

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. Owen^{1,2}, Adam C. Errington¹, Derek J. Blake², Daniel J. Whitcomb⁴, Andrew J. Pocklington^{2*}, Eunju Shin^{1,5*}

¹ Neuroscience and Mental Health Research Institute, Cardiff University, Cardiff CF24 4HQ, UK

² MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff CF24 4HQ, UK

³ Department of Biomedical Science, University of Sheffield, Sheffield S10 2TN, UK

⁴ Bristol Medical School, University of Bristol, Bristol BS1 3NY, UK

⁵ School of Life Sciences, Keele University, Keele ST5 5BG, UK

*corresponding authors

Correspondence: e.shin@keele.ac.uk, shine@cardiff.ac.uk, pocklingtonaj@cardiff.ac.uk

Abstract

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.

Introduction

Schizophrenia (SZ) is a highly heritable^{1,2} psychiatric disorder, with genetic variation ranging from common polymorphisms (SNPs) to rare mutations contributing to disease risk³⁻⁷. Rare

variant studies consistently implicate disruption of postsynaptic signalling complexes in SZ aetiology^{4,5,8-11}, 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 sets^{12,13}, 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 SZ^{11,15,16}. We hypothesised that a significant proportion of SZ common variants may contribute to disease via the disruption of neurodevelopmental pathways harbouring 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 foetal 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 foetal 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 foetal 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 patients^{8,22}. 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* (*Dlg*) is a core component of the *Scrib* signalling 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 cases^{4,5,8-11}. This raises the possibility that DLG2 may be required for the normal operation of both adult and developmental signalling 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 sister wild-type (WT) hESC lines were differentiated into cortical excitatory neurons and cells were characterised 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 analysed 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 foetal 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 protocol^{28,29}; 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 characterised 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 expressed at each timepoint,

~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 available¹², we utilised the competitive gene-set enrichment test implemented in MAGMA³¹. As expected for cells of neural lineage, the set of all genes expressed at one or more timepoint in *DLG2*^{-/-} or WT lines (*all*^{WT+KO}) was enriched for common variant association ($P = 8.03 \times 10^{-21}$, $N_{\text{gene}} = 14,274$). To investigate whether genes up-/down-regulated at each timepoint displayed SZ association above that seen for neurodevelopmentally-expressed genes in general, we tested them for association conditioning on *all*^{WT+KO}. This revealed enrichment solely for genes down-regulated at day 30 (*30*_{down}^{-/-}: $P_{\text{corrected}} = 9.5 \times 10^{-8}$ 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*_{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 aetiology.

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) (Fig. 2c-i, 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 identity³². 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 30_{down}^{-/-} 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, characterised 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 hrs (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 30_{down}^{-/-} should be captured by genes normally upregulated between days 20 and 30 in WT cultures (20-30_{up}^{WT}). Analysing differential expression between WT samples at successive timepoints, we found risk variant enrichment only in 20-30_{up}^{WT} (conditioning on all WT-expressed genes, Fig. 4a). Most 20-30_{up}^{WT} genes overlapped 30_{down}^{-/-} (3075 genes, 85%) and this overlap captured the signal in both sets ($P_{\text{overlap}} = 3.23 \times 10^{-10}$; 30_{down}^{-/-} only $P = 0.44$; 20-30_{up}^{WT} only $P = 0.62$). This was not simply due to the size of the overlap as the regression coefficient for this set ($\beta = 0.14$), which reflects magnitude of enrichment, was significantly greater than for genes unique to 30_{down}^{-/-} ($\beta = 0.006$, $P_{\text{greater}} = 0.00077$) or 20-30_{up}^{WT} ($\beta = -0.015$, $P_{\text{greater}} = 0.0023$). 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 20-30_{up}^{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 $30_{\text{down}}^{-/-}$ (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 foetal brain tissue³² (Methods). Early-stable^{-/-} and early-increasing^{-/-} remained highly associated when conditioning on foetal neuron-expressed genes (Supplementary Data 6). Furthermore, all^{WT+KO} displayed association that was not captured by foetal neuron-expressed genes. We therefore continued to condition genetic analyses on all^{WT+KO} , as this best captures the broad SZ signal from neuronal-lineage genes present in our dataset.

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 remodelling (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 CHD8³³⁻³⁷ from early-transient^{-/-}; transcription factor TCF4^{12,38-40} and translational regulator FMRP^{12,41,42} from early-stable^{-/-}; and transcription factors (and deep layer markers) TBR1^{42,43} and BCL11B (CTIP2)^{12,13,42,44} 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 hypothesised 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 down-regulated in *CHD8* knockdown cells³⁶. 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^{41,45}, and its function is known to be important for axon and dendrite growth⁴⁶. 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^{4,5,8-11}, 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. 5l-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 depolarised 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 depolarisation and repolarisation ($\delta V/\delta t$) (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) was comparable at days 50 and 60 (Fig. 5n) as was EPSC frequency and amplitude (Fig. 5l, m). Lack of effect on synaptic transmission may reflect compensation by DLG4, whose expression shows a trend towards an increase in synaptosomes from day 65 *DLG2*^{-/-} neurons (Fig. 5o). In summary, developing *DLG2*^{-/-} neurons have a reduced ability to fire and produce less mature APs.

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 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^{-/-}, localising 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 *all*^{WT+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 harbouring 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, analysing 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 IQ⁵². 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 SZ^{4,5,8-11}. Our data indicates that it also regulates early developmental pathways harbouring 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-types^{13,54}. 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 (pyramidal^{high}) we investigated their overlap with neurogenic programs. Pyramidal^{high} genes were over-represented in early-stable^{-/-}, early-increasing^{-/-} and late sets (Fig. 7a). This overlap captured GWAS association in pyramidal^{high}, but not early-stable^{-/-} or early-increasing^{-/-} (Fig. 7b). In contrast to the late program as a whole (Fig. 4c, 6a), genes in the late-pyramidal^{high} overlap were enriched for SZ association (Fig. 7b) and LoFi genes (OR = 1.43, P = 0.035). Late-pyramidal^{high} genes also displayed a pattern of enrichment for regulatory targets almost identical to that of early-increasing^{-/-} (Fig. 7c, 4d), linking this subset into the terminal phase of the hypothesised transcriptional cascade.

These analyses suggest that the SZ association seen in pyramidal excitatory neurons^{13,54} 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. Pyramidal^{high} genes overlapping early-stable^{-/-}, early-increasing^{-/-} and late sets were over-represented for GO terms linked to dendrite/spine development, calcium-mediated exocytosis, postsynaptic signalling 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}.

Neurogenic programs identified *in vitro* are expressed in human foetal 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 foetal brain³². After normalising 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 foetal 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-cycle⁵⁵.

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 corticoneurogenesis *in vitro* and in human foetal cortex (Fig. 7d). These programs are 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 harbour 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 aetiological 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*, *FMRP*, *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 potentials⁵⁶, but are also known to impact neuronal growth and migration^{57,58}. Further experimental work is required to more precisely delineate phenotypes associated with the disruption of individual programs and the risk genes they harbour, 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^{-/-}, Fig. 7d) 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 aetiological and neurodevelopmental continuum, the more severe the disorder the greater the contribution from rare, damaging mutations and the earlier their developmental impact⁵⁹⁻⁶¹ (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 harbouring pathogenic mutations are distributed across diverse pathways shaping pre-/post-natal 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 up-regulated during the NPC-neuron transition and persist as neurons develop (Fig. 7d), shaping their morphology, function and possibly connectivity (Fig. 8a).

SZ GWAS association previously noted in CA1 pyramidal neurons^{13,54} was captured by a subset of neurogenic genes contributing to dendrite/spine/synapse formation, signalling and plasticity. These processes underlie the establishment and maturation of neuronal circuits⁶² and their learning-dependent modification in adults⁶³. The most parsimonious explanation is that SZ common variants act largely via disruption of early brain development, as neurogenic programs harbour GWAS association beyond their overlap with pyramidal^{high} (Fig. 7b); and SZ shares extensive SNP heritability with early-onset disorders^{48,49}. However, effect sizes are greater for shared genes (Fig. 7b) – although not significantly so – and SZ onset extends from late childhood well into adulthood⁶⁴. We hypothesise that vulnerability to SZ is primarily established 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 signalling complexes²⁶. 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* signalling module²⁵, *DLG2* may link cell-surface receptors to signal transduction pathways regulating the activation of neurogenic programs (Supplementary Fig. 9). We hypothesise that stochastic signalling 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 behaviour.

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

Methods

hESC culture

H7 hESC line (WA07) is purchased from Wi Cell, USA. All hESC work were performed in accordance with Cardiff University's regulation 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% CO₂ in 6 well cell culture plates (Greiner) coated with 1% Matrigel hESC-Qualified Matrix (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 (E8, Thermo Fisher Scientific) and passaged at 80% confluency using Versene solution (Thermo Fisher Scientific) for 1.5 minutes 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)⁶⁵. 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 5µg of plasmids, FACS sorted on the following day and plated at a low density (~70 cells/cm²) for clonal isolation. 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 genic 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 protocol²⁸ with modifications (embryoid body to monolayer and replacement of KSR medium with N2B27 medium) suggested by Cambray et al., 2012²⁹. Prior to differentiation Versene treatment and mechanical dissociation was used to passage hESCs at approximately 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% CO₂ until 90% 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 Step Glutamine (Thermo Fisher Scientific) and 50 μ M 2-Mercaptoethanol (Thermo Fisher Scientific), which was supplemented with 100 nM LDN193189 (Cambridge Biosciences) and 10 μ M SB431542 (Strattech Scientific) 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 μ g/ml human plasma fibronectin (Merck) in Dulbecco's phosphate-buffered saline (DPBS, Thermo Fisher Scientific), passage was as previously described with the addition of a 1 hour incubation with 10 μ M Y27632 Dihydrochloride (ROCK inhibitor, Strattech Scientific) prior to Versene dissociation. During days 10 to 20 of differentiation cells were maintained in N2B27-RA (without LDN193189 or SB431542 supplementation) and passaged at day 20 in a 1:4 ratio into 24 well cell culture plates (Greiner) sequentially coated with 10 μ g/ml poly-d-lysine hydrobromide (PDL, Sigma) and 15 μ g/ml laminin (Sigma) in DPBS. Vitamin A was added to the differentiation media at day 26, standard 1x B27 Supplement (Thermo Fisher Scientific) replacing 1x B27 Supplement minus vitamin A, and cells 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 30, 1:2 onto PDL-laminin as previously described. In all cases cells maintained past day 30 were fed with N2B27+RA supplemented with 2 μ g/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 minutes at 4°C followed by a 1 hour 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 'Antibody concentration'), were diluted in blocking solution and incubated with cells overnight at 4°C. Following removal of primary antibody solution and 3 PBS washes, cells were incubated in the dark for 2 hours at room temperature with appropriate Alexa Fluor secondary antibodies (Thermo Fisher Scientific) diluted 1:500 with blocking solution. After an additional 2 PBS washes cells were counterstained with DAPI nucleic acid stain (Thermo Fisher Scientific), diluted 1:1000 with PBS, for 5 mins at room 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 minutes 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 minutes at 70°C before loading into Bolt 4-12% Bis-Tris Plus gels (Thermo Fisher Scientific). Gels were run at 120V for 2-3 hours 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 120V for 1 hour 45 minutes. Transfer was confirmed by visualising 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 hour 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 hour at room temperature with IRDye secondary antibodies (LI-COR) diluted 1:15000 with blocking solution. After 3 TBS washes staining was visualised using the Odyssey CLx Imaging System (LI-COR) and Image Studio Lite Version 5.2 (LI-COR).

Antibody concentration

Antibody	Company	Identifier Catalog No.	Dilution in ICC	Dilution in WB
Rabbit polyclonal anti-Calnexin	Abcam	ab22595	N/A	1:5000
Rat monoclonal anti-CTIP2 [25B6]	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 polyclonal anti-DLX1	AbCam	ab167575	1:100	N/A
Rabbit polyclonal anti-FOXP1	Abcam	ab18259	1:250	N/A
Mouse monoclonal anti-FOXP1 [JC12]	Abcam	ab32010	1:800	N/A
Rabbit polyclonal anti-GABA	Sigma	A2052	1:500	N/A
Rabbit polyclonal anti-GAPDH	Abcam	ab9485	N/A	1:5000
Goat polyclonal anti-GBX2	Antibodies.com	A84236	1:200	N/A
Mouse monoclonal anti-KI67 [B56]	BD Pharmingen	550609	1:150	1:1000
Mouse monoclonal anti-MAP2 [AP-20]	Sigma-Aldrich	M1406	1:250	N/A
Rabbit polyclonal anti-MAP2	Merck Millipore	AB5622	N/A	1:1000
Rabbit monoclonal anti-NANOG [D73G4]	Cell Signaling Technology	4903	1:400	N/A
Rabbit polyclonal anti-NEUN	Merck Millipore	ABN78	1:500	1:500
Rabbit monoclonal anti-NKX2.1 [EP1584Y]	Abcam	ab76013	1:1000	N/A
Rabbit monoclonal anti-OCT4 [C30A3]	Cell Signaling Technology	2840	1:400	N/A
Mouse monoclonal anti-OLIG3 [257934]	R&D Systems	MAB2456	1:500	N/A

Rabbit monoclonal anti-PAX6 [EPR15858]	Abcam	ab195045	1:300	N/A
Rabbit polyclonal anti-PSD-95 (DLG4)	Abcam	ab18258	N/A	1:1000
Mouse monoclonal anti-SATB2 [SATBA4B10]	Abcam	ab51502	1:25	1:100
Rabbit monoclonal anti-SOX2 [D6D9]	Cell Signaling Technology	3579	1:400	N/A
Goat polyclonal anti-SOX2	Santa Cruz	sc-17320	N/A	1:200
Rabbit polyclonal anti-TBR1	Abcam	ab31940	1:500	1:1000
Mouse monoclonal anti- β -Tubulin III (TUJ1) [SDL.3D10]	Sigma-Aldrich	T8660	1:1000	N/A
Donkey anti-Rabbit IgG (H+L) secondary antibody, Alexa Fluor 488	Thermo Fisher Scientific	A-21206	1:500	N/A
Donkey anti-Rat IgG (H+L) secondary antibody, Alexa Fluor 488	Thermo Fisher Scientific	A-21208	1:500	N/A
Donkey anti-Mouse IgG (H+L) secondary antibody, Alexa Fluor 594	Thermo Fisher Scientific	A-21203	1:500	N/A
Donkey anti-Rabbit IgG (H+L) secondary antibody, Alexa Fluor 594	Thermo Fisher Scientific	A-21207	1:500	N/A
IRDye 680RD goat anti-Rabbit	Li-Cor	926-68071	N/A	1:15000
IRDye 800CW goat anti-Mouse	Li-Cor	926-32210	N/A	1:15000
IRDye 800CW donkey anti-Goat	Li-Cor	925-68074	N/A	1:15000

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 (1,200g, 10 min, 4°C) the supernatant was centrifuged at high speed to pellet synaptosomes (15,000g, 20 min, 4°C) which were resuspended in fresh Syn-PER Reagent. Protein concentration was determined using a DC Protein Assay (BioRad) quantified with the CLARIOstar microplate reader (BMG Labtech).

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 synthesised (Pepceuticals) and fully dissolved in 90% v/v methanol + 1M 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 1M Tris pH9 (Sigma) with 2 hours room temp incubation on a roller mixer. The peptide bound resin was then washed 3 times with DOC buffer (1% w/v sodium deoxycholate; 50mM 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 hour on ice with regular vortexing, cell debris was pelleted by high speed centrifugation (21,300g, 2 hours, 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 5 times with ice cold DOC buffer and

the bound protein eluted by 15 minute 70°C incubation in 5% w/v sodium dodecyl sulphate (SDS, Sigma). The eluted protein was reduced with 10 mM TCEP and alkylated using 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 hour. 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 to ⁶⁶. Briefly, peptides were analysed 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 20mm (Thermo Fisher) and then separated using a 130-min gradient from 3 to 40% buffer B (0.5% formic acid in 80% acetonitrile) on an EASY-Spray column, 50 cm × 50 µm ID, PepMap C18, 2 µm particles, 100 Å pore size (Thermo Fisher). The Orbitrap Elite was operated with a cycle of one MS (in the Orbitrap) acquired at a resolution of 60,000 at m/z 400, with the top 20 most abundant multiply charged (2+ and higher) ions in a given chromatographic window subjected to MS/MS fragmentation in the linear ion trap. An FTMS target value of 1e6 and an ion trap MSn target value of 1e4 were used with the lock mass (445.120025) enabled. Maximum FTMS scan accumulation time of 500 ms and maximum ion trap MSn scan accumulation time of 100 ms were used. Dynamic exclusion was enabled with a repeat duration of 45 s with an exclusion list of 500 and an exclusion duration of 30 s. Raw mass spectrometry data were analysed 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 carbamidomethylation as a fixed modification, match between runs enabled with a match time window 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 neighbours of 3 and an average number of neighbours 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.

CNV analysis

Following manual dissociation of WT and *DLG2* KO hESC into DPBS, genomic DNA was extracted using the ISOLATE II Genomic DNA kit (Bioline). Following DNA amplification and fragmentation according to the associated Illumina HTS assay protocol samples were hybridized to an Infinium PsychArray v1.1 BeadChip (Illumina). The stained bead chip was imaged using the iScan System (Illumina) and Genome Studio v2.0 software (Illumina) subsequently used to normalise the raw signal intensity data and perform genotype clustering. Final analysis for Copy Number Variation (CNV) was carried out with PennCNV software⁶⁸.

