Most 20–30up WT genes overlapped 30down−/− (3075 genes, 85%) and this overlap captured the signal in both sets (Poverlap = 3.23 × 10−10, 30down−/− only P = 0.44; 20–30up WT only P = 0.62). This was not simply due to the size of the overlap as the regression coefficient for this set (β = 0.14), which reflects magnitude of enrichment, was significantly greater than for genes unique to 30down−/− (β = 0.006, Pgreater = 0.00077) or 20–30up WT (β = 0.015, Pgreater = 0.0023). Thus, it is neurogenic transcriptional programs that are typically upregulated in WT but down-regulated in DLG2−/− lines that are enriched for SZ common variants.

To more precisely identify SZ-associated transcriptional programs active during neurogenesis, we classified 20–30up WT genes based on their subsequent WT expression profiles (Fig. 4b, Methods). These included early-increasing genes, whose expression continues to rise between days 30 and 60; early-stable genes, whose expression stays at a relatively constant level; and early-transient genes, whose expression is later down-regulated. We also defined a set of late genes, whose expression only increases significantly after day 30. These were further partitioned into genes that were down-regulated at day 30 in DLG2−/− lines (e.g., early-stable−/−) and those that were not (e.g., early-stable WT only). The sole exception to this was the late set, which had minimal overlap with 30down−/− (62 out of 1399 genes) and was therefore left intact.
Early-stable−/− and early-increasing−/− sets were robustly enriched for SZ association (Fig. 4c). To more precisely control for association specifically in neuron-expressed genes, we identified genes expressed in newborn and developing cortical excitatory neurons from a recent single-cell RNAseq study of human fetal brain tissue32 (Methods). Early-stable−/− and early-increasing−/− remained highly associated when conditioning on fetal neuron-expressed genes (Supplementary Data 6). Furthermore, allWT+KO displayed association that was not captured by fetal neuron-expressed genes. We therefore continued to condition genetic analyses on allWT+KO, as this best captures the broad SZ signal from neuronal-lineage genes present in our dataset.

| GO term                                                                  | Nmiss | OR   | Pcorrected | Pcorrected |
|------------------------------------------------------------------------|-------|------|------------|-----------|
| DNA binding transcription factor activity RNA polymerase II specific    | 594   | 1.86 | 2.36x10^-19 | 2.36x10^-19 |
| axon                                                                   | 149   | 3.18 | 4.76x10^-18 | 1.97x10^-18 |
| regulation of cell morphogenesis                                       | 164   | 2.72 | 3.07x10^-14 | 3.07x10^-14 |
| positive regulation of neuron projection development                    | 96    | 3.93 | 3.30x10^-13 | 3.30x10^-13 |
| neuron projection morphogenesis                                         | 157   | 2.60 | 1.85x10^-12 | 1.85x10^-12 |
| synaptic membrane                                                       | 80    | 4.30 | 7.29x10^-12 | 7.29x10^-12 |
| peptidylserine phosphorylation                                          | 82    | 3.53 | 1.49x10^-10 | 1.49x10^-10 |
| regulation of small GTPase mediated signal transduction                | 104   | 2.39 | 2.85x10^-14 | 2.85x10^-14 |
| calcium ion regulated exocytosis                                        | 23    | 10.77| 6.90x10^-15 | 6.90x10^-15 |
| synapse assembly                                                        | 38    | 4.05 | 9.00x10^-03 | 9.00x10^-03 |
| positive regulation of axonogenesis                                      | 30    | 5.02 | 0.00079     | 0.00079     |
| membrane depolarization                                                 | 27    | 5.27 | 0.0020      | 0.0020      |
| microtubule end                                                         | 20    | 7.80 | 0.0044      | 0.0044      |
| neurotransmitter receptor complex                                       | 26    | 5.08 | 0.0047      | 0.0047      |
| neuron recognition                                                      | 22    | 6.44 | 0.0050      | 0.0050      |
| regulation of adherens junction organization                            | 32    | 3.95 | 0.0054      | 0.0054      |
| modification dependent protein binding                                  | 62    | 2.39 | 0.0066      | 0.0066      |
| protein import into nucleus                                             | 37    | 3.34 | 0.0084      | 0.0084      |
| ubiquitin like protein ligase activity                                   | 80    | 2.09 | 0.0088      | 0.0088      |
| protein phosphorylated amino acid binding                               | 23    | 5.38 | 0.012       | 0.012       |
| regulation of synapse assembly                                          | 28    | 4.10 | 8.00x10^-03 | 8.00x10^-03 |
| neuron migration                                                        | 27    | 3.95 | 0.036       | 0.036       |
| calcium ion transport into cytosol                                       | 27    | 3.95 | 8.68x10^-03 | 8.68x10^-03 |
In summary, SZ GWAS association during early neurogenesis is restricted to 2 transcriptional programs down-regulated in DLG2−/− lines.

Transcriptional cascade predicted to drive early neurogenesis.

We next investigated the biological function of early neurogenic programs dysregulated in DLG2−/− lines. Each was over-represented for a coherent set of GO terms indicating a distinct biological role (Supplementary Data 7): early-transient−/− for histone/chromatin binding and transcriptional regulation; early-stable−/− for signal transduction, transcriptional regulation, neurogenesis, cell projection development, migration and differentiation; and early-increasing−/− for axon guidance, dendrite morphology, components of pre- and post-synaptic compartments and electrophysiological properties. These functions suggest a linked, time-ordered cascade of transcriptional programs spanning early neurogenesis. This begins with an initial phase of chromatin remodeling (early-transient−/−) that establishes neuron sub-type identity and leads to activation of a longer-term program guiding the growth and migration of newborn neurons (early-stable−/−). This in turn promotes the development and fine-tuning of sub-type specific neuronal structure, function and connectivity as cells enter the terminal phase of differentiation (early-increasing−/−).

To test support for the existence of such a cascade and its disruption in disease, we identified disease-associated regulatory genes from each program whose downstream targets have been experimentally identified or computationally predicted (Methods). Reflecting our hypothesis that dysregulation of these pathways is likely to play a role in multiple neurodevelopmental disorders, we sought regulators linked to SZ, ASD and ID/NDD. This led us to chromatin modifier CHD8 from early-transient−/−; transcription factor TCF4 from early-transient−/−; and transcription factors (and deep layer markers) TBR1 and BCL11B from early-increasing−/−.

To link successive phases of the cascade, we predicted that each program would be enriched for direct targets of regulators in the immediately preceding program: early-stable−/− for CHD8 targets, early-increasing−/− for TCF4 and FMRP. TBR1 and BCL11B play important roles in the developmental expression of sub-type specific properties and would be predicted to regulate early-increasing−/− genes. Since early-transient−/− is hypothesized to initiate the cascade, we predicted that early-increasing−/− would be enriched for indirect targets of CHD8 – genes not directly regulated but whose expression is dysregulated in CHD8 knockout cells. We also predicted that genes in the earliest, most transitory phase of the cascade (i.e., early-transient−/−) would not be enriched for targets of terminal phase regulators (BCL11B and TBR1). FMRP represses the translation of its mRNA targets, facilitating their translocation to distal sites of protein synthesis and its function is known to be important for axon and dendrite growth. We therefore predicted that early-stable−/− (but not early-transient−/−) would also be enriched for FMRP targets. Over-representation tests confirmed these predictions, supporting the existence of a regulatory cascade driving early neurogenic programs disrupted in neuropsychiatric disorders (Fig. 4d). In addition, the targets of TCF4, FMRP, BCL11B and TBR1 were more highly enriched for SZ association than other genes in early-increasing−/− (Fig. 4d), highlighting specific pathways through which these known risk genes may contribute to disease.

Convergence of genetic risk on perturbed action potential generation.

We next tested whether biological processes over-represented in early-stable−/− or early-increasing−/− (Supplementary Data 7) captured more or less of the SZ association in these programs than expected (Methods). Iterative refinement identified 13 GO terms with independent evidence for over-representation in early-stable−/−. Genetic association for these terms did not differ substantially from early-stable−/− as a whole (Supplementary Data 8), indicating that risk factors are distributed relatively evenly between them. None of the 16 independent terms identified for early-increasing−/− showed evidence for depleted association, suggesting that diverse biological processes regulating neuronal differentiation, morphology and function are perturbed in SZ. However, somatodendritic compartment and membrane depolarization during action potential were more highly associated than early-increasing−/− as a whole (Fig. 4e). Enhanced enrichment in action potential (AP) related genes is noteworthy: while postsynaptic complexes
regulating synaptic plasticity are robustly implicated in SZ, this suggests that the molecular machinery underlying AP generation is also disrupted. We therefore sought to confirm the disruption of APs in DLG2−/− lines (Fig. 5a–j), also investigating the impact of DLG2 loss on synaptic transmission (Fig. 5k–n).

In line with the above, DLG2−/− neurons were found to be less excitable, with immature AP waveforms. Day 50 DLG2−/− neurons displayed a significantly more depolarized resting membrane potential (Fig. 5a). Upon stepped current injection, 80% of WT neurons but only 43% of DLG2−/− neurons showed AP firing (Fig. 5c). APs produced by DLG2−/− neurons were characteristic of less mature neurons (Fig. 5d), having smaller amplitude, longer half-width and a slower maximum rate of depolarization and repolarization ($\delta V/\delta t$) (Fig. 5e–h). We found no change in AP voltage threshold, rheobase current (Fig. 5i, j) or input resistance (Fig. 5b). In addition, the percentage of neurons displaying spontaneous excitatory postsynaptic currents (EPSCs)
Neurogenic programs capture genetic association in LoFi genes. Having identified neurodevelopmentally expressed pathways enriched for common SZ risk variants and investigated the phenotypic consequences of their dysregulation in DLG2−/− lines, we sought to test our hypothesis that these pathways capture a significant proportion of the SZ GWAS enrichment seen in...
One-sided tests were performed using MAGMA, conditioning on all expressed genes (early-transient tested. Bold indicates tests surviving Bonferroni correction. Source data are provided as a Source Data corrected p and all early-increasing (of neurogenesis were identified based upon WT differential expression between timepoints: early-increasing, genes significantly upregulated between days 20 and 30 (20–60upWT) and also days 30 and 60 (30–60upWT); early-stable genes, present in 20–30upWT and 20–60upWT but not 30–60upWT; early-transient (20–30upWT but not 20–60upWT); and late (30–60upWT but not 20–60upWT). Sz GWAS enrichment in each transcriptional program, further split into genes that are down-regulated in DLG2−/− lines at day 30 (e.g., early-stable−/−) and those that are not (e.g., early-stableWT only). One-sided tests were performed using MAGMA, conditioning on all expressed genes (allWT+KO); raw and Bonferroni-corrected p values are given (pGWAS, pGWAScorrected), where correction is for the 7 gene-sets tested. d A one-sided, Fisher’s Exact Test was used to identify programs over-represented for the targets of key regulators when compared to allWT+KO; both raw and Bonferroni-corrected p values are given (pGWAS, pGWAScorrected), where correction is for the 21 tests performed (3 programs x 7 regulators). All 10 program-regulator enrichments with corrected p < 0.05 were taken forward for genetic analysis. Two-sided tests were performed in MAGMA to investigate whether regulator targets were more highly enriched for Sz association than other genes in that program, conditioning on allWT+KO and the program as a whole. Raw and Bonferroni-corrected p values are given (pGWAS, pGWAScorrected), where correction is for the 10 sets tested. e Sz GWAS enrichment in GO terms with independent evidence of over-representation among early-increasing−/− genes, two-sided tests were performed using MAGMA, conditioning on all expressed and all early-increasing−/− genes. Raw and Bonferroni-corrected p values are given (pGWAS, pGWAScorrected), where correction is for the 16 terms tested. Bold indicates tests surviving Bonferroni correction. Source data are provided as a Source Data file.

