Synaptic protein DLG2 controls neurogenic transcriptional programs disrupted in schizophrenia and related disorders

Bret Sanders[1], Daniel D'Andrea[2], Mark O. Collins[3], Elliott Rees[2], Tom G. J. Steward[4], Ying Zhu[1], Gareth Chapman[1], Sophie E. Legge[2], Antonio F. Pardiñas[2], Adrian J. Harwood[1], William P. Gray[1], Michael C. O'Donovan[2], Michael J. Owen[1,2], Adam C. Errington[1], Derek J. Blake[2], Daniel J. Whitcomb[4], Andrew J. Pocklington[2,*], Eunju Shin[1, *]

[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

*corresponding authors
Correspondence: shine@cardiff.ac.uk, pocklingtonaj@cardiff.ac.uk

Abstract

Genetic studies robustly implicate perturbation of DLG2-scaffolded mature postsynaptic signalling complexes in schizophrenia. Here we study *in vitro* cortical differentiation of *DLG2*$^{-/-}$ human embryonic stem cells via integrated phenotypic, gene expression and disease genetic analyses. This uncovers a developmental role for DLG2 in the regulation of neural stem cell proliferation and adhesion, and the activation of transcriptional programs during early excitatory corticoneurogenesis. Down-regulation of these programs in *DLG2*$^{-/-}$ lines delays expression of cell-type identity and causes marked deficits in neuronal migration, morphology and active properties. Genetic risk factors for neuropsychiatric and neurodevelopmental disorders converge on these neurogenic programs, each disorder displaying a distinct pattern of enrichment. These data unveil an intimate link between neurodevelopmental and mature signalling deficits contributing to disease - suggesting a dual role for known synaptic risk genes - and reveal a common pathophysiological framework for studying the neurodevelopmental origins of Mendelian and genetically complex mental disorders.

Keywords

DLG2, neurogenesis, cortical differentiation, RNAseq, GWAS, rare variant, schizophrenia, autism, psychiatric genetics

Introduction

The aetiology of schizophrenia (SZ) has long been recognised to possess a developmental component based upon observational studies of anatomical differences (brain structural and cytoarchitectural changes and minor physical abnormalities); premorbid cognitive and motor

symptoms; and pre-natal environmental risk factors (Murray and Lewis, 1987; Weinberger, 1987; Harrison, 1997, 1999). More recently large-scale genotyping studies have provided additional support, revealing substantial overlap between variants contributing to SZ liability and those conferring risk for other neurodevelopmental disorders including intellectual disability/severe neurodevelopmental delay (ID/NDD), autism spectrum disorders (ASD) and attention-deficit/hyperactivity disorder (ADHD) (Purcell et al., 2009; Sebat et al., 2009; Williams et al., 2010; Girirajan et al., 2012; Rees et al., 2014; Anttila et al., 2018).

Despite compelling evidence for their existence, little is known concerning the pre-natal neurodevelopmental processes disrupted in SZ beyond the fact that they are likely to occur early in the second trimester of pregnancy (Harrison, 1997; Hill and Bray, 2012; Clifton et al., 2019). Recent evidence suggests that many common genetic risk factors may impact gene expression in the mid-foetal brain (Walker et al., 2019) and are enriched in cell-types at multiple stages of cortical excitatory neuron development (Polioudakis et al., 2019). This raises the question: do common SZ variants converge on specific gene expression (transcriptional) programs that are activated or repressed during cortical excitatory neuron development in the mid-foetal brain? Mutations disrupting key regulators of such programs would be expected to possess a higher contribution to disease risk, reflected in a larger effect size and lower allele frequency. We therefore sought rare single-gene mutations linked to SZ where the affected gene is expressed in mid-foetal brain and has the potential to regulate transcriptional changes. This led us to *DLG2*. Independent *de novo* deletion events within *DLG2* have been found in SZ patients (Kirov et al., 2012); recurrent deletions in the promoter of *DLG2* are also implicated in ASD (Ruzzo et al., 2019). In humans, *DLG2* mRNA is present from 8 weeks post-conception (Kang et al., 2011) as well as *in vitro* throughout all stages of differentiation from human embryonic stem cells (hESCs) to cortical projection neurons (van de Leemput et al., 2014). DLG2 is required for the formation of NMDA receptor complexes as synapses mature (Frank et al., 2016): these complexes regulate activation of transcriptional programs underlying long-term changes in synaptic function and are enriched for rare mutations found in SZ cases (Kirov et al., 2012; Fromer et al., 2014; Purcell et al., 2014; Szatkiewicz et al., 2014; Pocklington et al., 2015; Genovese et al., 2016). This raises the possibility that DLG2 may also be required for the formation of signalling complexes regulating early neurodevelopmental expression programs disrupted in SZ.

To investigate its role in neurodevelopment we engineered homozygous loss-of-function *DLG2* mutations into hESCs using the CRISPR-CAS9 system. Mutant ($DLG2^{-/-}$) and isogenic sister wild-type (WT) hESC lines were differentiated into cortical excitatory neurons using a modified dual SMAD inhibition protocol (Chambers et al., 2009; Cambray et al., 2012) (Figure 1A, B). Cells were extensively characterised at days 15, 20, 30 and 60 to identify phenotypes and expression programs altered in $DLG2^{-/-}$ lines and investigate their disease relevance (Figure 1A, B).

Results

Knockout generation and validation

Two *DLG2*$^{-/-}$ lines were created from H7 hESCs using the CRISPR/Cas9-D10A nickase system targeting the first PDZ domain (Figure S1). Sequencing revealed no off-target mutations (see Methods, Figure S2 & Table S1). A significant decrease in *DLG2* mRNA was observed for exons spanning the first PDZ domain, with a similar decrease inferred for PDZ-containing transcripts, indicating degradation of *DLG2*$^{-/-}$ transcripts via nonsense-mediated decay (Figure S3A, B). Quantitative mass spectrometry-based proteomic analysis of peptide-affinity pulldowns using the NMDA receptor NR2 subunit PDZ peptide ligand (Husi and Grant, 2001) confirmed the presence of DLG2 in pulldowns from WT but not *DLG2*$^{-/-}$ lines (Figure S3C-F & Table S2). Genotyping revealed no CNVs in either *DLG2*$^{-/-}$ line relative to WT (Figure S4A). Both *DLG2*$^{-/-}$ lines expressed pluripotency markers OCT4, SOX2 and NANOG at 100% of WT levels (Figure S4B-E). Cells were extensively characterised for their cortical identity using western blotting and immunocytochemistry from day 20 to 60 (Figure S5). Over 90% of day 20 cells were positive for FOXG1, PAX6 and SOX2 confirming their dorsal forebrain fate (Figure S5A-B).

DLG2 regulates neural stem-cell proliferation and adhesion

RNA was extracted in triplicate from each line at 4 timepoints spanning cortical excitatory neuron development (Figure 1A, B) and gene expression quantified. To robustly identify genes dysregulated by *DLG2* knockout, expression data from the 2 *DLG2*$^{-/-}$ lines were pooled and compared to a WT sister line at each timepoint (see Methods). Disruption of *DLG2* had a profound effect: of the >13,000 protein-coding genes expressed at each timepoint, ~7% were differentially expressed between *DLG2*$^{-/-}$ and WT at day 15, rising to 40-60% between days 20 and 30 then decreasing to ~25% by day 60 (Figure 1C). Strikingly, the 3 genes with the strongest evidence for differential expression at day 15 were from the same family of transcriptional regulators: *ID1*, *ID3* and *ID2* (Figure 1D). *ID1-3* expression increases proliferation of cortical neural stem cells (NSCs) while inhibiting differentiation and maintaining adhesion to extracellular matrix (ECM) proteins (Jung et al., 2010; Niola et al., 2012). Also highly perturbed at days 15 and 20 was expression of *COL1A1* (Figure 1D, Table S3): this encodes the pro-α1 chain of type I collagen, an important ECM component in developing cortex (Long et al., 2018). To evaluate whether the wider set of genes differentially expressed in early proliferating cell-types (day 15 NSCs, day 20 NPCs) also highlight biological processes related to proliferation and ECM adhesion, we performed annotation over-representation tests using Gene Ontology (GO) terms (see Methods). Compared to all genes expressed at one or more timepoint in *DLG2*$^{-/-}$ or WT lines (*all*$^{WT+KO}$), those up-regulated at days 15 and 20 were over-represented in terms related to proliferation, differentiation, ECM composition and the regulation of ECM adhesion (Table S4) supporting dysregulation of these processes. Reasoning that changes in protein level typically lag those in mRNA we sought to validate these phenotypes in day 26 neural precursor cells (NPCs), also assaying day 0 (i.e. non-neural) hESCs as a negative control. We predicted that *DLG2*$^{-/-}$ cell-lines (in which *ID1-3* mRNAs are increased, Figure 1D) would exhibit increased proliferation and increased

adhesion to ECM proteins. *DLG2*$^{-/-}$ cells displayed increased adhesion to ECM protein substrates (Figure 1E, S6A) including type I collagen (highlighted by our expression analysis, Figure 1D) and fibronectin, a known mediator of *ID*-regulated adhesion (Niola et al., 2012). Proliferation did not differ between WT and *DLG2*$^{-/-}$ lines maintained as hESCs (Figure 1F), while day 26 NPCs from *DLG2*$^{-/-}$ lines exhibited significantly increased proliferation (Figure 1G, S6B). These data robustly confirm our predictions, illustrating that expression analyses identify true cell-biological changes.

Common risk variants implicate disruption of neurogenesis in SZ

We next sought to identify timepoints at which genes differentially expressed in *DLG2*$^{-/-}$ lines are enriched for SZ common risk variants. Taking summary statistics from the largest available SZ GWAS (Pardiñas et al., 2018), we utilised the competitive gene-set enrichment test implemented in MAGMA (version 1.07) (de Leeuw et al., 2015). As expected, *all*$^{WT+KO}$ was highly enriched for common variant association ($P = 2.1 \times 10^{-17}$) reflecting the neural lineage of these cells. We therefore tested genes up- and down-regulated at each timepoint for genetic association conditioning on *all*$^{WT+KO}$ using the strict *condition-residualize* procedure (all subsequent GWAS enrichment tests were conditioned on *all*$^{WT+KO}$ in the same way). This revealed strong association enrichment solely in genes down-regulated at day 30 ($30_{down}^{-/-}$: $P_{corrected} = 1.9 \times 10^{-7}$, Figure 2A), coinciding with active neurogenesis (Figure 1B). This suggests that common variant perturbation of earlier (day 15-20) expression programs impacting NSC proliferation and ECM adhesion is not central to disease pathophysiology. $30_{down}^{-/-}$ genes were over-represented in GO terms related to neuronal development, function and migration (Table S5). Iterative refinement via conditional analyses identified 23 terms with independent evidence for over-representation (Figure 2B) (see Methods). This suggests that loss of *DLG2* dysregulates transcriptional programs underlying neurogenesis (neuronal growth, active properties and migration) and implicates these processes in SZ aetiology.

Dysregulation of neurogenesis in DLG2$^{-/-}$ *lines delays cell-fate determination*

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 (Figure 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 (Figure 2F), TUJ1 (Figure S5C) plus cortical deep layer markers TBR1 and CTIP2 and upper layer marker SATB2. Western blot of NEUN (Figure 2C) and MAP2 (Figure S5D) and quantification of NEUN$^+$ cells following ICC (Figure 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 to 60 in *DLG2*$^{-/-}$ and WT cultures indicated by a similar proportion of KI67$^+$ and SOX2$^+$ cells (Figure S5E-G). A similar analysis of layer markers TBR1, CTIP2 and SATB2 revealed a significant decrease in CTIP2$^+$ cells but a comparable proportion of TBR1$^+$ and SATB2$^+$ neurons for all timepoints investigated

(Figure 2D-E, G-I, Figure S5H-I). On average the proportion of CTIP2^{+} cells in *DLG2*$^{-/-}$ recovered from 15% of the WT level on day 30 to 50% by day 60, although there was notable variation between *DLG2*$^{-/-}$ lines (Figure S6C-E); total CTIP2 protein level also recovered to some extent, but at a slower rate (Figure S6F-H). Thus *DLG2*$^{-/-}$ does not affect the rate at which neurons are produced but delays the expression of subtype identity in new-born 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 (Figure 2B), we sought to experimentally validate these phenotypes. Immature (day 30) and mature (day 70) neurons were traced and their morphology quantified (Figure 3). At both timepoints *DLG2*$^{-/-}$ neurons displayed a simpler structure than WT, characterised by a similar number of primary neurites projecting from the soma (Figure 3A) but with greatly reduced branching (Figure 3B). Total neurite length did not differ (Figure 3C), leading to a clear *DLG2*$^{-/-}$ phenotype of longer and relatively unbranched primary neurites (Figure 3E). There was no significant difference in soma area (Figure 3D). Day 40 *DLG2*$^{-/-}$ neurons had a slower speed of migration (Figure 3F) and reduced displacement from their origin after 70 hrs (Figure 3G-H). In summary, *DLG2*$^{-/-}$ neurons show clear abnormalities in both morphology and migration, validating the GO term analysis.

Distinct transcriptional programs regulated by DLG2 are enriched for common SZ risk alleles

DLG2 knockout has a profound impact on neurogenesis (growth, migration and expression of cell-type identity), down-regulating the expression of genes enriched for common SZ risk variants. We postulated that loss of *DLG2* inhibits activation of transcriptional programs driving neurogenesis, which starts between days 20 and 30 and steadily increases thereafter. If this is the case, then SZ genetic enrichment in $30_{down}^{-/-}$ should be captured by genes normally upregulated between days 20 and 30 in WT cultures ($20\text{-}30_{up}^{WT}$). Analysing differential expression between WT samples at successive timepoints we found strong risk variant enrichment in $20\text{-}30_{up}^{WT}$ (Figure 4A). The overlap between $20\text{-}30_{up}^{WT}$ and $30_{down}^{-/-}$ captured the signal in both sets ($P_{overlap} = 3.23 \times 10^{-10}$; $30_{down}^{-/-}$ only P = 0.44; $20\text{-}30_{up}^{WT}$ only P = 0.62). This was not simply due to the size of the overlap (3075 genes, 85% of $20\text{-}30_{up}^{WT}$) as the regression coefficient for the set of overlapping genes ($\beta = 0.14$), which reflects magnitude of enrichment, was significantly greater than for genes unique to $30_{down}^{-/-}$ ($\beta = 0.006$, $P_{different} = 0.0015$) or $20\text{-}30_{up}^{WT}$ ($\beta = -0.015$, $P_{different} = 0.0045$). Thus, it is neurogenic transcriptional programs down-regulated in *DLG2*$^{-/-}$ lines that are enriched for SZ common variants.

To more precisely identify SZ-relevant transcriptional programs active during neurogenesis, we classified $20\text{-}30_{up}^{WT}$ genes based on their subsequent WT expression profiles (Figure 4B, see Methods): 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 downregulated. 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 $30_{down}^{-/-}$ (62 out of 1399 genes) and was therefore left intact. Early-stable$^{-/-}$ and early-increasing$^{-/-}$ sets were robustly enriched for SZ association (conditioning on *all*$^{WT+KO}$, Figure 4C), revealing that association is restricted to 2 distinct transcriptional programs normally activated during the onset of neurogenesis but down-regulated in *DLG2*$^{-/-}$ lines.

To confirm the existence of SZ-relevant transcriptional programs *in vivo* we repeated our derivation of early- and late programs using single-cell expression (scRNAseq) data from human foetal cortex (Nowakowski et al., 2017), mapping the principal cell-types present at each timepoint in culture (Figure 1B) onto sub-types identified *in vivo* (see Methods). Despite the notably reduced gene/transcript coverage of existing scRNAseq technologies, we confirmed that SZ association was restricted to early-stable and early-increasing programs. Effect sizes were closely comparable between *in vitro* and *in vivo* programs (Figure 4C-D).