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 lysing 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 analysed using the 2100 Bioanalyzer Eukaryote Total RNA Nano assay (Agilent). cDNA libraries for sequencing were produced 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 analysed 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 pair paired end reads and 8 base pair index reads.

All samples were modelled after the long-rna-seq-pipeline used 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)⁶⁹ and trimmed for adapter sequence and low base call quality (Phred score < 30 at ends) with cutadapt (v2.3)⁷⁰. The mapping of the trimmed reads was done using STAR (v2.7.0e)⁷¹ and the BAM files were produced in both genomic and transcriptomic coordinates and sorted using samtools (v1.9)⁷². The aligned and sorted BAM files were further assessed for quality using Picard tools (v2.20.2)⁷³. 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 KO1 and WT samples for this timepoint. GRCh38.p12 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)⁷⁴. Both STAR and RSEM executions were performed using the psychENCODE parameters.

RSEM gene and isoform level estimated counts were imported using the tximport package (v1.12.3)⁷⁵. 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)⁷⁶ and differentially expressed genes were considered significant if their p value after Bonferroni correction was < 0.05. Differential exon usage was analysed using the DEXSeq pipeline⁷⁷. Briefly, the GENCODE annotation .gtf file was translated into a .gff file with collapsed exon counting bins by using the dexseq_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 (0.11.2)⁷⁸. Finally, differential exon usage was evaluated using DEXSeq (v1.30)⁷⁷ and

significant differences identified using an FDR threshold of 0.05. All the differential analyses were performed by using R (v3.6.1).

When analysing differential gene expression in *DLG2*^{-/-} relative to WT, samples from KO1 and KO2 lines were combined i.e. for each timepoint a single differential gene expression analysis was performed, comparing expression in KO1 & KO2 samples against wild-type. To assess the impact of sample dropout at day 30, we investigated the similarity in gene expression between lines by clustering all KO1, KO2 and WT samples (Supplementary Fig. 3a). At all 4 timepoints, all replicates from KO1 and KO2 cluster together: while KO2 samples from day 30 are not of sufficient quality to be used with confidence in further analyses, they are clearly similar to KO1 day 30 samples. We also performed differential expression analyses separately for each line (i.e. KO1 v WT and KO2 v WT) at all other timepoints. The overlap in expressed genes accounted for >98% of the genes expressed in each line and gene expression fold change was highly correlated between KO1 v WT and KO2 v WT (Spearman's $\rho^{\text{day } 15} = 0.67$, $\rho^{\text{day } 20} = 0.95$, $\rho^{\text{day } 60} = 0.75$). Over-representation odds ratios for GO terms also remain well correlated for significantly up-regulated ($\rho^{\text{day } 15} = 0.70$, $\rho^{\text{day } 20} = 0.92$, $\rho^{\text{day } 60} = 0.67$) and down-regulated ($\rho^{\text{day } 15} = 0.55$, $\rho^{\text{day } 20} = 0.95$, $\rho^{\text{day } 60} = 0.56$) genes. We noted that agreement between lines was greatest for day 20, which also lies close to the onset of neurogenesis and displays a high level of differential expression between KO and WT lines, comparable to that for day 30 (Figure 1g). Further indicating a limited impact for sample dropout, phenotypes predicted by GO term analysis of differential expression at day 30 (deficits in neuron migration, morphology and action potential generation) were experimentally validated (Fig. 3 & 5); and all early neurogenic transcriptional programs identified in these data were shown to possess an identical profile of expression across human neurodevelopmental cell-types *in vivo* (Fig. 7).

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 (analysing WT data only): genes up-regulated in day 30 relative to day 20 (20-30_{up}^{WT}); genes up-regulated in day 60 relative to day 30 (30-60_{up}^{WT}); and genes up-regulated in day 60 relative to day 20 (20-60_{up}^{WT}). 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.

Human foetal cortex single-cell RNA sequencing data

Single-cell RNA-Seq gene expression data from Nowakowski et al., 2017³² were downloaded from bit.ly/cortexSingleCell. 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 clusters³² as follows:

Progenitors

RG (early): "RG-early"

RG: "RG-div1", "RG-div2", "oRG", "tRG", "vRG"

IPC: "IPC-div1", "IPC-div2"

Transitioning: "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 ($\text{TPM} \geq 1$) 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 normalised for each gene across all cells, then the average normalised expression score for each gene was calculated for each of the above cell-types. Over 80% of genes for each *in vitro* early neurogenic program (early-transient^{-/-}, early-stable^{-/-}, early-increasing^{-/-}) and 55% (101 out of 182 genes) of the late-pyramidal^{high} set were present in the *in vivo* data; all of these genes passed our filtering criteria. Taking the set of genes corresponding to each *in vitro* program, we calculated the mean and standard error of their gene-level averages in each *in vivo* cell-type (Fig. 7). For each program, the difference between successive neurodevelopmental cell-types/stages was calculated using a two-tailed Student's t-test and p-values Bonferroni-corrected for the 6 comparisons made: RG (early) v RG; RG v IPC; RG v Transitioning; IPC v Transitioning; Transitioning v Newborn neurons; and Newborn v Developing neurons.

To create a stringent set of foetal neuron expressed genes, we identified all genes expressed in at least 5% of Newborn cells or 5% of Developing cells ($N_{\text{gene}} = 9,332$). To create a more relaxed set, we identified all genes with non-zero expression in at least one Newborn or Developing cell ($N_{\text{gene}} = 16,431$).

Gene set construction

GO

The Gene Ontology (GO) ontology tree was downloaded from OBO:

<http://purl.obolibrary.org/obo/go/go-basic.obo>

Ontology trees were constructed separately for Molecular Function, Biological Process and Cellular Component using 'is_a' and 'part_of' relationships. GO annotations were downloaded from NCBI:

<ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/gene2go.gz>

Annotations containing the negative qualifier NOT were removed, as were all annotations with evidence codes IEA, NAS and RCA. Annotations were further restricted to protein-coding genes. Genes corresponding to each annotation term were then annotated with all 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 *TBR1* and *BCL11B* targets⁴²: 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 *TBR1* and *BCL11B* extracted (interaction file: INT-11_ElasticNet_Filtered_Cutoff_0.1_GRN_1.csv). Gene symbols were mapped to NCBI/Entrez ids using data from the NCBI gene_info file.

TCF4 targets⁴⁰: identifiers were updated using the gene_history file from NCBI.

FMRP targets⁴¹: NCBI/Entrez mouse gene identifiers were updated using the gene_history file from NCBI. Genes were then mapped from mouse to human using Homologene, restricting to protein-coding genes with a 1-1 mapping.

Direct *CHD8* mid-foetal promoter targets³⁷: symbols (NCBI gene_info file) and locations (NCBI Build37.3) were used to map genes to NCBI/Entrez gene ids.

Indirect *CHD8* targets³⁶: Ensembl ids were updated (Ensembl stable_id_event file) then mapped to current NCBI/Entrez ids using Ensembl and NCBI id cross-reference files. Taking genes with altered expression on *CHD8* shRNA knockdown, we removed those identified as direct *CHD8* targets in NPCs³⁶ or as *CHD8* mid-foetal promoter targets³⁷.

Pyramidal CA1 'specific' (pyramidal^{high})

Mouse genes with specificity score > 0 for pyramidal CA1 neurons were extracted from Supplementary Table 3 of Skene et al., 2018⁵⁴ and mapped to human using the gene orthology file from NCBI (https://ftp.ncbi.nlm.nih.gov/gene/DATA/gene_orthologs.gz). To construct the pyramidal^{high} set we took the top 10% of 1-1 mapped genes ranked by specificity score ($N_{\text{gene}} = 1,405$).

Functional over-representation test (gene set overlap)

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 *DLG2*^{-/-} lines (*all*^{WT+KO}). This was used for GO terms; the 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, 7), we used an iterative refinement procedure. Briefly, we selected the gene set with the largest enrichment odds ratio; removed all genes in this set from all other over-represented annotations; re-tested these reduced gene-sets for over-representation in 30_{down}^{-/-} genes; then discarded gene-sets with $P \geq 0.05$ (after Bonferroni-correction for the number of sets tested in that iteration). This process was repeated (with gene-sets being cumulatively depleted of genes at each iteration) until there were no remaining sets with a corrected $P < 0.05$.

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 or less of the SZ association in these programs than expected, a two-sided enrichment test was performed on term/target genes within the program, conditioning on all^{WT+KO} and on all genes in the program. All other GWAS enrichment tests were one-sided. To test whether common variant enrichment differed between two gene-sets, we took the regression coefficient β and its standard error $SE(\beta)$ for each gene-set from the MAGMA output file and compared $z = d/SE(d)$ to a standard normal distribution, where $d = \beta_1 - \beta_2$ and $SE(d) = \sqrt{[SE(\beta_1)^2 + SE(\beta_2)^2]}$. Gene-level association statistics for schizophrenia were taken from Pardiñas et al., 2018¹²; those for ADHD⁴⁹, bipolar disorder⁵⁰ and Alzheimer's disease⁵³ were calculated using the MAGMA multi model, with a fixed 20,000 permutations for each gene. Prior to analysis, SNPs with MAF < 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 analysed here are described in Rees et al., 2020¹⁶. *De novo* LoF mutations for NDD, ASD and unaffected siblings of individuals with ASD were taken from Satterstrom et al., 2020⁴⁷: these were re-annotated using Variant Effect Predictor⁷⁹ and mutations mapping to > 2 genes (once readthrough annotations had been discarded) were removed from the analysis. A two-sided Poisson rate 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 rates⁸⁰.

Migration assay

Cells were cultured and differentiated to cortical projection neurons as previously described. Neuronal migration was measured during a 70-hour period from day 40 by transferring cell culture plates to the IncuCyte Live Cell Analysis System (Sartorius). Cells were maintained at 37°C and 5% CO₂ with 20X magnification phase contrast images taken of whole wells in every 2 hours for the analysis period. The StackReg plugin⁸¹ 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 analysed 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-minute Accutase solution (Sigma) dissociation followed by plating at 100,000 cells per well on 24 well culture plates. 72 hours prior to morphology assessment cells were transfected with 500ng pmaxGFP (Lonza) per well using Lipofectamine 3000 Reagent (Thermo Fisher Scientific) and Opti-MEM Reduced Serum Media (Thermo Fisher Scientific) for the preparation of DNA-lipid complexes. At days 30 or 70, cells were fixed in 4% paraformaldehyde (PFA, Sigma) in PBS for 20 minutes at 4°C before mounting with Dako Fluorescence Mounting Medium (Agilent) and glass coverslips. Random fields were imaged using a DMI6000B Inverted microscope (Leica) and the morphology of GFP expressing cells with a clear neuronal phenotype quantified using the NeuroLucida 360 v2.00.2 (MBF Bioscience) neuron tracing and analysis software package. Data sets (Fig. 3a-d) were analysed by two-way ANOVA with post hoc comparisons using Bonferroni correction, comparing to WT controls.

Electrophysiology

Whole cell patch clamp electrophysiology was performed on cells cultured on 13mm 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 from WT and KO hESCs were dissociated and plated on each PDL-coated coverslip in 30µl diluted (20x) matrigel (Corning) together with 20,000 rat primary glial cells. Postnatal day 7-10 old Sprague-Dawley rats (Charles River) bred in-house were sacrificed via cervical dislocation and cortex was quickly dissected. Tissues were dissociated using 2mg/ml papain and plated in DMEM supplemented with 10% Foetal bovine serum and 1% penicillin/streptomycin/Amphotericin B and 1x Glutamax (all Thermo Fisher Scientific). Microglia and oligodendrocyte precursor cells were removed by shaking at 500 rpm for 24 hours 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), 10ng/ml BDNF (Cambridge Bioscience) and 200µM ascorbic acid (Sigma). To stop the proliferation of cells, 1x CultureOne (Thermo Fisher Scientific) was supplemented from day 21. For postsynaptic current experiment, coverslips were transferred to a recording chamber (RC-26G, Warner Instruments) and perfused with HEPES Buffered Saline (HBS) (119 mM NaCl; 5 mM KCl; 25 mM HEPES; 33 mM glucose; 2mM CaCl₂; 2mM MgCl₂; 1µM glycine; 100µM 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 Cs-based intracellular filling solution (130 mM CsMeSO₄; 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 osmolarity ~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 and analysed online and

offline using the WinLTP software⁸² (<http://www.winltp.com>). Only cells with series resistance <25 M Ω with a change in series resistance <10% from the start were included in this study. Data were analysed by importing Axon Binary Files into Clampfit (version 10.6; Molecular Devices). A threshold function of >12 pA was used to identify mEPSC events, which were then subject to manual confirmation. Results were outputted 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 20kHz with a 3 kHz Bessel filter with MultiClamp 700B amplifier. Coverslips were transferred into the recording chamber maintained at RT (20-21°C) on the stage of an Olympus BX61W (Olympus) differential interference contrast (DIC) microscope and perfused at 2.5ml/min with the external solution composed of 135mM NaCl, 3mM KCl, 1.2mM MgCl₂, 1.25mM CaCl₂, 15mM D-glucose, 5mM 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 117mM KCl, 10mM NaCl, 11mM HEPES, 2mM Na₂-ATP, 2mM Na-GTP, 1.2mM Na₂-phosphocreatine, 2mM MgCl₂, 1mM CaCl₂ and 11mM EGTA (all from Sigma), and pH was titrated to 7.3 by NaOH. The resistance of a patch pipette was 3–9 M Ω and the series resistance component was fully compensated using the bridge balance function of the instrument. The RMP of cells was recorded immediately after breaking into the cells in gap free mode. A systematic current injection protocol (duration, 1 s; increment, 20 pA; from -60pA to 120pA) was applied to the neurons held at -60mV to evoke APs. Input resistance (R_{in}) was calculated by $R_{in}=(V_i-V_m)/I$, where V_i is the potential recorded from -10pA current step. The AP properties are measured by the first over shooting AP. Further analysis for action potential characterization 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 differentiations. Within a given differentiation triplicate samples were used per cell line at each time point investigated.

References

1. Cardno, A. G. & Gottesman, I. I. Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx 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* **320**, 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. Pocklington, A. J. *et al.* Novel Findings from CNVs Implicate Inhibitory and Excitatory Signaling 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**, 194–13 (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 Informs Disease Mechanisms. *Cell* **179**, 750–771.e22 (2019).
21. Polioudakis, D. *et al.* A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. *Neuron* **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 the Regulation of Cell Signaling in Tissue Development and Tumorigenesis. *Journal of Molecular Biology* **430**, 3585–3612 (2018).
26. Frank, R. A. W. *et al.* NMDA receptors are selectively partitioned into complexes and supercomplexes during synapse maturation. *Nature Communications* **7**, –13 (2016).
27. Citri, A. & Malenka, R. C. Synaptic plasticity: multiple forms, functions, and mechanisms. *Neuropsychopharmacology* **33**, 18–41 (2008).
28. Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat. Biotechnol.* **27**, 275–280 (2009).
29. Cambray, S. I. *et al.* Activin induces cortical interneuron identity and differentiation in embryonic stem cell-derived telencephalic neural precursors. *Nature Communications* **3**, 841–11 (2012).
30. Husi, H. & Grant, S. Isolation of 2000-kDa complexes of N-methyl-D-aspartate receptor and postsynaptic density 95 from mouse brain. *Journal of Neurochemistry* **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. Nowakowski, 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 are strongly associated with autism. *Nature* **485**, 237–241 (2012).
36. Sugathan, A. *et al.* CHD8 regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E4468–77 (2014).
37. Cotney, J. *et al.* The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. *Nature Communications* **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. *Schizophrenia Bulletin* **44**, 1100–1110 (2018).
41. Darnell, J. C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247–261 (2011).
42. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, eaat8464 (2018).

43. Sanders, S. J. *et al.* Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87**, 1215–1233 (2015).
44. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
45. Krichevsky, A. M. & Kosik, K. S. Neuronal RNA granules: a link between RNA localization and stimulation-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. *Molecular and Cellular Neuroscience* **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 loci 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**, 668–681 (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. Telley, L. *et al.* Temporal patterning of apical progenitors and their daughter neurons in the developing neocortex. *Science* **364**, eaav2522 (2019).
56. Simms, 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 Channels 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. & Peterson, B. S. Normal development of brain circuits. *Neuropsychopharmacology* **35**, 147–168 (2010).
63. Runge, K., Cardoso, C. & de Chevigny, A. Dendritic Spine Plasticity: Function and Mechanisms. *Front Synaptic Neurosci* **12**, 36 (2020).