LoFi genes12. We predicted that LoFi genes would primarily be concentrated in earlier transcriptional programs where the impact of disruption is potentially more severe. LoFi genes were over-represented in all early neurogenic programs but notably depleted in the late set (Fig. 6a). LoFi Sz association was captured by the overlap with early-stable−/− and early-increasing−/−, localizing the GWAS signal to a fraction of LoFi genes (less than a third) located in specific neurogenic pathways (Fig. 6b).

Under our proposed model, early-transient−/− initiates activation of other early neurogenic programs, thus its dysregulation has the potential to cause more profound developmental deficits. We speculated that – while displaying no evidence for Sz GWAS association – LoFi genes in early-transient−/− would be enriched for rare mutations linked to Sz and/or more severe neurodevelopmental disorders. All early neurogenic programs displayed a markedly elevated rate of de novo LoF mutations relative to allWT+KO that was captured by LoFi genes: early-transient−/− for mutations identified in NDD and ASD cases47; early-stable−/− for NDD, ASD and SZ16; and early-increasing−/− for NDD (Fig. 6c). De novo LoF mutations from unaffected siblings of ASD cases47 showed no elevation. In all three programs, a clear gradient of effect was evident from NDD (largest elevation in rate) to ASD to Sz, visible only in LoFi genes (Fig. 6d). A modest gradient was also evident for LoFi genes lying outside early neurogenic programs (‘Other LoFi genes’, Fig. 6d), despite de novo rates not being robustly elevated here. This suggests the existence of additional biological pathways harboring disease-associated LoFi genes.

Given the robust rare variant enrichment across multiple disorders, we investigated whether neurogenic programs are also enriched for common variants contributing to disorders other than Sz, analyzing a range of conditions with which Sz is known to share heritability: ASD48, attention-deficit/hyperactivity disorder (ADHD)49, bipolar disorder (BP)50; and major depressive disorder (MDD)51. Since altered cognitive function is a feature of all these disorders, we also tested enrichment for common variants linked to IQ22. All disorders showed evidence for common variant enrichment in one or more early neurogenic program that was again captured by LoFi genes (Fig. 6e, f). In contrast, common variants conferring risk for the neurodegenerative disorder Alzheimer’s disease (AD)53 were not enriched. Whereas rare variant enrichment was concentrated towards the initial stages of the transcriptional cascade, GWAS association was confined to later stages (Fig. 6c–f). Dysregulation of transcriptional programs underlying cortical excitatory neurogenesis thus contributes to a wide spectrum of neuropsychiatric disorders. Furthermore, robust enrichment of early-stable−/− and early-increasing−/− for IQ association (Fig. 6e, f) suggests that perturbation of neurogenic programs may contribute to the emergence of cognitive symptoms in these disorders.

Sz association seen in mature neurons captured by neurogenic programs. DLG2 plays an essential role26 in scaffolding mature postsynaptic complexes implicated in Sz4-8-11. Our data indicates that it also regulates early developmental pathways harboring Sz genetic risk. To further explore the relationship between developmental and adult disease mechanisms, we investigated the extent to which genes from neurogenic programs contribute to Sz-associated biology in mature excitatory neurons. GWAS association has previously been noted in genes with relatively high expression in CA1 pyramidal neurons compared to other brain cell-types13,54. Although different to the sub-types generated by our in vitro protocol, we reasoned that developmental processes shared between these two dorsal forebrain-derived neuronal types are likely to account for a substantial proportion of the neurogenic programs we have identified. Taking the 10% of genes with the highest CA1 pyramidal neuron specificity score54 (pyramidalhigh) we investigated their overlap with neurogenic programs. Pyramidalhigh genes were over-represented in early-stable−/−, early-increasing−/− and late sets (Fig. 7a). This overlap captured GWAS association in pyramidalhigh, but not early-stable−/− or early-increasing−/− (Fig. 7b). In contrast to the late program as a whole (Figs. 4c and 6a), genes in the late-pyramidalhigh overlap were enriched for Sz association (Fig. 7b) and LoFi genes (OR = 1.43, P = 0.035). Late-pyramidalhigh genes also displayed a pattern of enrichment for regulatory targets almost identical to that of early-increasing−/− (Figs. 4d and 7c), linking this subset into the terminal phase of the hypothesized transcriptional cascade.

These analyses suggest that the Sz association seen in pyramidal excitatory neurons13,34 may arise from molecular pathways contributing to early neurogenesis that remain active in post-natal life. To investigate the nature of these pathways, we performed a functional analysis of Sz-associated gene-sets from Fig. 7b. Pyramidalhigh genes overlapping early-stable−/−, early-increasing−/− and late sets were over-represented for GO path-
terms linked to dendrite/spine development, calcium-mediated exocytosis, postsynaptic signaling and synaptic plasticity (Supplementary Data 9). In contrast, genes unique to early neurogenic programs were over-represented for terms linked to the regulation of transcription/neurogenesis, axonogenesis, axon guidance, pre-synaptic function and sodium channel activity (Supplementary Data 9). Thus it appears to be primarily postsynaptic processes regulating the formation, function and plasticity of synaptic connections throughout development (from the pre-natal period into adulthood) that underlies the mature neuronal contribution to SZ encapsulated by pyramidal high.

![Image](155x83 to 496x397)

**a** Resting membrane potential  
**b** Input resistance  
**c** Percentage of cells firing APs  
**d** Depolarisation δV/δt maximum  
**e** Repolarisation δV/δt maximum  
**f** Spike threshold voltage  
**g** Rheobase current  
**h** Patch example

| DLG2+/+ | DLG2-/− |
|---|---|
| mV | mV |
| 0 | 0 |
| 20 | 20 |
| 50 | 50 |
| 80 | 80 |

| DLG2+/+ | DLG2-/− |
|---|---|
| mV | mV |
| 0 | 0 |
| 20 | 20 |
| 50 | 50 |
| 80 | 80 |

| DLG2+/+ | DLG2-/− |
|---|---|
| pA | pA |
| 0 | 0 |
| 50 | 50 |
| 100 | 100 |
| 150 | 150 |

| DLG2+/+ | DLG2-/− |
|---|---|
| Inter-Event Interval (sec) | Inter-Event Interval (sec) |
| 0.0 | 0.0 |
| 0.1 | 0.1 |
| 0.2 | 0.2 |
| 0.3 | 0.3 |
| 0.4 | 0.4 |

**l** D50 DLG2+/+  
**m** D60 DLG2+/+  

| DLG2+/+ | DLG2-/− |
|---|---|
| Inter-Event Interval (sec) | Inter-Event Interval (sec) |
| 0.0 | 0.0 |
| 0.1 | 0.1 |
| 0.2 | 0.2 |
| 0.3 | 0.3 |
| 0.4 | 0.4 |

**o** DLG4 in D65 synaptosomes

| DLG2+/+ | DLG2-/− |
|---|---|
| pA | pA |
| 0 | 0 |
| 50 | 50 |
| 100 | 100 |
| 150 | 150 |

| DLG2+/+ | DLG2-/− |
|---|---|
| Cumulative probability | Cumulative probability |
| 0.0 | 0.0 |
| 0.1 | 0.1 |
| 0.2 | 0.2 |
| 0.3 | 0.3 |
| 0.4 | 0.4 |

| DLG2+/+ | DLG2-/− |
|---|---|
| Cumulative probability | Cumulative probability |
| 0.0 | 0.0 |
| 0.1 | 0.1 |
| 0.2 | 0.2 |
| 0.3 | 0.3 |
| 0.4 | 0.4 |

| DLG2+/+ | DLG2-/− |
|---|---|
| Cumulative probability | Cumulative probability |
| 0.0 | 0.0 |
| 0.1 | 0.1 |
| 0.2 | 0.2 |
| 0.3 | 0.3 |
| 0.4 | 0.4 |

| DLG2+/+ | DLG2-/− |
|---|---|
| Cumulative probability | Cumulative probability |
| 0.0 | 0.0 |
| 0.1 | 0.1 |
| 0.2 | 0.2 |
| 0.3 | 0.3 |
| 0.4 | 0.4 |

| DLG2+/+ | DLG2-/− |
|---|---|
| Cumulative probability | Cumulative probability |
| 0.0 | 0.0 |
| 0.1 | 0.1 |
| 0.2 | 0.2 |
| 0.3 | 0.3 |
| 0.4 | 0.4 |

| DLG2+/+ | DLG2-/− |
|---|---|
| Cumulative probability | Cumulative probability |
| 0.0 | 0.0 |
| 0.1 | 0.1 |
| 0.2 | 0.2 |
| 0.3 | 0.3 |
| 0.4 | 0.4 |

| DLG2+/+ | DLG2-/− |
|---|---|
| Frequency (Hz) | Frequency (Hz) |
| 0.0 | 0.0 |
| 0.1 | 0.1 |
| 0.2 | 0.2 |
| 0.3 | 0.3 |
| 0.4 | 0.4 |
Neurogenic programs identified in vitro are expressed in human fetal cortex. Early neurogenic programs are enriched for variants contributing to cognitive function and the pathogenesis of neuropsychiatric disorders. However, these programs were identified from bulk RNAseq data in vitro and it remains to be shown that their constituent genes are actively co-expressed in the appropriate cell-types during cortical excitatory neurogenesis in vivo. To address this, we extracted gene expression data for cell-types spanning cortical excitatory neurodevelopment from a single-cell RNAseq study of human fetal brain.32 After normalizing the expression for each gene across all cells, we calculated the average expression for each gene in each cell-type/stage of development available: early radial glia (early RG), RG, intermediate progenitor cells (IPCs), transitioning cells (intermediate between progenitors and neurons), newborn and developing neurons (Methods). We then plotted the expression of each program (mean and standard error of gene-level averages) in each cell-type/stage and tested for differences in expression between successive types/stages (Fig. 7d, Supplementary Data 10). The expression profile seen for each program in vitro was recapitulated across neurodevelopmental cell-types from human fetal cortex (Fig. 7d). Notably, while other programs in the cascade were significantly upregulated during the transition from progenitors to neurons, early-transient expression was found to be low in early RG, rising in more mature neural progenitor cells (NPCs) then declining in neurons. This is consistent with its predicted role in shaping neuronal sub-type identity, which recent evidence indicates is determined by the internal state of NPCs immediately prior to their exit from the cell-cycle.