A cascade of transcriptional programs drives neurogenesis & differentiation

We next investigated the biological function of the 4 main programs identified *in vitro*: early-transient$^{-/-}$, early-stable$^{-/-}$, early-increasing$^{-/-}$ and late. Each was over-represented for a coherent set of GO terms indicating a distinct biological role (Table S6): early-transient$^{-/-}$ for histone/chromatin binding and transcriptional regulation; early-stable$^{-/-}$ for signal transduction, transcriptional regulation, neurogenesis, cell projection development, migration and differentiation; early-increasing$^{-/-}$ for axon guidance, dendrite morphology and the establishment of synaptic connections and active properties; and late for mitochondrial energy production, peptide synthesis/transport, synaptic signalling and the regulation of ion homeostasis linked to synaptic transmission.

The distinct biological roles of the 3 early- programs prompted us to hypothesise that they form a linked, time-ordered cascade of transcriptional programs triggered at the onset of neurogenesis. This begins with an initial phase of chromatin remodelling as NPCs exit the cell-cycle and commit to a neuronal fate (early-transient$^{-/-}$), activating a longer-term program guiding cell growth and migration (early-stable$^{-/-}$). This, in turn, promotes the fine-tuning of cell-type specific neuronal structure, function and connectivity as cells enter the terminal phase of differentiation (early-increasing$^{-/-}$).

To investigate support for this regulatory cascade we identified key regulators from each program whose downstream targets have been experimentally identified or computationally predicted (see Methods): chromatin modifier *CHD8* from early-transient (Sugathan et al., 2014; Cotney et al., 2015); transcription factor *TCF4* and translational regulator FMRP from early-stable$^{-/-}$ (Forrest et al., 2018, Darnell et al., 2011); and transcription factors (and deep layer markers) *TBR1* and *BCL11B* (CTIP2) from early-increasing$^{-/-}$ (Wang et al., 2018). We

predicted that a substantial proportion of early-stable$^{-/-}$ genes would be directly regulated by *CHD8*, while early-increasing$^{-/-}$ would be enriched for genes directly regulated by *TCF4*, *TBR1* and *BCL11B*. Since early-transient$^{-/-}$ is responsible for activating later programs, we predicted that early-increasing$^{-/-}$ would be enriched for indirect targets of *CHD8* (genes not directly regulated but whose expression is altered when *CHD8* is perturbed) that are down-regulated in *CHD8* knockdown cells (Sugathan et al., 2014). We also predicted that early-transient$^{-/-}$ genes 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 (Krichevsky and Kosik, 2001; Darnell et al., 2011), and its function is known to be important for axon and dendrite growth (Antar et al., 2006). We therefore predicted that early-stable and -increasing (but not -transient) genes would be enriched for FMRP targets. Over-representation tests emphatically confirmed these predictions (Figure 4E). In addition, the targets of *TCF4*, FMRP, *BCL11B* and *TBR1* were more highly enriched for SZ association than other genes in early-increasing$^{-/-}$ (Figure 4E), indicating an important role for these regulators in disease.

Convergence of genetic risk on perturbed action potential generation

We next tested whether biological processes over-represented in early-stable$^{-/-}$ or early-increasing$^{-/-}$ (Table S6) captured more or less of the SZ association in these programs than expected (see Methods). None of the 13 semi-independent GO term subsets identified in early-stable$^{-/-}$ differed substantially from early-stable$^{-/-}$ as a whole (Table S7), indicating that risk factors are distributed relatively evenly between them. Of the 16 subsets for early-increasing$^{-/-}$, *somatodendritic compartment* and *membrane depolarization during action potential* displayed evidence for excess enrichment relative to the program as a whole (Figure 4F). Again, no single term showed evidence for depletion, suggesting that diverse biological processes regulating neuronal growth, morphology and function are perturbed in SZ. The enhanced enrichment in action potential (AP) related genes is particularly striking: while complexes regulating postsynaptic information processing have been robustly implicated in disease, this represents the first evidence that electrical properties underlying information transmission are also disrupted. We therefore sought to confirm the disruption of APs in *DLG2*$^{-/-}$ lines (Figure 5A-K), also investigating the impact of *DLG2* loss on synaptic transmission (Figure 5L-N).

Day 50 *DLG2*$^{-/-}$ neurons were less excitable, with a significantly more depolarised resting membrane potential (Figure 5A). Stepped current injection evoked AP firing in 80% WT but only 43% *DLG2*$^{-/-}$ neurons (Figure 5C). APs produced by *DLG2*$^{-/-}$ cells were characteristic of less mature neurons (Figure 5D), having smaller amplitude, longer half-width and a slower maximum rate of depolarisation and repolarisation ($\delta V/\delta t$) (Figure 5E-H). We found no change in AP voltage threshold, rheobase current (Figure 5I-J) or input resistance (Figure 5B). The percentage of neurons displaying spontaneous excitatory postsynaptic currents (EPSCs) was comparable at days 50 and 60 (Figure 5N) as was EPSC frequency and amplitude (Figure

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 (Figure 5O). In summary, *DLG2*$^{-/-}$ neurons have a reduced ability to fire APs and produce less mature APs.

Loss-of-function intolerant genes enriched for common risk variants localise to developmental transcriptional programs

Genes intolerant to loss-of-function mutations (LoFi genes) are highly enriched for SZ GWAS association (Pardiñas et al., 2018). However, it is not known when or where in development they exert their effect. We predicted that LoFi genes would primarily be concentrated in earlier transcriptional programs (early-transient$^{-/-}$, early-stable$^{-/-}$) where the impact of disruption is potentially more severe. LoFi genes were over-represented in both of these and also early-increasing$^{-/-}$ (Figure 6A). LoFi GWAS association (conditioned on *all*$^{WT+KO}$) was largely captured by the overlap with early-stable$^{-/-}$ and early-increasing$^{-/-}$ sets, localising genetic risk factors to roughly one third of LoFi genes and providing a clear biological context for their action (Figure 6B).

Rare and common SZ risk variants converge on common neurodevelopmental pathways

We next tested each program for enrichment in rare SZ mutations. Compared to all expressed genes (*all*$^{WT+KO}$), both *de novo* and singleton LoF mutations were enriched in early-stable$^{-/-}$. Early-increasing$^{-/-}$ displayed a trend towards enrichment in LoF singleton mutations only (Figure 6C-D). This provides independent genetic evidence for disruption of transcriptional programs regulating neurogenesis in SZ.

Risk variants for other neuropsychiatric disorders display distinct patterns of enrichment

Finally, we interrogated transcriptional programs for enrichment (relative to *all*$^{WT+KO}$) in variants conferring risk for different disorders. Early-transient$^{-/-}$, -stable$^{-/-}$ and -increasing$^{-/-}$ sets had an increased rate of *de novo* LoF mutations contributing to NDD and ASD (Satterstrom et al., 2019) (Figure 6E). Comparing the magnitude of the increase, a clear gradient was evident from NDD to ASD to SZ (Figure 6F). LoF mutations from unaffected siblings of individuals with ASD (Satterstrom et al., 2019) showed no evidence for an increased rate in any of the sets tested (Figure 6E-F). ADHD common variants (Demontis et al., 2019) were enriched in early-increasing$^{-/-}$ and there was a modest enrichment for bipolar disorder (BIP) common variants (Stahl et al., 2019) in early-stable$^{-/-}$ (Figure 6G-H). In contrast, common variants conferring risk for neurodegenerative disorder Alzheimer's disease (AD) (Lambert et al., 2013) were not enriched in any program (Figure 6G-H). Dysregulation of neurogenesis thus contributes to a wide spectrum of neuro-developmental/-psychiatric disorders, each displaying a distinct profile of risk variant enrichment across neurogenic transcriptional programs (Figure 6F,H).

Discussion

A complex choreography of cell proliferation, specialisation, growth, migration and network formation underlies brain development. To date, limited progress has been made pinpointing aspects of this process disrupted in neurodevelopmental disorders. Here we demonstrate that DLG2 - in addition to its role as a postsynaptic scaffold protein - regulates neural stem cell proliferation and adhesion and the activation of transcriptional programs underlying early corticoneurogenesis that are enriched for genetic risk factors contributing to a wide range of disorders (Figure 7). 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 impact (Craddock and Owen, 2010; Owen et al., 2011; Owen and O'Donovan, 2017)(Figure 7B). Our data support this model and ground it in developmental neurobiology (Figure 6E-H), embedding genetic risk for multiple disorders in a common pathophysiological framework.

Novel role for DLG2 in early corticoneurogenesis

The birth of neurons during CNS development is a sequential process involving the generation of progressively more specialised cell-types. Proliferating neuroectodermal cells (NSCs) in the neural tube give rise to neural precursors (NPCs): radial glia (RG) and intermediate progenitors (iPCs). Both RG and iPCs have a limited ability to proliferate, with RG giving rise to neurons either directly or via iPCs. In the cortex, excitatory neurons in layers VI-II are born in a sequential, inside-out fashion (from deep to upper layers), with early-born post-mitotic cells becoming deep layer (VI, V) neurons. This process is broadly recapitulated by *in vitro* differentiation protocols (e.g. Figure 1B), with neurons produced between our day 30 and 60 timepoints predominantly displaying a deep layer identity.

NPCs from *DLG2*$^{-/-}$ lines exhibit increased proliferation but do not significantly differ in terms of the differentiation timepoint at which they start to commit to a neuronal fate or the subsequent rate at which neurons are generated. This indicates that loss of DLG2 leads to more rapid cell-cycling of NSCs/NPCs but does not impair the onset of neurogenesis. The increased expression of *ID1-3* in *DLG2*$^{-/-}$ NSCs is consistent with a more rapid cell-cycle, as knockout of *ID1-3* has been shown to increase cell-cycle length (Niola et al., 2012). *ID1-3* knockout also decreases adhesion to the ventricular surface of embryonic cortex *in vivo* and to laminin and fibronectin *in vitro* (Niola et al., 2012). In line with this, *DLG2*$^{-/-}$ NPCs displayed increased adhesion to multiple ECM components including fibronectin.

DLG2$^{-/-}$ lines give rise to fewer neurons expressing deep layer marker CTIP2 (*BCL11B*), although bulk mRNA levels recover fully by day 60 (Table S3, Figure 2E). The proportion of CTIP2^{+} cells also significantly recovers over time (Figure S6C-E). Neuronal identity is thought to be determined by the internal state of NPCs immediately prior to cell-cycle exit, with changes in this state over time leading to the progressive generation of layer VI-II cells (Telley et al., 2019). This would suggest that *DLG2* loss causes a delay in new-born neurons exhibiting

their latent, deep layer identity. This also matches the observed pattern of TBR1 expression. TBR1 is expressed by virtually all post-mitotic glutamatergic neurons *in vivo* but has markedly higher expression in early-born neurons (Hevner et al., 2001; Englund et al., 2005) i.e. while basal levels of TBR1 are a feature of all new-born neurons, higher levels are associated with neurons manifesting deep layer properties. The proportion of $TBR1^{+}$ neurons does not differ between *DLG2*$^{-/-}$ and wild-type (Figure 2D, G), consistent with no change in the rate of neurogenesis. However, total TBR1 mRNA is significantly decreased in *DLG2*$^{-/-}$ lines at day 30 (Table S3) indicating a reduced level of expression in individual new-born neurons (predominantly deep layer subtypes, as noted earlier). This reduction in TBR1 mRNA recovers by day 60 (Table S3), consistent with *DLG2* loss delaying the manifestation of deep layer identity. The wider changes in gene expression and neuronal physiology seen in *DLG2*$^{-/-}$ lines are also consistent with such a delay. Loss of *DLG2* suppresses expression of many genes (85%) normally up-regulated between days 20 and 30 in wild-type, coinciding with the onset of neurogenesis. Prominent amongst those suppressed are genes contributing to the development of cell morphology, connectivity and active properties (Figure 2B, 7A) – key features of neuronal identity – leading to developing neurons with long, sparsely branched primary neurites and immature action potentials (Figure 3, 5). Gene expression recovers by day 60, at which point none of the above processes (morphology, connectivity, active property development) are over-represented amongst genes dysregulated in *DLG2*$^{-/-}$ lines (data not shown). Morphological deficits persist at least until day 70, although the extent to which delayed upregulation of gene expression impacts mature neuronal identity is unclear: further work will need to investigate the persistence of deficits in hESC-derived *DLG2*$^{-/-}$ neurons following transplantation into rodent brain.

Genes down-regulated at day 30 in *DLG2*$^{-/-}$ lines are enriched for rare and common variants conferring risk to a wide range of disorders including SZ and ASD (Figure 6E-H), linking *DLG2*-dependent neuronal deficits to disease. Consistent with this, multiple independent deletion events at the *DLG2* locus have been identified in SZ (Kirov et al., 2012) and ASD cases (Ruzzo et al., 2019). Previous studies also support a role for SZ and ASD risk factors during corticoneurogenesis (Willsey et al., 2013; Polioudakis et al., 2019; Walker et al., 2019). Abnormalities in neuronal migration and morphology have been reported in post-mortem studies of both disorders (Harrison, 1999; Reiner et al., 2016; Martínez-Cerdeño, 2017) although caveats over sample size, reproducibility and confounding via reverse causation remain. Here we localise genetic risk to specific gene expression programs regulating these phenotypes, supporting their direct involvement in disease pathophysiology.

Disruption of neurogenic transcriptional programs in neurodevelopmental disorders

The commitment of precursors to a neuronal fate triggers an orchestrated sequence of events requiring the timely activation and repression of multiple transcriptional programs. Here we identify a cascade of transcriptional programs active during early corticoneurogenesis. While much further work is required to validate and refine this model it provides a useful conceptual

framework that locates many genetic risk factors in a coherent sequence of cellular events, generating insight into the overlapping aetiologies of Mendelian and genetically complex neurodevelopmental disorders (Figure 6, 7A).

Major changes in identity require alteration of chromatin structure to reflect cell-type differences in gene transcription (Yadav et al., 2018). This is evident in early-transient$^{-/-}$, in which histone- and chromatin-binding genes are highly over-represented (Table S6, Figure 7A). Known risk genes include chromatin modifier *CHD8* (ASD, Neale *et al.*, 2012; O'Roak *et al.*, 2012; Sanders *et al.*, 2012) and histone methyltransferase *SETD1A* (NDD & SZ, Singh *et al.*, 2016). Re-organisation of the transcriptional landscape leads to activation of a second expression program (early-stable$^{-/-}$) which oversees a prolonged phase of cell growth and differentiation, accompanied *in vivo* by radial migration into the appropriate cortical layer (Figure 7A). Targets of *CHD8* are highly over-represented in early-stable$^{-/-}$ (Figure 4E) indicating an important role in the activation of this program. These targets include transcription factor *TCF4*, another major regulator of early-stable$^{-/-}$ genes (Figure 4E) and a known interactor of pro-neural genes during differentiation (Flora et al., 2007). Rare mutations in *TCF4* lead to Pitt-Hopkins syndrome, whose features include ID/NDD (Amiel *et al.*, 2007); genome-wide significant (GWS) SZ association has also been identified at the *TCF4* locus (Ripke *et al.*, 2013, Pardiñas *et al.*, 2018). Also present are *TRIO* (SZ, Katrancha *et al.*, 2017), which promotes reorganization of the actin cytoskeleton during neurite outgrowth (Bateman et al., 2000; van Haren et al., 2014), and *FMR1* (FMRP) (fragile-X syndrome, Verkerk *et al.*, 1991; ASD, Darnell *et al.*, 2011; SZ, Pardiñas *et al.*, 2018). While discussion of FMRP in disease has tended to focus on its role in plasticity at mature synapses (Bagni and Zukin, 2019, Darnell et al., 2011, Fromer et al., 2014), it is also present in the axons and dendrites of developing neurons where it controls growth cone morphology and motility (Antar et al., 2006). Reflecting this role, targets of FMRP are highly over-represented in early-stable$^{-/-}$ (Figure 4E).