64. Solmi, M. *et al.* Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. *Mol. Psychiatry* 1–15 (2021). doi:10.1038/s41380-021-01161-7
65. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
66. Woodley, K. T. & Collins, M. O. S-acylated Golga7b stabilises DHH5 at the plasma membrane to regulate cell adhesion. *EMBO Rep* **17**, 813–19 (2019).
67. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
68. Wang, K. *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* **17**, 1665–1674 (2007).
69. Andrews, S. FastQC: A Quality Control tool for High Throughput Sequence Data. <https://www.bioinformatics.babraham.ac.uk/projects/> (2010).
70. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
71. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
72. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
73. Broad Institute. Picard Tools. <https://broadinstitute.github.io/picard/> (2019).
74. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323–16 (2011).
75. Sonesson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* **4**, 1521 (2015).
76. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 31–21 (2014).
77. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res.* **22**, 2008–2017 (2012).
78. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
79. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122–14 (2016).
80. Ware, J. S., Samocha, K. E., Homsy, J. & Daly, M. J. Interpreting de novo Variation in Human Disease Using denovolyzeR. *Curr Protoc Hum Genet* **87**, 7.25.1–7.25.15 (2015).
81. Thévenaz, P. StackReg: An ImageJ plugin for the recursive alignment of a stack images. <http://bigwww.epfl.ch/thevenaz/stackreg/> (2010).
82. Anderson, W. W. & Collingridge, G. L. Capabilities of the WinLTP data acquisition program extending beyond basic LTP experimental functions. *Journal of Neuroscience Methods* **162**, 346–356 (2007).
83. D'Andrea, D. & Pocklington, A. J. DLG2 knockout in human embryonic stem cells impacts neurodevelopmental transcriptional programs harbouring genetic risk for Schizophrenia and related disorders. *GitHub*, <http://doi.org/10.5281/zenodo.5729267> (2021).

84. Singh, T. *et al.* Exome sequencing identifies rare coding variants in 10 genes which confer substantial risk for schizophrenia. *MedRxiv*, <http://doi.org/10.1101/2020.09.18.20192815> (2020).

Acknowledgments

This work was supported by Wellcome Trust Strategic Award (100202/Z/12/Z, AH & MO), MRC programme grant (G08005009, MOD, MO & AP), MRC Centre grant (MR/L010305/1, MOD, MO & AP), Waterloo Foundation 'Changing Minds' programme (506926, ES) and start-up funding (AH1132S012, AH1132S204, ES) from the Neuroscience and Mental Health Research Institute, Cardiff University. We acknowledge excellent technical support for RNA sequencing from Joanne Morgan (MRC Centre) and MS analysis from Lydia Kiesel (University of Sheffield) and assistance in morphology tracing from Sophie Pocklington. We appreciate excellent general lab support from Emma Dalton, Trudy Workman and Olena Petter. We thank Prof. Meng Li for her advice and Dr. Claudia Tamburini for technical support in the initial stages of the project and Profs. Yves Barde and Lesley Jones for helpful comments on the manuscript and Emily Adair for providing rat primary glial cells.

We thank the International Genomics of Alzheimer's Project (IGAP) for providing summary results data for AD common variant analysis. The investigators within IGAP contributed to the design and implementation of IGAP and/or provided data but did not participate in analysis or writing of this report. IGAP was made possible by the generous participation of the control subjects, the patients, and their families. The i-Select chips was funded by the French National Foundation on Alzheimer's disease and related disorders. EADI was supported by the LABEX (laboratory of excellence program investment for the future) DISTALZ grant, Inserm, Institut Pasteur de Lille, Université de Lille 2 and the Lille University Hospital. GERAD was supported by the Medical Research Council (Grant no 503480), Alzheimer's Research UK (Grant no 503176), the Wellcome Trust (Grant no 082604/2/07/Z) and German Federal Ministry of Education and Research (BMBF): Competence Network Dementia (CND) grant no 01GI0102, 01GI0711, 01GI0420. CHARGE was partly 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 NIH/NIA grants: U01 AG032984, U24 AG021886, 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, AP, ES

Design of Methodology, MC, AE, DB, AP, ES

Software/Data curation, BS, DDA, MC, AP, ES

Formal analysis/Investigation, BS, DDA, MC, TS, ER, YZ, GC, SL, AFP, DW, AP, ES

Writing – Original Draft, BS, DDA, MC, YZ, DW, AP, ES

Writing – Review & Editing, BS, DDA, MC, ER, SL, AFP, WG, MOD, MO, AE, DB, DW, AP, ES

Visualization, BS, DDA, MC, YZ, DW, AP, ES

Supervision, AP, ES

Funding acquisition, AH, MOD, MO, AP, ES

Competing Interests statement

DDA, YZ, AH, WG, MOD, MO, AP 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.

Correspondence and requests for materials should be addressed to E.S or A.J.P.

Data availability

RNAseq data generated by this study have been deposited in the Gene Expression Omnibus (GEO) archive with accession number GSE172199 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=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 (<http://www.ebi.ac.uk/pride/archive/projects/PXD029526>). Human proteome reference sequences are available in UniProt (<https://www.uniprot.org/>). Processed gene annotation data (GO terms, LoFi genes, BCL11B, TBR1, FMRP, TCF4, CHD8 target genes and Pyramidal^{high} set) are available in the accompanying Supplementary Software zip file. The human foetal single cell RNAseq data used in this study are available in the UCSC Cell Browser repository (bit.ly/cortexSingleCell). GWAS data used in this study are available in the Psychiatric Genomics Consortium download database (<https://www.med.unc.edu/pgc/download-results/>) under accession codes: 29483656 (SZ, <https://figshare.com/articles/dataset/scz2018clozuk/14681220>), 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 MDD 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.unc.edu/pgc/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.cncr.nl/software/summary_statistics/) as a compressed file (SavageJansen_IntMeta_sumstats.zip). The AD GWAS data used in this study are available from the authors of the original study (<https://www.nature.com/articles/ng.2802>). The SZ *de novo* rare variant data used in this study are available from the cited study (<https://www.nature.com/articles/s41593-019-0565-2>). The ASD and NDD *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/10.1016/j.cell.2019.12.036/attachment/44aca411-6be3-4158-b1d6-6b339d60136d/mmc1.xlsx>, <https://www.cell.com/cms/10.1016/j.cell.2019.12.036/attachment/45f1baed-a2ec-4128-be82-a50ae08adc38/mmc4.xlsx>). The GRCh38.p12 genome reference sequence and gene annotation data used for the bulk RNAseq analysis are available from GENCODE (https://ftp.ebi.ac.uk/pub/databases/genocode/Genocode_human/release_31/genocode.v31.primary_assembly.annotation.gtf.gz, https://ftp.ebi.ac.uk/pub/databases/genocode/Genocode_human/release_31/GRCh38.primary_assembly.genome.fa.gz).

Code availability

All publicly available software utilised is noted in Methods. Analysis scripts are provided with this paper in Supplementary Software and as a citeable GitHub repository⁸³.

Biological 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.

Figure Legends

Figure 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 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 late) programs. LoFi genes in early neurogenic programs were enriched for common/rare variants contributing to mental disorders (Autism spectrum disorder, ASD; attention-deficit hyperactivity disorder, ADHD; Bipolar disorder, BP; major depression, MDD) 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 (pyramidal^{high}). The expression profile seen for each disease-associated neurogenic program *in vitro* was recapitulated across neurodevelopmental cell-types from human foetal 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 **c** ($F_{2,9}=45.54$, $P=1.96 \times 10^{-5}$) and **e** ($F_{2,9}=172.9$, $P=6.59 \times 10^{-8}$). Kruskal-Wallis test with Dunn's multiple comparison correction applied to **d** ($H(2)=10.46$, $p=0.0061$) 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.

Figure 2. Common risk variants implicate disruption of neurogenesis in schizophrenia. a, Schizophrenia GWAS association in genes up-/down-regulated at each timepoint (*DLG2*^{-/-} relative to WT), conditioning on all expressed genes. **b**, Gene ontology (GO) terms over-represented amongst genes down-regulated at day 30 in *DLG2*^{-/-} lines relative to all expressed genes (one-sided Fisher's Exact Test). Raw and Bonferroni-corrected p-values are given (p_{overlap} , $p_{\text{overlap}}^{\text{corrected}}$), where correction is for the total number of GO terms tested. **c**, NEUN western blot protein quantification. Neither genotype ($F_{1,90}=0.1852$; $P=0.6680$) nor time ($F_{3,90}=0.5382$; $P=0.6573$) had significant effects on NEUN expression (WT n=10, 9, 10, 9; KO n=14, 14, 16, 16 biological repeats from 2 independent experiments for each genotype, days 30, 40, 50 & 60 respectively). **d**, TBR1 western blot protein quantification. Neither genotype ($F_{1,95}=0.3899$; $P=0.5338$) nor time ($F_{3,35}=0.5052$; $P=0.6793$) had significant effects on TBR1 expression (WT n=10, 10, 9, 10 biological repeats, days 30, 40, 50 & 60 respectively; KO n=16 biological repeats at each timepoint; from 3 independent experiments for each genotype). **e**, CTIP2 western blot protein quantification. Genotype ($F_{1,86}=39.89$; $P=1.14 \times 10^{-8}$) and time ($F_{3,86}=5.262$; $P=0.0022$) had significant effects on CTIP2 expression (WT n=10, 7, 7, 10; KO n=17, 12, 13, 18 biological repeats; from 2 independent experiments for each genotype, days 30, 40, 50 & 60 respectively). **f**, ICC quantification of NEUN⁺ cells. Time ($F_{3,52}=7.018$, $P=0.0005$) had a significant effect on NEUN expression; genotype ($F_{1,52}=1.687$; $P=0.1998$) did not (WT n=6 biological repeats; KO n=9 biological repeats from 2 independent experiments at each timepoint for each genotype). **g**, ICC quantification of TBR1⁺ cells. Time ($F_{3,58}=4.738$, $P=0.0050$) had a significant effect on TBR1 expression; genotype ($F_{1,58}=1.664$; $P=0.2022$) did not (WT n=7, 6, 6, 7; KO n=11, 9, 9, 11 biological repeats from 3, 2, 2, 3 independent experiments for each genotype, days 30, 40, 50 & 60 respectively). **h**, ICC quantification of CTIP2⁺ cells. Genotype ($F_{1,67}=101.8$; $P=4.46 \times 10^{-15}$) and time ($F_{3,67}=18.93$; $P=5.33 \times 10^{-9}$) had significant effects on CTIP2 expression (WT n=10, 6, 6, 7; KO n=17, 9, 9, 11 biological repeats from 3, 2, 2, 3 independent experiments for each genotype, days 30, 40, 50 & 60 respectively). **i**, Representative ICC images of NEUN, TBR1 and CTIP2 with DAPI nuclear counterstain. Western blotting (**c-e**) and ICC data (**f-h**) were analysed by two-way ANOVA with post hoc comparisons using Bonferroni correction, comparing to WT controls. Stars indicate Bonferroni-corrected p-values, * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$ vs. WT control. Source data are provided as a Source Data file.

Figure 3. *DLG2*^{-/-} lines display deficits in neuron morphology & migration. a, The number of primary neurites (projecting from the soma). Neither genotype ($F_{1,126}=1.591$; $P=0.2095$) nor time ($F_{1,126}=2.278$; $P=0.1337$) had significant effects the numbers of primary neurites. **b**, The number of secondary neurites (projecting from primary neurites). Genotype ($F_{1,126}=18.78$, $P=2.97 \times 10^{-5}$) had a significant effect on number of secondary neurites, while time ($F_{1,126}=1.082$, $P=0.3003$) did not. **c**, The total neurite length. Both genotype ($F_{1,126}=4.568$; $P=0.0345$) and time ($F_{1,126}=26.33$; $P=1.06 \times 10^{-6}$) had significant effects on total neurite length. However, post hoc analysis showed no significant differences at individual timepoints. **d**, The soma area. Neither genotype ($F_{1,136}=$; $P=0.9170$) nor time ($F_{1,136}=1.399$; $P=0.2390$) had a significant effect on soma area. For **a-d**, WT n=32, 38 cells from 3 independent experiments, days 30 & 70 respectively; KO n=32, 28 cells from 3 independent experiments, days 30 & 70

respectively. **e**, Representative traces showing the neuronal morphology. **f**, The average speed of neuronal migration over 70 hours, from day 40 of cortical differentiation. *DLG2*^{-/-} neurons showed significantly decreased average migration speed compared to WT ($t_{52}=6.1175$; $P=1.26\times 10^{-7}$; $n=27$). For **f-g**, both WT & KO $n=27$ cells from 3 independent experiments. **g**, The displacement of neurons at 70 hours migration. *DLG2*^{-/-} neurons showed significantly decreased displacement compared to WT ($t_{52}=3.244$; $P=0.0021$; $n=27$). **h**, Representative traces of neuronal migration from a given origin over 70 hours. Morphology data sets (**a-d**) were analysed by two-way ANOVA with post hoc comparisons using Bonferroni correction, comparing to WT controls. Migration data sets (**f,g**) were analysed by unpaired two-tailed Student's t-test. Stars above bars represent, ** $P<0.01$; **** $P<0.0001$ vs. WT control (Bonferroni-corrected for morphology analyses). Source data are provided as a Source Data file.

Figure 4. *DLG2* regulates a cascade of transcriptional programs driving neurogenesis & differentiation. **a**, Enrichment for common schizophrenia risk variants in genes up- & down-regulated between each successive pair of timepoints in wild-type (WT), conditioning on all WT-expressed genes. Dotted line indicates $P_{\text{corrected}}=0.05$ following Bonferroni correction for 6 tests. **b**, Four discrete transcriptional programs initiated following the onset of neurogenesis were identified based upon WT differential expression between timepoints: *early-increasing*, genes significantly upregulated between days 20 and 30 ($20-30_{\text{up}}^{\text{WT}}$) and also days 30 and 60 ($30-60_{\text{up}}^{\text{WT}}$); *early-stable* genes, present in $20-30_{\text{up}}^{\text{WT}}$ and $20-60_{\text{up}}^{\text{WT}}$ but not $30-60_{\text{up}}^{\text{WT}}$; *early-transient* ($20-30_{\text{up}}^{\text{WT}}$ but not $20-60_{\text{up}}^{\text{WT}}$); and *late* ($30-60_{\text{up}}^{\text{WT}}$ but not $20-30_{\text{up}}^{\text{WT}}$). **c**, 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-stable*^{WT only}). One-sided tests were performed using MAGMA, conditioning on all expressed genes (*all*^{WT+KO}); raw and Bonferroni-corrected p-values are given (P^{GWAS} , $P^{\text{GWAS}}_{\text{corrected}}$), 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 *all*^{WT+KO}; both raw and Bonferroni-corrected p-values are given (P^{overlap} , $P^{\text{overlap}}_{\text{corrected}}$), 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 *all*^{WT+KO} and the program as a whole. Raw and Bonferroni-corrected p-values are given (P^{GWAS} , $P^{\text{GWAS}}_{\text{corrected}}$), where correction is for the 10 sets tested. **e**, SZ GWAS enrichment in GO terms with independent evidence of over-representation amongst *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 (P^{GWAS} , $P^{\text{GWAS}}_{\text{corrected}}$), where correction is for the 16 terms tested. Bold indicates tests surviving Bonferroni correction. Source data are provided as a Source Data file.

Figure 5. Electrophysiological properties of *DLG2*^{-/-} neurons. **a**, Resting membrane potential ($t_{27}=2.151$, $P=0.0406$) and **b**, input resistance ($t_{27}=0.3366$, $P=0.7390$) of day 50 WT and *DLG2*^{-/-} neurons ($n=15$ and 14 , respectively, unpaired two-tailed t test). **c**, Percentages of cells firing action potentials (APs) upon current step injection. **d**, Example traces of first overshooting AP and APs evoked by current step injection (-60pA to 120pA, increment 20pA, duration 1s). **e**, AP height ($t_{16}=3.661$, $P=0.0021$), **f**, AP half-width ($t_{16}=4.462$, $P=0.0004$), **g**, AP maximum

depolarising speed ($t_{16}=2.463$, $P=0.0255$), **h**, AP maximum repolarising speed ($t_{16}=3.728$, $P=0.0018$), **i**, spike threshold voltage ($t_{16}=0.004093$, $P=0.9968$) and **j**, rheobase current ($t_{16}=0.4061$, $P=0.6900$) of day 50 WT and *DLG2*^{-/-} neurons ($n=12$ and 6 , respectively, unpaired two-tailed t test) are shown. **k**, Example of day 50 neuron being whole-cell patch clamped with fluorescent dye injection. Spontaneous excitatory postsynaptic current (sEPSC) examples from day 50 (**l**) and 60 (**m**) neurons. Both the amplitude and frequency of day 50 and 60 neurons from WT and *DLG2*^{-/-} neurons were comparable by unpaired two-tailed t test (day 50, amplitude: $t_{68}=0.6974$, $P=0.4879$, $n=34$ and 36 for WT and KO; frequency: $t_{66}=0.5467$, $P=0.5865$, $n=33$ and 35 for WT and KO; day 60 amplitude: $t_{59}=1.021$, $P=0.3114$, $n=31$ and 30 for WT and KO; frequency: $t_{58}=0.7671$, $P=0.4464$, $n=30$ each for WT and KO). **n**, Percentage of cells displaying sEPSCs. **o**, Western blot analysis of DLG4 in synaptosomes of day 65 WT and *DLG2*^{-/-} neurons, displaying trend towards increased DLG4 expression in *DLG2*^{-/-} neurons ($t_2=2.157$, $P=0.1637$, $n=2$, unpaired two-tailed t test). * $p<0.05$; ** $p<0.01$; ***, $p<0.001$. All data presented as mean \pm SEM and are from two independent experiments. Source data are provided as a Source Data file.