Discussion
A complex choreography of cell proliferation, specification, growth, migration and network formation underlies brain development. To date, limited progress has been made pinpointing aspects of this process disrupted in neuropsychiatric disorders. Here we uncover distinct gene expression programs expressed during early excitatory corticogenesis in vitro and in human fetal cortex (Fig. 7d). These programs enriched for variants contributing to a wide spectrum of disorders and cognitive function (Fig. 6). The consistency of these enrichments is noteworthy, with multiple associations identified for each early neurogenic program. These programs harbor well-supported risk genes for complex and Mendelian disorders, some of which are highlighted in Fig. 8a. This convergence of genetic evidence suggests that these programs play an etiological role in a wide range of psychiatric disorders.

Each program has a unique gene expression profile and molecular composition, indicating a distinct functional role during neurogenesis. Based on our findings we propose that they form a transcriptional cascade regulating neuronal growth, migration, differentiation and network formation (Fig. 8a). Computational analyses of gene/mRNA regulatory interactions implicate known neurodevelopmental disorder risk genes (CHD8, TCF4, FMR1, BCL11B and TBR1) as regulators of this cascade and reveal pathways through which they may contribute to disease (Fig. 4d). Supporting this model, down-regulation of neurogenic programs in DLG2/lines is accompanied by deficits that match their predicted function: impaired migration; simplified neuronal morphology; immature action potential generation; and delayed expression of cell-type identity (reduced expression of CTIP2 protein (Fig. 2e, h, i) and TBR1 mRNA (Supplementary Data 3)). Interestingly, voltage-gated sodium and L-type calcium channels present in early neurogenic programs are not only involved in the generation and control of action potentials, but are also known to impact neuronal growth and migration.

Further experimental work is required to more precisely delineate phenotypes associated with the disruption of individual programs and the risk genes they harbor, and to map out regulatory interactions shaping their expression and activity, testing predictions (Fig. 8a). Here we focus on phenotypes expressed by individual newborn excitatory neurons; in future studies it will be important to investigate the persistence of these phenotypes and explore longer-term effects on neuronal circuit formation and function.

A clear pattern of enrichment was evident across early neurogenic programs (Fig. 6c–f). Rare damaging mutations contributing to more severe disorders were concentrated in initial stages of the cascade, impacting both progenitors (early-transient) and neurons. Common variant association was restricted to neuronally expressed pathways (early-stable, early-increasing). It has been proposed that adult and childhood disorders lie on an etiological and neurodevelopmental continuum, the more severe the disorder the greater the contribution from rare, damaging mutations and the earlier their developmental impact (Fig. 8b). Our data support this model and ground it in developmental neurobiology, embedding genetic risk for multiple disorders in a common pathophysiological framework.

Genetic risk for all disorders was concentrated in LoFi genes, indicating wider relevance for these genes than previously appreciated and providing insight into their pathophysiological roles. Being under high selective constraint, LoFi genes profoundly impact development through to sexual maturity. It has not been clear whether LoFi genes harboring pathogenic mutations are distributed across diverse pathways shaping pre-/postnatal growth or are concentrated in specific pathways and/or stages of development. Our analyses reveal that not all neurodevelopmental pathways are enriched for LoFi genes (Fig. 6a), and that the subset of LoFi genes (~40%) concentrated in early development of neuropsychiatric disorders. However, these programs were identified from bulk RNAseq data in vitro and it remains to be shown that their constituent genes are actively co-expressed in the appropriate cell-types during cortical excitatory neurogenesis in vivo. To address this, we extracted gene expression data for cell-types spanning cortical excitatory neurodevelopment from a single-cell RNAseq study of human fetal brain. After normalizing the expression for each gene across all cells, we calculated the average expression for each gene in each cell-type/stage of development available: early radial glia (early RG), RG, intermediate progenitor cells (IPCs), transitioning cells (intermediate between progenitors and neurons), newborn and developing neurons (Methods). We then plotted the expression of each program (mean and standard error of gene-level averages) in each cell-type/stage and tested for differences in expression between successive types/stages (Fig. 7d, Supplementary Data 10). The expression profile seen for each program in vitro was recapitulated across neurodevelopmental cell-types from human fetal cortex (Fig. 7d). Notably, while other programs in the cascade were significantly upregulated during the transition from progenitors to neurons, early-transient expression was found to be low in early RG, rising in more mature neural progenitor cells (NPCs) then declining in neurons. This is consistent with its predicted role in shaping neuronal sub-type identity, which recent evidence indicates is determined by the internal state of NPCs immediately prior to their exit from the cell-cycle.

Each program has a unique gene expression profile and molecular composition, indicating a distinct functional role during neurogenesis. Based on our findings we propose that they form a transcriptional cascade regulating neuronal growth, migration, differentiation and network formation (Fig. 8a). Computational analyses of gene/mRNA regulatory interactions implicate known neurodevelopmental disorder risk genes (CHD8, TCF4, FMR1, BCL11B and TBR1) as regulators of this cascade and reveal pathways through which they may contribute to disease (Fig. 4d). Supporting this model, down-regulation of neurogenic programs in DLG2/lines is accompanied by deficits that match their predicted function: impaired migration; simplified neuronal morphology; immature action potential generation; and delayed expression of cell-type identity (reduced expression of CTIP2 protein (Fig. 2e, h, i) and TBR1 mRNA (Supplementary Data 3)). Interestingly, voltage-gated sodium and L-type calcium channels present in early neurogenic programs are not only involved in the generation and control of action potentials, but are also known to impact neuronal growth and migration.

Further experimental work is required to more precisely delineate phenotypes associated with the disruption of individual programs and the risk genes they harbor, and to map out regulatory interactions shaping their expression and activity, testing predictions (Fig. 8a). Here we focus on phenotypes expressed by individual newborn excitatory neurons; in future studies it will be important to investigate the persistence of these phenotypes and explore longer-term effects on neuronal circuit formation and function.

A clear pattern of enrichment was evident across early neurogenic programs (Fig. 6c–f). Rare damaging mutations contributing to more severe disorders were concentrated in initial stages of the cascade, impacting both progenitors (early-transient) and neurons. Common variant association was restricted to neuronally expressed pathways (early-stable, early-increasing). It has been proposed that adult and childhood disorders lie on an etiological and neurodevelopmental continuum, the more severe the disorder the greater the contribution from rare, damaging mutations and the earlier their developmental impact (Fig. 8b). Our data support this model and ground it in developmental neurobiology, embedding genetic risk for multiple disorders in a common pathophysiological framework.

Genetic risk for all disorders was concentrated in LoFi genes, indicating wider relevance for these genes than previously appreciated and providing insight into their pathophysiological roles. Being under high selective constraint, LoFi genes profoundly impact development through to sexual maturity. It has not been clear whether LoFi genes harboring pathogenic mutations are distributed across diverse pathways shaping pre-/postnatal growth or are concentrated in specific pathways and/or stages of development. Our analyses reveal that not all neurodevelopmental pathways are enriched for LoFi genes (Fig. 6a), and that the subset of LoFi genes (~40%) concentrated in early...
neurogenic programs capture virtually all common and rare variant LoFi association across a wide spectrum of disorders (Fig. 6). While early-transient−/− expression is limited to initial stages of neurogenesis (peaking as RG mature, Fig. 7d), other programs are upregulated during the NPC-neuron transition and persist as neurons develop (Fig. 7d), shaping their morphology, function and possibly connectivity (Fig. 8a).

**b** Schizophrenia GWAS

|                  | N overlap | OR       | P overlap | P overlap corrected |
|------------------|-----------|----------|-----------|--------------------|
| Late (1399)      | 221       | 0.68     | 2.96x10^-7| 1.19x10^-4         |
| Early-increasing+ (536) | 216 | 2.66     | 4.14x10^-30| 1.66x10^-24       |
| Early-stable+ (1713) | 756 | 3.64     | 4.27x10^-30| 1.71x10^-24       |
| Early-transient+ (826) | 289 | 2.14     | 9.42x10^-20| 3.77x10^-11       |

SZ GWAS association previously noted in CA1 pyramidal neurons\(^1\),\(^3\) was captured by a subset of neurogenic genes contributing to dendrite/spine/synapse formation, signaling and plasticity. These processes underlie the establishment and maturation of neuronal circuits\(^6\) and their learning-dependent modification in adults\(^6\). The most parsimonious explanation is that SZ common variants act largely via disruption of early brain...
development, as neurogenic programs harbor GWAS associations beyond their overlap with pyramidal\(^2^b\) (Fig. 7b), and SZ shares extensive SNP heritability with early-onset disorders\(^4^b,^8^a\). However, effect sizes are greater for shared genes (Fig. 7b) – although not significantly so – and SZ onset extends from late childhood well into adulthood\(^6^4\). We hypothesize that vulnerability to SZ is primarily driven during early neurodevelopment, and that this is subsequently compounded by a gradual accumulation of deficits during circuit maturation due to both external stressors and the impaired function of neurogenic pathways that remain operant throughout childhood and into adulthood.