TCF4 and FMRP also regulate components of a third program (early-increasing$^{-/-}$) that shapes the emergence of cell-type specific properties: dendrite morphology, synaptic connectivity and neuronal activity (Figure 4E, 7A). This program contains transcriptional regulators *TBR1* and *BCL11B* (CTIP2), markers of deep layer neurons discussed above. *TBR1* is enriched for rare coding variants in ASD cases (Sanders et al., 2015), while *BCL11B* lies in a single-gene GWS SZ locus (Ripke et al., 2014; Pardiñas et al., 2018). Rare coding variation in genes controlling AP propagation has previously been implicated in ASD (De Rubeis et al., 2014); here we find that early-increasing$^{-/-}$ genes regulating action potential (AP) firing display stronger evidence for SZ association than other elements of this program. These 8 genes encode: voltage-gated sodium channels *SCN1A-SCN3A* ($Na_v1.1$-1.3) and *SCN3B* ($Na_v\beta3$); voltage-gated calcium channels *CACNA1C* ($Ca_v1.2$), *CACNA1I* ($Ca_v3.3$), *CACNA2D1* ($Ca_v\alpha2\delta$); and hyper-polarization activated cyclic-nucleotide-gated channel *HCN2*. Na^+ influx through $Na_v1.1$-1.3 drives membrane depolarisation during AP firing, with K^+ efflux driving repolarisation. Membrane

hyperpolarisation activates HCN channels, allowing mixed Na^+/K^+ currents which help restore the membrane potential to its resting level (Robinson and Siegelbaum, 2003). T-type ($Ca_v3.3$) and L-type ($Ca_v1.2$) calcium channels contribute to neuronal firing by initiating rebound bursting and controlling calcium-sensitive potassium channels respectively (Simms and Zamponi, 2014). Mutations in $Na_v1.2$ (*SCN2A*) lead to a wide spectrum of epileptic disorders, ID and ASD (Wolff et al., 2017). An excess of rare variants in Na_v genes has also recently been reported for SZ (Rees et al., 2019a). Mutations in $Ca_v1.2$ (*CACNA1C*) lead to Timothy syndrome (Splawski et al., 2004) and Brugada syndrome (Watanabe and Minamino, 2016); both *CACNA1C* and *CACNA1I* are located in single-gene GWS SZ loci (Pardiñas et al., 2018). Intriguingly, enhancement of spontaneous calcium transients by Na_v activity impedes migration and induces premature dendritic branching in developing neurons (Bando et al., 2016). $Ca_v1.2$ (*CACNA1C*) plays a critical role in spontaneous calcium transient generation in the axons and dendrites of developing cortical neurons, with loss of *CACNA1C* reducing neurite growth and a Timothy syndrome gain-of-function mutation impairing radial migration (Kamijo et al., 2018). This suggests that, in addition to their role in AP firing, early-increasing $^{-/-}$ Na_v and Ca_v expression may drive an increase in calcium transients that acts as a cue for neurons to end migration and initiate terminal differentiation (dendritic branching and synaptogenesis). Deficits in this process could thus alter cell-type specific morphology and connectivity, perturbing network formation. Consequently, reduced Na_v and Ca_v gene expression (Table S3) may contribute to the reduced neurite branching seen in developing *DLG2*$^{-/-}$ neurons.

One further transcriptional program (late), initiated towards day 60 as developing neurons start to become active, upregulates mitochondrial energy production and ion transport/homeostatic pathways supporting neuronal signalling (Figure 7A). Peptide synthesis and transport are also upregulated, possibly indicating the production of co-released neuropeptides. Strikingly, this program did not display any evidence for disease-genetic association. It was however strongly depleted for *de novo* LoF mutations in NDD cases (Figure 6E-F). While there is long-standing interest in the role of mitochondria in disorders such as SZ, our data suggest that mitochondrial deficits do not play a major role in the developmental aetiology of those disorders studied here.

DLG2 complexes in disease and development

Disease-relevant genes and physiological processes contributing to mature neuronal function also play important (and potentially unexpected e.g. Na_v/Ca_v) roles during development. For SZ, disease models with robust genetic support centre upon postsynaptic signalling and the regulation of synaptic plasticity by NMDAR/ARC complexes and FMRP (Kirov et al., 2012; Fromer et al., 2014; Purcell et al., 2014; Szatkiewicz et al., 2014; Pocklington et al., 2015; Genovese et al., 2016). Scaffold protein DLG2 is required for assembly of NMDAR complexes at mature synapses (Frank et al., 2016) and we propose a similar mode of action during development (Figure 7C). Supporting this, the single invertebrate orthologue of vertebrate

genes *DLG1-DLG4 (Dlg)* is a core component of the *Scrib* signalling module regulating cell polarity, differentiation and migration (Stephens et al., 2018). We propose that DLG2 links cell-surface receptors to signal transduction molecules driving the activation of neurogenic transcriptional programs. This is normally tightly coupled to cell-cycle exit – unaffected by *DLG2* loss (Figure 2) indicating its initiation via independent pathways (Figure 7C). *DLG2* knockout results in stochastic signalling that delays and impairs transcriptional activation, disrupting the orchestration of events required for normal development and the precise specification of neuronal properties. Proliferation and adhesion deficits observed in *DLG2*$^{-/-}$ NPCs, prior to the onset of neurogenesis, may be driven by similar mechanisms. Aberrant chromatin structure in NPCs has been linked to altered neuronal maturation in ASD with macrocephaly (Schafer et al., 2019). While common variant analyses did not implicate NPC biology in SZ, we do see evidence for rare variant enrichment in genes differentially expressed between NSCs and NPCs, indicating a wider role for DLG2 in cell-type transitions relevant to disease (unpublished data).

Individuals with SZ, ASD and other disorders carry a highly heterogeneous burden of rare and common variants impacting neurogenesis. Whether this results in more or less severe disruption than that observed in *DLG2*$^{-/-}$ cells is not known. Precise timing is crucial during brain development, where the correct dendritic morphology, axonal length and electrical properties are required for normal cortical and subcortical circuit formation. Consequently, even transient perturbation of neurogenesis may have a profound impact on fine-grained neuronal wiring, network function and ultimately perception, cognition and behaviour.

Acknowledgments

This work was supported by Wellcome Trust Strategic Award (100202/Z/12/Z), MRC programme grant (G08005009), MRC Centre grant (MR/L010305/1), Waterloo Foundation 'Changing Minds' programme and start-up funding from the Neuroscience and Mental Health Research Institute, Cardiff University. We acknowledge excellent technical support for RNA sequencing from Joanne Morgan (MRC Centre) 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, Lesley Jones and James Walters for helpful comments on the manuscript and Emily Adair for providing rat primary glial cells.

Data usage acknowledgements

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 n° 503480), Alzheimer's Research UK (Grant n° 503176), the Wellcome Trust (Grant n° 082604/2/07/Z) and German Federal Ministry of Education and Research (BMBF): Competence Network Dementia (CND) grant n° 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

Loss of function singleton data used for the analysis described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/gap through dbGaP accession number phs000473.v2.p2. Samples used for data analysis were provided by the Swedish Cohort Collection supported by the NIMH Grant No. R01MH077139, the Sylvan C. Herman Foundation, the Stanley Medical Research Institute and The Swedish Research Council (Grant Nos. 2009-4959 and 2011-4659). Support for the exome sequencing was provided by the NIMH Grand Opportunity Grant No. RCMH089905, the Sylvan C. Herman Foundation, a grant from the Stanley Medical Research Institute and multiple gifts to the Stanley Center for Psychiatric Research at the Broad Institute of MIT and Harvard.

The Schizophrenia Exome Sequencing Meta-analysis (SCHEMA) consortium is a large multi-site collaboration dedicated to aggregating, generating, and analyzing high-throughput sequencing data of schizophrenia patients to improve our understanding of disease architecture and advance gene discovery. The consortium was formed in mid-2017 with a deep commitment of data sharing, diversity, and inclusivity. The SCHEMA consortium is made possible by the generosity of many funders, including the Stanley Foundation, and NIH, and the leadership of its members. We would also like to thank the many tens of thousands of patients and families who generously contributed to this effort.

Author contributions

Conceptualization, AP, ES
Methodology, AE, DB, AP, ES
Software/Data curation, DDA, AP
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, MOD, MO, AE, DB, DW, AP, ES
Visualization, BS, DDA, MC, YZ, DW, AP, ES
Supervision, AP, ES

Declaration of Interests

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 remaining authors report no financial relationships with commercial interests.

Figure 1. Study design and initial characterisation of *DLG2*$^{-/-}$ phenotypes. (A) Study summary. *DLG2*$^{-/-}$ hESCs were generated via the CRISPR/CAS9 system; a sister hESC line that went through the gene editing protocol but remained as WT served as a control. hESCs were differentiated into cortical excitatory neurons and RNA collected from day 15, 20, 30 and 60 cultures where the major cell types are neural stem cells (NSCs), neural precursor cells (NPCs), new-born and mature neurons respectively. RNA sequencing analysis and subsequent experimental validation revealed *DLG2*$^{-/-}$ phenotypes: increased NSC/NPC proliferation and adhesion to extracellular molecules; and decreased expression of deep layer markers, reduced morphological complexity, migration and action potential maturity in human cortical neurons *in vitro*. Transcriptional programs active during neurogenesis were identified based upon differential gene expression across successive timepoints. SZ common variant risk was concentrated in two early programs downregulated in *DLG2*$^{-/-}$ cells (early-stable$^{-/-}$ and early-increasing$^{-/-}$). Loss of function intolerant (LoFi) genes were highly over-represented in all 3 early expression programs; only LoFi genes in early-stable$^{-/-}$ and -increasing$^{-/-}$ were enriched for SZ common risk variants. Early transcriptional programs were enriched for rare LoF variants contributing to a wide range of neurodevelopmental/neuropsychiatric disorders; no enrichment was seen in rare mutations from unaffected siblings or common risk variants for (neurodegenerative disorder) Alzheimer's disease. **(B)** Overview of cortical differentiation protocol with approximate timings of key developmental processes and predominant cell types present in culture. Asterisks indicate time point selected for RNA sequencing. **(C)** Number of protein coding genes differentially expressed in *DLG2*$^{-/-}$ cells relative to WT at each timepoint **(D)** Top 5 differentially expressed genes at day 15 of cortical differentiation. Fold change is for *DLG2*$^{-/-}$ expression level relative to WT **(E)** Adhesion of *DLG2*$^{-/-}$ and WT cells to various ECM protein substrates at day 25 of cortical differentiation. Both genotype ($F_{1,128}$=38.08; P<0.0001; n≥6) and ECM substrate ($F_{7,128}$=16.20; P<0.0001; n≥6) had significant effects on adhesion. **(F)** Proliferation of *DLG2*$^{-/-}$ and WT lines maintained as hESCs for 5 days. While time ($F_{4,35}$=1924, P<0.0001; n≥3) had a significant effect on hESC proliferation, genotype ($F_{1,35}$=0.3780, P=0.8470; n≥3) did not. **(G)** Proliferation of *DLG2*$^{-/-}$ and WT lines from days 26 to 36 of cortical differentiation. Both genotype ($F_{1,80}$=738.5; P<0.0001; n≥6) and time ($F_{4,80}$=126.1; P<0.0001; n≥6) had significant effects on proliferation and there was a significant interaction between these factors ($F_{4,80}$=59.36; P<0.0001; n≥6). Adhesion and proliferation data sets (E-G) were analysed by two-way ANOVA with post hoc comparisons using Bonferroni correction, comparing to WT controls. Stars above bars represent Bonferroni-corrected post hoc tests, *P<0.05; **P<0.01; ****P<0.0001 vs. WT control. All data presented as mean ± SEM.

Figure 2. Common risk variants implicate disruption of neurogenesis in schizophrenia. (A) Enrichment for common schizophrenia risk variants in genes up- & down-regulated at each timepoint (*DLG2*$^{-/-}$ relative to WT), conditioning on all expressed genes. Dotted line indicates $P_{corrected}$ = 0.05 following Bonferroni correction for 8 tests. **(B)** Gene ontology (GO) terms over-represented amongst genes down-regulated at day 30 in *DLG2*$^{-/-}$ lines relative to all expressed genes. **(C)** NEUN western blot protein bands and histograms of expression normalised to GAPDH for *DLG2*$^{-/-}$ and WT cells at days 30, 40, 50 and 60 of cortical differentiation. Neither genotype ($F_{1,90}$=0.1852; P=0.6680; n≥7) nor time ($F_{3,90}$=0.5382; P=0.6573; n≥7) had significant effects on NEUN expression. **(D)** TBR1 western blot protein bands and histograms of expression normalised to GAPDH for *DLG2*$^{-/-}$ and WT cells at days 30, 40, 50 and 60 of cortical differentiation. Neither genotype ($F_{1,95}$=0.3899; P=0.5338; n≥9) nor time ($F_{3,35}$=0.5052; P=0.6793; n≥9) had significant effects on TBR1 expression. **(E)** CTIP2 western blot protein bands and histograms of expression normalised to GAPDH for *DLG2*$^{-/-}$ and WT cells at days 30, 40, 50 and 60 of cortical differentiation. Both genotype ($F_{1,86}$=39.89; P<0.0001; n≥7) and time ($F_{3,86}$=5.262; P=0.0022; n≥7) had significant effects on CTIP2 expression. **(F)** ICC quantification of NEUN expressing nuclei for *DLG2*$^{-/-}$ and WT cells at 4 time points of cortical differentiation. Time ($F_{3,52}$=7.018, P=0.0005; n≥6) had a significant effect on NEUN expression, while genotype ($F_{1,52}$=1.687; P=0.1998; n≥6) did not. **(G)** ICC quantification of TBR1 expressing nuclei for *DLG2*$^{-/-}$ and WT cells at 4 time points of cortical differentiation. Time ($F_{3,58}$=4.738, P=0.0050; n≥6) did have a significant effect on TBR1 expression, while genotype ($F_{1,58}$=1.664; P=0.2022; n≥6) did not. **(H)** ICC quantification of CTIP2 expressing nuclei for *DLG2*$^{-/-}$ and WT cells at 4 time points of cortical differentiation. Both genotype ($F_{1,67}$=101.8; P<0.0001; n≥6) and time ($F_{3,67}$=18.93; P<0.0001; n≥6) had significant effects on CTIP2 expression. **(I)** Representative ICC images of NEUN, TBR1 and CTIP2 with DAPI nuclear counterstain for 2 *DLG2*$^{-/-}$ lines and WT controls at days 30 and 60 of cortical differentiation. Western blotting (C-E) and ICC data sets (F-H) were analysed by two-way ANOVA with post hoc comparisons using Bonferroni correction, comparing to WT controls. Stars above bars represent Bonferroni-corrected post hoc tests, *P<0.05; **P<0.01; ***P<0.001; ****P<0.0001 vs. WT control. All data presented as mean ± SEM.