Figure 6. Neuropsychiatric disorder risk variants localise to LoF intolerant genes in early neurogenic transcriptional programs. **a**, Identification of programs enriched/depleted for LoFi genes (P^{overlap}) relative to all expressed genes (two-sided Fisher's Exact Test). Raw and Bonferroni-corrected p-values given (P^{overlap} , $P^{\text{overlap}}_{\text{corrected}}$), where correction is for the 4 tests performed. **b**, LoFi genes were partitioned based on their overlap with early 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 schizophrenia (SZ) common variant enrichment (one-sided MAGMA test), conditioning on *all*^{WT+KO}. Bold indicates enrichments surviving Bonferroni correction for 7 tests. **c**, A two-sided Poisson rate ratio test was used to identify programs (partitioned by LoFi gene overlap) enriched for *de novo* LoF mutations from individuals with intellectual disability/severe neurodevelopmental delay (ID/NDD), autism spectrum disorder (ASD) and SZ when compared to all other expressed genes. Unaffected siblings (SIB) of ASD cases were analysed as a control. Data points lying above the x axis indicate an increased rate (rate ratio > 1), those below indicate a reduced rate (rate ratio < 1). Dotted lines indicate $P_{\text{corrected}} = 0.05$ following Bonferroni correction for 7 tests (4 LoFi + 3 non-LoFi sets tested for each disorder). **d**, Rate ratios (genes in set versus all other expressed genes) from tests shown in **c**. Dotted line shows rate ratio of 1 (i.e. rate of mutations in set equals that in all other genes). **e**, Programs (partitioned by LoFi gene overlap) were tested for enrichment (one-sided MAGMA test) in common variants contributing to ASD, attention-deficit/hyperactivity disorder (ADHD), bipolar disorder (BP), major depressive disorder (MDD) and Alzheimer's disease (AD). Values for SZ (Fig. 6b) included for comparison. Dotted lines indicate $P_{\text{corrected}} = 0.05$ following Bonferroni correction for 7 tests (4 LoFi + 3 non-LoFi sets tested for each disorder). **f**, Effect sizes (β coefficient from MAGMA gene-set enrichment test) for tests shown in **e**. Bold indicates tests surviving Bonferroni correction. Source data are provided as a Source Data file.

Figure 7. In vivo expression of early neurogenic programs in neurodevelopmental cell-types from human foetal cortex.

a, Identification of programs enriched/depleted for genes with high expression in CA1 pyramidal neurons relative to other brain cell-types (pyramidal^{high}); enrichment compared to all expressed genes (two-sided Fisher's Exact Test). Raw and Bonferroni-corrected p-values

given (p_{overlap} , $p_{\text{overlap}_{\text{corrected}}}$), correction is for the 4 tests performed. **b**, Pyramidal^{high} 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 pyramidal^{high} genes overlapping the late neurogenic program when compared to all expressed genes; both raw and Bonferroni-corrected p-values are given (p_{overlap} , $p_{\text{overlap}_{\text{corrected}}}$), where correction is for the 7 regulator sets tested. **d**, Cells corresponding to distinct neurodevelopmental cell-types (radial glia, RG; Intermediate progenitor cell, IPC), including cell-types at different stages of maturity, were identified and extracted from a previously published single-cell RNA-seq study of human foetal cortex across peak stages of neurogenesis³² (Methods). >80% of genes belonging to each *in vitro*-defined early neurogenic program and 55% of the late-pyramidal^{high} set (top RHS) were present in the foetal data (early-transient^{-/-} n=665 genes; early-stable^{-/-} n=1484 genes; early-increasing^{-/-} n=431 genes; late-pyramidal^{high} n=101 genes). For each program, the mean and standard error of gene-level expression averages (see main text/Methods) were calculated for each foetal cell-type/developmental stage. The foetal 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.

Figure 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 from⁸⁴ and rare mutations causing Mendelian neurodevelopmental syndromes are denoted. **b**, Neurodevelopmental continuum/gradient model. Disorders shown are: intellectual disability/severe neurodevelopmental delay (ID/NDD); autism spectrum disorders (ASD); attention-deficit/hyperactivity disorder (ADHD); schizophrenia (SZ); bipolar disorder (BP); and major depressive disorder (MDD).

Figure 1

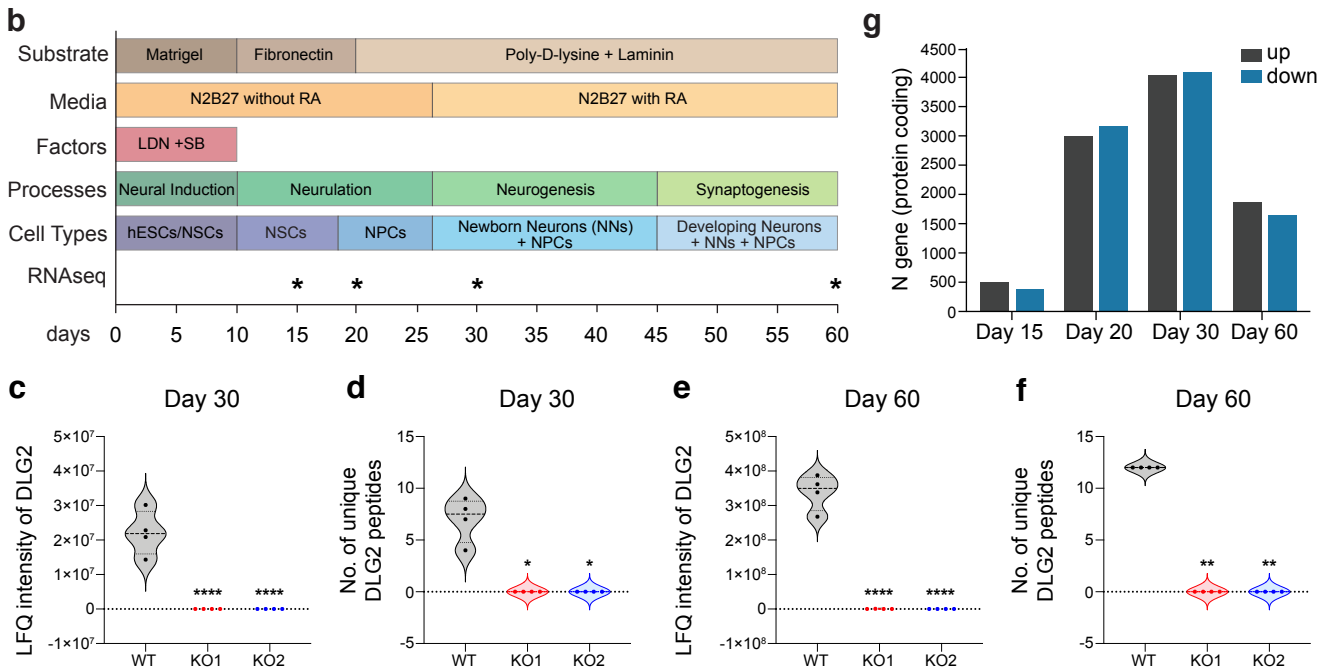
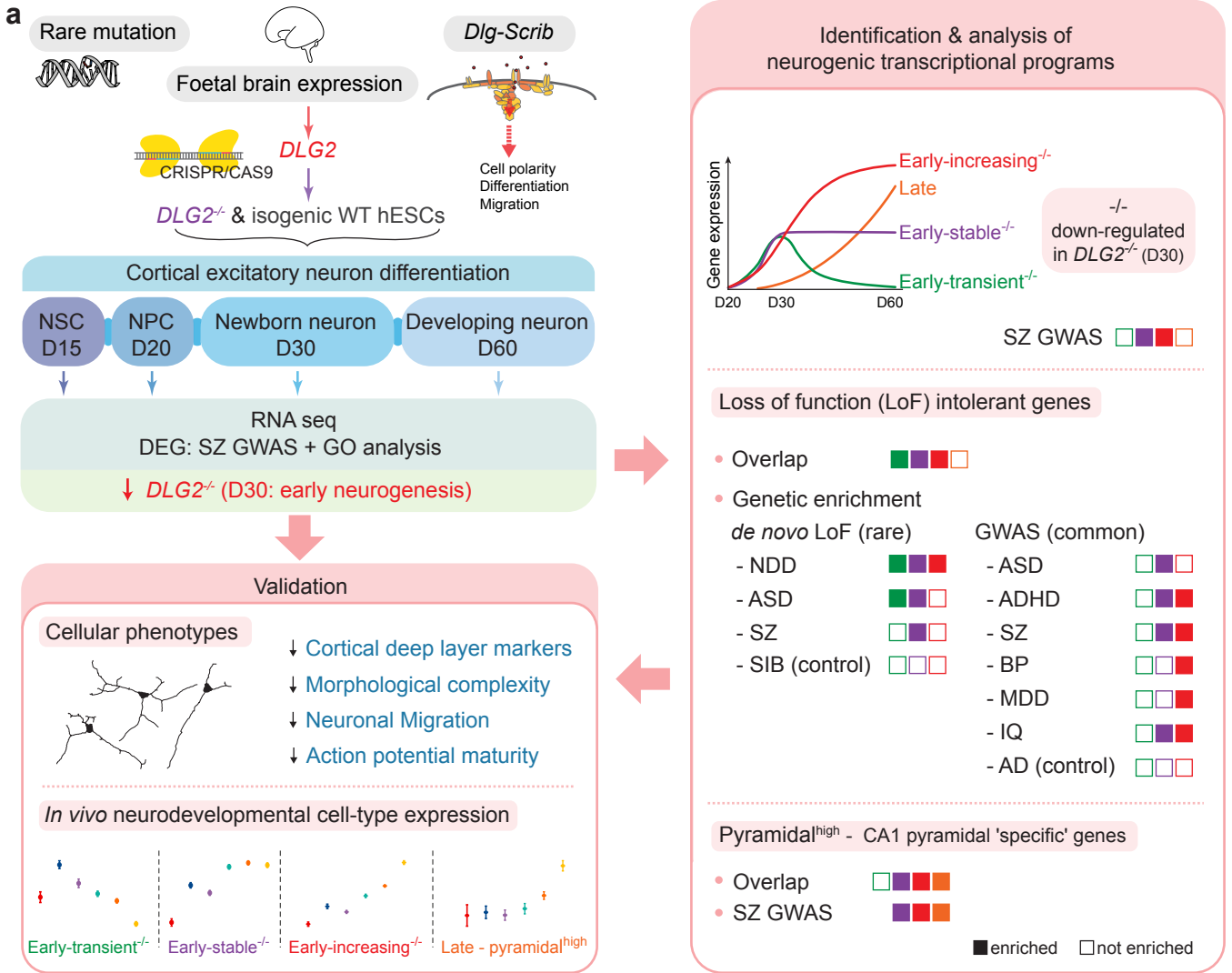