While DLG2 knockout led to the identification of disease-associated programs and allowed us to investigate cellular phenotypes associated with their dysregulation, DLG2 itself has yet to reach the status of a canonical SZ/ASD risk gene. DLG2 is primarily known for its role as a postsynaptic scaffold protein in mature neurons, where it is required for normal formation of NMDA receptor signaling complexes\(^2^6\). We show that DLG2 expression is also important for cortical excitatory neurodevelopment, but the mechanism by which it operates remains to be determined. Based on its known function and the involvement of invertebrate Dlg in the developmental Scrib signaling module\(^2^5\), DLG2 may link cell-surface receptors to signal transduction pathways regulating the activation of neurogenic programs (Supplementary Fig. 9). We hypothesize that stochastic signaling in DLG2\(^-/-\) lines due to impaired complex formation could delay and impair transcriptional activation, disrupting the orchestration of events required for normal development and the specification of neuronal properties. Precise timing is crucial during brain development, where the correct dendritic morphology, axonal length and electrical properties are required for normal circuit formation and function. Consequently, even transient perturbation of neurogenesis may have a profound impact on fine-grained neuronal wiring, network activity and ultimately perception, cognition and behavior.

Although much remains to be uncovered, our findings sketch the foundations for an integrated etiological model of psychiatric genetic disorders and their developmental origins.

Methods

**hESCs culture.** H7 hESC line (WA07) was purchased from WiCell, USA. All hESC work was performed in accordance with Cardiff University’s regulations under Health and Safety Executive approval (GM130/14.3) and WiCell’s MTA and SLA. All hESC lines were maintained at 37 °C and 5% CO\(_2\) in 6 well cell culture plates (Greiner) coated with 1% Matrigel hESC-qualified Matrigel (Corning) prepared in Dulbecco’s Modified Eagle Medium: Nutrient Mixture F-12 (DMEM/F12, Thermo Fisher Scientific). Cells were fed daily with Essential 8 medium (ES, Thermo Fisher Scientific) and passed at 80% confluency using Versene solution (Thermo Fisher Scientific) for 1.5 min at 37 °C followed by manual dissociation with a serological pipette. All cells were kept below passage 25 and confirmed as negative for mycoplasma infection.

**DLG2 Knockout hESC line generation.** Two guide RNAs targeting exon 22 (Supplementary Fig. 1) of the human DLG2 gene, covering the first PDZ domain, were designed using a web-based tool (crispr.mit.edu) and cloned into two plasmids containing D10A nickase mutant Cas9 with GFP (PX461) or Puromycin resistant gene (PX462)\(^5^c\). pSpCas9n(BB)-2A-GFP (PX461) and pSpCas9n(BB)-2A-Puro (PX462) was a gift from Feng Zhang (for PX461, Addgene plasmid#48140; http://n2t.net/addgene/48140; RRID:Addgene_48140; for PX462, Addgene plasmid #48141; http://n2t.net/addgene/48141; RRID:Addgene_48141). H7 hESCs (WiCell) were nucleofected using P4 solution and CB150 programme (Lonza) with 5 µg of plasmids. FACS sorted on the following day and plated at a low density (~70 cells/cm\(^2\)) for cloning. 19 clonal populations were established with 6 WT and 13 mutant lines after targeted sequencing of the exon 22. One WT and two homozygous knockout lines were chosen for study: our WT and KO lines therefore originate from the same H7 parental line and have gone through the same process of nucleofection and FACS sorting together.

**Genetic validation.** The gRNA pair had zero predicted off-target nickase sites (Supplementary Fig. 2). Even though we did not use a wild-type Cas9 nuclease (where only a single gRNA is required to create a double-stranded break), we further checked genomic predicted off-target sites for each individual gRNA by PCR and Sanger sequencing (GATC & LGC). Out of 30 sites identified, we randomly selected 14 (7 for each gRNA) for validation. No mutations relative to WT were present at any site (Supplementary Data 1). In addition, genotyping on the Illumina PsychArray v1.1 revealed no CNV insertions/deletions in either DLG2\(^-/-\) line relative to WT (Supplementary Fig. 5).

**Cortical differentiation.** Differentiation to cortical projection neurons (Fig. 1b) was achieved using the dual SMAD inhibition protocol\(^2^8\) with modifications (embryoid body to monolayer and replacement of KSR medium with N2B27 medium) suggested by Cambray et al\(^3^9\). Prior to differentiation Versene treatment and mechanical dissociation was used to passage hESCs at ~100,000 cells per well into 12 well cell culture plates (Greiner) coated with 1% Matrigel Growth Factor Reduced (GFR) Basement Membrane matrix (Corning) in DMEM/F12, cells were maintained in E8 medium at 37 °C and 5% CO\(_2\) until confluent. At day 0 of the differentiation E8 media was replaced with N2B27-RA neuronal differentiation media consisting of: 2/3 DMEM/F12, 1/3 Neurobasal (Thermo Fisher Scientific), 1x N-2 Supplement (Thermo Fisher Scientific), 1x B27 Supplement minus vitamin A (Thermo Fisher Scientific), 1x Pen Stead Glutamine (Thermo Fisher Scientific) and 50 µM 2-Mercaptoethanol (Thermo Fisher Scientific), which was supplemented with 100 nM LDN193189 (Cambridge Biosciences) and 10 µM SB431542 (Stratagene) for the first 10 days only (the neural induction period). At day 10 cells were passaged at a 2:3 ratio into 12 well cell culture plates coated with 15 µg/ml human plasma fibronectin (Merk) in Dulbecco’s phosphate-buffered saline (DPBS, Thermo Fisher Scientific), passage was as previously described with the addition of a 1 h incubation with 10 µM Y27632 Dihydrochloride (ROCK inhibitor, Stratagene) prior to Versene dissociation. During days 10–20 of differentiation cells were maintained in N2B27-RA (without LDN193189 or SB431542 supplementation) and passed at day 20 in a 1:4 ratio into 24 well cell culture plates (Greiner) sequentially coated with 10 µg/ml poly-d-lysine hydrobromide (PDL, Sigma) and 15 µg/ml laminin (Sigma) in DPBS. Vitamin A was added to the differentiation media at day 26, standard 1x B27 Supplement (Thermo Fisher Scientific), the non-essential amino acids, 50 µM 2-Mercaptoethanol, 1x N-2 Supplement were maintained in the resulting N2B27-RA media for the remainder of the differentiation. Cells maintained to day 40 received no additional passage beyond
passage 2 at day 20 while cells kept beyond day 40 received a third passage at day 40, 1:2 onto PDL-laminin as previously described. In all cases cells maintained past day 40 were fed with N2B27 + RA supplemented with 2 µg/ml laminin once weekly to prevent cell detachment from the culture plates.

**Immunocytochemistry.** Cells were fixed in 4% paraformaldehyde (PFA, Sigma) in PBS for 20 min at 4 °C followed by a 1 h room temperature incubation in blocking solution of 5% donkey serum (Biosera) in 0.3% Triton-X-100 (Sigma) in PBS (0.3% PBST). Primary antibodies, used at an assay dependent concentration (see main text/Methods) were diluted in blocking solution and incubated with cells overnight. Secondary antibodies (Thermo Fisher Scientific) diluted 1:500 with blocking solution. After an additional 2 PBS washes cells were counterstained with DAPI nuclei acid stain (Thermo Fisher Scientific), diluted 1:1000 with PBS, for 5 mins at room temperature.

**Fig. 7 Expression of early neurogenic programs in neurodevelopmental cell-types from human fetal cortex.** a Identification of programs enriched/depleted for genes with high expression in CA1 pyramidal neurons relative to other brain cell-types (pyramidalhigh); enrichment compared to all expressed genes (two-sided Fisher’s Exact Test). Raw and Bonferroni-corrected p values given (poverlap, poverlapcorrected), correction is for the 4 tests performed. b Pyramidalhigh genes were partitioned based on their overlap with neurogenic programs. Each segment of the Venn diagram shows the number of genes in each subset and the regression coefficient (β) and uncorrected p value (P) for SZ common variant enrichment (one-sided MAGMA test), conditioning on all expressed genes. Bold indicates enrichments surviving correction for 7 tests. c A one-sided Fisher’s Exact Test was used to identify key regulators whose known/predicted targets are over-represented amongst pyramidalhigh genes overlapping the late neurogenic program when compared to all expressed genes; both raw and Bonferroni-corrected p values are given (poverlap, poverlapcorrected), where correction is for the 7 regulator sets tested. d Cells corresponding to distinct neurodevelopmental cell-types (RG radial glia; IPC Intermediate progenitor cell), including cell-types at different stages of maturity, were identified and extracted from a previously published single-cell RNA-seq study of human fetal cortex across peak stages of neurogenesis (Methods). >80% of genes belonging to each in vitro-defined early neurogenic program and 55% of the late-pyramidalhigh set (top RHS) were present in the fetal data (early-transient/-/- n = 665 genes; early-stable/-/- n = 1484 genes; early-increasing/-/- n = 431 genes; late-pyramidalhigh n = 101 genes). For each program, the mean and standard error of gene-level expression averages (see main text/Methods) were calculated for each fetal cell-type/developmental stage. The fetal data captures both direct (RG neuron) and indirect (RG IPC neuron) neurogenic pathways. The (deep layer) neurons present in our day 30-60 cultures in vitro are predominantly born via the direct neurogenic pathway. For each program, differences between gene-level averages for successive cell-types/stages were evaluated using a two-tailed Student’s t test and p values Bonferroni-corrected for 6 pairwise comparisons. *p < 0.05; **p < 0.01; ***p < 0.0001. All data presented as mean ± SEM. Bold indicates tests surviving Bonferroni correction. The exact p values are provided in Supplementary Data 10.