Figure 3. *DLG2*$^{-/-}$ lines display deficits in neuron morphology & migration. (A) The number of primary neurites (projecting from the soma) in *DLG2*$^{-/-}$ and WT neurons at days 30 and 70 of cortical differentiation. Neither genotype ($F_{1,126}$=1.591; P=0.2095; n≥28) nor time ($F_{1,126}$=2.278; P=0.1337; n≥28) had significant effects the numbers of primary neurites. **(B)** The number of secondary neurites (projecting from primary neurites) in *DLG2*$^{-/-}$ and WT neurons at days 30 and 70 of cortical differentiation. Genotype ($F_{1,126}$=18.78, P<0.0001; n≥28) had a significant effect on number of secondary neurites, while time ($F_{1,126}$=1.082, P=0.3003; n≥28) did not. **(C)** The total neurite length in *DLG2*$^{-/-}$ and WT neurons at days 30 and 70 of cortical differentiation. Both genotype ($F_{1,126}$=4.568; P=0.0345; n≥28) and time ($F_{1,126}$=26.33; P<0.0001; n≥28) had significant effects on total neurite length. However, post hoc analysis

showed no significant differences at individual timepoints. **(D)** The soma area of *DLG2*$^{-/-}$ and WT neurons at days 30 and 70 of cortical differentiation. Neither genotype ($F_{1,136}$=; P=0.9170; n≥28) nor time ($F_{1,136}$=1.399; P=0.2390; n≥28) had a significant effect on soma area. **(E)** Representative traces showing the morphology of *DLG2*$^{-/-}$ and WT neurons at days 30 and 70 of cortical differentiation, scale bars are 100 µm and apply to both cell lines at a given time point. **(F)** The average speed of migration for *DLG2*$^{-/-}$ and WT neurons 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<0.0001; n=27). **(G)** The displacement of *DLG2*$^{-/-}$ and WT neurons at 70 hours migration, from day 40 of cortical differentiation. *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, from day 40 of cortical differentiation. 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). All data presented as mean ± SEM.

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 WT, conditioning on all expressed genes. Dotted line indicates $P_{corrected}$ = 0.05 following Bonferroni correction for 6 tests. **(B)** Four discrete transcriptional programs initiated following the onset of neurogenesis were identified based upon differential expression between WT timepoints: *early-increasing*, genes significantly upregulated between days 20 and 30 (20-30$_{up}^{WT}$) and also days 30 and 60 (30-60$_{up}^{WT}$); *early-stable* genes, present in 20-30$_{up}^{WT}$ and 20-60$_{up}^{WT}$ but not 30-60$_{up}^{WT}$; *early-transient* (20-30$_{up}^{WT}$ but not 20-60$_{up}^{WT}$); and *late* (30-60$_{up}^{WT}$ but not 20-30$_{up}^{WT}$). **(C)** Enrichment for common schizophrenia risk variants 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$^{WT\ only}$). Tests condition on all expressed genes; Bonferroni p-value correction is for 7 tests. **(D)** SZ GWAS enrichment in programs derived from scRNAseq of human foetal cortex **(E)** Identification of programs over-represented for the targets of key regulators ($P^{overlap}$) when compared to all expressed genes; all program-regulator enrichments identified were taken forward for genetic analysis, testing whether regulator targets were more highly enriched for SZ association than other genes in that program (P^{GWAS}). **(F)** Enrichment for schizophrenia common variant association in GO terms over-represented amongst early-increasing$^{-/-}$ genes, conditioning on all expressed and all early-increasing$^{-/-}$ genes.

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). **(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) 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 (day 50, amplitude: t_{68}=0.6974, P=0.4879, frequency: t_{66}=0.5467, P=0.5865; day 60 amplitude: t_{59}=1.021, P=0.3114, frequency: t_{58}=0.7671, P=0.4464). **(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). All data presented as mean ± SEM.

Figure 6. LoF intolerant genes & neurodevelopmental/neuropsychiatric disorder risk variants localise to early neurogenic transcriptional programs. (A) Identification of programs enriched for LoF intolerant (LoFi) genes ($P^{overlap}$) when compared to all expressed genes. **(B)** LoFi genes were partitioned based on their overlap with early-transient$^{-/-}$, -stable$^{-/-}$ and -increasing$^{-/-}$ sets. 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, conditioning on all expressed genes. **(C)** A two-sided Poisson rate ratio test was used to identify programs enriched for LoF *de novo* mutations identified in individuals with SZ (Rees et al. in press), comparing each program to all other expressed genes **(D)** Firth's penalized regression test was used to identify programs with an increased rate of LoF singleton mutations identified in SZ cases compared to the rate in controls (Genovese et al., 2016), conditioning on the rate observed in other expressed genes. **(E)** A two-sided Poisson rate ratio test was used to identify programs enriched for LoF *de novo* mutations identified in individuals with ID/NDD and ASD (Satterstrom et al., 2019) when compared to all other expressed genes. Unaffected siblings (SIB) of ASD individuals (Satterstrom et al., 2019) were analysed as a control; values for SZ (C) included for comparison. Data points lying above the x axis indicate an increased rate (rate ratio > 1), those below indicate a reduced rate (rate ratio < 1). Dotted line indicates $P_{corrected}$ = 0.05 following Bonferroni correction for 4 tests (each disorder being tested in 4 programs). **(F)** Rate ratios (genes in program versus all other expressed genes) from tests shown in (E). Dotted line shows rate ratio of 1 (i.e. rate of mutations in program equals that in all other genes). **(G)** Programs were tested for enrichment in common risk variants contributing to ADHD (Demontis et al., 2019) and BIP (Stahl et al., 2019) and Alzheimer's disease (AD) (Lambert et al., 2013). Values for SZ (Figure 4C) included for comparison. **(H)** Effect sizes (β) from tests shown in (G).

Figure 7. Model of disease pathophysiology in early corticoneurogenesis (A) Summary of main GO term enrichments for each transcriptional program (Table S6) and identified regulatory interactions between them. **(B)** Summary of 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); and bipolar disorder (BIP) **(C)** Proposed model for DLG2 action in neurodevelopment: External cues transduced by DLG2-scaffolded complexes activate neurogenic programs underlying cell growth, migration and development of active signalling properties; this is normally tightly coupled to cell cycle exit, controlled via a DLG2-independent signalling pathway. *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.

Methods

hESC culture

All hESC lines were maintained at 37°C and 5% CO_2 in 6 well cell culture plates (Greiner) coated with 1% Matrigel hESC-Qualified 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 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) (Ran et al., 2013). 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/cm^2) 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.

Genetic validation

The gRNA pair had zero predicted off-target nickase sites (Figure S2). 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 (Table S1). In addition, genotyping on the Illumina PsychArray v1.1 revealed no CNV insertions/deletions in either *DLG2*$^{-/-}$ line relative to WT (Table S2).

Cortical differentiation

Differentiation to cortical projection neurons (Figure 1A) was achieved using the dual SMAD inhibition protocol (Chambers et al., 2009) with modifications (embryoid body to monolayer and replacement of KSR medium with N2B27 medium) suggested by (Cambray et al., 2012). 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% CO_2 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 µM 2-Mercaptoethanol (Thermo Fisher Scientific), which was supplemented with 100 nM LDN193189 (Cambridge Biosciences) and 10 µM SB431542 (Stratech 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 µg/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 µM Y27632 Dihydrochloride (ROCK inhibitor, Stratech 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 µg/ml poly-d-lysine hydrobromide (PDL, Sigma) and 15 µg/ml laminin (Sigma) in DPBS. Vitamin A was added to the differentiation media at day 26, standard 1x B27 Supplement (Thermo Fisher Scientific) 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 2µg/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) PBS (0.3% PBST). Primary antibodies, used at an assay dependent 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) or Cellinsight Cx7 High-Content Screening Platform (Thermo Fisher Scientific) with HCS Studio Cell Analysis software (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 H_2O 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).

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 time 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. Samples were analysed by LC-MS/MS on an Orbitrap Elite and data was processed and quantified as described previously (Woodley and Collins, 2019).

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 (Illuminia) 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 with PennCNV software (Wang et al., 2007).

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) (Andrews, 2010) and trimmed for adapter sequence and low base call quality (Phred score < 30 at ends) with cutadapt (v2.3) (Martin, 2011). The mapping of the trimmed reads was done using STAR (v2.7.0e) (Dobin et al., 2013) and the BAM files were produced in both genomic and transcriptomic coordinates and sorted using samtools (v1.9) (Li et al., 2009). The aligned and sorted BAM files were further assessed for quality using Picard tools (v2.20.2) (Broad Institue, 2019). 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. GRCh38.p13 was used as the reference genome and the comprehensive gene annotations on the primary assembly from Gencode (release 32) used as gene annotation. Gene and transcript-level quantifications were calculated using RSEM (v1.3.1) (Li and Dewey, 2011). 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) (Soneson et al., 2015). 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) (Love et al., 2014) 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 pipeline (Anders et al., 2012). 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) (Anders et al., 2015). Finally, differential exon usage was evaluated using DEXSeq (v1.30) (Anders et al., 2012) 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 for day 30, we compared differential expression results when performed separately for each line (i.e. KO1 v WT and KO2 v WT) at day 20: the timepoint closest to day 30 and with a comparable number of differentially expressed genes (Figure 1B). The overlap in expressed genes accounted for 99% of the genes expressed in each line and the overlap in differentially expressed genes (those with Bonferroni P < 0.05) was 85%. Fold

change and differential expression p-values were highly correlated (Spearman's $\rho = 0.95$ and 0.85 respectively) between lines. GO term enrichment analyses were also extremely robust. GO terms significantly enriched (Bonferroni $P < 0.05$) in up- and down-regulated genes (denoted ^{up} and ^{down} respectively) displayed a high degree of overlap between the 2 lines; overlapping terms accounted for 68% $KO1^{up}$, 94% $KO2^{up}$, 90% $KO1^{down}$ and 90% $KO2^{down}$ enriched GO terms. GO term p-values were strongly correlated (Spearman's $\rho^{up} = 0.85$, $\rho^{down} = 0.87$), as were GO term odds ratios – used in our iterative refinement procedure to identify semi-independent terms (Spearman's $\rho^{up} = 0.92$, $\rho^{down} = 0.95$). Differential expression data from each line is thus highly comparable and the potential impact of sample dropout is limited.

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 ($20\text{-}30_{up}^{WT}$); genes up-regulated in day 60 relative to day 30 ($30\text{-}60_{up}^{WT}$); and genes up-regulated in day 60 relative to day 20 ($20\text{-}60_{up}^{WT}$). Early-transient, early-stable, early-increasing and late programs were then defined based upon the intersections of these gene-sets as shown in Figure 4B.

Single-cell RNA sequencing data

Single-cell RNA-Seq gene expression data from (Nowakowski et al., 2017) were downloaded from bit.ly/cortexSingleCell. The principal cell-types present at each timepoint in culture (Figure 1B) were mapped onto sub-types identified *in vivo* as follows:

day 20 NPCs: "RG-div1", "RG-div2", "oRG", "tRG", "vRG", "IPC-div1", "IPC-div2", "IPC-nEN1", "IPC-nEN2", and "IPC-nEN3" clusters
day 30 neurons (new-born): "nEN-early1", "nEN-early2", and "nEN-late" clusters
day 60 neurons: "EN-PFC1","EN-PFC2","EN-PFC3","EN-V1-1","EN-V1-2", and "EN-V1-3" clusters

Differential expression analysis was performed on protein coding genes using Seurat (v 3.1.1) (Butler et al., 2018) with a Wilcoxon Rank Sum test and without any internal filtering. Analysis was restricted to cells with at least 5% of their genes expressed (tpm>0) and genes expressed in at least 5% of cells in the smallest group of the comparison. Differentially expressed genes (Bonferroni $P < 0.05$) were identified between pairs of timepoints and used to define transcriptional programs as above.

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* targets (Wang et al., 2018): 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 targets (Forrest et al., 2018): identifiers were updated using the gene_history file from NCBI.

FMRP targets (Darnell et al., 2011): 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 targets (Cotney et al., 2015): symbols (NCBI gene_info file) and locations (NCBI Build37.3) were used to map genes to NCBI/Entrez gene ids. Note: using (Sugathan et al., 2014) rather than (Cotney et al., 2015) to define direct CHD8 targets did not alter the observed pattern of overlap with transcriptional programs (Figure 4)(data not shown).

Indirect *CHD8* targets (Sugathan et al., 2014): 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 NPCs (Sugathan et al., 2014) or as *CHD8* mid-foetal promoter targets (Cotney et al., 2015). Using a stricter definition of genes with altered expression on *CHD8* shRNA knockdown than that taken by (Sugathan et al., 2014) - genes with a Bonferroni corrected differential expression $P < 0.05$, rather than genes with a nominal $P < 0.05$ - did not alter the pattern of overlap with transcriptional programs (data not shown).

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 (Figure 4E) and the overlap between LoFi genes and transcriptional programs (Figure 6A). In order to identify a semi-independent subset of functionally enriched annotations from the output of functional enrichment tests (Figure 4F, Table S5-6), 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 enriched annotations; re-tested these reduced gene-sets for enrichment 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

All 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 (Figure 4F, Table S7) or regulator targets (Figure 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(β) 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., 2018); those for ADHD (Demontis et al., 2019), bipolar disorder (Stahl et al., 2019) and AD (Lambert et al., 2013) were calculated using the MAGMA multi model, with a fixed 20,000 permutations for each gene.

Rare variant association

The *de novo* LoF mutations for SZ analysed here are described in (Rees et al., 2019b). *De novo* LoF mutations for NDD, ASD and unaffected siblings of individuals with ASD were taken from (Satterstrom et al., 2019): these were re-annotated using VEP (McLaren et al., 2016) 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. The expected rate of *de novo* LoF mutations in a set of genes was estimated using individual gene mutation rates (Ware et al., 2015). Firth's penalized regression was used to test gene-sets for an increased rate of LoF singleton mutations identified in cases compared to the rate of LoF singleton mutations identified in controls (case-control data from Genovese et al., 2016). The set of all expressed genes (all^{WT+KO}) was included as a covariate and all tests were two-sided.

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% CO_2 with 20X magnification phase contrast images taken of whole wells in every 2 hours for the analysis period. The StackReg plugin (Thévenaz, 2011) 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 (Figure 3F, G) were analysed by unpaired two-tailed Student's t-test.

Adhesion assay

Cells were differentiated towards cortical projection neurons as previously described and dissociated to single cells at day 25 by 15-minute incubation with Accutase solution (Sigma). 100,000 cells per well were plated onto various ECM substrates (Collagen I, Collagen II, Collagen IV, Fibronectin, Laminin, Tenascin, Vitronectin) and the ECM540 Cell Adhesion Array Kit (EMD Millipore) was used to assess cell adhesion. The degree of adhesion to each ECM substrate was quantified following cell staining by absorbance at 560 nm, using the CLARIOstar microplate reader (BMG Labtech). Data sets (Figure 1E) were analysed by two-way ANOVA with post hoc comparisons using Bonferroni correction, comparing to WT controls.

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 Lipofectaime 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 (MBF Bioscience) neuron tracing and analysis software package. Data sets (Figure 3A-D) were analysed by two-way ANOVA with post hoc comparisons using Bonferroni correction, comparing to WT controls.

Proliferation assay

hESCs were either maintained in E8 medium or differentiated towards cortical projection neurons as required for the assay. At the time point of interest single-cell suspensions were produced by 15-minute Accutase solution (Sigma) incubation and cells plated at a density of

300,000 cells per well into 5 wells of a 24-well. Cells were maintained in appropriate media and allowed to proliferate from this point for either a 5-10 day period, with the total number of cells present within a well quantified colorimetrically at 1 or 2 day intervals. Using components of the ECM540 Cell Adhesion Array Kit (EMD Millipore) all cells within a well were stained using Cell Stain Solution (EMD Millipore) and following repeated H_2O washes to remove excess, cells were lysed to release bound stain using Extraction Buffer (EMD Millipore). The absorbance at 560 nm of the Extraction Buffer, being relative to the total number of cells within the stained well, was quantified using a CLARIOstar microplate reader (BMG Labtech). Data sets (Figure 1F, G) 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. 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 $CaCl_2$; 2mM $MgCl_2$; 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 $CsMeSO_4$; 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 software (Anderson and Collingridge, 2007; http://www.winltp.com). Only cells with series resistance <25 MΩ with a change in series resistance <10% from the start were included in this study. Data were 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 $MgCl_2$, 1.25mM $CaCl_2$, 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 Na_2-ATP, 2mM Na-GTP, 1.2mM Na_2-phosphocreatine, 2mM $MgCl_2$, 1mM $CaCl_2$ and 11mM EGTA (all from Sigma), and pH was titrated to 7.3 by NaOH. The resistance of a patch pipette was 3–9 MΩ and the series resistance component was fully compensated using the bridge balance function of the instrument. The resting membrane potential (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).