Figure 2

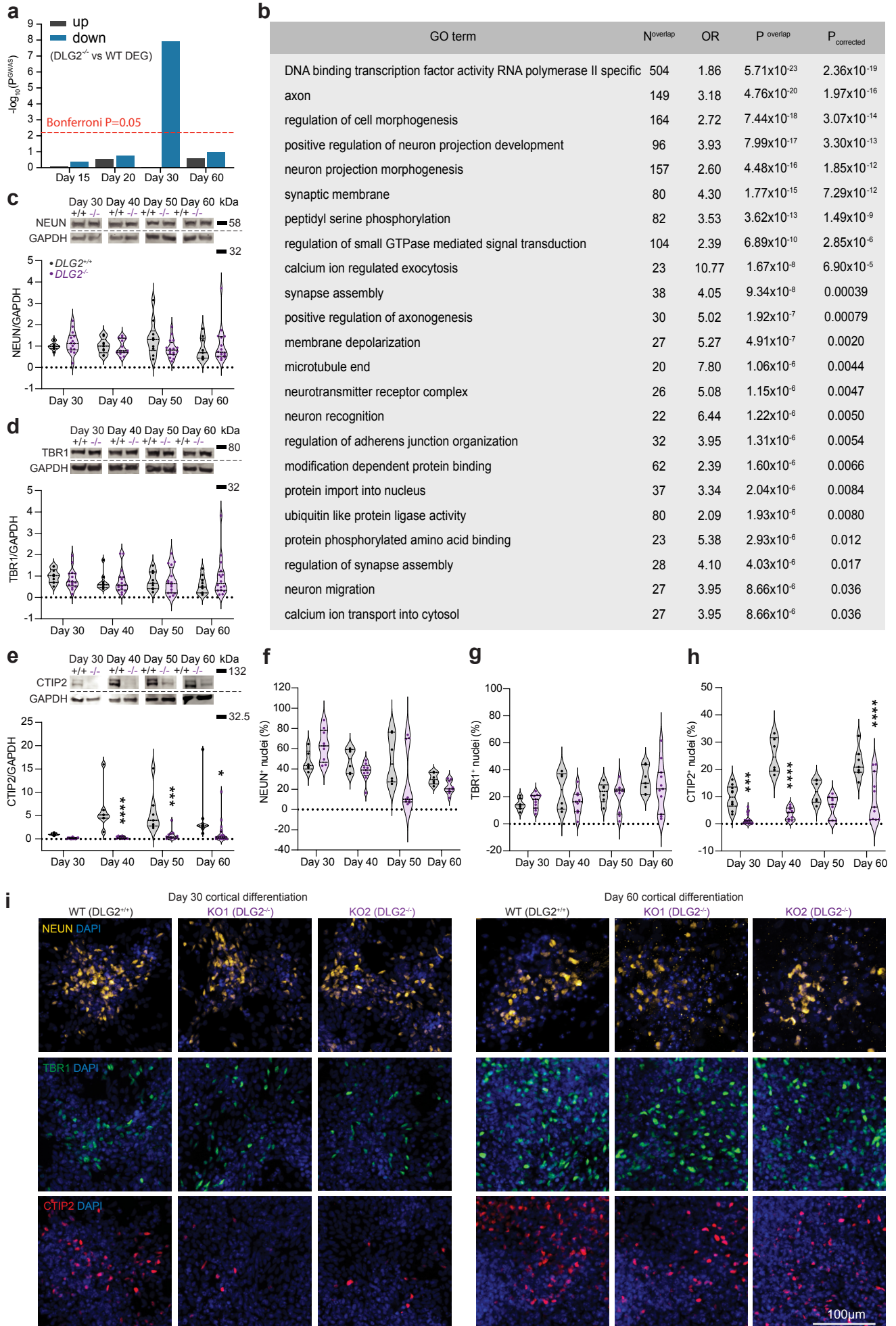


Figure 3

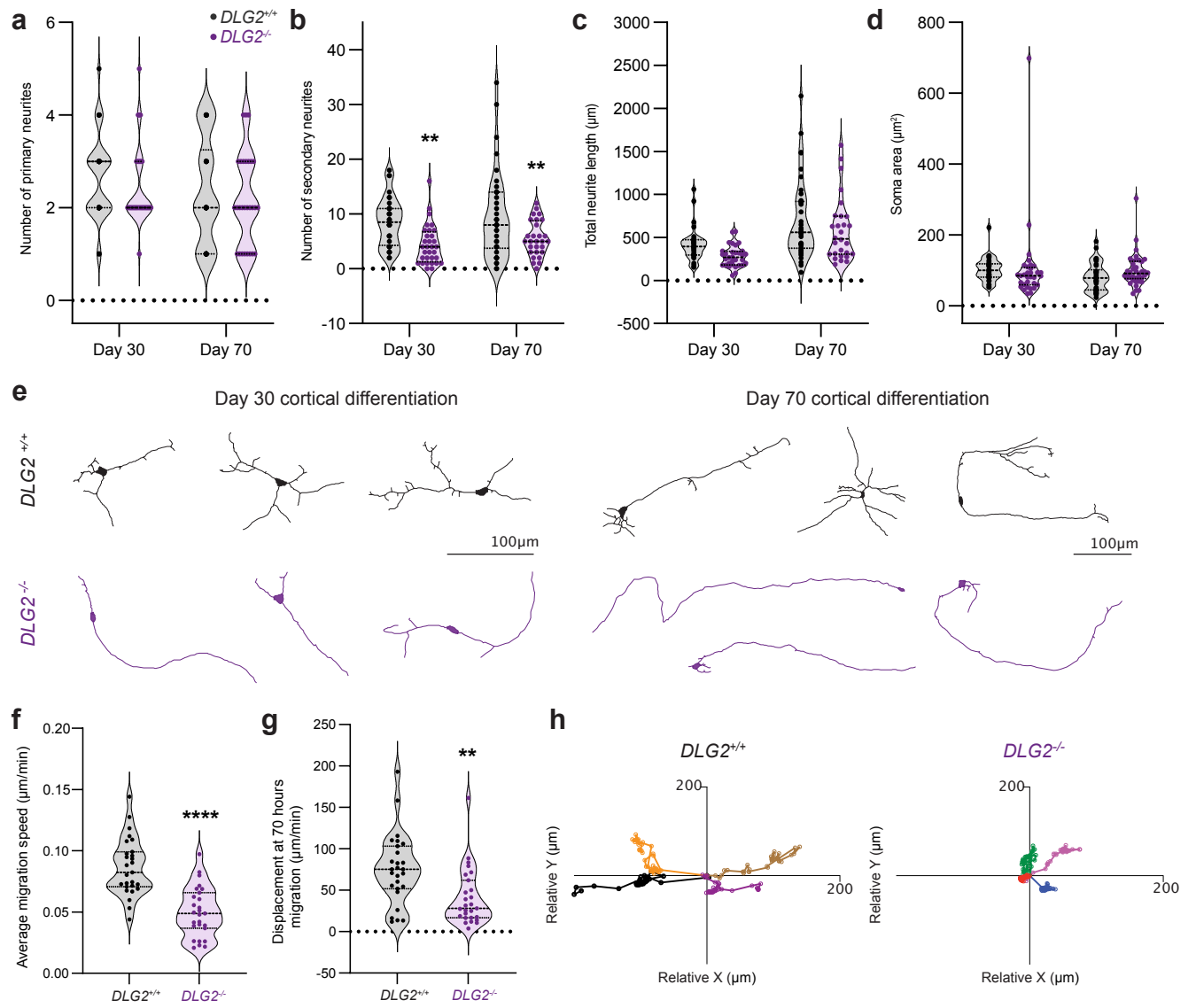


Figure 4

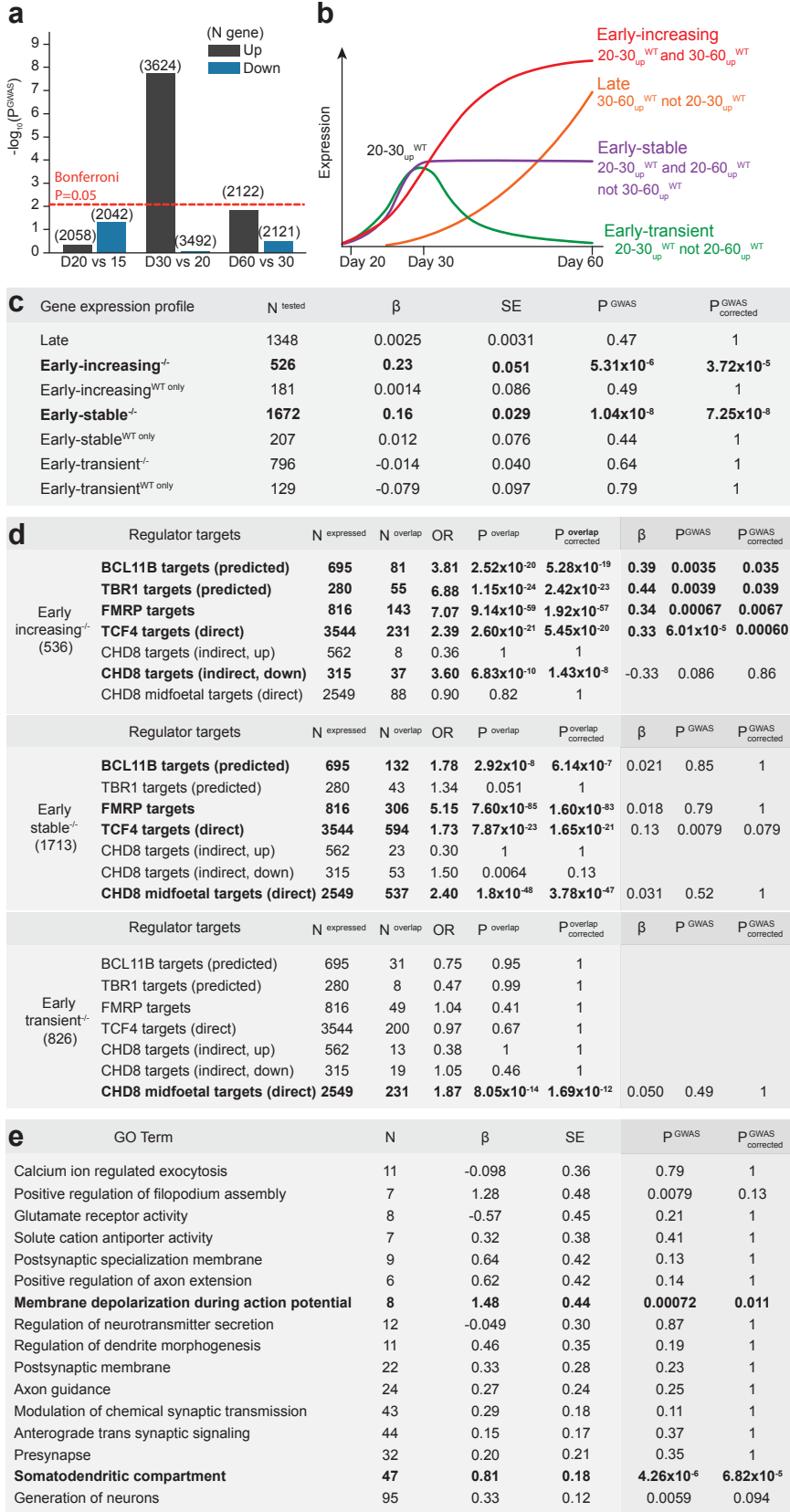


Figure 5

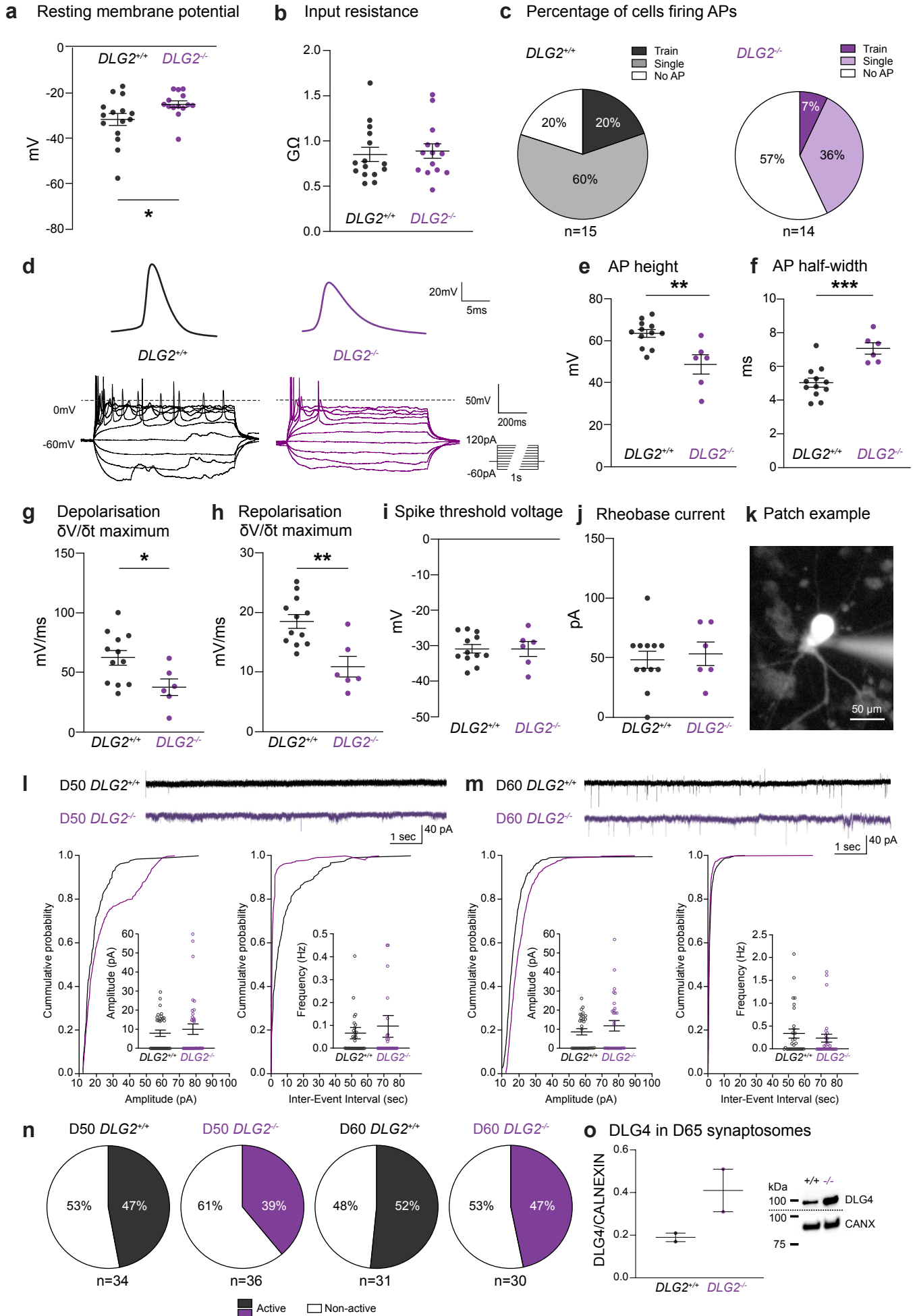
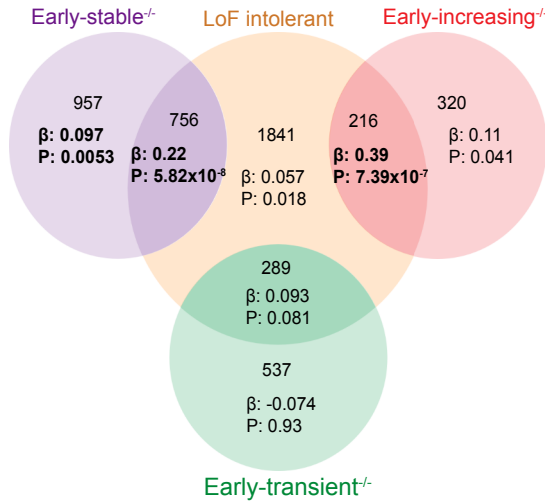


Figure 6

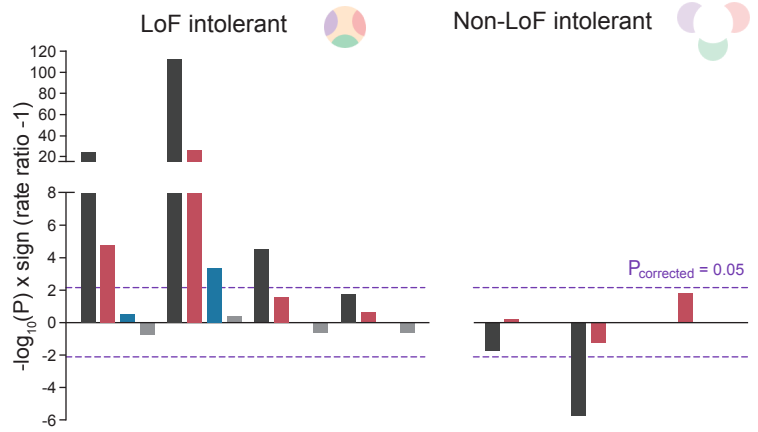
a LoF intolerant gene overlap

	N overlap	OR	P overlap	P overlap corrected
Late (1399)	221	0.68	2.96×10^{-7}	1.19×10^{-6}
Early-increasing ^{+/−} (536)	216	2.66	4.14×10^{-25}	1.66×10^{-24}
Early-stable ^{+/−} (1713)	756	3.64	4.27×10^{-119}	1.71×10^{-118}
Early-transient ^{+/−} (826)	289	2.14	9.42×10^{-22}	3.77×10^{-21}

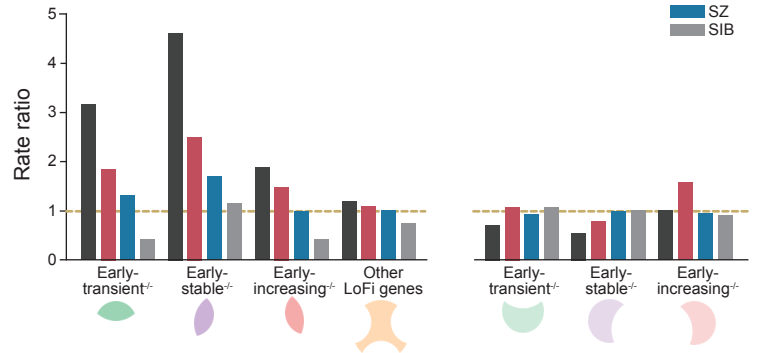
b Schizophrenia GWAS



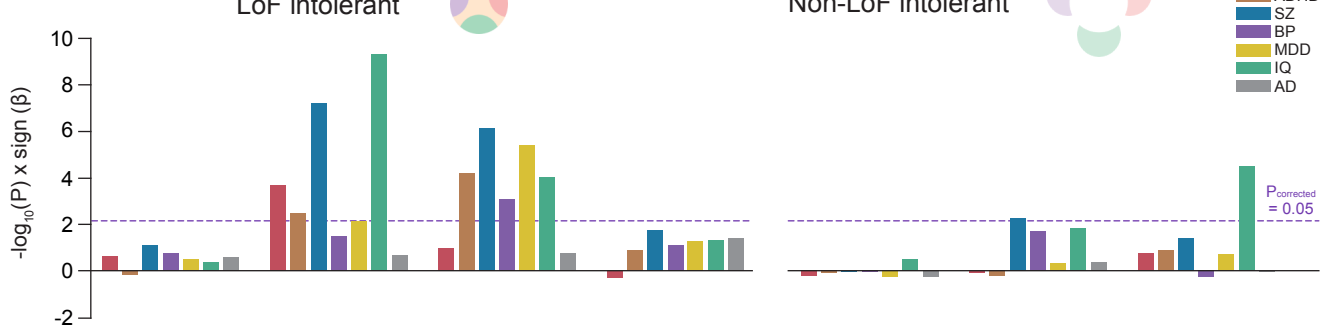
c De novo LoF mutation (P)



d De novo LoF mutation (Effect size)



e GWAS (P)



f GWAS (Effect size)

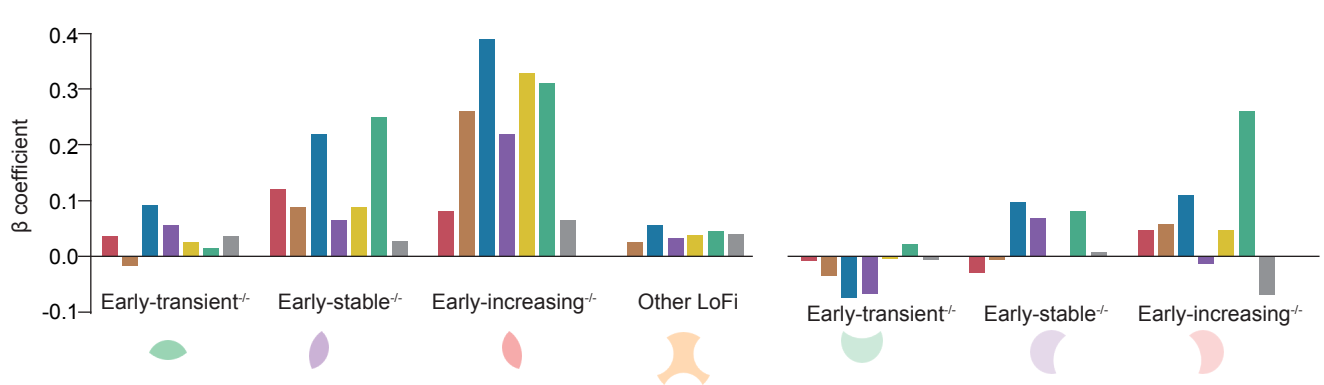
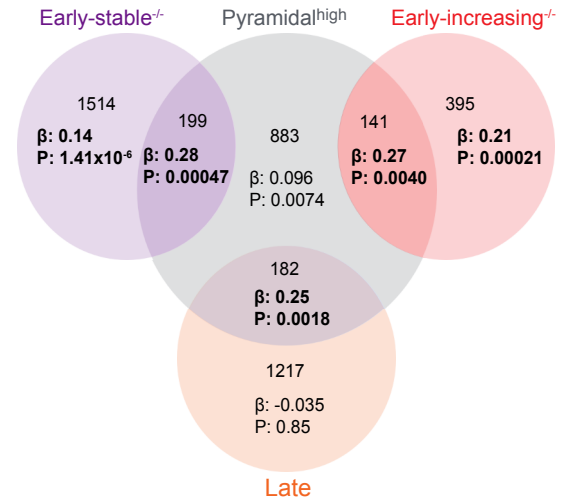


Figure 7

a Pyramidal^{high} gene overlap

	N overlap	OR	P overlap	P overlap corrected
Late (1399)	182	1.71	3.05x10⁻⁹	1.22x10⁻⁸
Early-increasing ^{-/-} (536)	141	4.19	1.01x10 ⁻³⁵	4.02x10 ⁻³⁵
Early-stable ^{-/-} (1713)	199	1.49	3.05x10 ⁻⁶	1.22x10 ⁻⁵
Early-transient ^{-/-} (826)	58	0.80	0.12	0.49

b Schizophrenia GWAS



c Late - pyramidal^{high} (182 genes)