**Table 1.** 

| Regulator targets | Noverlap | poverlapcorrected |
|-------------------|----------|--------------------|
| BCL11B targets (predicted) | 695 | 26 | 3.35 | 7.46x10^-7 | 5.22x10^-4 |
| TBR1 targets (predicted) | 280 | 20 | 6.58 | 4.18x10^-6 | 3.14x10^-3 |
| FMRP targets | 816 | 32 | 3.63 | 1.00x10^-4 | 7.42x10^-1 |
| TCF4 targets (direct) | 3544 | 70 | 1.91 | 2.60x10^-1 | 5.45x10^-8 |
| CHD8 targets (indirect, up) | 562 | 4 | 0.55 | 0.93 | 1 |
| CHD8 targets (indirect, down) | 315 | 10 | 2.63 | 0.0071 | 0.05 |
| CHD8 midfetal targets (direct) | 2549 | 21 | 0.60 | 0.99 | 1 |

**Table 2.** 

| Stage | Expression (z-score mean ± s.e.) |
|-------|----------------------------------|
| Early-transient/-/- | Early-stable/-/- | Early-increasing/-/- | Late-pyramidalhigh |
| Day 20 | Day 30 | Day 60 |
| Early-transient/-/- | 0.8058 ± 0.12 | 0.8554 ± 0.09 | 0.7170 ± 0.07 |
| Early-stable/-/- | 4.1914 ± 0.10 | 3.8418 ± 0.09 | 3.5110 ± 0.08 |
| Early-increasing/-/- | 1.7118 ± 0.10 | 1.4718 ± 0.09 | 1.1818 ± 0.08 |
| Late-pyramidalhigh | 1.5318 ± 0.10 | 1.2918 ± 0.09 | 1.0518 ± 0.08 |
temperature and following a final PBS wash, mounted using Dako Fluorescence Mounting Medium (Agilent) and glass coverslips. Imaging was with either the LSM710 confocal microscope (Zeiss) using Zen 2012 SP2 (black) v11.02.190 (Zeiss), LAS X for DMI6000B inverted microscope (Leica) or Cellinsight Cx7 High-Content Screening Platform (Thermo Fisher Scientific) with HCS Studio Cell Analysis software v6.6.0 (Thermo Fisher Scientific) used for quantification.

**Western blotting.** Total protein was extracted from dissociated cultured cells by incubating in 1x RIPA buffer (New England Biolabs) with added MS-SAFE Protease and Phosphatase Inhibitor (Sigma) for 30 min on ice with regular vortexing. Concentration was determined using a DC Protein Assay (BioRad) quantified with the CLARIOstar microplate reader (BMG Labtech). Proteins for western blotting were incubated with Bolt LDS sample buffer (Thermo Fisher Scientific) and Bolt Sample Reducing Agent (Thermo Fisher Scientific) for 10 min at 70 °C before loading into Bolt 4–12% Bis-Tris Plus gels (Thermo Fisher Scientific). Gels were run at 120 V for 2–3 h in Bolt MES SDS Running Buffer (Thermo Fisher Scientific) prior to protein transfer to Amersham Protran nitrocellulose blotting membrane (GE Healthcare) using a Mini Trans-Blot Cell (BioRad) and Bolt Transfer Buffer (Thermo Fisher Scientific) run at 120 V for 1 h 45 min. Transfer was confirmed by visualizing protein bands with 0.1% Ponceau S (Sigma) in 5% acetic acid (Sigma) followed by repeated H2O washes to remove the stain.

Following transfer, membranes were incubated in a blocking solution of 5% milk in TBST, 0.1% TWEEN 20 (Sigma) in TBS (Formedium), for 1 h at room temperature. Primary antibodies, used at an assay dependent concentration, were diluted with blocking solution prior to incubation with membranes overnight at 4 °C. Following 3 TBST washes, membranes were incubated in the dark for 1 h at room temperature with IRDye secondary antibodies (LI-COR) diluted 1:15000 with blocking solution. After 3 TBS washes staining was visualized using the Odyssey CLx Imaging System (LI-COR) and Image Studio Lite Version 5.2 (LI-COR).

**Genome-wide significant (P<2.2x10^{-6}) in recent SCHEMA rare variant analyses**

**Rare variation causes Mendelian neurodevelopmental syndromes**

---

**Fig. 8 Model of disease pathophysiology in early corticoneurogenesis.**

**a** Summary of main GO term enrichments for each disease-associated transcriptional program (Supplementary Data 7) and significant regulatory interactions between them. Key regulators, genome-wide significant rare coding variants from\textsuperscript{8,4} and rare mutations causing Mendelian neurodevelopmental syndromes are denoted. **b** Neurodevelopmental continuum/gradient model. Disorders shown are: ID/NDD intellectual disability/severe neurodevelopmental delay; ASD autism spectrum disorders; ADHD attention-deficit/hyperactivity disorder; SZ schizophrenia; BP bipolar disorder; MDD major depressive disorder.
Antibody concentration.

| Antibody | Company       | Identifier catalog no. | Dilution in ICC | Dilution in WB |
|----------|---------------|-------------------------|-----------------|----------------|
| Rabbit polyclonal anti-Calnexin | Abcam | ab22395 | N/A | 1:5000 |
| Rat monoclonal anti-CTIP2 [25S6] | Abcam | ab18465 | 1:500 | N/A |
| Rabbit polyclonal anti-CTIP2 | Abcam | ab70453 | N/A | 1:1000 |
| Rabbit monoclonal anti-DARPP32 [EP721Y] | AbCam | ab40802 | 1:500 | N/A |
| Mouse monoclonal anti-DLX1 | Abcam | ab32010 | 1:800 | N/A |
| Mouse monoclonal anti-FOXG1 | Abcam | ab9485 | N/A | 1:5000 |
| Goat polyclonal anti-GABA | Antibodies.com | A84236 | 1:200 | N/A |
| Mouse monoclonal anti-FOX1P1 [JC12] | BD Pharmingen | 550690 | 1:150 | 1:1000 |
| Mouse monoclonal anti-Ki67 [B56] | Sigma-Aldrich | M1406 | 1:250 | N/A |
| Mouse monoclonal anti-MAP2 [AP-20] | Millipore | AB5622 | N/A | 1:1000 |
| Rabbit polyclonal anti-MAP2 | Cell Signaling Technology | 4903 | 1:400 | N/A |
| Rabbit monoclonal anti-NANOG [D73G4] | Millipore | ABN78 | 1:500 | 1:500 |
| Rabbit polyclonal anti-NEUN | Merck | M30A3 | 1:1000 | N/A |
| Rabbit monoclonal anti-NEUN [EP1584Y] | Cell Signaling Technology | 2840 | 1:400 | N/A |
| Rabbit monoclonal anti-Oct4 [C30A3] | R&D Systems | MAB2456 | 1:500 | N/A |
| Mouse monoclonal anti-OLIG3 [257934] | Abcam | ab195045 | 1:300 | N/A |
| Rabbit monoclonal anti-PAX6 [EP15858] | Abcam | ab18258 | N/A | 1:1000 |
| Rabbit polyclonal anti-PSD-95 (DLG4) | Abcam | ab51502 | 1:25 | 1:100 |
| Mouse monoclonal anti-SATB2 [SATRA4B10] | Cell Signaling Technology | 3579 | 1:400 | N/A |

Synaptosomal preparation. Synaptic protein was extracted by manually dissociating cultured cells in 1x Syn-PER Reagent (Thermo Fisher Scientific) with added MS-SAFE Protease and Phosphatase Inhibitor (Sigma). Following low speed centrifugation to pellet cell debris (1200 g, 10 min, 4 °C) the supernatant was added MS-SAFE Protease and Phosphatase Inhibitor, all Sigma) and stored on ice until required. Total protein was extracted from dissociated cultured cells by incubating for 1 h on ice with regular vortexing, cell debris was pelleted by high speed centrifugation (21,300 × g, 2 h, 4 °C) and the supernatant added to the previously prepared “SIESDV” peptide bound resin. After overnight 4 °C incubation on a roller mixer, the resin was washed 3 times with ice cold DOC buffer and the bound protein eluted by 15 min 70 °C incubation in 5% w/v sodium dodecyl sulfate (SDS, Sigma). The eluted protein was reduced with 10 mM TCEP, trapped and washed on an S-trap micro spin column (PD-10, Bio-Rad) according to the manufacturer’s instructions and protein digested using trypsin sequence grade (Pierce) at 47 °C for 1 h. Eluted peptides were dried in a vacuum concentrator and resuspended in 0.5% formic acid for MS analysis.

Peptide affinity purification. PDZ domain containing proteins were enriched from total protein extracts by peptide affinity purification. NMDA receptor subunit 2 C-terminal peptide “SIESDV” was synthesized (Pepceuticals) and fully dissolved in 90% v/v methanol + 1 M HEPES pH7 (both Sigma). Dissolved peptide was coupled to Affi-Gel 10 resin (Bio-Rad) that had been washed 3 times in methanol, followed by overnight room temperature incubation on a roller mixer. Unreacted NHS groups were subsequently blocked using 1 M Tris pH9 (Sigma) with 2 h room temperature incubation on a roller mixer. The peptide bound resin was then washed 3 times with DOC buffer (1% w/v sodium deoxycholate; 50 mM Tris pH9; 1X MS-SAFE Protease and Phosphatase Inhibitor, all Sigma) and stored on ice until required. Total protein was extracted from dissociated cultured cells by incubating in DOC buffer for 1 h on ice with regular vortexing, cell debris was pelleted by high speed centrifugation (21,300 × g, 2 h, 4 °C) and the supernatant added to the previously prepared “SIESDV” peptide bound resin. After overnight 4 °C incubation on a roller mixer, the resin was washed 3 times with ice cold DOC buffer and the bound protein eluted by 15 min 70 °C incubation in 5% w/v sodium dodecyl sulfate (SDS, Sigma). The eluted protein was reduced with 10 mM TCEP and alkylated following 20 mM Iodoacetamide, trapped and washed on an S-trap micro spin column (ProtiFi, LLC) according to the manufacturer’s instructions and protein digested using trypsin sequence grade (Pierce) at 47 °C for 1 h. Eluted peptides were dried in a vacuum concentrator and resuspended in 0.5% formic acid for MS analysis.