Statistical analysis and data presentation

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<0.0001 vs. WT control. In panels for phenotypic assays in Figures 1 & 2, n indicates the number of replicates used (from at least two independent differentiations). All data presented as mean ± SEM. Dotted lines in the p value graphs (Figure 2A, 4A, 6E and 6G) indicate $P_{corrected}$ = 0.05 following Bonferroni correction.

Bonferroni test correction

GO term analyses were corrected for the ~4,200 terms tested.

Figure 2A – corrected for 8 tests (up & down regulated x 4 timepoints)

Figure 4A – corrected for 6 tests (up & down regulated x 3 pairs of timepoints)

Figure 4C – corrected for 7 tests (7 gene expression programs)

Figure 4D – corrected for 4 tests (4 gene expression programs)

Figure 4E ($P^{overlap}$) – corrected for 21 tests (7 regulator target sets x 3 programs)

Figure 4E (P^{GWAS}) – corrected for 10 tests (10 over-represented sets taken forward for genetic analysis)

Figure 4F – corrected for 16 tests (16 semi-independent over-represented GO terms)

Table S7– corrected for 13 tests (13 semi-independent over-represented GO terms)
Figure 6A – corrected for 4 tests (4 programs)
Figure 6B – uncorrected (secondary, exploratory analysis)
Figure 6C – corrected for 4 tests (4 programs)
Figure 6D – corrected for 4 tests (4 programs)
Figure 6E – each disorder (& SIB) corrected for 4 tests (4 programs)
Figure 6G – each disorder corrected for 4 tests (4 programs)

Supplemental Information

Figure S1. Generation of *DLG2*$^{-/-}$ human embryonic stem cell lines, Related to Figure 1
Figure S2. Quality of two gRNAs used in the study, Related to Figure 1
Figure S3. Validation of *DLG2*$^{-/-}$ human embryonic stem cells, Related to Figure 1
Figure S4. Quality control of *DLG2*$^{-/-}$ human embryonic stem cells, Related to Figure 1
Figure S5. Characterisation of cortical differentiation of *DLG2*$^{-/-}$ and WT hESCs, Related to Figure 1
Figure S6. The phenotype of two *DLG2* knockout cell lines, Related to Figure 1 & 2
Table S1. CRISPR/Cas9 off-target validation, Related to Figure 1, S1 & S2
Table S2. DLG2 and SCRIB unique peptides (LC-MS/MS of day 30 samples), Related to Figure 1 & S3
Table S3. Differential gene expression (KO v WT and successive WT timepoints), Related to Figure 1 & 4
Table S4. GO over-representation analysis (KO V WT days 15 and 20, up- and down-regulated genes), Related to Figure 1
Table S5. GO over-representation analysis (KO V WT day 30 down-regulated genes), Related to Figure 2 & 3
Table S6. GO over-representation analysis (neurogenic transcriptional programs), Related to Figure 4 & 7
Table S7. Schizophrenia GWAS enrichment (GO terms over-represented in early-stable$^{-/-}$), Related to Figure 4

References

Amiel, J., Rio, M., de Pontual, L., Redon, R., Malan, V., Boddaert, N., Plouin, P., Carter, N.P., Lyonnet, S., Munnich, A., et al. (2007). 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.
Anders, S., Reyes, A., and Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. Genome Res. *22*, 2008–2017.
Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.
Anderson, W.W., and Collingridge, G.L. (2007). Capabilities of the WinLTP data acquisition program extending beyond basic LTP experimental functions. J. Neurosci. Methods *162*, 346–

356.

Andrews, S. (2010). FastQC: A Quality Control tool for High Throughput Sequence Data. Available online at: *https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.*

Antar, L.N., Li, C., Zhang, H., Carroll, R.C., and Bassell, G.J. (2006). Local functions for FMRP in axon growth cone motility and activity-dependent regulation of filopodia and spine synapses. Mol. Cell. Neurosci. *32*, 37–48.

Anttila, V., Bulik-Sullivan, B., Finucane, H.K., Walters, R.K., Bras, J., Duncan, L., Escott-Price, V., Falcone, G.J., Gormley, P., Malik, R., et al. (2018). Analysis of shared heritability in common disorders of the brain. Science *360*, eaap8757.

Bagni, C., and Zukin, R.S. (2019). A Synaptic Perspective of Fragile X Syndrome and Autism Spectrum Disorders. Neuron *101*, 1070–1088.

Bando, Y., Irie, K., Shimomura, T., Umeshima, H., Kushida, Y., Kengaku, M., Fujiyoshi, Y., Hirano, T., and Tagawa, Y. (2016). Control of Spontaneous Ca $^{2+}$ Transients Is Critical for Neuronal Maturation in the Developing Neocortex. Cereb. Cortex *26*, 106–117.

Bateman, J., Shu, H., and Van Vactor, D. (2000). The Guanine Nucleotide Exchange Factor Trio Mediates Axonal Development in the Drosophila Embryo. Neuron *26*, 93–106.

von Bohlen und Halbach, O. (2007). Immunohistological markers for staging neurogenesis in adult hippocampus. Cell Tissue Res. *329*, 409–420.

Broad Institue (2019). Picard Tools. Available online at: *https://broadinstitute.github.io/picard/.*

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

Cambray, S., Arber, C., Little, G., Dougalis, A.G., De Paola, V., Ungless, M.A., Li, M., and Rodríguez, T.A. (2012). Activin induces cortical interneuron identity and differentiation in embryonic stem cell-derived telencephalic neural precursors. Nat. Commun. *3*, 841.

Chambers, S.M., Fasano, C.A., Papapetrou, E.P., Tomishima, M., Sadelain, M., and Studer, L. (2009). Highly Efficient neural conversion of ES/iPS by SMAD Inhibition. Nat. Biotechnol. *27*, 275–280.

Clifton, N.E., Hannon, E., Harwood, J.C., Di Florio, A., Thomas, K.L., Holmans, P.A., Walters, J.T.R., O'Donovan, M.C., Owen, M.J., Pocklington, A.J., et al. (2019). Dynamic expression of genes associated with schizophrenia and bipolar disorder across development. Transl. Psychiatry *9*, 74.

Cotney, J., Muhle, R.A., Sanders, S.J., Liu, L., Willsey, A.J., Niu, W., Liu, W., Klei, L., Lei, J., Yin, J., et al. (2015). The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. Nat. Commun. *6*, 6404.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

Craddock, N., and Owen, M.J. (2010). The Kraepelinian dichotomy - going, going... but still not gone. Br. J. Psychiatry *196*, 92–95.

Darnell, J.C., Van Driesche, S.J., Zhang, C., Hung, K.Y.S., Mele, A., Fraser, C.E., Stone, E.F., Chen, C., Fak, J.J., Chi, S.W., et al. (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell *146*, 247–261.

Demontis, D., Walters, R.K., Martin, J., Mattheisen, M., Als, T.D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., et al. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat. Genet. *51*, 63–75.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Englund, C., Fink, A., Lau, C., Pham, D., Daza, R.A.M., Bulfone, A., Kowalczyk, T., and Hevner, R.F. (2005). Pax6, Tbr2, and Tbr1 are expressed sequentially by radial glia, intermediate progenitor cells, and postmitotic neurons in developing neocortex. J. Neurosci. *25*, 247–251.

Flora, A., Garcia, J.J., Thaller, C., and Zoghbi, H.Y. (2007). The E-protein Tcf4 interacts with Math1 to regulate differentiation of a specific subset of neuronal progenitors. Proc. Natl. Acad. Sci. U. S. A. *104*, 15382–15387.

Forrest, M.P., Hill, M.J., Kavanagh, D.H., Tansey, K.E., Waite, A.J., and Blake, D.J. (2018). The Psychiatric Risk Gene Transcription Factor 4 (TCF4) Regulates Neurodevelopmental Pathways Associated With Schizophrenia, Autism, and Intellectual Disability. Schizophr. Bull. *44*, 1100–1110.

Frank, R.A.W., Komiyama, N.H., Ryan, T.J., Zhu, F., O'Dell, T.J., and Grant, S.G.N. (2016). NMDA receptors are selectively partitioned into complexes and supercomplexes during synapse maturation. Nat. Commun. *7*, 11264.

Fromer, M., Pocklington, A.J., Kavanagh, D.H., Williams, H.J., Dwyer, S., Gormley, P., Georgieva, L., Rees, E., Palta, P., Ruderfer, D.M., et al. (2014). De novo mutations in schizophrenia implicate synaptic networks. Nature *506*, 179–184.

Genovese, G., Fromer, M., Stahl, E.A., Ruderfer, D.M., Chambert, K., Landén, M., Moran, J.L., Purcell, S.M., Sklar, P., Sullivan, P.F., et al. (2016). Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. Nat. Neurosci. *19*, 1433–1441.

Girirajan, S., Rosenfeld, J.A., Coe, B.P., Parikh, S., Friedman, N., Goldstein, A., Filipink, R.A., McConnell, J.S., Angle, B., Meschino, W.S., et al. (2012). Phenotypic heterogeneity of genomic disorders and rare copy-number variants. N. Engl. J. Med. *367*, 1321–1331.

Harrison, P.J. (1997). Schizophrenia: a disorder of neurodevelopment? Curr. Opin. Neurobiol. *7*, 285–289.

Harrison, P.J. (1999). The neuropathology of schizophrenia A critical review of the data and their interpretation. Brain *122*, 593–624.

Hevner, R.F., Shi, L., Justice, N., Hsueh, Y.-P., Sheng, M., Smiga, S., Bulfone, A., Goffinet, A.M., Campagnoni, A.T., and Rubenstein, J.L.. (2001). Tbr1 Regulates Differentiation of the Preplate and Layer 6. Neuron *29*, 353–366.

Hill, M.J., and Bray, N.J. (2012). Evidence That Schizophrenia Risk Variation in the *ZNF804A* Gene Exerts Its Effects During Fetal Brain Development. Am. J. Psychiatry *169*, 1301–1308.

Husi, H., and Grant, S.G.N. (2001). Isolation of 2000-kDa complexes of N-methyl-D-aspartate receptor and postsynaptic density 95 from mouse brain. J. Neurochem. *77*, 281–291.

Jung, S., Park, R.-H., Kim, S., Jeon, Y.-J., Ham, D.-S., Jung, M.-Y., Kim, S.-S., Lee, Y.-D., Park, C.-H., and Suh-Kim, H. (2010). Id Proteins Facilitate Self-Renewal and Proliferation of Neural Stem Cells. Stem Cells Dev. *19*, 831–841.

Kamijo, S., Ishii, Y., Horigane, S.-I., Suzuki, K., Ohkura, M., Nakai, J., Fujii, H., Takemoto-Kimura, S., and Bito, H. (2018). A Critical Neurodevelopmental Role for L-Type Voltage-Gated Calcium Channels in Neurite Extension and Radial Migration. J. Neurosci. *38*, 5551–5566.

Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A.M.M., Pletikos, M., Meyer, K.A., Sedmak, G., et al. (2011). Spatio-temporal transcriptome of the human brain. Nature *478*, 483–489.

Katrancha, S.M., Wu, Y., Zhu, M., Eipper, B.A., Koleske, A.J., and Mains, R.E. (2017). Neurodevelopmental disease-associated de novo mutations and rare sequence variants affect TRIO GDP/GTP exchange factor activity. Hum. Mol. Genet. *26*, 4728–4740.

Katrancha, S.M., Shaw, J.E., Zhao, A.Y., Myers, S.A., Cocco, A.R., Jeng, A.T., Zhu, M., Pittenger, C., Greer, C.A., Carr, S.A., et al. (2019). Trio Haploinsufficiency Causes Neurodevelopmental Disease-Associated Deficits. Cell Rep. *26*, 2805-2817.e9.

Kirov, G., Pocklington, A.J., Holmans, P., Ivanov, D., Ikeda, M., Ruderfer, D., Moran, J., Chambert, K., Toncheva, D., Georgieva, L., et al. (2012). De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Mol. Psychiatry *17*, 142–153.

Krichevsky, A.M., and Kosik, K.S. (2001). Neuronal RNA Granules: A Link between RNA Localization and Stimulation-Dependent Translation. Neuron *32*, 683–696.

Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., DeStafano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., et al. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat. Genet. *45*, 1452–1458.

van de Leemput, J., Boles, N.C., Kiehl, T.R., Corneo, B., Lederman, P., Menon, V., Lee, C., Martinez, R.A., Levi, B.P., Thompson, C.L., et al. (2014). CORTECON: A temporal transcriptome analysis of in vitro human cerebral cortex development from human embryonic stem cells. Neuron *83*, 51–68.

de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol. *11*, e1004219.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup, 1000 Genome Project Data Processing (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Long, K.R., Newland, B., Florio, M., Kalebic, N., Langen, B., Kolterer, A., Wimberger, P., and Huttner, W.B. (2018). Extracellular Matrix Components HAPLN1, Lumican, and Collagen I

Cause Hyaluronic Acid-Dependent Folding of the Developing Human Neocortex. Neuron *99*, 702-719.e6.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal *17*, 10–12.

Martínez-Cerdeño, V. (2017). Dendrite and spine modifications in autism and related neurodevelopmental disorders in patients and animal models. Dev. Neurobiol. *77*, 393–404.

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. *17*, 122.

Murray, R.M., and Lewis, S.W. (1987). Is schizophrenia a neurodevelopmental disorder? Br. Med. J. (Clin. Res. Ed). *295*, 681–682.

Neale, B.M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K.E., Sabo, A., Lin, C.-F., Stevens, C., Wang, L.-S., Makarov, V., et al. (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature *485*, 242–245.

Niola, F., Zhao, X., Singh, D., Castano, A., Sullivan, R., Lauria, M., Nam, H., Zhuang, Y., Benezra, R., Di Bernardo, D., et al. (2012). Id proteins synchronize stemness and anchorage to the niche of neural stem cells. Nat. Cell Biol. *14*, 477–487.

Nowakowski, T.J., Bhaduri, A., Pollen, A.A., Alvarado, B., Mostajo-Radji, M.A., Di Lullo, E., Haeussler, M., Sandoval-Espinosa, C., Liu, S.J., Velmeshev, D., et al. (2017). Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. Science *358*, 1318–1323.

O'Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D., et al. (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature *485*, 246–250.

Owen, M.J., and O'Donovan, M.C. (2017). Schizophrenia and the neurodevelopmental continuum:evidence from genomics. World Psychiatry *16*, 227–235.

Owen, M.J., O'Donovan, M.C., Thapar, A., and Craddock, N. (2011). Neurodevelopmental hypothesis of schizophrenia. Br. J. Psychiatry *198*, 173–175.

Pardiñas, A.F., Holmans, P., Pocklington, A.J., Escott-Price, V., Ripke, S., Carrera, N., Legge, S.E., Bishop, S., Cameron, D., Hamshere, M.L., et al. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat. Genet. *50*, 381–389.

Pocklington, A.J., Rees, E., Walters, J.T.R., Han, J., Kavanagh, D.H., Chambert, K.D., Holmans, P., Moran, J.L., McCarroll, S.A., Kirov, G., et al. (2015). Novel Findings from CNVs Implicate Inhibitory and Excitatory Signaling Complexes in Schizophrenia. Neuron *86*, 1203–1214.