Regulator targets	N expressed	N overlap	OR	P overlap	P overlap corrected
BCL11B targets (predicted)	695	26	3.35	7.45x10 ⁻⁷	5.22x10 ⁻⁶
TBR1 targets (predicted)	280	20	6.58	4.48x10 ⁻¹⁰	3.14x10 ⁻⁹
FMRP targets	816	32	3.63	1.06x10 ⁻⁸	7.42x10 ⁻⁸
TCF4 targets (direct)	3544	70	1.91	2.60x10 ⁻²¹	5.45x10 ⁻²⁰
CHD8 targets (indirect, up)	562	4	0.55	0.93	1
CHD8 targets (indirect, down)	315	10	2.63	0.0071	0.05
CHD8 midfoetal targets (direct)	2549	21	0.60	0.99	1

d Neurogenic programs: expression in human foetal cortex

- RG (early)
- RG
- IPC
- Transitioning
- Newborn neuron
- Developing neuron

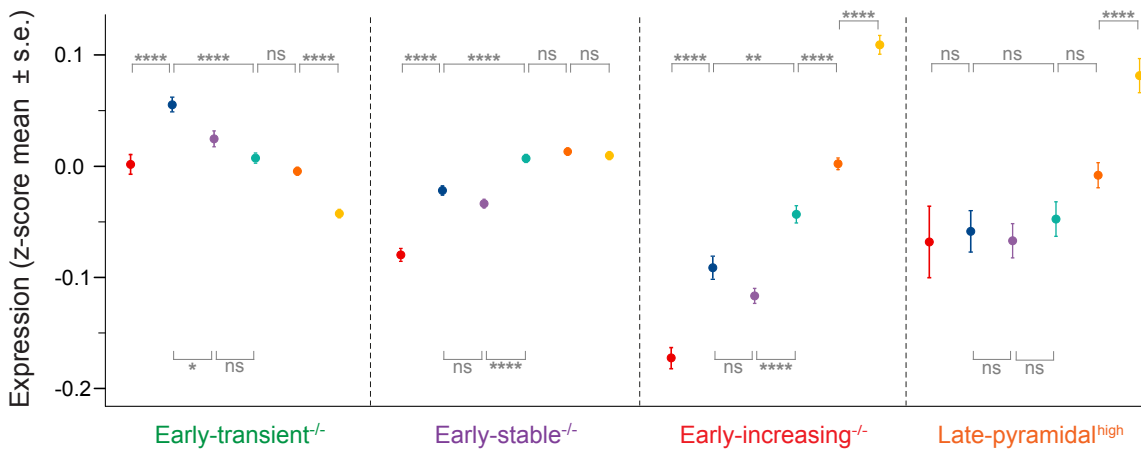
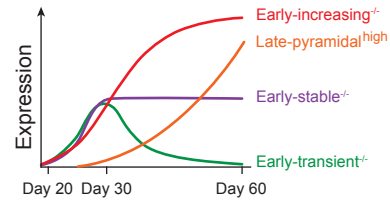
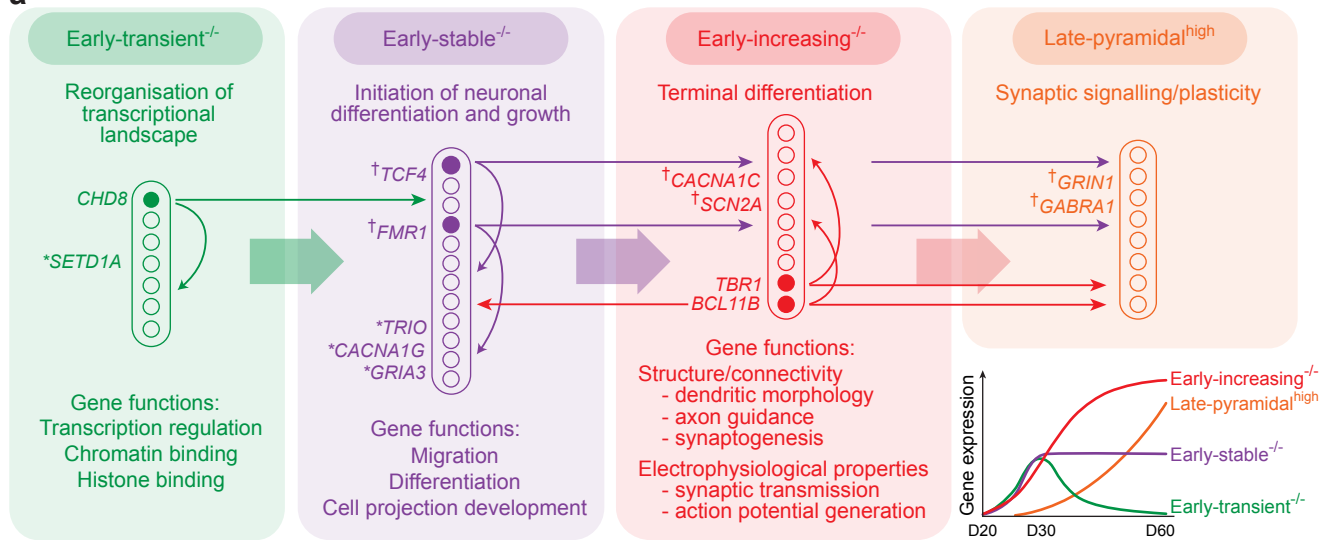
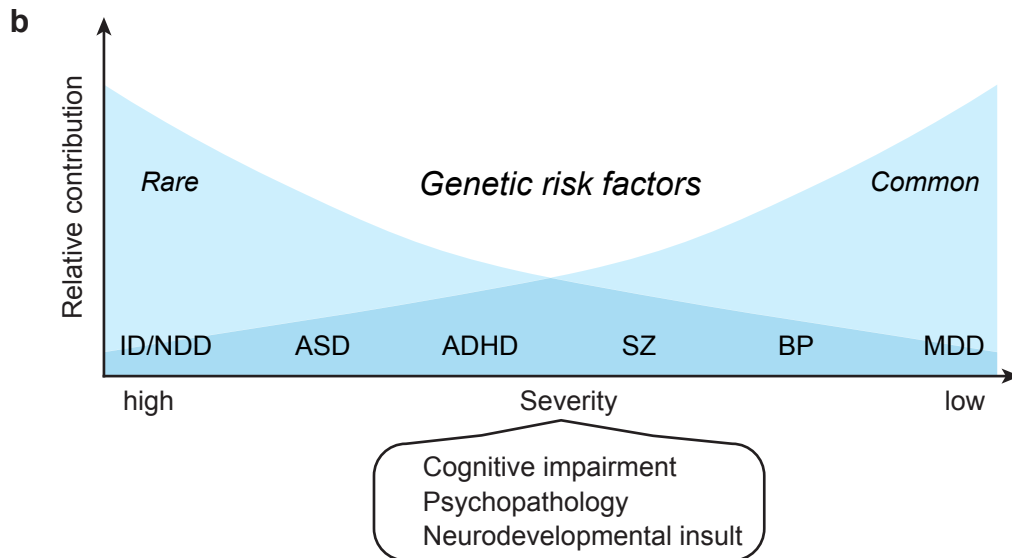


Figure 8

a



- Key regulators (arrows denote programs over-represented for known/predicted targets)
- * Genome-wide significant ($P < 2.2 \times 10^{-6}$) in recent SCHEMA rare variant analyses
- † Rare variation causes Mendelian neurodevelopmental syndromes



Supplementary Information

Supplementary Figure 1

Supplementary Figure 2

Supplementary Figure 3

Supplementary Figure 4

Supplementary Figure 5

Supplementary Figure 6

Supplementary Figure 7

Supplementary Figure 8

Supplementary Figure 9

Supplementary Data 1

Supplementary Data 2

Supplementary Data 3

Supplementary Data 4

Supplementary Data 5

Supplementary Data 6

Supplementary Data 7

Supplementary Data 8

Supplementary Data 9

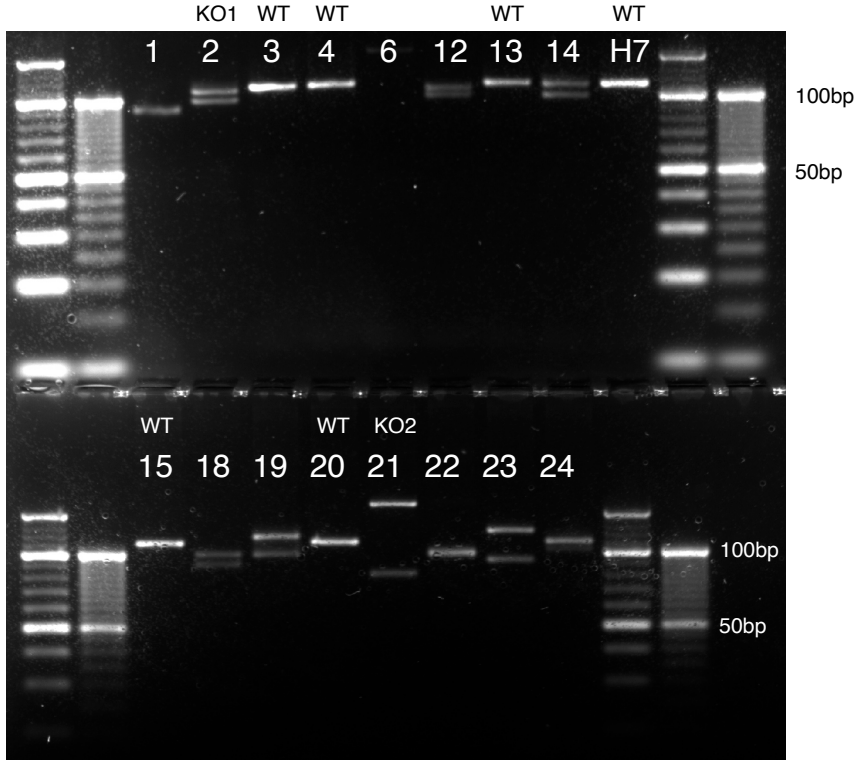
Supplementary Data 10

Supplementary Software

Supplementary Software file contains scripts for RNA sequencing data analysis and downstream bioinformatic/human genetic analyses, plus files containing annotated gene sets. Further details can be found in enclosed README files. For access to human genetic data please see Data availability statement.

* Supplementary Data (excel file) and Supplementary Software (zip file) are attached separately.

- a** 1st gRNA sequence - 5' GGCTTCCACCGCTTTACTGT 3'
 2nd gRNA sequence - 5' AGGAAGCAGGGTCTATCGTT 3'
- b** Forward primer: 5' ATCTTGCGGGTGAATGAGGTTG
 Reverse primer: 5' GTCGTCTTCTACGCACATACAG



c

Cut by Cas9 nickase

WT 5' GAGGTTGATGTGTCAGAGGTTTCCACAGTAAAGCGGTGGAAGCCCTGAAGGAAGCAGGGTCTATCGTTCGG
 CTC CAACTACACAGTCTCCAAAGGGTGTCAATTCGCCACCT TCGGGACT TCCT TCGTCCCAGATAGCAAGCC5'

Cut by Cas9 nickase

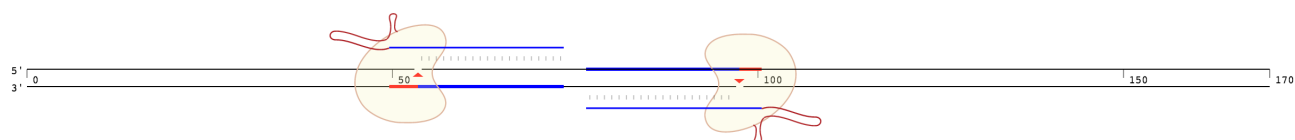
KO1 allele 1 5' GAGGTTGATGTGTCAGAGGTTTCCACAGTAAAGCGGTGGAAGCCCTGAAGGAAGCAGGGTCTATCGTTCGG
 KO1 allele 2 5' GAGGTTGATGTGTCAGAGGTTTCCACAGTAAAGCGGTGGAAGCCCTGAAGGAAGCAGGGTCTATCGTTCGG
 KO2 allele 1 5' GAGGTTGATGTGTTGGAAACCTCTGAACAGATTATCTGAGGTTCTCCTTGAACCTTTGAAATCTGAACTTCAG
 TATTATAATATTATAAGATGAATCGAGGTGGAGCAGAGGTTTCCACAGTAAAGCGGTGGAAGCCCTGAAGGA
 AGCAGGGTCTATCGTTCGG
 KO2 allele 2 5' GAGGTTGATGTGTCAGAGGTTTCCACAGTAAAGCGGTGGAAGCCCTGAAGGAAGCAGGGTCTATCGTTCGG

d

WT VDVSEV SHSKAVEALKEAGSIVRLYVRRRRPILETVVEIKLFGPKGLG
 KO1 allele 1 VDVSEV SHSKAKP*
 KO1 allele 2 VDVSEVSHSKARKQGLSFGCMVEDDLFWRPLWKSNC SKALKV*
 KO2 allele 1 VDVWETSEQDYLRFSLNFNLSIYNIIR*
 KO2 allele 2 VDVSEVSHRVYRSVCA*

Supplementary Figure 1. Generation of *DLG2*^{-/-} human embryonic stem cell lines

a, Two guide RNA sequences used to target exon 22 of *DLG2* which is translated into part of the 1st PDZ domain. **b**, A primer set used for screening successful indels on targeted DNA. PCR and gel electrophoresis shows several hESC clones with indels which were confirmed in three independent experiments. Parental H7 lines shows a single 109bp band whereas several clones, such as 2 and 21, show two different sized bands suggesting generation of indels on the targeted exon. **c**, PCR amplicons were sequenced and clones no.2 and 21 found to be homozygous knockouts (named KO1 and KO2). Red letters indicate deleted bases while blue indicate inserted DNA. **d**, Indels in KOs generated premature stop codons on the targeted exon. Green indicates different amino acid sequences from WT.



Overhang region

```

GGTCAATGATTGTATCTTCCGGG
TGAATGAGGTTGATGTGCAGAG
GTTTCCACAGTAAAGCGGTGGA
AGCCCTGAAGGAAGCAGGGTCTA
TCGTTCCGCTGTATGTGCGTAGA
AGACGACCTATTTGGAGACCGT
TGTGGAAATCAAACGTTCAAAG
GCCCTAAAG

```

Guide A

81

quality
high

cuts after position
53 in query

sequence
GGCTTCCACCGCTTACTGTGGG

offtargets
89

genic offtargets
18

[Export off-targets to .csv](#)

Guide B

90

quality
high

cuts after position
97 in query

sequence
AGGAAGCAGGTCATCGTTCGG

offtargets
125

genic offtargets
14

[Export off-targets to .csv](#)

Pair score, A & B

74

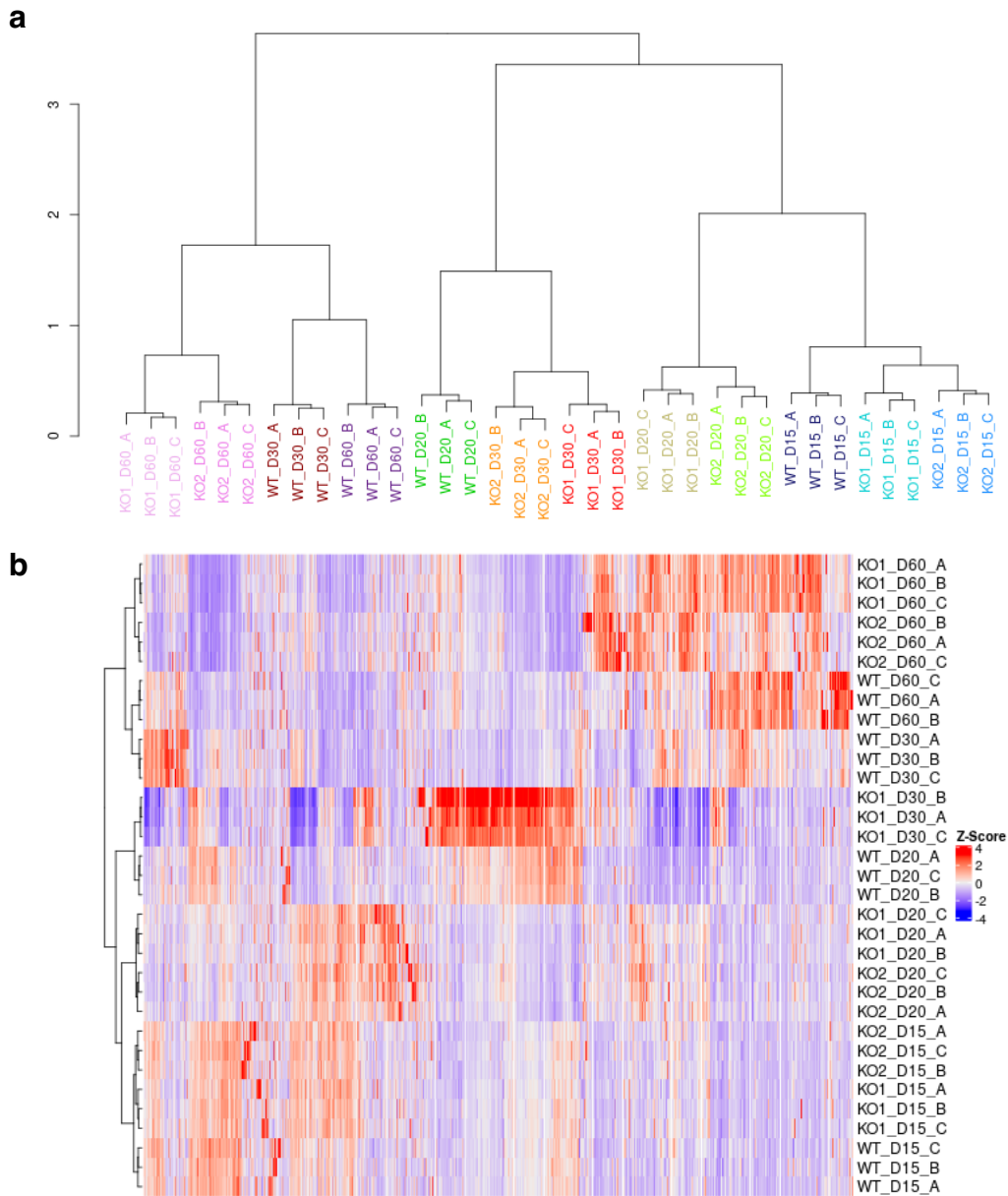
quality
high

offtarget pairs of A & B
0

genic OT pairs of A & B
0

Supplementary Figure 2. Quality of two gRNAs used in the study

Separately the gRNAs had predicted off-targets of 89 and 125 loci each; however, the predicted off-targets when used as a pair was none. Captured from crispr.mit.edu.