Mass spectrometry analysis. LC-MS/MS analysis was performed and data was processed and quantified according to66. Briefly, peptides were analyzed by nanoflow LC-MS/MS using an Orbitrap Elite (Thermo Fisher) hybrid mass spectrometer equipped with a nanospray source, coupled to an Ultimate RSLCnano LC System (Dionex) and Tune Plus for Orbitrap Elite (Thermo Fisher). Peptides were desalted on-line using a nano trap column, 75 μm I.D. X 20 mm (Thermo Fisher) and then separated using a 130-min gradient from 3 to 40% buffer B (0.5% formic acid containing 0.1% formic acid in water) over 130 min using a nano-LC column, 75 μm I.D. X 150 mm (Thermo Fisher). The gradient was started at 3% buffer B (0.5% formic acid in water) and ramped to 40% buffer B over 130 min. The flow rate was maintained at 300 nL/min and the samples were analyzed using a linear gradient from 0 to 10% buffer A (0.1% formic acid in water) over 15 min. The samples were then eluted with 100% buffer B for 5 min and re-equilibrated with 0% buffer B for 5 min. The data were analyzed using the Xcalibur software (Thermo Fisher) and the protein identification was performed using the Mascot search engine (Matrix Science) with the following parameters: trypsin cleavage, one missed cleavage allowed, variable modifications, and the mass accuracy is ±0.05 Da. The peptide confidence threshold was set to 95% (Xcorr).
NATURAL COMMUNICATIONS | https://doi.org/10.1038/s41467-021-27601-0

RNA sequencing. WT and DLG2 KO cells were cultured to days 15, 20, 30 and 60 of cortical differentiation as described above (See ‘Cortical differentiation’). Total transcriptome RNA was isolated from triplicate wells for all cell lines at each time point by lysis cells in TRIzol Reagent (Thermo Fisher Scientific) followed by purification with the PureLink RNA Mini Kit (Thermo Fisher Scientific). RNA quality control (QC) was performed with the RNA 6000 Nano kit analyzed using the 2100 Bioanalyzer (Agilent). RNA libraries were sequenced for producing the sequenced using the KAPA mRNA HyperPrep Kit for Illumina Platforms (Kapa Biosystems) and indexed with KAPA Single-Indexed Adapter Set A + B (Kapa Biosystems). Library quantification was by Qubit 1x dsDNA HS Assay kit (Thermo Fisher Scientific) and QC by High Sensitivity DNA kit analyzed using the 2100 Bioanalyzer High Sensitivity DNA assay (Agilent). Sequencing was performed using the HiSeq4000 Sequencing System (Illumina) with libraries split into 2 equimolar pools, each of which was run over 2 flow cell lanes with 75 base pairs paired end reads and 8 base pair index read.

All samples were modeled after the long-rna-seq pipeline developed by the PsychENCODE Consortium and available at https://www.synapse.org/

Synapse syn12026837. Briefly, the fastq files from Illumina HiSeq4000 were assessed for quality by using FastQC tool (v0.11.8) and trimmed for adapter sequence and low base call quality (Phred score <30 at ends) with cutadapt (v2.3). The mapping of the trimmed reads was done using STAR (v2.7.0e) and the BAM files were produced in both genomic and transcriptomic coordinates and sorted using samtools (v1.9). The aligned and sorted BAM files were further assessed for quality using Picard tools (v2.20.2). This revealed a high level of duplicate reads in day 30 KO2 samples (~72% compared to an average of 23% for other samples). These samples were removed prior to further analyses, which were thus performed on K01 and WT samples for this timepoint. GRC38.p2 was used as the reference genome and the comprehensive gene annotations on the primary assembly from Gencode (release 31) used as gene annotation. Gene and transcript-level quantifications were calculated using RSEM (v1.3.1) and STAR. Both RSEM and STAR executions were performed using the psychENCODE parameters.

RSEM gene annotated level estimated counts were imported using the tximport package (v1.2.3). Protein coding genes expressed (cpm ≥ 1) in at least 1/3 of the samples were taken forward for differential analyses of genes, transcripts and exons. Differential gene expression analysis was performed using the DESeq2 package (v1.24.0) and differentially expressed genes could be considered significant if their p-value after Benferroni correction was <0.05. Differential exon usage was analyzed using the DESeq pipeline. Briefly, the GENCODE annotation.gff file was translated into a .gff file with collapsed exon counting bins by using the dsexq_prepare_annotation.py script. Mapped reads overlapping each of the exon counting bins were then counted using the python_count.py script and the HTSeq software (0.11.2r1). Finally, differential exon usage was evaluated using DESeqX (v1.30) and significant differences identified using an FDR threshold of 0.05. All the differential analyses were performed by using R (v3.6.1).

When analyzing differential gene expression in DLG2KO relative to WT, samples from KO1 and KO2 lines were combined i.e., for each timepoint a single differential gene expression analysis was performed, with all KO1 and KO2 samples combined into 2 equimolar pools, each of which was run over 2 using the 2100 Bioanalyzer High Sensitivity DNA assay (Agilent). Sequencing was performed on paired end reads and 8 base pair index reads.

Functional annotation with the PureLink RNA Mini Kit (Thermo Fisher Scientific) was translated into a .gff (v1.30) and significantly overrepresented GO terms were identified using the GOseq package (v1.3.1) and DAVID gene annotation database (v6.8) with an exclusion list of 500 and an exclusion duration of 30 s. Raw mass spectrometry data were analyzed with MaxQuant version 1.6.10.4367. Data were searched against a human UniProt (https://www.uniprot.org/) sequence database (downloaded December 2019) using the following search parameters: digestion set to Trypsin/P, methionine oxidation and N-terminal protein acetylation as variable modifications, cysteine carboxymethylation as a fixed modification, match between runs with a match window time of 0.7 min and a 20-min alignment time window, label-free quantification enabled with a minimum ratio count of 2, minimum number of neighbors of 3 and an average number of neighbors of 6. PSM and protein match thresholds were set at 0.1 ppm. A protein FDR of 0.01 and a peptide FDR of 0.01 were used for identification level cut-offs.

Human fetal cortex single-cell RNA sequencing data. Single-cell RNAseq gene expression data from Nowakowski et al. were downloaded from https://www.ncbi.nlm.nih.gov/geo/and cells corresponding to distinct neurodevelopmental cell-types (including cell-types at different stages of maturity) were identified and extracted, collating all cells from the corresponding in vivo cell clusters as follows:

Progenitors.

RG: (early): “RG-early”

RG: “RG-div1”, “RG-div2”, “oRG”, “iRG”, “vRG”

IPC: “IPC-div1”, “IPC-div2”

Transitoning.

“IPC-nEN1”, “IPC-nEN2”, “IPC-nEN3”

Cortical excitatory neurons.

Newborn: “nEN-early1”, “nEN-early2”, “nEN-late”

Developing: “EN-PFC1”, “EN-PFC2”, “EN-PFC3”, “EN-V1-1”, “EN-V1-2”, “EN-V1-3”

Cells with less than 5% of all protein-coding genes expressed (TPM ≥ 1) and genes expressed in less than 5% of cells were filtered out. The remaining dataset consisted of 2318 cells and 9239 protein-coding genes. Gene expression counts (TPM) were z-score normalized for each gene across all cells, then the average of the TPM was calculated for each gene upregulated in day 60 relative to day 20 and 4000 genes were upregulated in day 60 relative to day 20 (20–600WT). Early-transient, early-stable, early-increasing and late programs were then defined based upon the intersections of these gene-sets as shown in Fig. 4b.

Transcriptional programs. Genes were partitioned based upon their WT expression profiles as follows. Differentially expressed genes (Bonferroni P < 0.05) were first identified between pairs of timepoints (analyzing WT data only): genes upregulated in day 30 relative to day 20 (20–30Δ), genes upregulated in day 60 relative to day 20 (20–600Δ) and genes upregulated in day 60 relative to day 20 (20–60ΔWT). Early-transient, early-stable, early-increasing and late programs were then defined based upon the intersections of these gene-sets as shown in Fig. 4b.
parents of that term, identified using the appropriate ontology tree. Finally, terms containing between 20 and 2000 genes were extracted for analysis.

**Regulator targets.** Predicted TRB1 and BCL11B targets. Transcription factor-target gene interactions identified by elastic net regression were downloaded from the PsychENCODE resource website (http://resource.psychencode.org/#Derived) and predicted targets for TRB1 and BCL11B interacted (interaction file: INT-11_ElasticNet_Filtered_Cutoff_0.1_GRN_L.csv). Gene symbols were mapped to NCBI Entrez gene ids from the NCBIsift directory. TCFl4 targets were identified using the gene history file from NCBI. FMRP targets[^22] NCBI/Entrez mouse gene identifiers were updated using the gene history file from NCBI. Genes were then mapped from mouse to human using Ensembl, removing genes that do not map to human. Remaining sets with a corrected city score >0 for each gene-set from the analysis phase (Leica) and files from NCBI. Genes were then mapped from mouse to human using Ensembl and NCBI id cross-reference files. Taking genes with altered expression on ChDH8 shRNA knockdown, we removed those identified as direct ChDH8 targets in NPG[^26] or as ChDH8 mid-fetal promoter targets[^27].

**Pyramidal CA1 ‘specific’ (pyramidal[^28]) Mouse genes with specificity score >0 for pyramidal CA1 neurons were extracted from Supplementary Table 3 of Skene et al.,[^28] and mapped to human using the gene orthology file from NCBI (https://ftp.ncbi.nlm.nih.gov/gene/DATA/gene_orthologs.gtf). To construct the pyramidal[^28] set we took the top 10% of 1-1 mapped genes ranked by specificity score (Ngene = 14055).

**Functional over-representation test (gene set over lap).** The degree of overlap between pairs of gene sets was evaluated using Fisher’s Exact test, where the background set consisted of all genes expressed in either WT or KO forms. (allWT+KO). This was used for GO term analysis of regulator targets (Fig. 4e); and the overlap between LoFi genes and transcriptional programs (Fig. 6a). In order to identify a semi-independent subset of over-represented annotations from the output of GO term tests (Fig. 4f, Supplementary Data 5 and 7), we used an iterative refinement procedure. Briefly, we found the set with the largest enrichment odds ratio; removed all genes in this set from all other over-represented sets; used the remaining sets with a corrected city score >0 for each gene-set from the analysis phase (Leica) and files from NCBI. Genes were then mapped from mouse to human using Ensembl and NCBI id cross-reference files. Taking genes with altered expression on ChDH8 shRNA knockdown, we removed those identified as direct ChDH8 targets in NPG[^26] or as ChDH8 mid-fetal promoter targets[^27].

**Common variant association.** Common variant gene-set enrichment analyses were performed using the competitive gene-set enrichment test implemented in MAGMA version 1.07, conditioning on all[WT+KO] using the condition-residualize function. To test whether GO terms (Fig. 4f, Supplementary Data 8) or regulator targets (Fig. 4e) enriched in a specific program captured more of the variance in the gene-set from the MAGMA output file and compared z = d − S(d) to a standard normal distribution, where d = β1 − β2, and SE(d) = √SE(β1)^2 + SE(β2)^2. Gene-level association statistics for schizophrenia were taken from Pardinas et al.,[^12] those for ADHD[^18], bipolar disorder[^19] and Alzheimer’s disease[^20] were calculated using the MAGMA multi model, briefly, we selected 20,000 permutations for each gene. Prior to analysis, SNPs with MAP < 0.01 or INFO score < 0.6 were removed from the bipolar GWAS, bringing it into line with the other datasets.