Polioudakis, D., de la Torre-Ubieta, L., Langerman, J., Elkins, A.G., Shi, X., Stein, J.L., Vuong, C.K., Nichterwitz, S., Gevorgian, M., Opland, C.K., et al. (2019). A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. Neuron *103*, 785-801.e8.

Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., and Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar

disorder. Nature *460*, 748–752.

Purcell, S.M., Moran, J.L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., O'Dushlaine, C., Chambert, K., Bergen, S.E., Kähler, A., et al. (2014). A polygenic burden of rare disruptive mutations in schizophrenia. Nature *506*, 185–190.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308.

Rees, E., Walters, J.T.R., Georgieva, L., Isles, A.R., Chambert, K.D., Richards, A.L., Mahoney-Davies, G., Legge, S.E., Moran, J.L., McCarroll, S.A., et al. (2014). Analysis of copy number variations at 15 schizophrenia-associated loci. Br. J. Psychiatry *204*, 108–114.

Rees, E., Carrera, N., Morgan, J., Hambridge, K., Escott-Price, V., Pocklington, A.J., Richards, A.L., Pardiñas, A.F., GROUP Investigators, G., McDonald, C., et al. (2019a). Targeted Sequencing of 10,198 Samples Confirms Abnormalities in Neuronal Activity and Implicates Voltage-Gated Sodium Channels in Schizophrenia Pathogenesis. Biol. Psychiatry *85*, 554–562.

Rees, E., Han, J., Morgan, J., Carrera, N., Escott-Price, V., Pocklington, A.J., Duffield, M., Hall, L., Legge, S.E., Pardiñas, A.F., et al. (2019b). De novo mutations identified by exome sequencing implicate rare missense variants in SLC6A1 in schizophrenia. Nat. Neurosci. *In press*.

Reiner, O., Karzbrun, E., Kshirsagar, A., and Kaibuchi, K. (2016). Regulation of neuronal migration, an emerging topic in autism spectrum disorders. J. Neurochem. *136*, 440–456.

Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J.L., Kähler, A.K., Akterin, S., Bergen, S.E., Collins, A.L., Crowley, J.J., Fromer, M., et al. (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat. Genet. *45*, 1150–1159.

Ripke, S., Neale, B.M., Corvin, A., Walters, J.T., Farh, K.-H., Holmans, P.A., Lee, P., Bulik-Sullivan, B., Collier, D.A., Huang, H., et al. (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421–427.

Robinson, R.B., and Siegelbaum, S.A. (2003). Hyperpolarization-Activated Cation Currents: From Molecules to Physiological Function. Annu. Rev. Physiol. *65*, 453–480.

De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., Cicek, A.E., Kou, Y., Liu, L., Fromer, M., Walker, S., et al. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. Nature *515*, 209–215.

Ruzzo, E.K., Pérez-Cano, L., Jung, J.-Y., Wang, L., Kashef-Haghighi, D., Hartl, C., Singh, C., Xu, J., Hoekstra, J.N., Leventhal, O., et al. (2019). Inherited and De Novo Genetic Risk for Autism Impacts Shared Networks. Cell *178*, 850-866.e26.

Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L., et al. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature *485*, 237–241.

Sanders, S.J., He, X., Willsey, A.J., Ercan-Sencicek, A.G., Samocha, K.E., Cicek, A.E., Murtha, M.T., Bal, V.H., Bishop, S.L., Dong, S., et al. (2015). Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron *87*, 1215–1233.

Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., Rubeis, S. De, An, J.-Y., Peng, M., Collins,

R., Grove, J., Klei, L., et al. (2019). Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. BioRxiv 484113.

Schafer, S.T., Paquola, A.C.M., Stern, S., Gosselin, D., Ku, M., Pena, M., Kuret, T.J.M., Liyanage, M., Mansour, A.A.F., Jaeger, B.N., et al. (2019). Pathological priming causes developmental gene network heterochronicity in autistic subject-derived neurons. Nat. Neurosci. *22*, 243–255.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of Image Analysis. Nat. Methods *9*, 671–675.

Sebat, J., Levy, D.L., and McCarthy, S.E. (2009). Rare structural variants in schizophrenia: one disorder, multiple mutations; one mutation, multiple disorders. Trends Genet. *25*, 528–535.

Simms, B.A., and Zamponi, G.W. (2014). Neuronal Voltage-Gated Calcium Channels: Structure, Function, and Dysfunction. Neuron *82*, 24–45.

Singh, T., Kurki, M.I., Curtis, D., Purcell, S.M., Crooks, L., McRae, J., Suvisaari, J., Chheda, H., Blackwood, D., Breen, G., et al. (2016). Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. Nat. Neurosci. *19*, 571–577.

Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Research *4*, 1521.

Splawski, I., Timothy, K.W., Sharpe, L.M., Decher, N., Kumar, P., Bloise, R., Napolitano, C., Schwartz, P.J., Joseph, R.M., Condouris, K., et al. (2004). CaV1.2 Calcium Channel Dysfunction Causes a Multisystem Disorder Including Arrhythmia and Autism. Cell *119*, 19–31.

Stahl, E.A., Breen, G., Forstner, A.J., McQuillin, A., Ripke, S., Trubetskoy, V., Mattheisen, M., Wang, Y., Coleman, J.R.I., Gaspar, H.A., et al. (2019). Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat. Genet. *51*, 793–803.

Stephens, R., Lim, K., Portela, M., Humbert, P.O., and Richardson, H.E. (2018). The Scribble Cell Polarity Module in the Regulation of Cell Signaling in Tissue Development and Tumorigenesis. J. Mol. Biol. *430*, 3585–3612.

Sugathan, A., Biagioli, M., Golzio, C., Erdin, S., Blumenthal, I., Manavalan, P., Ragavendran, A., Brand, H., Lucente, D., Miles, J., et al. (2014). CHD8 regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors. Proc. Natl. Acad. Sci. U. S. A. *111*, E4468-77.

Szatkiewicz, J.P., O'Dushlaine, C., Chen, G., Chambert, K., Moran, J.L., Neale, B.M., Fromer, M., Ruderfer, D., Akterin, S., Bergen, S.E., et al. (2014). Copy number variation in schizophrenia in Sweden. Mol. Psychiatry *19*, 762–773.

Telley, L., Agirman, G., Prados, J., Amberg, N., Fièvre, S., Oberst, P., Bartolini, G., Vitali, I., Cadilhac, C., Hippenmeyer, S., et al. (2019). Temporal patterning of apical progenitors and their daughter neurons in the developing neocortex. Science *364*, eaav2522.

Thévenaz, P. (2011). StackReg - An ImageJ plugin for the recursive alignment of a stack of images.

van Haren, J., Boudeau, J., Schmidt, S., Basu, S., Liu, Z., Lammers, D., Demmers, J., Benhari, J., Grosveld, F., Debant, A., et al. (2014). Dynamic Microtubules Catalyze Formation of Navigator-TRIO Complexes to Regulate Neurite Extension. Curr. Biol. *24*, 1778–1785.

Verkerk, A.J.M.H., Pieretti, M., Sutcliffe, J.S., Fu, Y.-H., Kuhl, D.P.A., Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F., Zhang, F., et al. (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell *65*, 905–914.

Walker, R.L., Ramaswami, G., Hartl, C., Mancuso, N., Gandal, M.J., de la Torre-Ubieta, L., Pasaniuc, B., Stein, J.L., and Geschwind, D.H. (2019). Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. Cell *179*, 750-771.e22.

Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F.C.P., Clarke, D., Gu, M., Emani, P., Yang, Y.T., et al. (2018). Comprehensive functional genomic resource and integrative model for the human brain. Science *362*, eaat8464.

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F.A., Hakonarson, H., and Bucan, M. (2007). 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.

Ware, J.S., Samocha, K.E., Homsy, J., and Daly, M.J. (2015). Interpreting de novo Variation in Human Disease Using denovolyzeR. Curr. Protoc. Hum. Genet. *87*, 7.25.1-7.25.15.

Watanabe, H., and Minamino, T. (2016). Genetics of Brugada syndrome. J. Hum. Genet. *61*, 57–60.

Weinberger, D.R. (1987). Implications of Normal Brain Development for the Pathogenesis of Schizophrenia. Arch. Gen. Psychiatry *44*, 660–669.

Williams, N.M., Zaharieva, I., Martin, A., Langley, K., Mantripragada, K., Fossdal, R., Stefansson, H., Stefansson, K., Magnusson, P., Gudmundsson, O.O., et al. (2010). Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. Lancet (London, England) *376*, 1401–1408.

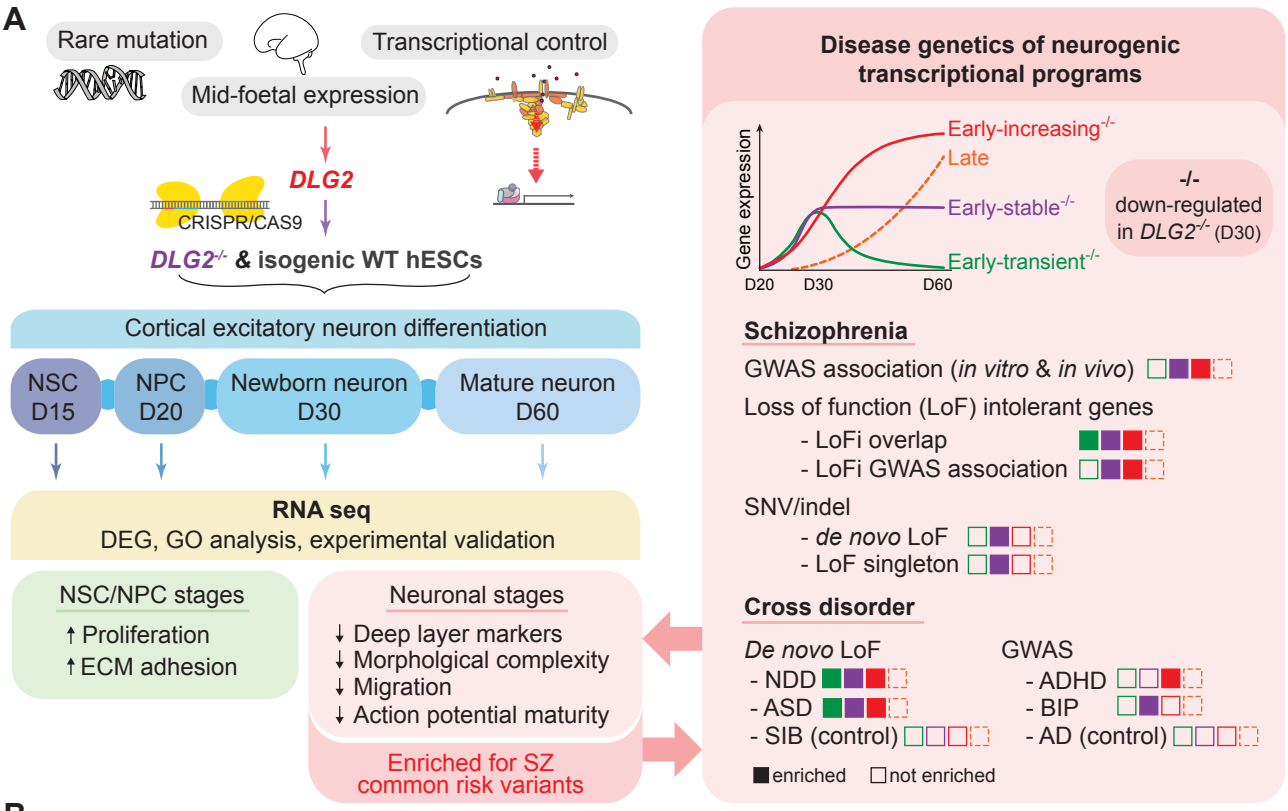
Willsey, A.J., Sanders, S.J., Li, M., Dong, S., Tebbenkamp, A.T., Muhle, R.A., Reilly, S.K., Lin, L., Fertuzinhos, S., Miller, J.A., et al. (2013). Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell *155*, 997–1007.

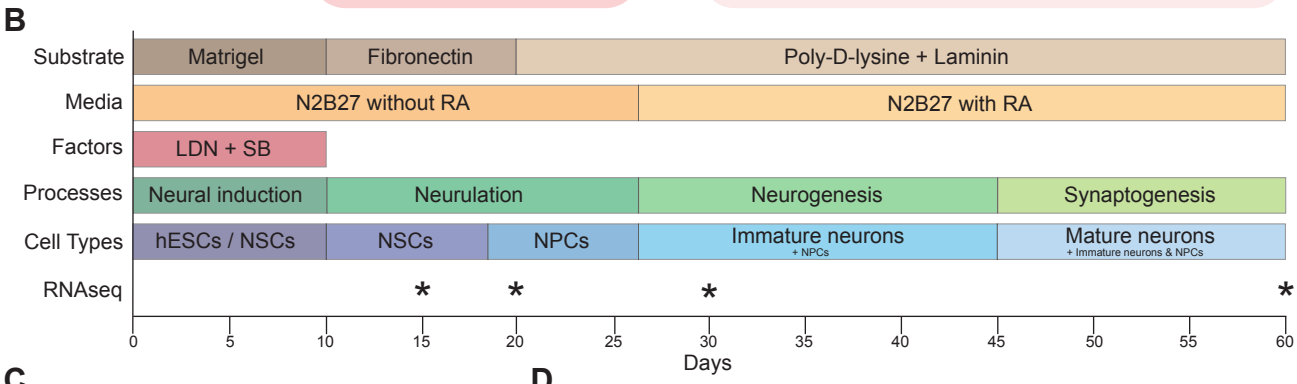
Wolff, M., Johannesen, K.M., Hedrich, U.B.S., Masnada, S., Rubboli, G., Gardella, E., Lesca, G., Ville, D., Milh, M., Villard, L., et al. (2017). Genetic and phenotypic heterogeneity suggest therapeutic implications in SCN2A-related disorders. Brain *140*, 1316–1336.

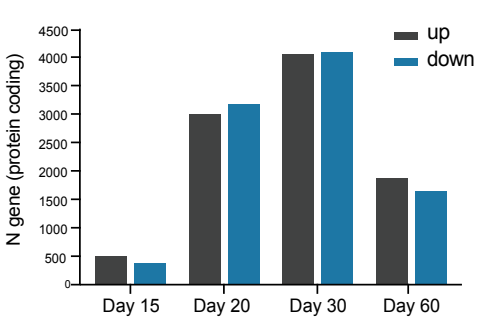
Woodley, K.T., and Collins, M.O. (2019). S-acylated Golga7b stabilises DHHC 5 at the plasma membrane to regulate cell adhesion. EMBO Rep. *20*, e47472.

Yadav, T., Quivy, J.P., and Almouzni, G. (2018). Chromatin plasticity: A versatile landscape that underlies cell fate and identity. Science (80-.). *361*, 1332–1336.