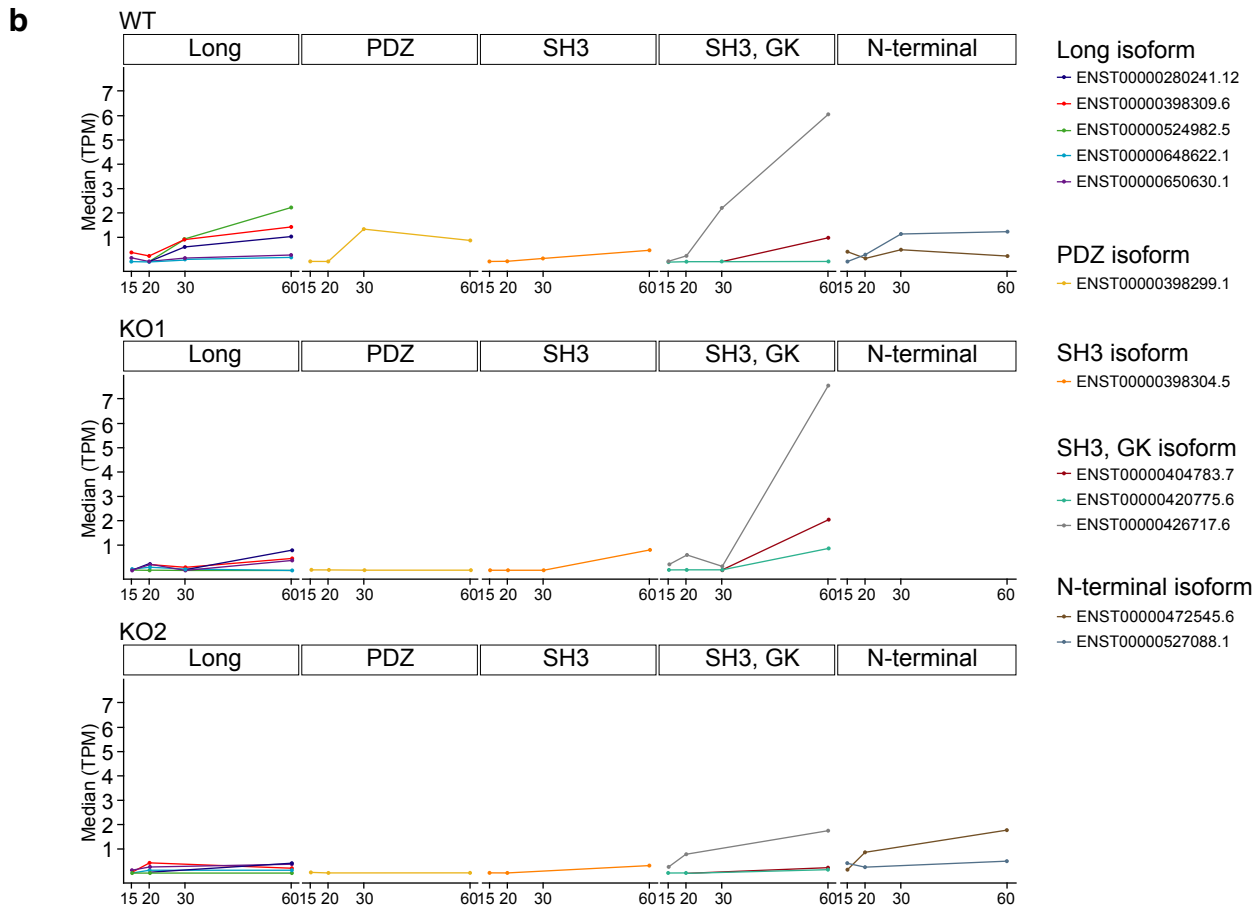


Supplementary Figure 3. Clustering and visualisation of KO and WT sample gene expression. **a**, Hierarchical clustering of KO and WT samples based on relative expression of protein-coding genes (20,004 genes). Clustering was performed on a Pearson's correlation matrix of the z-scored gene expression, using Ward's linkage criterion (R hclust option 'ward.D2') and 1-r as the distance function, where r is the pairwise Pearson's correlation coefficient. **b**, Heatmap showing transcriptomic profiles of KO and WT samples (low-quality KO2 samples from day 30 removed) based on the z-scored expression levels of variably expressed protein-coding genes (coefficient of variation > 25%; 15,003 genes). Heatmap clustering was performed as in **a**, using Pearson's correlation matrix and Ward's linkage criterion. Clustering and heatmap both illustrate the similarity between KO lines and their delayed development relative to WT.

a

Day 30	Start	End	Domain	Exon ID	KO	WT	log ₂ fold-KO-WT	P	P _{corrected}
	84059484	84059435	PDZ1	ENSE00003484205	0.023	4.60	-15.29	0.0018	0.0016
	84098994	84098923	PDZ1	ENSE00003586389	0.023	4.54	-15.26	0.0020	0.0019

Day 60	Start	End	Domain	Exon ID	KO	WT	log ₂ fold-KO-WT	P	P _{corrected}
	83965468	83965324	PDZ2	ENSE00003483868	7.64	13.05	-1.56	6.38x10 ⁻¹⁷	1.03x10 ⁻¹⁴
	83980642	83980561	PDZ3	ENSE00003595386	6.17	10.62	-1.57	1.02x10 ⁻¹²	9.56x10 ⁻¹¹
	84059434	84059315	PDZ2	ENSE00003484205	4.89	10.47	-2.21	2.66x10 ⁻¹²	2.34x10 ⁻¹⁰
	83980560	83980506	PDZ1	ENSE00003595386	6.28	10.57	-1.51	9.00x10 ⁻¹²	7.26x10 ⁻¹⁰
	83963023	83962963	PDZ2~3	ENSE00003607494	6.43	11.10	-1.58	6.63x10 ⁻¹⁰	3.96x10 ⁻⁸
	84059484	84059435	PDZ1	ENSE00003484205	4.63	8.79	-1.85	1.42x10 ⁻⁹	8.05x10 ⁻⁸
	84098994	84098923	PDZ1	ENSE00003586389	5.15	8.81	-1.56	3.47x10 ⁻⁷	1.23x10 ⁻⁵
	83786792	83786690	PDZ3	ENSE00003542004	6.61	8.83	-0.84	0.00041	0.0062
	83962962	83962885	PDZ2~3	ENSE00003607494	6.51	11.20	-1.57	0.00048	0.0071
	84163511	84163461	PDZ1	ENSE00003616457	4.41	8.05	-1.74	0.0012	0.015
	83833770	83833614	PDZ3	ENSE00003544320	6.48	8.32	-0.73	0.0022	0.024

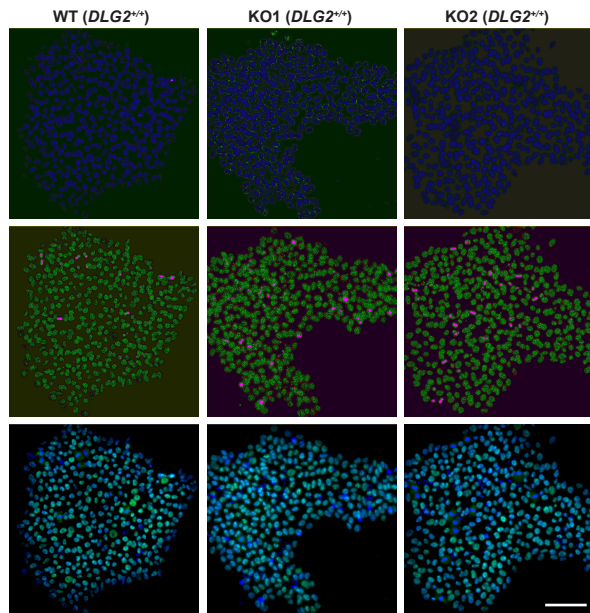
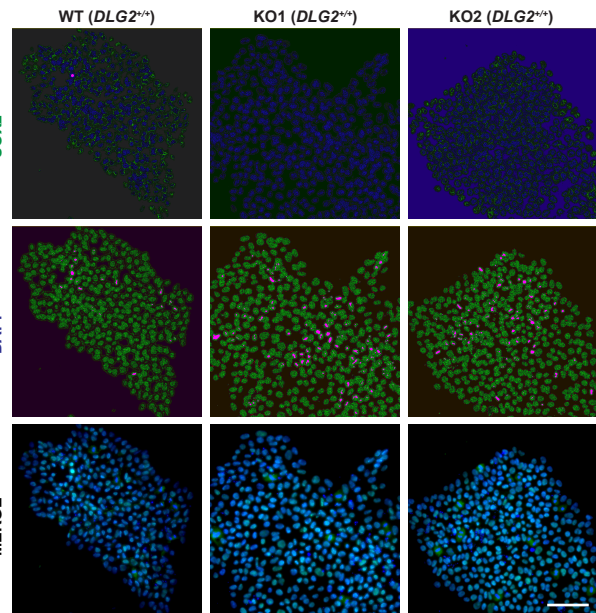
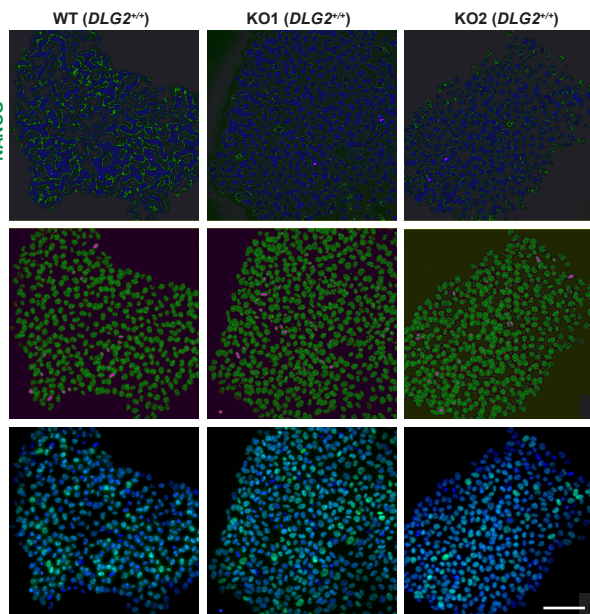
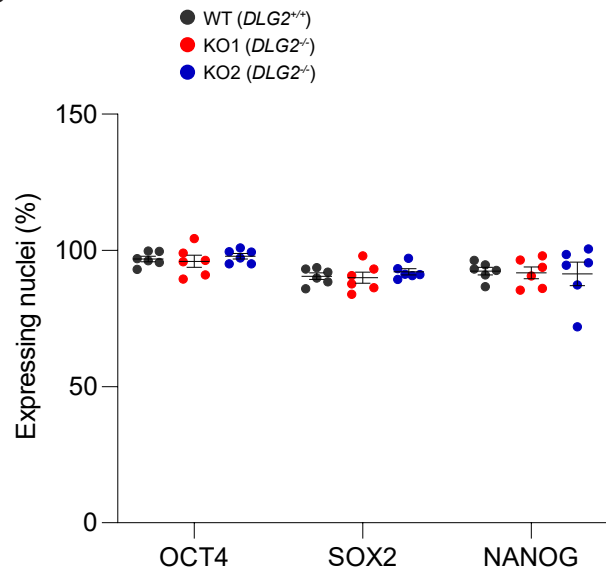


Supplementary Figure 4. Validation of *DLG2*^{-/-} human embryonic stem cells

a, Differential exon usage (days 30 and 60) for exons encoding *DLG2* PDZ domains. Analysis was performed using DEXseq. Shown are all exonic regions significant at $P_{\text{corrected}} < 0.05$ following FRD correction for multiple testing. **b**, *DLG2* predicted transcript expression in WT, KO1 and KO2 lines at each time point. Transcript-level counts in TPM (Transcripts Per Million) were imported from RSEM using tximport (v1.12.3) (Soneson et al., 2015). Shown are all protein-coding transcripts consistently identified (TPM > 0 in all replicates) in at least one time point. Transcripts were classified according to the functional domains they contain and the median TPM plotted for each time point. Long transcripts contain all 3PDZs, SH3 and GK domains.

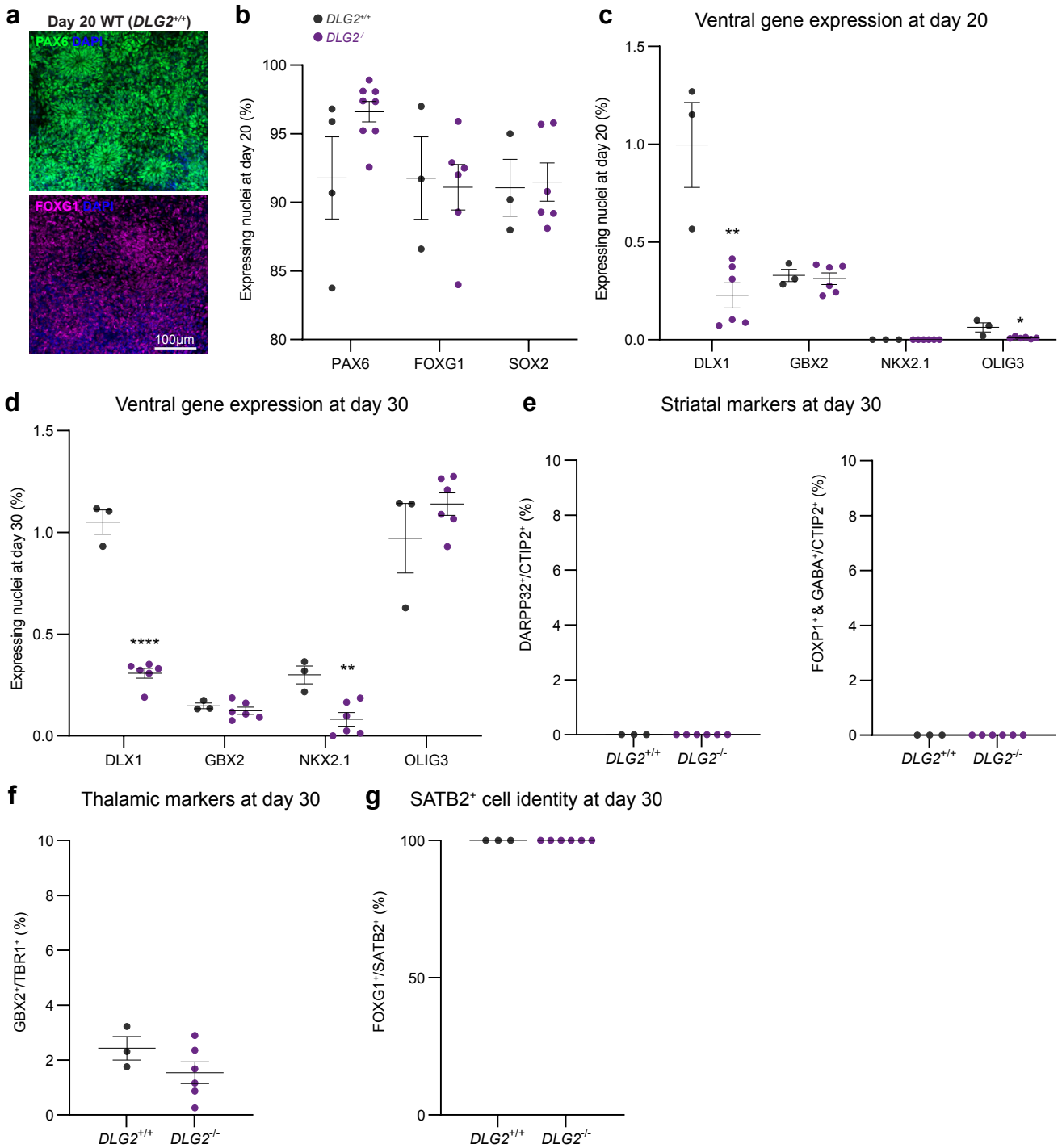
a

Cell line	CNV location	Duplication/Deletions	Position	Length
WT, KO1, KO2	2q37.3	Deletion	chr2:242900000-243100000	200Kb
WT, KO1, KO2	3q26.1	Duplication	chr3:166000000-1664000000	400Kb
WT, KO1, KO2	20p12-p11.1	Deletion	chr20:91000-25889300	25.8Mb
WT, KO1, KO2	20q11.21-q13.33	Duplication	chr20:26088528-62707975	36.6Mb