**Rare variant association.** The de novo LoF mutations for SZ analyzed here are described in Rees et al.,[^15] de novo LoF mutations for NDD, ASD and unaffected siblings of individuals with ASD were taken from Satterstrom et al.,[^14] these were re-annotated using Variant Effect Prediction[^16] and mutations mapping to >2 genes (once readthrough annotations had been discarded) were removed from the analysis. A two-sided Poisson ratio test was used to evaluate whether the enrichment of de novo LoF mutations in specific gene-sets was significantly greater than that observed for all other expressed genes (using all[WT+KO]). The expected rate of de novo LoF mutations in a set of genes was estimated using individual gene mutation rates[^21].

**Migration assay.** Cells were cultured and differentiated to cortical projection neurons as previously described. Neuronal migration was measured during a 70-h period from day 40 by transferring cell culture plates to the IncuCyte Live Cell Imaging System (Sartorius). Cells were maintained at 37 °C and 5% CO2 with 20% humidity. Image stacks were taken continuously (10–20 minute intervals) and digitized using a rate of 3 frames/minute. The StepStack plugin for ImageJ was used to fully align the resulting stacks of time lapse images after which the cartesian coordinates of individual neuronal soma were recorded over the course of the experiment, enabling the distance and speed of neuronal migration to be calculated. Data sets (Fig. 3f, g) were analyzed by unpaired two-tailed Student’s t test.

**Morphology analysis.** Cells were differentiated to cortical projection neurons essentially as described and neuronal morphology assessed at days 30 and 70. To generate low density cultures for analysis, cells were passaged at either day 25 or 50 using 15-min Accutase solution (Sigma) dissociation followed by plating at 100,000 cells per well on 24 well culture plates. 72 h prior to morphology assessment cells were stained with 500 ng/ml DiI (Invitrogen) per well. Coverslips containing between 20 and 2000 genes were extracted for analysis.

**Electrophysiology.** Whole-cell patch clamp electrophysiology was performed on cells cultured on 13 mm round coverslips and the most morphologically mature neurons were patched in each culture; hence the most comparable subpopulation of cells from each genotype was compared. On day 20 of hESC differentiation, 250,000 human neural precursor cells were sampled from WT and KO hESCs were dissociated and plated on each PDL-covered coverslip in 30 µl diluted (20x) matrigel (Corning) targeting at least 10,000 cells/well. Postnatal day 10–12 mouse P1-P3 Newborn rats (Charles River) bred in-house were sacrificed by cervical dislocation and cortex was quickly dissected. Tissues were dissociated using 2 mg/ml papain and plated in DMEM supplemented with 10% Fetal bovine serum and 1% penicillin/ streptomycin/Ampthoterin B and 1x Glutamax (all Thermo Fisher Scientific). Microglia and oligodendrocyte precursor cells were removed by shaking at 500 rpm for 24 h at 37 °C. All animal procedures were performed in accordance with Cardiff University’s animal care committee’s regulations and the European Directive 2010/ 63/EU on the protection of animals used for scientific purposes. Plated cells were fed with BrainPhys medium (stem cell technologies) supplemented with 1X B27 (Thermo Fisher Scientific), 10 ng/ml BDNF (Cambridge Bioscience) and 250 µM ascorbic acid (Sigma). To stop the proliferation of cells, 1X CultureOne (Thermo Fisher Scientific) was supplemented from day 21. For postynaptic current experiment, coverslips were transferred to a recording chamber (RC-26G, Warner Instruments) and perfused with HEPES Buffered saline (HBS) (119 mM NaCl; 6 mM KCl; 25 mM HEPES; 33 mM glucose; 2 mM CaCl2; 2 mM MgCl2; 1 µM glycine; 100 µM picrotoxin; pH 7.4), at a flow rate of 2–3 ml per minute. Recordings were made using pipettes pulled from borosilicate glass capillaries (1.5 mm OD, 0.86 mm ID, Harvard Apparatus), and experiments were performed at room temperature (~20 °C). mEPSC recordings were made using recording electrodes filled with a CsCl-based internal solution (130 mM CsCl; 8 mM NaCl; 4 mM Mg-ATP; 0.3 mM Na-GTP; 0.5 mM EGTA; 10 mM HEPES; 6 mM QX-314; with pH 7.3 and osmolality ~295 mosm). Cells were voltage clamped at −60 mV using a Multiclamp 700B amplifier (Axon Instruments). Continuous current acquisition, series resistance and input resistance were monitored online and data was corrected offline (http://www.wiiltp.com). Only cells with series resistance <25 MΩ with a change in series resistance <10% from the start were included in this study. Data were analyzed by importing Axon Binary Files into Clampfit (version 10.6; Molecular Devices). A threshold function of >12 pA was used to identify mEPSC events, and then subject to further confirmation. Results were output to Sigmaplot (version 12.5, Systat Software), where analysis of peak amplitude and frequency of events was performed.

The current clamp was used to record resting membrane potential (RMP) and action potentials (AP). Data were sampled at 20 kHz with a 3 kHz Bessel filter with Multiclamp 700B amplifier. Coverslips were transferred to a recording chamber maintained at RT (20–21 °C) on the stage of an Olympus BX61W (Olympus) differential interference contrast (DIC) microscope and perfused at 2.5 ml/min with the external solution composed of 135 mM NaCl, 3 mM KCl, 1.2 mM MgCl2, 1.25 mM CaCl2, 15 mM D-glucose, 5 mM HEPES (all from Sigma), and pH was titrated to 7.4 by NaOH. The internal solution used to fill the patch pipettes was composed of 117 mM KCl, 10 mM NaCl, 11 mM HEPES, 2 mM Na2-ATP, 2 mM Na-GTP, 1.2 mM Na2-phosphocreatine, 2 mM MgCl2, 1 mM CaCl2 and 11 mM EGTA (all from Sigma), and pH was titrated to 7.3 by NaOH. The resistance of a patch pipette was ~3–9 MΩ and the series resistance component was further evaluated using the balance function of the instrument. The RMP of the cell was recorded immediately after breaking into the cells in gap free mode. A systematic current injection protocol (duration, 1 s; increment, 20 pA; from −60 pA to 120 pA) was applied to the neurons held at −60 mV to evoke APs. Input resistance (Rin) was calculated by Rin = (V – Vm)/I, where V is the potential recorded from the described positions in each cell, recorded in steady state. All data were analyzed for action potential parameterization was carried out by Clampfit 10.7 software (Molecular Devices).
Statistics and reproducibility. Unless specifically stated in each methodology section, GraphPad Prism (version 8.3.0) was used to test the statistical significance of the data and to produce the graphs. Stars above bars in each graph represents Bonferroni-corrected post hoc tests. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001 vs. WT control. All phenotypic validation results were from a minimum of two independent determinations. Within a given differentiation triplicate samples were used per cell line at each time point investigated.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
RNAseq data generated by this study have been deposited in the Gene Expression Omnibus (GEO) archive with accession number GSE172199. The mass spectrometry data generated in this study have been deposited in the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD029526. Human proteome reference sequences are available in UniProt (https://www.uniprot.org/). Processed gene annotation data (GO terms, LoFi genes, CBL11B, TBR1, FMRF, TCFA4, CHDS target genes and Pyramidal high set) are available in the accompanying Supplementary Software zip file. The human fetal single cell RNAseq data used in this study are available in the UCSC Cell Browser repository (https://cell.ucsc.edu/id=/ cortex-dev). GWAS data used in this study are available in the Psychiatric Genomics Consortium download database (https://www.med.yale.edu/ncg/download-results/) under accession codes: 29483656 (SZ, https://figshare.com/articles/dataset/scr2018conkor_16481220/), 31043756 (BP, https://figshare.com/articles/dataset/bip2019/14671998), 30478444 (ADHD, https://figshare.com/articles/dataset/adhd2019/14671965), 30804558 (ASD, https://figshare.com/articles/dataset/asd2019/14671989). The MDE GWAS data are available under restricted access due to inclusion of data from 23andMe, for which permission must be obtained separately. Access can be obtained by qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants; see cited study for details (https:// www.nature.com/articles/s41588-018-0090-3). Summary statistics for the MDD GWAS excluding the 23andMe data are available from the Psychiatric Genomics Consortium download database (https://www.med.yale.edu/ncg/download-results/) under accession code 29700475 (https://figshare.com/articles/dataset/mdd2018/14672085). The IQ GWAS data used in this study are available from (https://ctg.cnrc.nl/software/summary_statistics/) as a compressed file (Savagelians_MetaMuts_stats.zip). The AD GWAS data used in this study are available from the authors of the original study (https://www.nature.com/articles/ang2002). The SZ de novo rare variant data used in this study are available in Supplementary Tables 1 and 4 of the cited study (https://www.cell.com/cms/10106/;cell.2019.12.036 of attachment/ 44a4c411-6be3-4158-bd1d-6b339d60136edmcmcl.xds, https://www.cell.com/cms/10106/ jcell.2019.12.036/attachment/45f1b3ed-a2c2-4bc8-8e2c-a05ef8ad383edmcmc.xds). The GRC38p12 genome reference sequence and gene annotation data used for the bulk RNAseq analysis are available from GENCODE (ftp.ensembl.org/pub/databases/ genome/Gencode/human/release_31/genome/v31.primary_assembly.annotation.gtf.gz, ftp://ftp.ensembl.org/pub/databases/genome/Gencode/human/release_31/ GRC38p12_primarysembly_genome.fa.gz). Source data are provided with this paper.

Material availability
H7 hESC line was purchased from WiCell. DLG2 knockouts and WT sister line created from H7 are available from the corresponding author upon request, provided the purpose of the transfer is within the MOU and SLA from WiCell.

Code availability
All publicly available software utilized is noted in Methods. Analysis scripts are provided with this paper in Supplementary Software and as a citable GitHub repository (https:// zenodo.org/record/5729267#.YaeoXtDP2Uk)83.