Figure 1







D

NCBI	Symbol	log_2 fold change	SE	P	$P_{corrected}$
3397	ID1	3.52	0.11	$5.73x10^{-216}$	$7.57x10^{-212}$
3399	ID3	2.67	0.093	$4.85x10^{-183}$	$6.41x10^{-179}$
3398	ID2	2.29	0.083	$7.88x10^{-168}$	$1.04x10^{-163}$
151449	GDF7	3.70	0.15	$3.49x10^{-135}$	$4.61x10^{-131}$
1277	COL1A1	1.43	0.059	$5.12x10^{-130}$	$6.77x10^{-126}$

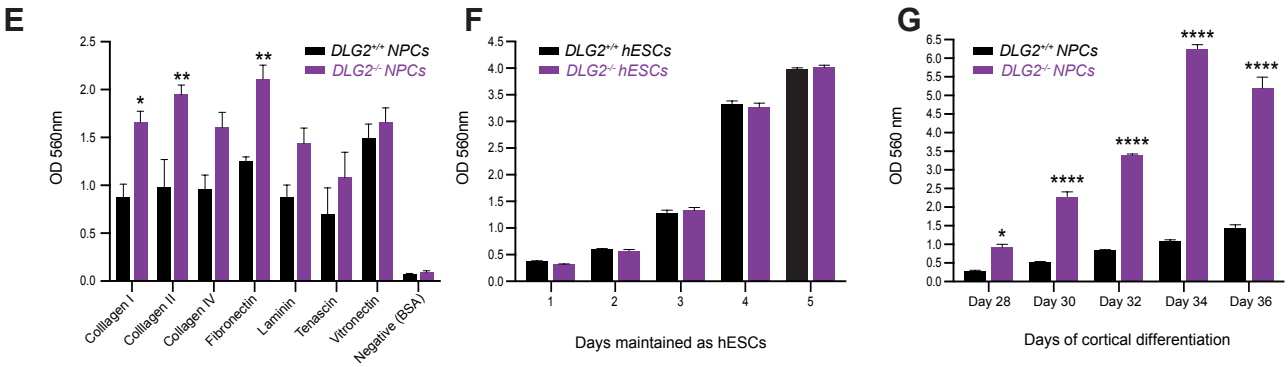
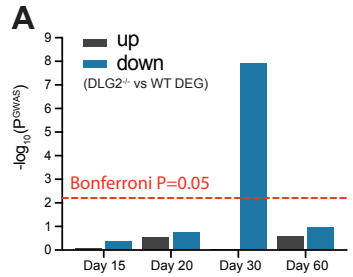
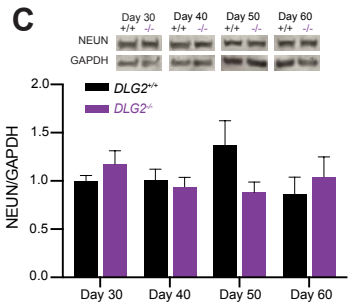


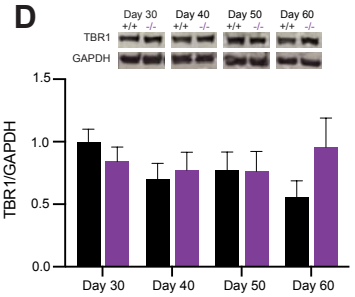
Figure 2

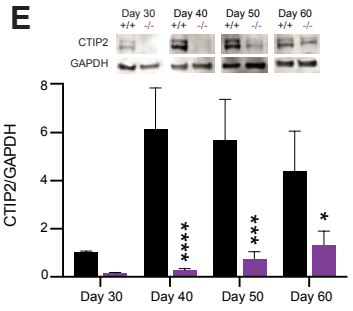


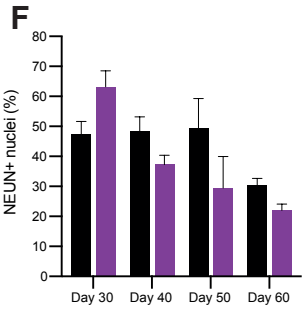
B

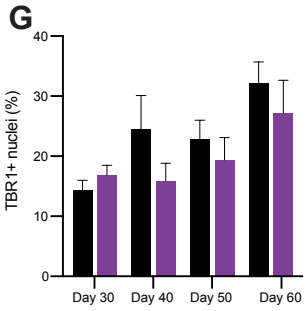
GO term	$N^{overlap}$	OR	$P^{overlap}$	$P^{overlap}_{corrected}$
DNA binding transcription factor activity RNA polymerase II specific	504	1.86	$5.71x10^{-23}$	$2.36x10^{-19}$
axon	149	3.18	$4.76x10^{-20}$	$1.97x10^{-16}$
regulation of cell morphogenesis	164	2.72	$7.44x10^{-18}$	$3.07x10^{-14}$
positive regulation of neuron projection development	96	3.93	$7.99x10^{-17}$	$3.30x10^{-13}$
neuron projection morphogenesis	157	2.60	$4.48x10^{-16}$	$1.85x10^{-12}$
synaptic membrane	80	4.30	$1.77x10^{-15}$	$7.29x10^{-12}$
peptidyl serine phosphorylation	82	3.53	$3.62x10^{-13}$	$1.49x10^{-9}$
regulation of small GTPase mediated signal transduction	104	2.39	$6.89x10^{-10}$	$2.85x10^{-6}$
calcium ion regulated exocytosis	23	10.77	$1.67x10^{-8}$	$6.90x10^{-5}$
synapse assembly	38	4.05	$9.34x10^{-8}$	0.00039
positive regulation of axonogenesis	30	5.02	$1.92x10^{-7}$	0.00079
membrane depolarization	27	5.27	$4.91x10^{-7}$	0.0020
microtubule end	20	7.80	$1.06x10^{-6}$	0.0044
neurotransmitter receptor complex	26	5.08	$1.15x10^{-6}$	0.0047
neuron recognition	22	6.44	$1.22x10^{-6}$	0.0050
regulation of adherens junction organization	32	3.95	$1.31x10^{-6}$	0.0054
modification dependent protein binding	62	2.39	$1.60x10^{-6}$	0.0066
protein import into nucleus	37	3.34	$2.04x10^{-6}$	0.0084
ubiquitin like protein ligase activity	80	2.09	$1.93x10^{-6}$	0.0080
protein phosphorylated amino acid binding	23	5.38	$2.93x10^{-6}$	0.012
regulation of synapse assembly	28	4.10	$4.03x10^{-6}$	0.017
neuron migration	27	3.95	$8.66x10^{-6}$	0.036
calcium ion transport into cytosol	27	3.95	$8.66x10^{-6}$	0.036

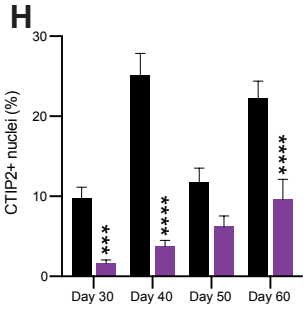












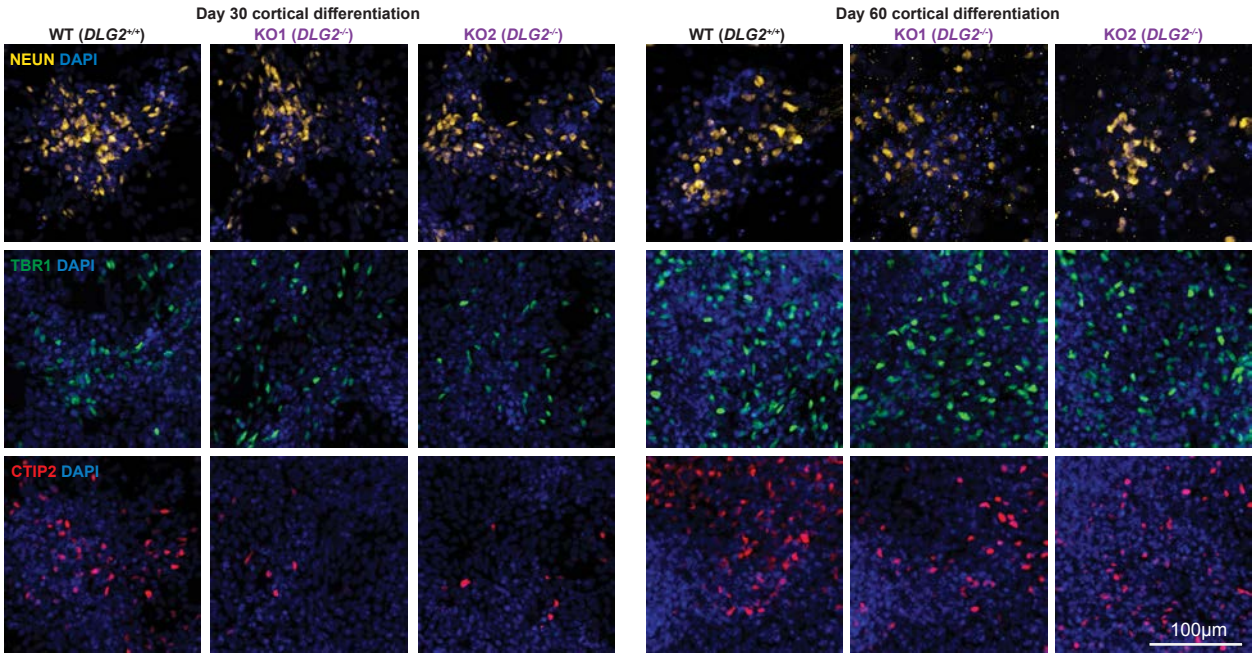


Figure 3

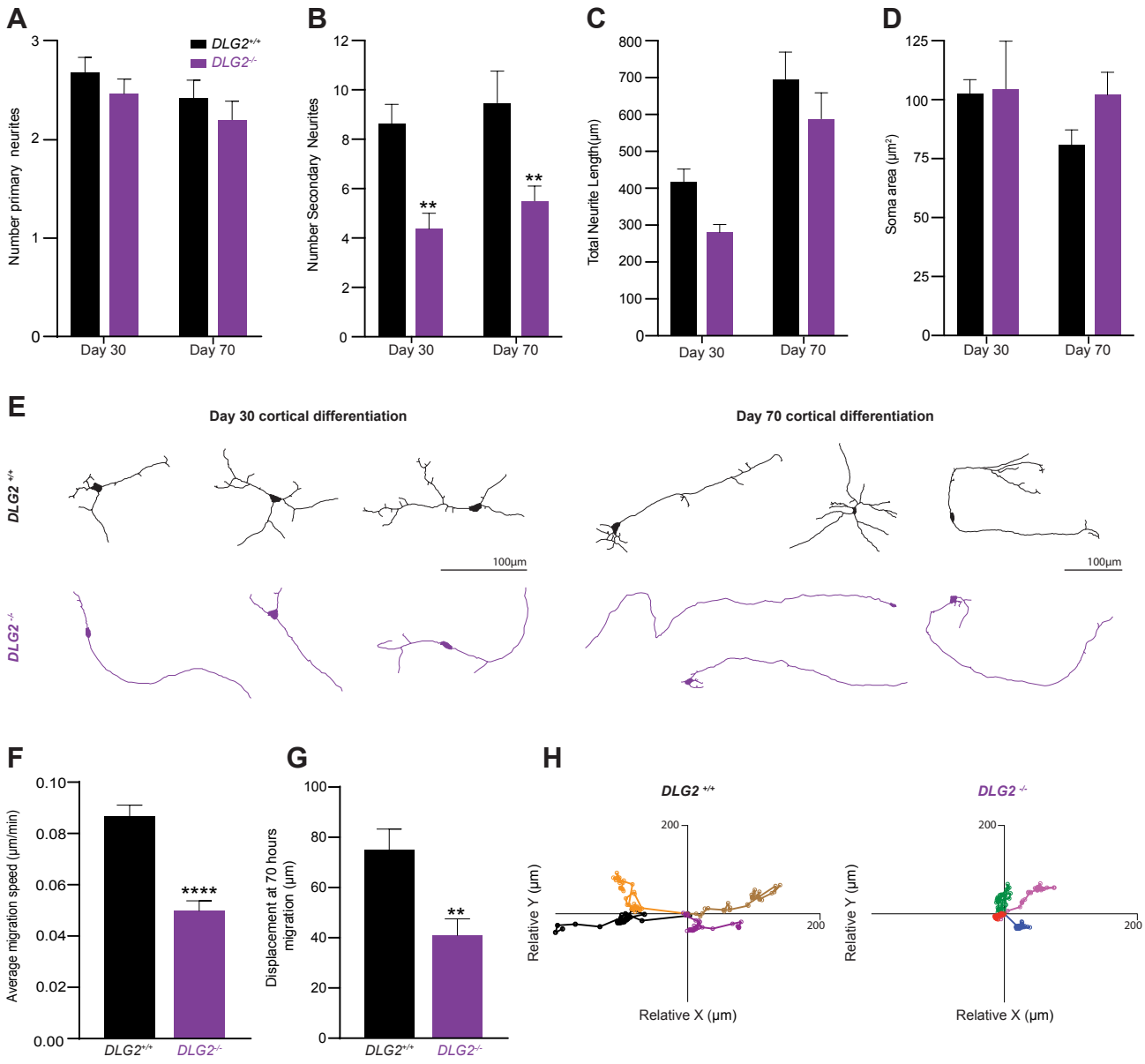
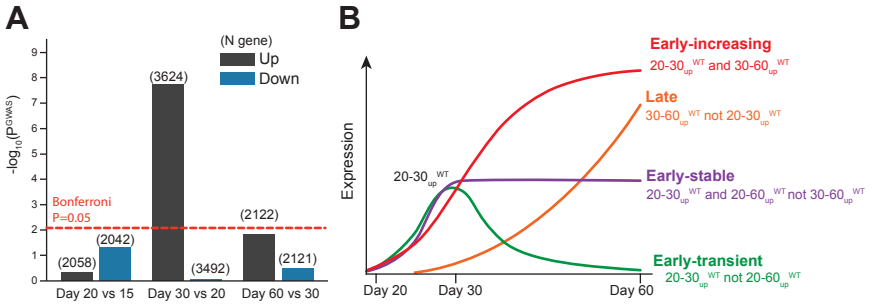


Figure 4



C

Gene expression profile	N	β	SE	P GWAS	P GWAS corrected
Late	1348	0.0025	0.0031	0.47	1
Early-increasing$^{-/-}$	**526**	**0.23**	**0.051**	**5.31x10^{-6}**	**3.72x10^{-5}**
Early-increasing$^{WT\ only}$	181	0.0014	0.086	0.49	1
Early-stable$^{-/-}$	**1672**	**0.16**	**0.029**	**1.04x10^{-8}**	**7.25x10^{-8}**
Early-stable$^{WT\ only}$	207	0.012	0.076	0.44	1
Early-transient$^{-/-}$	796	-0.014	0.040	0.64	1
Early-transient$^{WT\ only}$	145	-0.079	0.097	0.79	1

D

Nowakowski et al., 2017	N	β	SE	P GWAS	P GWAS corrected
Late$^{in\ vivo}$	1555	-0.0059	0.030	0.58	1
Early-increasing$^{in\ vivo}$	**252**	**0.25**	**0.073**	**0.00024**	**0.00095**
Early-stable$^{in\ vivo}$	**234**	**0.19**	**0.077**	**0.0059**	**0.023**
Early-transient$^{in\ vivo}$	750	0.047	0.040	0.12	0.47

E

	Regulator targets	N	N overlap	OR	P overlap	P overlap corrected	β	P GWAS	P GWAS corrected
Early increasing$^{-/-}$ (536)	**BCL11B targets (predicted)**	**695**	**81**	**3.81**	**2.52x10^{-20}**	**5.28x10^{-19}**	**0.39**	**0.0035**	**0.035**
	TBR1 targets (predicted)	**280**	**55**	**6.88**	**1.15x10^{-24}**	**2.42x10^{-23}**	**0.44**	**0.0039**	**0.039**
	FMRP targets	**816**	**143**	**7.10**	**9.14x10^{-59}**	**1.92x10^{-57}**	**0.34**	**0.00067**	**0.0067**
	TCF4 targets (direct)	**3544**	**231**	**2.39**	**2.60x10^{-21}**	**5.45x10^{-20}**	**0.33**	**6.01x10^{-5}**	**0.00060**
	CHD8 targets (indirect, up)	562	8	0.40	1	1			
	CHD8 targets (indirect, down)	**315**	**37**	**3.60**	**6.83x10^{-10}**	**1.43x10^{-8}**	-0.33	0.086	0.86
	CHD8 midfoetal targets (direct)	2549	88	0.90	0.82	1			