b**c****d****e**

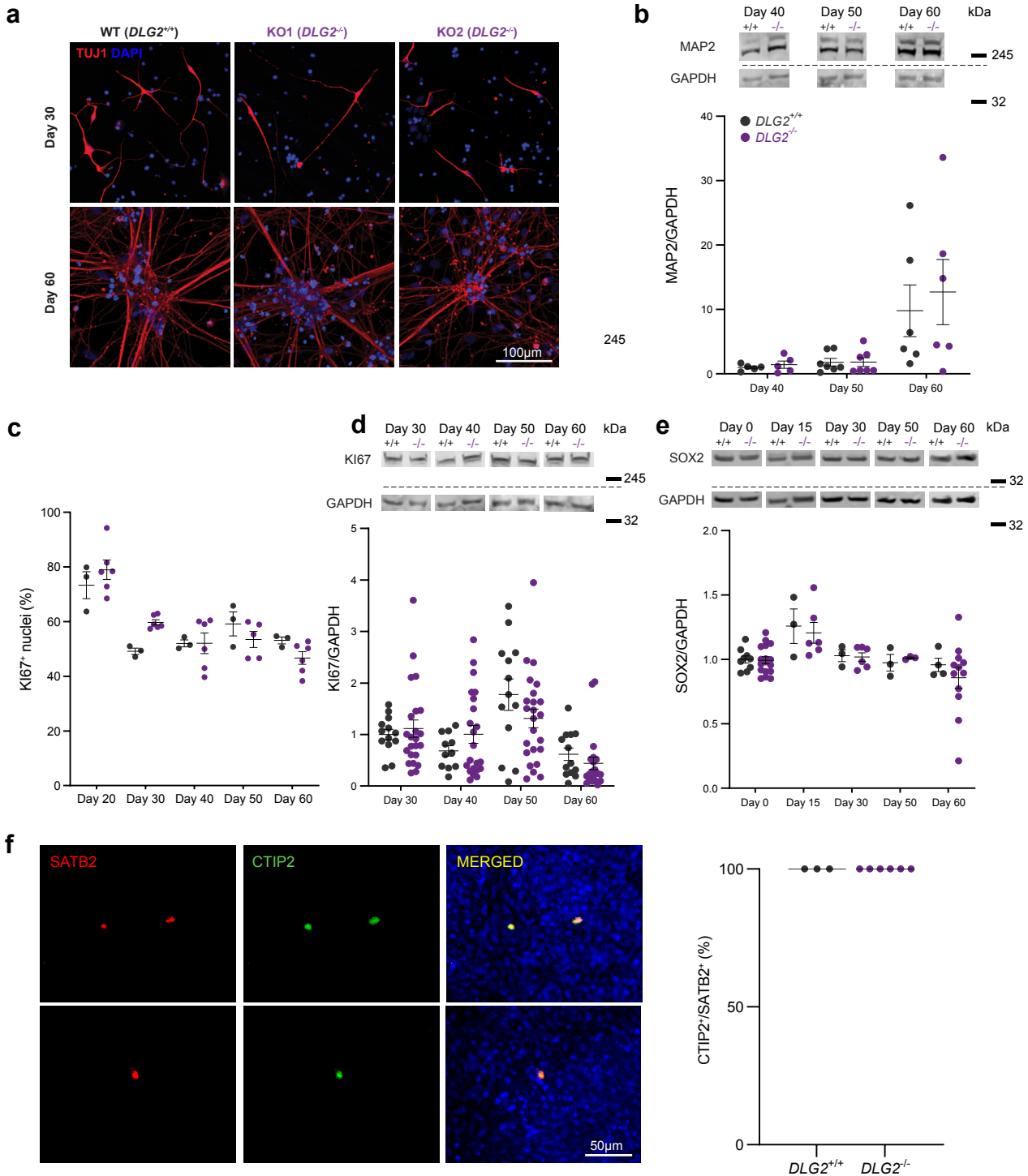
Supplementary Figure 5. Quality control of $DLG2^{-/-}$ human embryonic stem cells

a, Copy number variant analysis of $DLG2^{-/-}$ hESCs. WT, KO1 and KO2 genomic DNAs were analysed using Illumina PsychArray v1.1 and data analysed using PennCNV. CNVs smaller than 100Kb and containing less than 10 SNPs were filtered out in the PennCNV QC. No additional CNVs are identified in both KO lines in comparison to WT line. **b-e**, Pluripotency marker expression of $DLG2^{-/-}$ hESCs. Key pluripotency transcription factors were expressed both in 2 $DLG2^{-/-}$ hESC lines and WT controls. Representative ICC images of OCT4 (**b**), SOX2 (**c**) and NANOG (**d**) expression with DAPI nuclear counterstain for 2 $DLG2^{-/-}$ hESC lines and WT controls. **e**, ICC quantification of nuclei expressing either OCT4, SOX2 or NANOG in 2 $DLG2^{-/-}$ hESC lines and WT controls, the genotype had no significant effect on the expression of these markers (OCT4: $F_{2,15}=0.3780$, $P=0.9616$, $n=6$; SOX2: $F_{2,15}=0.5383$, $P=0.5946$, $n=6$; NANOG: $F_{2,15}=0.03433$, $P=0.9663$, $n=6$). Analysis was by one-way ANOVA for each pluripotency marker. All data presented as mean \pm SEM and all scale bars are 100 μ m.



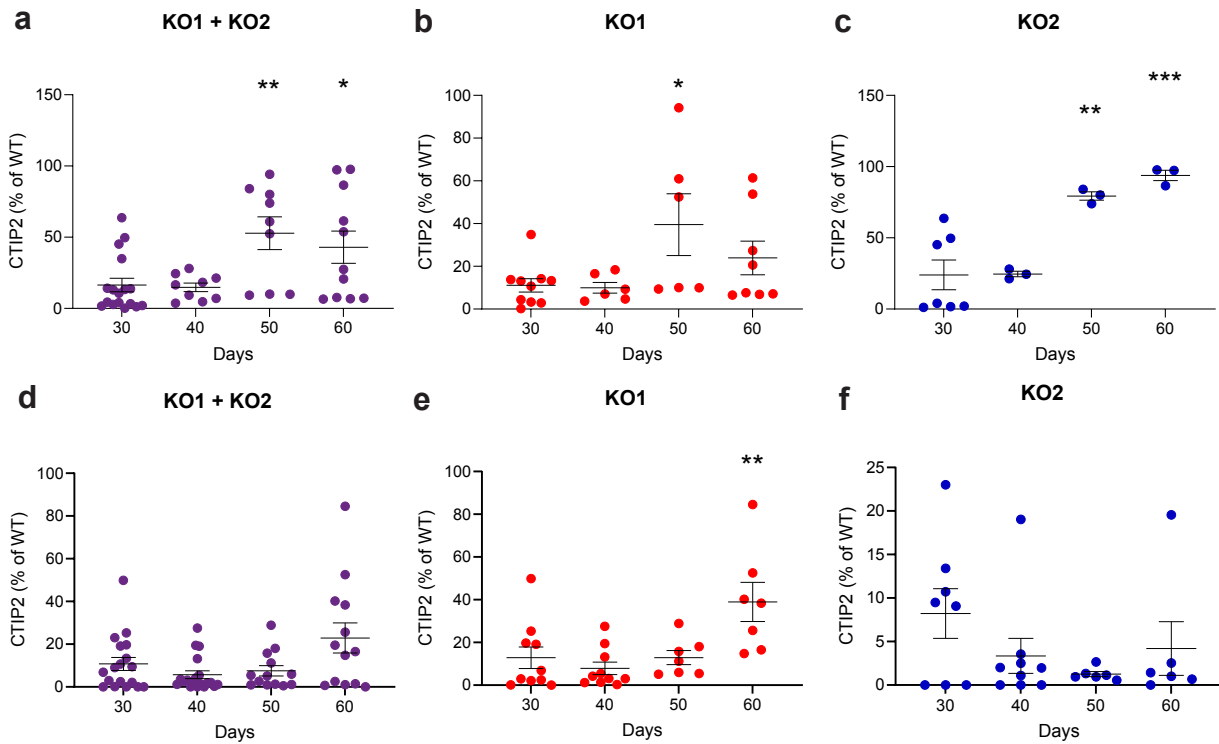
Supplementary Figure 6. Cell fate characterisation of WT and *DLG2*^{-/-} cells during cortical differentiation

a, Representative ICC images showing expression of forebrain progenitor markers PAX6 and FOXG1 at day 20 of cortical differentiation. **b**, ICC quantification of nuclei expressing either PAX6, FOXG1 or SOX2 in *DLG2*^{-/-} and WT cells at day 20 of cortical differentiation, the genotype had no significant effect on the expression of these markers (PAX6: $t_{10}=2.113$, $P=0.0607$, $n=4, 8$ for WT and KO; FOXG1: $t_7=0.2133$, $P=0.8372$, $n=3, 6$ for WT and KO; SOX2: $t_7=0.1702$, $P=0.8697$, $n=3, 6$ for WT and KO). **c**, ICC quantification of nuclei expressing either DLX1, GBX2, NKX2.1 or OLIG3 in *DLG2*^{-/-} and WT cells at day 20 of cortical differentiation. Cells expressing these markers were minor, less than 1%. $n=4, 6$ for WT and KO for each staining. DLX1: $t_7=4.521$, $P=0.0027$; GBX2: $t_7=0.3386$, $P=0.7448$; OLIG3: $t_7=3.414$, $P=0.0112$. **d**, ICC quantification of nuclei expressing either DLX1, GBX2, NKX2.1 or OLIG3 in *DLG2*^{-/-} and WT cells at day 30 of cortical differentiation. Cells expressing these markers were minor. $n=4, 6$ for WT and KO for each staining. DLX1: $t_7=14.05$, $P=2.19 \times 10^{-6}$; GBX2: $t_7=0.8917$, $P=0.4021$; NKX2.1: $t_7=3.875$, $P=0.0061$; OLIG3: $t_7=1.219$, $P=0.2623$. **e**, None of the CTIP2⁺ cells co-expressed striatal markers such as DARPP32 or FOXP1 and GABA. $n=3, 6$ for WT and KO. **f**, Minority of TBR1⁺ cells expressed the thalamic marker GBX2. $t_7=1.379$, $P=0.2105$, $n=3, 6$ for WT and KO. **g**, All SATB2⁺ cells expressed the telencephalic marker FOXG1. $n=3, 6$ for WT and KO. All data set except for **e** and **g** was analysed using unpaired two-tailed Student's t-tests for each marker investigated. All data presented as mean \pm SEM where possible. *, $p < 0.05$; **, $p < 0.01$; ****, $p < 0.0001$ compared to WT.



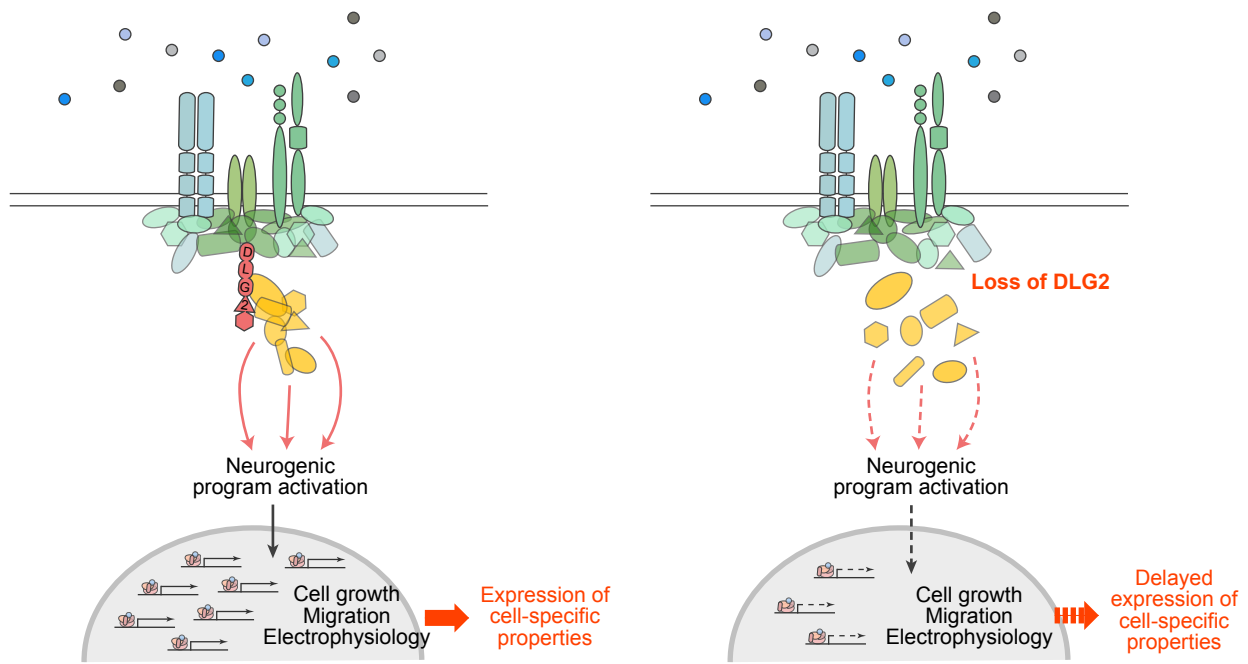
Supplementary Figure 7. Further characterisation of WT and *DLG2^{-/-}* cells during cortical differentiation

a, Representative ICC images showing the expression of the neuronal marker TUJ1 (β -Tubulin III) in 2 *DLG2^{-/-}* lines and WT controls at days 30 and 60 of cortical differentiation. **b**, MAP2 western blot protein bands and histograms of expression normalised to GAPDH and day 40 WT expression for *DLG2^{-/-}* and WT cells at 3 time points of cortical differentiation. $n=5, 7, 6$ from day 40 to 60 for WT and KO. Time ($F_{2,30}=8.721, P=0.0010$) did have a significant effect on MAP2 expression, while genotype ($F_{1,30}=0.2673, P=0.6157$) did not. **c**, ICC quantification of KI67 expressing nuclei for *DLG2^{-/-}* and WT cells at 5 time points of cortical differentiation. $n=3$ for WT at all time points, $n=6$ for KO at all timepoints except $n=5$ at day 50. Time ($F_{4,34}=20.71, P=1.00 \times 10^{-8}$) did have a significant effect on KI67 expression, while genotype ($F_{1,34}=0.1535, P=0.6976$) did not. **d**, KI67 western blot protein bands and histograms of expression normalised to GAPDH and day 30 WT expression for *DLG2^{-/-}* and WT cells at 4 time points of cortical differentiation. $n=13, 11, 13, 13$ for WT and $n=23, 22, 24, 21$ for KO from day 30 to 60. Time ($F_{3,132}=10.94, P=1.83 \times 10^{-6}$) did have a significant effect on MAP2 expression, while genotype ($F_{1,132}=0.1536, P=0.6957$) did not. **e**, SOX2 western blot protein bands and histograms of expression normalised to GAPDH and day 0 WT expression for *DLG2^{-/-}* and WT cells at 5 time points of cortical differentiation. $n=9, 3, 3, 3, 4$ for WT and $n=17, 6, 6, 3, 12$ for KO from day 0 to 60. Time ($F_{4,56}=4.856, P=0.0020$) did have a significant effect on SOX2 expression, while genotype ($F_{1,56}=0.2928, P=0.5906$) did not. **f**, Examples of SATB2⁺ cells co-expressing CTIP2 and quantification. $n=3, 6$ for WT and KO. All data sets except those in **f** were analysed by two-way ANOVA with post hoc comparisons using Bonferroni correction, comparing to WT controls. All data presented as mean \pm SEM.



Supplementary Figure 8. CTIP2 analysis of two *DLG2*^{-/-} cell lines

a, The proportion of CTIP2⁺ cells in both *DLG2*^{-/-} lines relative to WT level from ICC analysis. One-way ANOVA $F_{3,42}=5.391$, $P=0.0031$, $n=17, 9, 9, 11$ from day 30 to 60. **b**, The proportion of CTIP2⁺ cells in KO1 *DLG2*^{-/-} line relative to WT level from ICC analysis. One-way ANOVA $F_{3,26}=3.053$, $P=0.0462$, $n=10, 6, 6, 8$ from day 30 to 60. **c**, The proportion of CTIP2⁺ cells in KO2 *DLG2*^{-/-} line relative to WT level from ICC analysis. One-way ANOVA $F_{3,12}=12.61$, $P=0.0005$, $n=7, 3, 3, 3$ from day 30 to 60. **d**, The level of CTIP2 protein expression in both *DLG2*^{-/-} lines relative to WT level from Western blot analysis. One-way ANOVA $F_{3,59}=3.939$, $P=0.0125$, $n=18, 19, 13, 13$ from day 30 to 60. **e**, The level of CTIP2 protein expression in KO1 *DLG2*^{-/-} line relative to WT level from Western blot analysis. One-way ANOVA $F_{3,30}=6.375$, $P=0.0018$, $n=10, 10, 7, 7$ from day 30 to 60. **f**, The level of CTIP2 protein expression in KO2 *DLG2*^{-/-} line relative to WT level from Western blot analysis. One-way ANOVA $F_{3,25}=1.493$, $P=0.2407$, $n=8, 9, 6, 6$ from day 30 to 60. Data sets were analysed by one-way ANOVA, with post hoc comparisons using Bonferroni correction in all cases. Stars above bars represent Bonferroni-corrected post hoc tests, * $P<0.05$; ** $P<0.01$; *** $P<0.001$ vs. day 30. All data presented as mean \pm SEM.



Supplementary Figure 9. Proposed model for DLG2 action in neurodevelopment

External cues transduced by DLG2-scaffolded complexes regulate transcriptional accessibility and/or activation of neurogenic programs underlying cell growth, migration and development of electrophysiological signalling properties. DLG2 knockout impairs signal transduction, disrupting the orchestration of events required for normal development and leading to stochastic, imprecise signaling that delays expression of cell-specific properties.