Received: 25 May 2020; Accepted: 1 December 2021; Published online: 14 January 2022

References
1. Cardno, A. G. & Gottesman, I. I. Twin studies of schizophrenia: from bow- and-arrow controversies to star wars Mix and functional genomics. Am. J. Med. Genet. 97, 12–17 (2000).
2. Sullivan, P. F., Kendler, K. S. & Neale, M. C. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Arch. Gen. Psychiatry 60, 1187–1192 (2003).
3. International Schizophrenia Consortium. et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460, 748–752 (2009).
4. Fromer, M. et al. De novo mutations in schizophrenia implicate synaptic networks. Nature 506, 179–184 (2014).
5. Purcell, S. M. et al. A polygenic burden of rare disruptive mutations in schizophrenia. Nature 506, 185–190 (2014).
6. International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature 455, 237–241 (2008).
7. Walsh, T. et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science 339, 539–543 (2008).
8. Kirov, G. et al. De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Mol. Psychiatry 17, 142–153 (2012).
9. Szatkiewicz, J. P. et al. Copy number variation in schizophrenia in Sweden. Mol. Psychiatry 19, 762–773 (2014).
10. Poklington, A. J. et al. Novel findings from CNVs implicate inhibitory and excitatory signalling complexes in schizophrenia. Neuron 86, 1203–1214 (2015).
11. Genovese, G. et al. Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. Nat. Neurosci. 19, 1433–1441 (2016).
12. Pardiñas, A. F. et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat. Genet. 50, 381–389 (2018).
13. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. MedRxiv, https://doi.org/10.1101/2020.09.12.20192922 (2020).
14. Kosmicki, J. A. et al. Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. Nat. Genet. 49, 504–510 (2017).
15. Singh, T. et al. The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. Nat. Genet. 49, 1167–1173 (2017).
16. Rees, E. et al. De novo mutations identified by exome sequencing implicate rare missense variants in SLC6A1 in schizophrenia. Nat. Neurosci. 23, 179–184 (2020).
17. Hill, M. J. & Bray, N. J. Evidence that schizophrenia risk variation in the ZNF804A gene exerts its effects during fetal brain development. Am. J. Psychiatry 169, 1301–1308 (2012).
18. O’Brien, H. E. et al. Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders. Genome Biol. 19, 94–13 (2018).
19. Clifton, N. E. et al. Dynamic expression of genes associated with schizophrenia and bipolar disorder across development. Translational Psychiatry 9, 9 (2019).
20. Walker, R. L. et al. Genetic control of expression and splicing in developing human brain impairs disease mechanisms. Cell 179, 759–771.e22 (2019).
21. Polioudakis, D. et al. A single-cell transcriptomic atlas of human neocortical development during mid-gestation. Neurobiol. 103, 785–801.e8 (2019).
22. Ruzzo, E. K. et al. Inherited and de novo genetic risk for autism impacts shared networks. Cell 178, 850–866.e26 (2019).
23. Kang, H. J. et al. Spatio-temporal transcriptome of the human brain. Nature 478, 483–489 (2011).
24. van de Leemput, J. et al. CORTECON: a temporal transcriptome analysis of in vitro human cerebral cortex development from human embryonic stem cells. Neuron 83, 51–68 (2014).
25. Stephens, R. et al. The scribble cell polarity module in tissue development and tumorigenesis. J. Mol. Cell. Biol. 430, 3585–3612 (2018).
26. Frank, R. A. W. et al. NMDA receptors are selectively partitioned into complexes and supercomplexes during synapse maturation. Nat. Commun. 7, 13 (2016).
27. Citri, A. & Malenka, R. C. Synaptic plasticity: multiple forms, functions, and mechanisms. Neureuphychopharmacology 33, 18–41 (2008).
28. Chambers, S. M. et al. Highly efficient neural conversion of human ES and iPSCs by dual inhibition of SMAD signaling. Nat. Biotechnol. 27, 275–280 (2009).
29. Cambry, S. I. et al. Activin induces cortical interneuron identity and differentiation in embryonic stem cell-derived telencephalic neural precursors. Nat. Commun. 3, 841–841 (2012).
30. Husi, H. & Grant, S. Isolation of 2000-kDa complexes of N-methyl-D-aspartate receptor and postsynaptic density 95 from mouse brain. J. Neurochem. 77, 281–291 (2001).
31. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol. 11, e1004219 (2015).
32. Szpakowski, T. J. et al. Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. Science 358, 1318–1323 (2017).
33. Neale, B. M. et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 485, 242–245 (2012).
34. O’Roak, B. J. et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 485, 246–250 (2012).
35. Sanders, S. J. et al. De novo mutations revealed by whole-exome sequencing in 5,266 individuals. Science 334, 587–590 (2011).
36. Sanders, S. J. et al. De novo mutations revealed by whole-exome sequencing in 5,266 individuals. Nature 485, 246–250 (2012).
37. Cotonne, J. et al. The autism-associated chromosome modifie HD8 regulates other autism risk genes during human neurodevelopment. Nat. Commun. 6, 6404 (2015).
38. Amiel, J. et al. Mutations in TCF4, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. Am. J. Hum. Genet. 80, 988–993 (2007).
39. Ripke, S. et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat. Genet. 45, 1150–1159 (2013).
40. Forrest, M. P. et al. The psychiatric risk gene transcription factor 4 (TCF4) regulates neurodevelopmental pathways associated with schizophrenia, autism, and intellectual disability. Schizophr. Bull. 44, 1100–1110 (2018).
41. Darnell, J. C. et al. FMRP stalls ribosomal translation on mRNAs linked to synaptic function and autism. Cell 146, 247–261 (2011).
42. Wang, D. et al. Comprehensive functional genomic resource and integrative variation in Human. Science 362, eaaf864 (2018).
43. Sanders, S. J. et al. Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. NeuroReport 87, 1213–1215 (2013).
44. schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature 511, 421–427 (2014).
45. Shcherbikova, A. M. & Kosik, K. S. Neuronal RNA granules: a link between local RNA localization and stabilization-dependent translation. Neuron 32, 683–696 (2001).
46. Antar, L. N., Li, C., Zhang, H., Carroll, R. C. & Bassell, G. J. Local functions for FMRP in axon growth cone motility and activity-dependent regulation of filopodia and spine synapses. Mol. Cell. Neurosci. 32, 37–48 (2006).
47. Satterstrom, F. K. et al. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. Cell 180, 568–584.e23 (2020).
48. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. Nat. Genet. 51, 431–444 (2019).
49. Demontis, D. et al. Discovery of the first genome-wide significant risk locus for attention deficit/hyperactivity disorder. Nat. Genet. 51, 63–75 (2019).
50. Stahl, E. A. et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat. Genet. 51, 793–803 (2019).
51. Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat. Genet. 50, 668–671 (2018).
52. Savage, J. E. et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nat. Genet. 50, 912–919 (2018).
53. Lambert, J. C. et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer’s disease. Nat. Genet. 45, 1452–1458 (2013).
54. Skene, N. G. et al. Genetic identification of brain cell types underlying schizophrenia. Nat. Genet. 50, 825–833 (2018).
55. Teller, L. et al. Temporal patterning of apical progenitors and their daughter neurons in the developing neocortex. Science 364, eaay2522 (2019).
56. Simmons, B. A. & Zamponi, G. W. Neuronal voltage-gated calcium channels: structure, function, and dysfunction. Neuron 82, 24–45 (2014).
57. Bando, Y. et al. Control of spontaneous Ca2+ transients is critical for neuronal maturation in the developing neocortex. Cereb. Cortex 26, 106–117 (2016).
58. Kamijo, S. et al. A critical neurodevelopmental role for L-type voltage-gated calcium channel Cav1.4 in neurite extension and radial migration. J. Neurosci. 38, 5551–5566 (2018).
59. Craddock, N. & Owen, M. J. The Kraepelinian dichotomy - going, going... but still not gone. Br. J. Psychiatry 196, 92–95 (2010).
60. Owen, M. J., O’Donovan, M. C., Thapar, A. & Craddock, N. Neurodevelopmental hypothesis of schizophrenia. Br. J. Psychiatry 198, 173–175 (2011).
61. Owen, M. J. & O’Donovan, M. C. Schizophrenia and the neurodevelopmental continuum: evidence from genomics. World Psychiatry 16, 227–235 (2017).
62. Tau, G. Z. & Petersen, B. S. Normal development of brain circuits. Neуroporphysiology 35, 147–168 (2010).
63. Rungé, K. Cardoso, C. & de Chevigny, A. Dendritic spine plasticity: function and mechanisms. Front. Synaptic Neurosci. 12, 36 (2020).
supported by the NIH/NIA grant R01 AG033193 and the NIA AG081220 and AGES contract N01–AG–12100, the NHLBI grant R01 HL105756, the Icelandic Heart Association, and the Erasmus Medical Center and Erasmus University. ADGC was supported by the NIE/NIA grants: U01 AG032984, U24 AG201886, U01 AG016976, and the Alzheimer’s Association grant ADGC–10–19672. We thank the research participants and employees of 23andMe for the sharing of summary statistics for MDD common variant analysis.

Author contributions
Conceptualization, A.P., E.S. Design of Methodology, M.C., A.E., D.B., A.P., E.S. Software/Data curation, B.S., D.D.A., M.C., A.P., E.S. Formal analysis/Investigation, B.S., D.D.A., M.C., T.S., E.R., Y.Z., G.C., S.L., A.F.P., D.W., A.P., E.S. Writing – Original Draft, B.S., D.D.A., M.C., Y.Z., D.W., A.P., E.S. Writing – Review & Editing, B.S., D.D.A., M.C., E.R., S.L., A.F.P., W.G., M.O.D., M.O., A.E., D.B., D.W., A.P., E.S. Visualization, B.S., D.D.A., M.C., Y.Z., D.W., A.P., E.S. Supervision, A.P., E.S. Funding acquisition, A.H., M.O.D., M.O., A.P., E.S.

Competing interests
D.D.A., Y.Z., A.H., W.G., M.O.D., M.O., A.P. are supported by a collaborative research grant from Takeda (Takeda played no part in the conception, design, implementation, or interpretation of this study). The other authors report no financial relationships with commercial interests. The remaining authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-27601-0.

Correspondence and requests for materials should be addressed to Andrew J. Pocklington or Eunju Shin.

Peer review information Nature Communications thanks Zhexing Wen and the other anonymous reviewer(s) for their contribution to the peer review this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022