	Regulator targets	N	N overlap	OR	P overlap	P overlap corrected	β	P GWAS	P GWAS corrected
Early stable$^{-/-}$ (1713)	**BCL11B targets (predicted)**	**695**	**132**	**1.78**	**2.92x10^{-8}**	**6.14x10^{-7}**	0.021	0.85	1
	TBR1 targets (predicted)	280	43	1.34	0.051	1			
	FMRP targets	**816**	**306**	**5.10**	**7.60x10^{-85}**	**1.60x10^{-83}**	0.018	0.79	1
	TCF4 targets (direct)	**3544**	**594**	**1.73**	**7.87x10^{-23}**	**1.65x10^{-21}**	0.13	0.0079	0.079
	CHD8 targets (indirect, up)	562	23	0.30	1	1			
	CHD8 targets (indirect, down)	315	53	1.50	0.0064	0.13			
	CHD8 midfoetal targets (direct)	**2549**	**537**	**2.40**	**1.8x10^{-48}**	**3.78x10^{-47}**	0.031	0.52	1

	Regulator targets	N	N overlap	OR	P overlap	P overlap corrected	β	P GWAS	P GWAS corrected
Early transient$^{-/-}$ (826)	BCL11B targets (predicted)	695	31	0.75	0.95	1			
	TBR1 targets (predicted)	280	8	0.47	0.99	1			
	FMRP targets	816	49	1	0.41	1			
	TCF4 targets (direct)	3544	200	0.97	0.67	1			
	CHD8 targets (indirect, up)	562	13	0.40	1	1			
	CHD8 targets (indirect, down)	315	19	1	0.46	1			
	CHD8 midfoetal targets (direct)	**2549**	**231**	**1.87**	**8.05x10^{-14}**	**1.69x10^{-12}**	0.050	0.49	1

F

GO Term	N	β	SE	P GWAS	P GWAS corrected
Calcium ion regulated exocytosis	11	-0.098	0.36	0.79	1
Positive regulation of filopodium assembly	7	1.28	0.48	0.0079	0.13
Glutamate receptor activity	8	-0.57	0.45	0.21	1
Solute cation antiporter activity	7	0.32	0.38	0.41	1
Postsynaptic specialization membrane	9	0.64	0.42	0.13	1
Positive regulation of axon extension	6	0.62	0.42	0.14	1
Membrane depolarization during action potential	**8**	**1.48**	**0.44**	**0.00072**	**0.011**
Regulation of neurotransmitter secretion	12	-0.049	0.30	0.87	1
Regulation of dendrite morphogenesis	11	0.46	0.35	0.19	1
Postsynaptic membrane	22	0.33	0.28	0.23	1
Axon guidance	24	0.27	0.24	0.25	1
Modulation of chemical synaptic transmission	43	0.29	0.18	0.11	1
Anterograde trans synaptic signaling	44	0.15	0.17	0.37	1
Presynapse	32	0.20	0.21	0.35	1
Somatodendritic compartment	**47**	**0.81**	**0.18**	**4.26x10^{-6}**	**6.82x10^{-5}**
Generation of neurons	95	0.33	0.12	0.0059	0.094

Figure 5

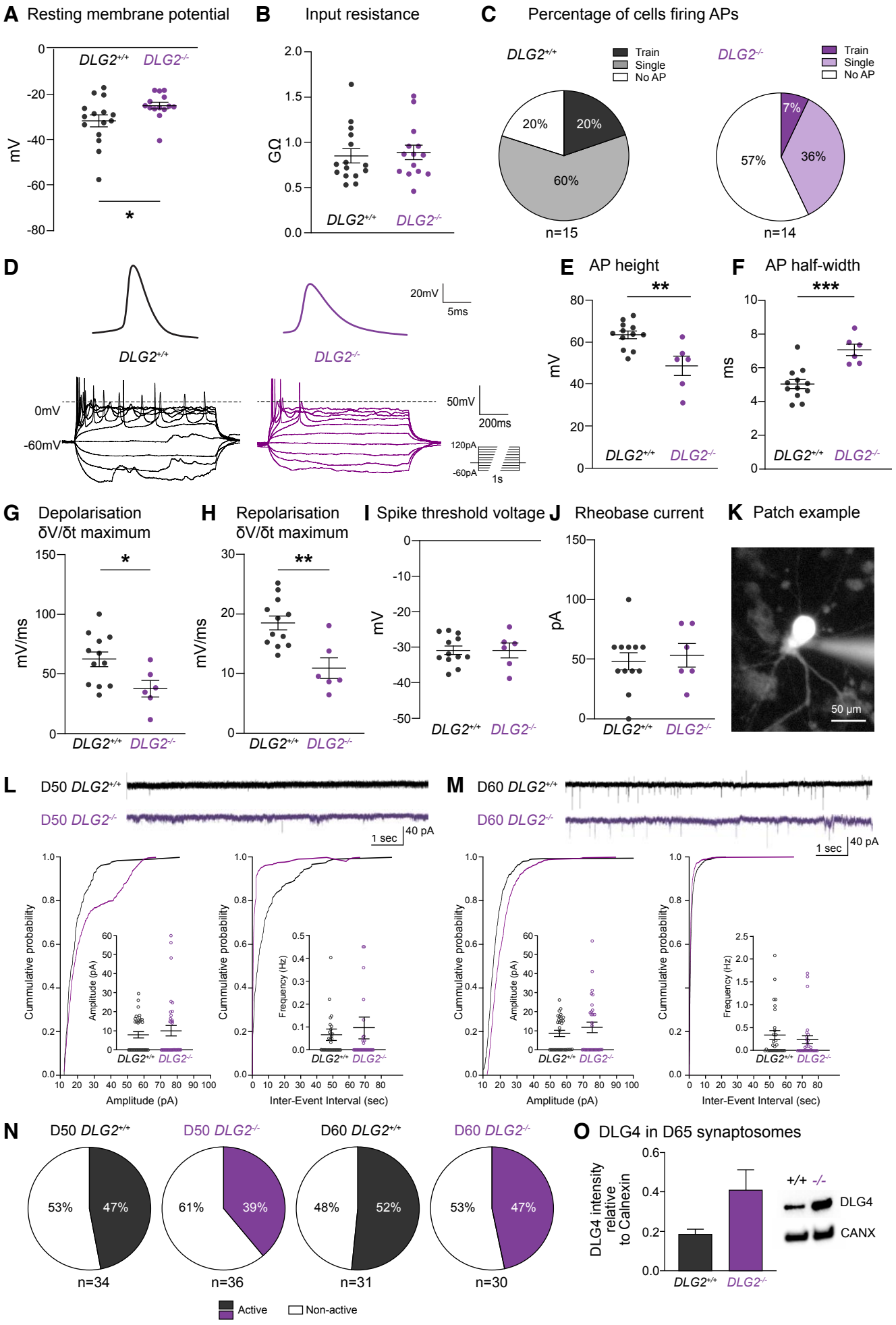
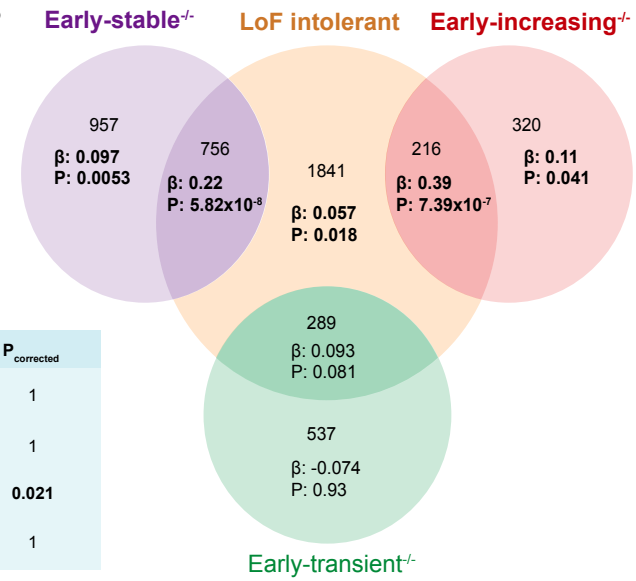


Figure 6

A N LoF intolerant = 3102

	N overlap	OR	P overlap	P overlap corrected
Late (1399)	221	0.68	1	1
Early-increasing$^{-/-}$ (536)	**216**	**2.66**	**2.84×10^{-25}**	**1.14×10^{-24}**
Early-stable$^{-/-}$ (1713)	**756**	**3.64**	**2.38×10^{-119}**	**9.54×10^{-119}**
Early-transient$^{-/-}$ (826)	**289**	**2.14**	**5.60×10^{-22}**	**2.24×10^{-21}**

B



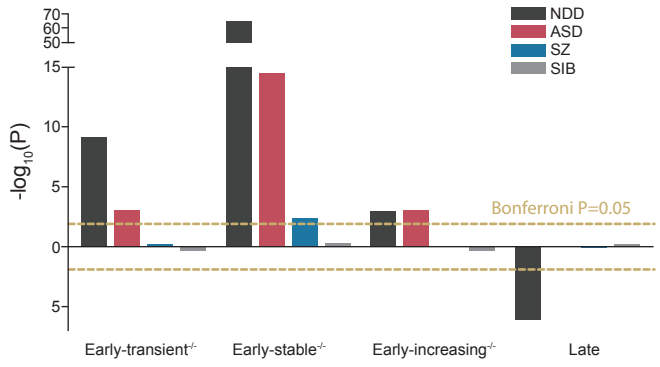
C SZ LoF (*de novo*)

	N observed	Rate ratio (95% CI)	P	P corrected
Late	22	0.89 (0.55-1.38)	0.75	1
Early-increasing$^{-/-}$	15	0.97 (0.54-1.62)	1	1
Early-stable$^{-/-}$	**81**	**1.45 (1.11-1.87)**	**0.005**	**0.021**
Early-transient$^{-/-}$	30	1.10 (0.73-1.60)	0.62	1

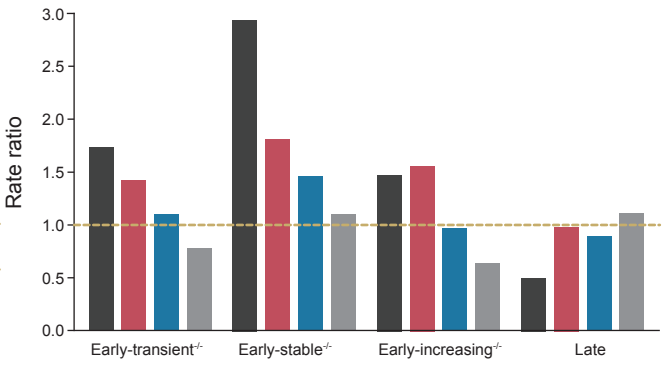
D SZ LoF singleton (case-control)

	N case	Rate case	N control	Rate control	OR (95% CI)	P	P corrected
Late	386	0.095	545	0.095	0.91 (0.80-1.05)	0.19	0.77
Early-increasing$^{-/-}$	214	0.052	219	0.038	1.28 (1.05-1.55)	0.013	0.052
Early-stable$^{-/-}$	**795**	**0.19**	**836**	**0.15**	**1.26 (1.13-1.40)**	**3.09×10^{-5}**	**0.00012**
Early-transient$^{-/-}$	403	0.099	484	0.085	1.09 (0.94-1.25)	0.25	1

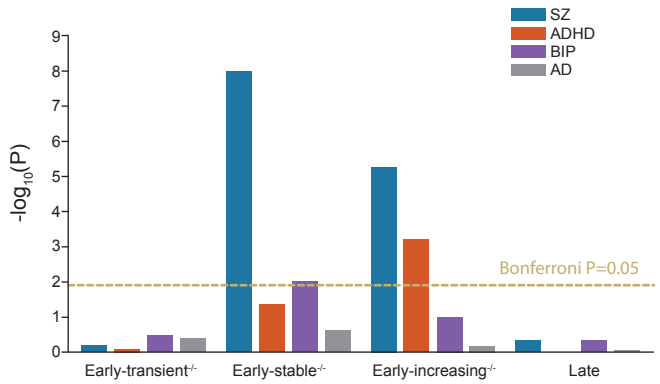
E *De novo* LoF (P)



F *De novo* LoF (Effect size)



G GWAS (P)



H GWAS (Effect size)

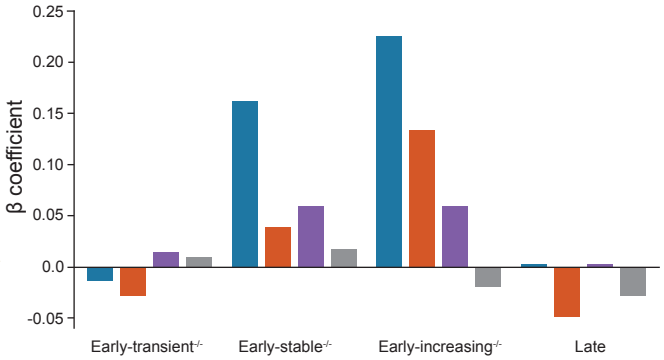
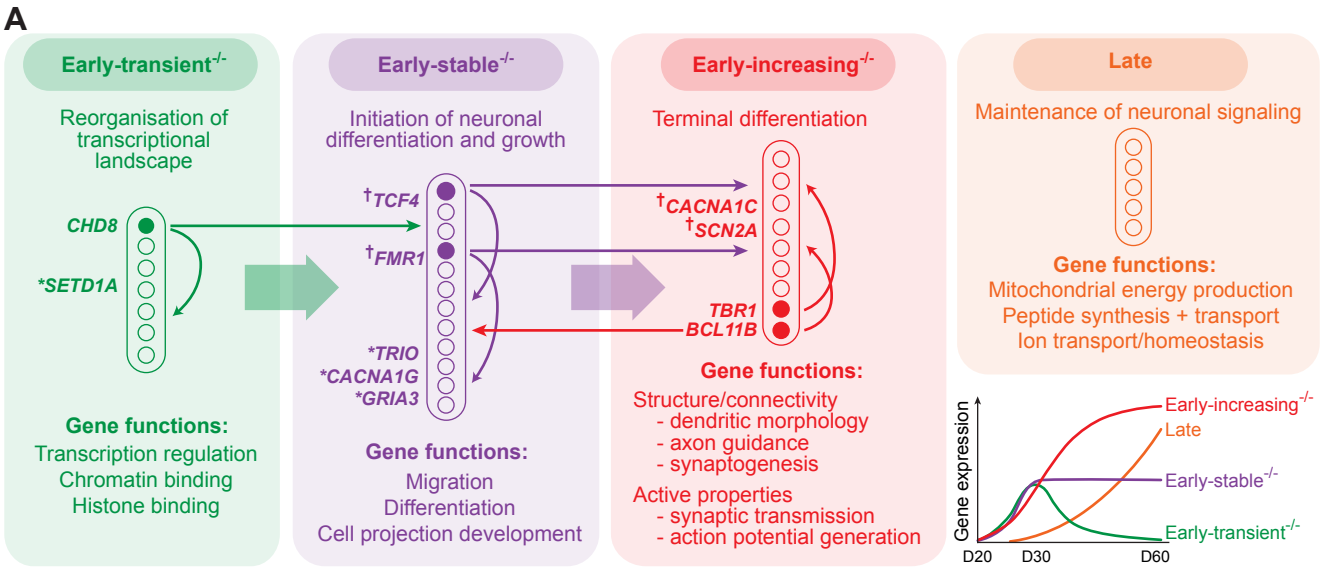


Figure 7



● 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 (http://schema.broadinstitute.org/)
† Rare variation causes Mendelian neurodevelopmental syndromes